eRAD-Fe: Emotion Recognition-Assisted Deep Learning Framework

Sun-Hee Kim, Ngoc Anh Thi Nguyen, Hyung-Jeong Yang, and Seong-Whan Lee, *Fellow, IEEE*

***Abstract*— With recent advancements in artificial intelligence technologies and human–computer interaction, strategies to identify the inner emotional states of humans through physiological signals such as electroencephalography (EEG) have been actively investigated and applied in various fields. Thus, there is an increasing demand for emotion analysis and recognition via EEG signals in real time. In this article, we proposed a new framework, emotion recognition-assisted deep learning framework from eeg signal (eRAD-Fe), to achieve the best recognition result from EEG signals. eRAD-Fe integrates three aspects by exploiting sliding-window segmentation method to enlarge the size of the training dataset, configuring the energy threshold-based multiclass common spatial patterns to extract the prominent features, and improving emotional state recognition performance based on a long short-term memory model. With our proposed recognition-assisted framework, the emotional classification accuracies were 82%, 72%, and 81% on three publicly available EEG datasets, such as SEED, DEAP, and DREAMER, respectively.**

***Index Terms*— Deep learning, electroencephalogram, emotion recognition, energy threshold, feature extraction, long short-term memory, multiclass common spatial pattern (CSP).**

I. Introduction

EMOTION refers to a change in the mind and mood state that manifests as psychological and physiological changes after a person recognizes an external stimulus generated by a distinct situation, event, or target [1]. Emotion is the most significant factor in daily interactions between humans when learning or making decisions. Recently, in various application technology fields, the demand for intelligent systems for interactions between computer systems and humans has increased [2]. Consequently, several studies have been conducted to automatically understand and recognize human intentions, states of mind, and emotions. Emotion recognition studies have been conducted on facial expression recognition, speech recognition, gesture recognition, and biometrics using the core technologies of artificial intelligence, such as deep learning, neural networks, and big data [3].

Emotion recognition technology identifies human emotions and internal states through video (facial expression), voice (tone), and biosignals (electroencephalography (EEG), pulse, and similar methods). Video-based emotion detection technology judges human emotional states by using feature points on the face or by classifying behaviors through human gestures [4]. Vocal emotion recognition technology analyzes patterns such as trembling, speed, tempo, and intonation of a human voice to understand the emotional state [5]. However, methods for recognizing emotions through facial expressions, motions, or voices lead to poor immediacy and reliability because users can artificially distort their emotional expressions. Conversely, biosignals are not directly controlled by the user and they contain information that is highly correlated with the internal emotional states [6]. Hence, recognizing emotions through biosignals, such as pulse, heart rate, temperature, EEGs, electromyographs, and electrocardiograms, has more recently been considered a promising strategy. Among these biosignals, EEG is advantageous because it captures the relevance of mentalization and the emotional change in real time during the process of recording a neurophysiological activity. As a result, many researchers have focused on EEGs to develop emotion detection systems [7], [8].

EEG captures electrical signals generated in the brain. Typically, electrodes are placed on the scalp to detect, amplify, and record the electrical activity generated in the brain cells of the target region [9], [10]. Emotion recognition research using EEGs has been conducted using various approaches. To extract features related to emotional states from EEGs, methodologies, such as approximate entropy, differential entropy (DE), higher order crossings (HOCs), common spatial patterns (CSPs), short-time Fourier transform (STFT), wavelet transform (WT), power spectral density (PSD), and fractal dimensions, including the Higuchi method [11], have often been used. Moreover, empirical mode decomposition (EMD) [12] and variational mode decomposition (VMD) [13] using intrinsic mode functions (IMFs) have recently been

Manuscript received August 9, 2021; accepted September 12, 2021. Date of publication September 23, 2021; date of current version October 7, 2021. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2021R1I1A1A01048455, in part by Vietnam National Foundation for Science and Technology Development (NAFOS-TED) under Grant 102.01-2020.27, in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under Grant NRF-2020R1A4A1019191, and in part by the Institute for Information and Communications Technology Promotion (IITP) Grant funded by the Government of South Korea (Development of BCI-Based Brain and Cognitive Computing Technology for Recognizing User's Intentions using Deep Learning under Grant 2017-0-00451; Artificial Intelligence Graduate School Program, Korea University, under Grant 2019-0- 00079). The Associate Editor coordinating the review process was Dr. Mohamad Forouzanfar. *(Corresponding author: Seong-Whan Lee.)*

Sun-Hee Kim and Seong-Whan Lee are with the Department of the Artificial Intelligence, Korea University, Seoul 02841, South Korea (e-mail: sunheekim@korea.ac.kr; sw.lee@korea.ac.kr).

Ngoc Anh Thi Nguyen is with the Faculty of Information Technology, The University of Danang—University of Science and Education, Da Nang 550000, Vietnam (e-mail: ngocanhnt@ued.udn.vn).

Hyung-Jeong Yang is with the Department of Computer Science, Chonnam National University, Gwangju 61186, South Korea (e-mail: hjyang@jnu.ac.kr).

Digital Object Identifier 10.1109/TIM.2021.3115195

proposed to provide meaningful time–frequency information about EEG signals and extract features from them. In addition, machine learning classifiers, such as support vector machine (SVM), K-nearest neighbor (K-NN), and artificial neural networks (ANNs), have been used to assess emotion recognition performance. In particular, deep learning-based classifiers have primarily been used in recent times [14].

In this article, we propose emotion recognition-assisted deep learning framework from eeg signal (eRAD-Fe), a novel emotion recognition framework, for enhancing the accuracy of emotional state classification based on EEG. The contributions of the proposed system are as follows.

1) *Favorable Framework:* eRAD-Fe is a new framework, which facilitates EEG-based emotion recognition by improving the classification accuracy rate through increment of the training datasets, optimal feature extraction, and rapid processing speed.
2) *Effectiveness:* The proposed framework provides a good performance compared to state-of-the-art methods in extracting features and classifying emotional states on publicly available EEG datasets.
3) *Applicability:* The proposed framework is a promising approach with a high classification rate of various emotional states based on EEG data. In addition, it works well on real-time data mechanisms comprising multiclasses.

The remainder of this article is organized as follows. Section II discusses related work on emotion recognition using EEGs. Section III presents the proposed framework. Section IV describes the experimental results and evaluation. Section V provides a discussion of this work. Finally, Section VI presents the conclusions drawn from the study.

II. Related Works

Emotion recognition research using EEG signals primarily consists of three main categories: preprocessing, feature extraction, and classification [15]. First, preprocessing eliminates noise and artifacts from the original data often by using bandpass filters or independent component analysis [16]. Preprocessing is easily and conveniently performed using an open toolbox, EEGLAB [17]. Second, feature extraction is an important stage in emotion recognition research. Various methods for extracting features have been proposed. For example, Duan *et al.* [18] extracted features using the DE method, which measured the complexity of a preprocessed EEG signal. Ali *et al.* [19] extracted features by combining the wavelet energy, modified energy, and wavelet entropy features for EEG-based emotion recognition for ambient-assisted living in the field of e-healthcare, and Ang *et al.* [20] extracted the features of emotional states using DWT method. Ackermann *et al.* [21] employed STFT, HOC, and Hilbert–Huang spectra to identify prominent features. Petrantonakis and Hadjileontiadis [22] extracted emotional state features by applying HOC, which captured patterns in the EEG sequence, to classify six basic emotional states. Thereafter, they improved the classification accuracy by approximately 2% using an enhanced HAF-HOC method, a modified version of HOC [23]. Basar *et al.* [24] extracted features relating to two emotional states by applying the CSP method, which maximized the difference in class variance, for emotion recognition.

However, existing methods for feature extraction, such as the Fourier transform (FT) method of frequency analysis and the WT time–frequency analysis method, assume that the signal is linear or that the basis functions are fixed. Given this drawback, if the signal is complex, the basis functions are not mapped to various attributes of the signal itself. Because the EEG signal is a complex nonlinear/nonstationary signal, not all significant attributes can be extracted effectively by FT or WT, and due to this, the performance of emotion recognition is reduced. In addition, because the entropy-based method is an analysis approach that is applied to a single time scale and measures the irregularity of time series, the quantification of the time-series complexity may not be accurate [25]. To overcome this weakness, multivariate multiscale entropy was proposed. However, this technique has the disadvantage of discarding high-frequency constituents and capturing only low-frequency constituents [26]. To overcome this weakness, EMD and VMD methods were recently proposed to extract features from an emotion dataset [12], [13].

Third, for classification, machine learning classifiers, such as SVM and KNN, have primarily been used in EEG-based emotion recognition studies [27]. Recently, with developments in deep learning technology, researchers have focused on deep learning mechanisms to boost the classification accuracy. In particular, Huang *et al.* [28] identified patterns in EEG data by combining emotional patches with a deep belief network model to explore the characteristics of temporal information. As a result, the authors proposed the DEEP model to recognize emotions with high classification accuracy. In addition, Wang *et al.* [29] proposed a long short-term memory (LSTM)-based classification framework for the classification learning of motor imagery EEG signals, and Alhagry *et al.* [30] applied the LSTM method for EEG-based emotion recognition using an end-to-end deep learning approach.

Li *et al.* [31] combined convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to detect emotional states at the trial level, and Li *et al.* [32] extracted features from each EEG channel using PSD. Subsequently, a hybrid deep neural network was proposed by combining a CNN with an LSTM-RNN to classify emotional states. Similarly, Tang *et al.* [33] used a bimodal-LSTM model to classify emotional states. Chen *et al.* [34] used a deep CNN (DCNN) model to improve the performance of their emotion recognition system by automatically learning the temporal and frequential features extracted from the DEAP data. However, most CNN-based approaches require complex preprocessing to transform a raw EEG signal into an image [35]. In contrast, the RNN-based approach is a useful method for the prediction and analysis of time-series data using deep learning algorithms without an additional data transformation process [32]. Because of its advantages, we used the RNN-based LSTM algorithm for the classification learning of time-series EEG data.

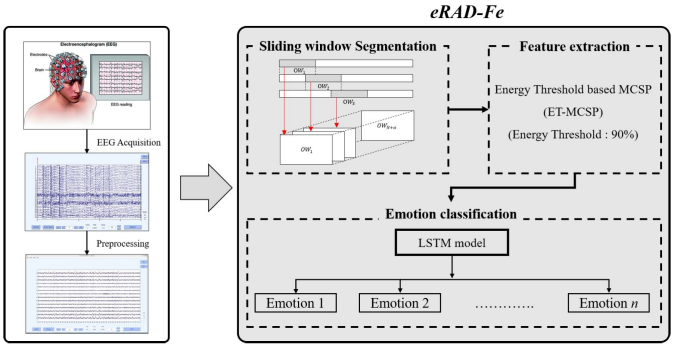

Fig. 1. Emotion recognition-assisted deep learning framework.

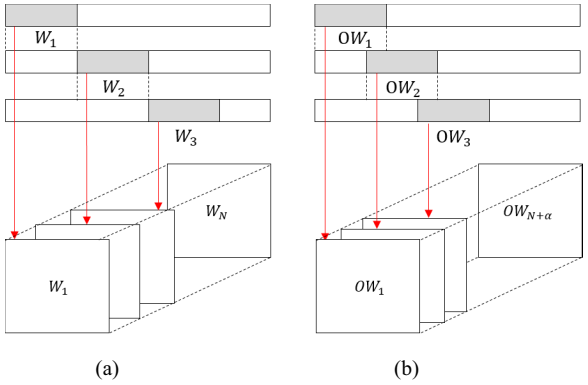

Fig. 2. Sliding-window segmentation with (a) nonoverlapped and (b) overlapped data.

III. Materials and Methods

In this section, a newly proposed framework for EEG-based emotion recognition is presented. Fig. 1 shows the eRAD-Fe system. To improve the performance of the framework, we increased the number of limited data points using the sliding-window segmentation method in the preprocessing phase. In the next phase, the optimal features were automatically extracted by applying an energy threshold-based multicommon spatial pattern (ET-MCSP). Finally, these extracted features were used as inputs for the deep learning-based LSTM classifier for emotional state recognition.

A. Sliding-Window Segmentation

The size of the EEG dataset was expanded using a sliding-window segmentation technique [36], as shown in Fig. 2. Because a deep learning model can process massive volumes of data effectively, the obtained sequences were used as inputs for the deep learning classifier. Sliding-window segmentation increases the number of trials (sample number) by segmenting the data to a specified window length. We selected a window size of 10 s and applied window overlap by moving the window at an interval of 7 s. In other words, as shown in Fig. 2, the data of the original windows, W_1 and W_2 [Fig. 2(a)] overlapped by 30% to generate OW_1 and OW_2 [Fig. 2(b)]. Fig. 2(b) shows the sliding operation with an overlap of 30% used in the proposed framework. We applied ET-MCSP to extract the optimal features with respect to the emotional states from the data obtained from the sliding-window segmentation in each EEG trial.

B. Energy Threshold-Based Multicommon Spatial Pattern

In this article, we applied ET-MCSP to extract emotional state features from the EEG data [37]. ET-MCSP automatically extracts the optimal features with regard to the emotional states by combining the energy-threshold and traditional MCSP methods [38]. MCSP is an expanded multiclass CSP algorithm that identifies a spatial matrix that maximizes the difference in class variance [24]. To extract the CSP from the composed multiclass data, MCPS calculates the composite covariance matrix, R, which is the sum of the covariance matrix of each class, given by the following equation:

$$R = R^i + R^2 + \cdots + R^i \tag{1}$$

where i indicates the number of classes. The eigen decomposition of R is given by the following equation:

$$R = Q\Lambda Q^T \tag{2}$$

where Q is a matrix of eigenvectors, Λ is a matrix of eigenvalue of R, and ^{T} is a transpose operator

$$P = \sqrt{\Lambda^{-1}}Q^T. \tag{3}$$

Using Q and Λ, obtained from the eigendecomposition of R, the whitening transformation matrix, P, is calculated using (3). P makes the covariance matrix of the transformed signal into a unit matrix and simultaneously makes variance into 1 to normalize the values of the transformed data. MCSP uses a one-versus-the-rest (OVR) [39] method to extract the CSP of each emotional state from the multiclass data. OVR assumes that one class is R^1 and the remainder of the classes are $\acute{R}^I$, where $\acute{R}^I = R^2 + \cdots + R^i$

$$PRP^T = PR^1P^T + P\acute{R}^IP^T = S^1 + S^{2+\cdots+i} = I. \tag{4}$$

The unit matrix requirement is satisfied by (4) when applying P in R. S is the transformed covariance matrix, and S^1 and $S^{2+\cdots+i}$ share a common eigenvector. The transformed covariance matrices (S^1, $S^{2+\cdots+i}$) can be decomposed as follows:

$$S^1 = \bar{Q}\Lambda^1\bar{Q}^T, \quad S^{2+\cdots+i} = \bar{Q}(1-\Lambda^1)\bar{Q}^T. \tag{5}$$

Meanwhile, the spatial matrix SM_{n*n} can be composed according to each emotional state as (6) through the common eigenvector, $\bar{Q}$, and P from (5). Afterward, when the spatial matrix is applied to the original data, X^i, and the transformed EEG signal, Z^i, can be obtained using (7)

$$\mathrm{SM}_{n*n} = \bar{Q}^T P \tag{6}$$

$$Z^i = \mathrm{SM}_{n*n}X^i \tag{7}$$

where X is the EEG data in an $n * t$ matrix with n channels and time t.

To extract the features, the MCSP method generally uses the maximum and minimum variance values of the data for each class, and they are transformed by a spatial matrix usually by employing data from the features [40], for example, doubling the number of classes ($2*i$) or $(i\,(i-1))/2$. However, the optimal coefficients for extracting the emotional states features are

unknown. Therefore, ET-MCSP uses an energy threshold [41], [42] to determine the number of spatial filter coefficients needed to classify the emotion states. In other words, as shown in (8), to obtain the spatial filter matrix, SF_{m*n}, for extracting the features according to the emotional states, m eigenvectors $\bar{Q}_m = (\bar{Q}_1, \ldots, \bar{Q}_m, \bar{Q}_{N-m+1}, \ldots, \bar{Q}_N)$ were selected based on the energy threshold, with $m \ll N$

$$\mathrm{SF}_{m*n} = \bar{Q}^T_{n*m} P_{n*n}. \tag{8}$$

The energy threshold of ET-MCSP was set to 90% to determine m. In this article, we determined the energy-threshold value based on the experimental results obtained from three EEG data.

C. LSTM-Based Deep Learning Model for Emotion Classification

In this article, we employed the RNN-based LSTM algorithm to recognize EEG-based emotional states. LSTM is widely used for data with temporal characteristics to resolve the long-term dependence problem [43], [44]. It is a structure in which the cell states, namely the forget gate, input gate, and output gate, are added to the existing RNN. The first stage of the LSTM is the forget gate, which determines the information to be discarded from the cell state using a sigmoid layer, as shown in (9). This phase receives h_{t-1} and x_t and sends a value between 0 and 1 to the cell state C_{t-1}. If the value is equal to 1, all the information is preserved; if it is 0, all the information is discarded

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{9}$$

where t is time, W is the weighted value, x is the input value, h is the value of the previous node, and b is the bias.

The next stage is the input gate (hidden state) as shown in (10), which determines parts of the new information to be stored in the cell state, and the sigmoid layer decides the value to update. Then, the tanh layer creates the $\bar{C}_t$ vector, which comprises new candidate values as shown in (11), and $\bar{C}_t$ adds on new cell state, C_t, as shown in (12)

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{10}$$

$$\bar{C}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right) \tag{11}$$

$$C_t = f_t * C_{t-1} + i_t * \bar{C}_t. \tag{12}$$

The final stage is the output gate, which determines the output. The output gate first receives the input data in the sigmoid layer and determines which part of the cell state is sent to the output, as shown in (13). Afterward, the cell state is placed on the tanh layer and is assigned a value between -1 and 1, which is multiplied by the output of the sigmoid layer, to generate the output, as given by (14). Thus, only the part that is sent as an output can be let out

$$O_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{13}$$

$$h_t = O_t * \tanh(C_t). \tag{14}$$

In this article, we employed multiple LSTM layers to classify the EEG emotion data. The composition of each layer of the classification model is shown in Fig. 3. The model comprised two LSTM layers with hidden units set to 125 and 100. Other parameters were set as follows: the dropout was 0.5, batch size was 30, and optimizer was "adam." In addition, we set the learning rate as 0.001, the maximum number of epochs as 50, and the activation function on the LSTM layer was "than."

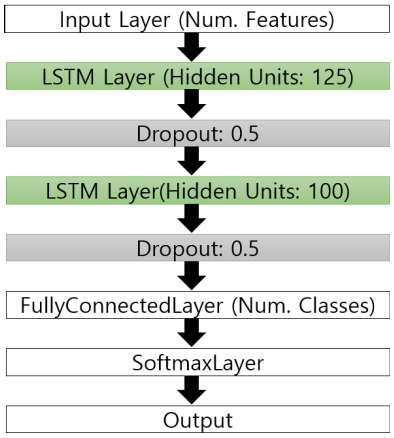

Fig. 3. LSTM structure for classification learning.

IV. Experimental Results

A. Data Description

To verify the effectiveness of the proposed eRAD-Fe, three publicly available datasets frequently used in emotion recognition research were selected: the SEED database [45], DEAP database [46], and wireless-based DREAMER database [47]. In the SEED data, 15 subjects (seven males and eight females) who participated in the experiment were showed approximately 15 Chinese film clips that evoked positive, neutral, and negative emotions, and the EEGs of the subjects were recorded for 4 min, while the clips were played. The data for each subject comprised 45 trials downsampled to 200 Hz, using a total of 62 channels (15 subjects $\times$ 45 trials $\times$ 62 channels $\times$ 4 min). In the DEAP dataset, 32 subjects who participated in the experiment were shown 40 music videos, and the physiological signals of the subjects were recorded. In addition, the subjects specified rating values according to four emotional states (valence, arousal, liking, and dominance) using consecutive numbers between 1 and 9. The length of the DEAP data was 63 s sampled at 128 Hz in 32 channels (32 subjects $\times$ 40 trials $\times$ 32 channels $\times$ 8064 data length). In the DREAMER dataset, 23 subjects were shown 18 video clips with audiovisual stimuli, and their physiological signals were measured using wireless Emotiv EPOC+ equipment with 14 channels for EEG and ECG recording. The length of each video clip was between 1 and 10 min, and the EEG data were sampled at 128 Hz. The subjects specified rating values according to three emotional states (valence, arousal, and dominance) using the consecutive numbers between 1 and 5.

A summary of the data used in this study is shown in Table I. As shown in Table I, all the subjects' data in the SEED dataset were used in the experiment. However, the DEAP and

TABLE I
NUMBER OF EEG EMOTION DATA FOR EXPERIMENTS

Emotion EEG data set	Num. subjects	Num. total trials	Num. channels	Rating (High, Low)	Num classes
SEED	15	675	62	-	3
DEAP	5	200	32	High:$\geq$7 Low: $\leq$3	4
DREAMER	10	180	14	High:$\geq$4 Low: $\leq$2	4

DREAMER datasets were divided into emotion classes based on the rating values of the subjects. For the DEAP data, the gap between the classes was handled by discarding records with ratings between 4 and 6 on the valence–arousal plane, and the remainder were divided into four classes: HVHA, LVHA, HVLA, and LVLA. For instance, the valence and arousal ratings in HVHA were greater than or equal to 7. Moreover, in HVLA, the valence rating was greater than or equal to 7, whereas the arousal rating was less than or equal to 3. The DREAMER data were categorized similar to the DEAP data using valence–arousal plane ratings greater than or equal to 4 as high and valence-arousal plane ratings less than or equal to 2 as low to divide the data into four classes: HVHA, LVHA, HVLA, and LVLA. We used the datasets composed of subjects within the range of the rating conditions and included all four classes for the DEAP and DREAMER data, as presented in Table I. In addition, we used an equal number of trials per class to solve the class imbalance problem.

B. Emotion Recognition-Assisted Framework

1) Sliding-Window Segmentation: The eRAD-Fe proposed in this article utilizes sliding-window segmentation with overlap to increase the processing speed and increase the number of training data for emotion classification based on deep learning model. Table II lists the experimental results obtained to determine the optimal window size and overlapping rate. We used a five-subject dataset chosen randomly from each dataset for this experiment, considering the number of subjects in the DEAP dataset. In addition, we used equal trials per class considering the class imbalance problem of each subject's data. We performed an experiment that used lengths of 1, 3, 5, 10, and 20 s to find the optimal window size for segmentation and the best overlapping rate according to 10%, 20%, 30%, 40%, and 50% and nonoverlapping (0%), that is, we compared and analyzed the emotional states classification accuracy by experiments in which the data were segmented according to each window size with or without overlapping, features extracted using ET-MCSP, and the emotional states classified by the LSTM model, which uses the extracted features as input. The classification accuracy was measured using tenfold cross validation.

To perform tenfold cross validation, we divided the total number of trials of each subject into ten groups. One group was used as the testing data and the remaining groups were used as training data for the classifiers (one group includes all classes). We repeated the trial ten times until each K-fold became a test set, and the obtained score from this process presented the classification accuracy of each subject. As shown in Table II, the accuracy according to each data point indicates the average classification accuracy obtained from five subjects. As a result, the SEED dataset indicated the highest average accuracy when 10-s window size and 30% overlapping were used. This is a 4% improved result than without overlapping and 2% improved results than when 5-s window size was used. In addition, the SEED dataset had the second highest average accuracy when 20 s window size and 30% overlapping were used. Meanwhile, the SEED data likewise DEAP data showed the highest classification accuracy as 81% when data segmented with 10-s window size and 30% overlapping. This is a 5% improvement compared to the case with nonoverlapping. The DREAMER data exhibited the highest accuracy of 83% in the 10-s window size when 30% overlapping was used, and it had an approximately 4% improvement over the case with nonoverlapping. Therefore, based on these results, we applied 10-s window size with 30% overlapping in the sliding-window segmentation for all experiments regarding the proposed emotion recognition-assisted framework.

TABLE II
CLASSIFICATION ACCURACY ACCORDING TO THE SLIDING-WINDOW SEGMENTATION BY EACH WINDOW SIZE AND OVERLAPPING RATE

Dataset	Overlapping Rate	Window Size				
		20 s	10 s	5 s	3 s	1 s
SEED	0%	77.31	81.24	70.06	71.28	61.25
	10%	81.92	82.89	79.60	71.85	63.97
	20%	84.02	83.87	81.58	74.65	67.53
	30%	85.43	**85.63**	83.48	76.31	71.40
	40%	81.67	82.22	72.89	71.11	58.33
	50%	79.11	72.22	75.00	68.33	63.34
DEAP	0%	70.85	76.31	75.12	65.64	53.19
	10%	74.58	79.94	78.25	67.30	62.61
	20%	75.08	76.33	78.15	70.98	64.46
	30%	78.85	**81.00**	78.95	73.79	69.66
	40%	63.89	69.44	65.22	46.79	43.06
	50%	75.83	76.08	68.61	63.06	64.44
DREAMER	0%	78.85	79.28	78.71	69.42	63.15
	10%	80.02	80.76	78.87	71.47	63.22
	20%	82.13	79.16	80.78	73.84	66.10
	30%	82.21	**83.01**	81.93	75.27	72.39
	40%	72.22	79.17	69.61	65.28	65.10
	50%	79.14	78.15	78.13	76.86	72.92

2) Automatic Feature Extraction: We used the ET-MCSP to extract the prominent features according to emotional states in the EEG dataset. The energy threshold was determined based on performance comparisons using various energy values. Fig. 4 shows the average number of features and the average classification accuracy for energy values between 75% and 95%, with a 5% difference between the values. As shown in Fig. 4(a), with 15 subjects in the SEED dataset, when an energy threshold of 95% was used, the average number of extracted features was 19.85, and the average classification accuracy was 83%. When the energy threshold was 90%, the average number of extracted features was 13.71, and the average accuracy was 82.54%. Thus, the number of features was reduced by approximately 30% when the

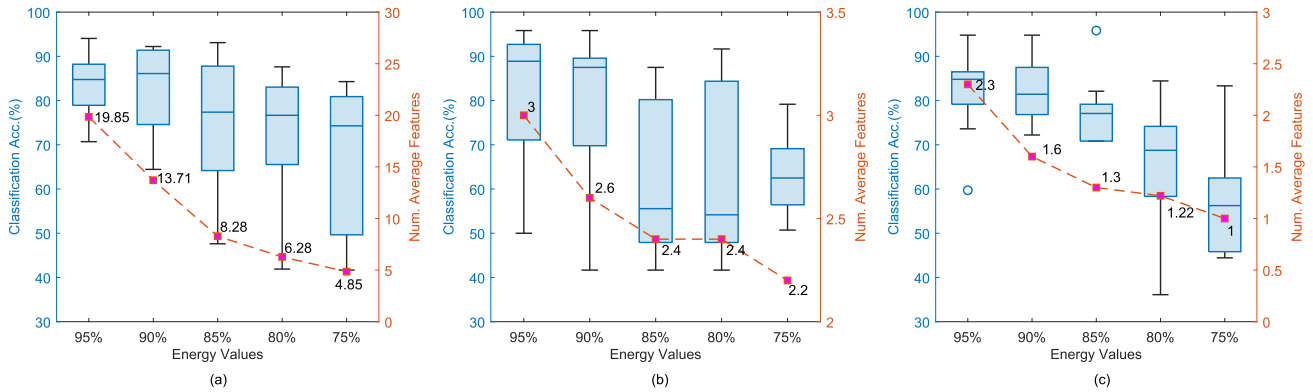

Fig. 4. Average number of features and classification accuracy according to energy value. (a) SEED data. (b) DEAP data. (c) DREAMER data.

TABLE III
SEED DATA CLASSIFICATION ACCURACY USING FEATURES EXTRACTION METHODS AND LSTM

No.Subject	MEMD				IVMD				ET-MCSP	
	Num. IMFs	ACC(%)	Num. IMFs	ACC(%)	Num. IMFs	ACC(%)	Num. IMFs	ACC(%)	Num. Feature	ACC(%)
Subject 1	3	85.0	6	81.7	3	71.7	6	78.3	5	71.1
Subject 2	3	38.3	6	50.0	3	45.0	6	41.7	4	50.0
Subject 3	3	51.6	6	83.3	3	60.0	6	71.7	14	73.9
Subject 4	3	51.6	6	38.3	3	53.3	6	60.0	15	76.1
Subject 5	3	65.0	6	53.3	3	63.3	6	68.3	19	81.7
Subject 6	3	75.0	6	60.0	3	81.7	6	66.7	17	89.4
Subject 7	3	55.0	6	76.7	3	56.7	6	60.0	8	94.4
Subject 8	3	73.3	6	73.3	3	80.0	6	73.3	12	80.0
Subject 9	3	76.6	6	73.3	3	78.3	6	86.7	18	98.9
Subject 10	3	40.0	6	56.7	3	41.7	6	55.0	8	68.9
Subject 11	3	76.6	6	70.0	3	78.3	6	80.0	10	95.0
Subject 12	3	48.3	6	53.3	3	50.0	6	70.0	17	82.2
Subject 13	3	85.0	6	86.7	3	76.7	6	71.7	13	86.1
Subject 14	3	78.3	6	75.0	3	65.0	6	71.7	19	70.0
Subject 15	3	66.8	6	75.9	3	83.3	6	85.4	8	98.3
Average	3	64.4(±15.78)	6	67.1(±14.23)	3	65.6(±14.06)	6	69.3(±11.76)	12.5	81.1(±13.35)

energy threshold was reduced from 95% to 90%, whereas the classification accuracy was similar. For the remaining energy values tested, the average number of features and average accuracy decreased with lower energy threshold. The results are shown in Fig. 4(b) and (c).

As shown in Fig. 4(b), in the DEAP dataset, an average of three features was extracted with an average classification accuracy of 80% when an energy threshold of 95% was used. When an energy threshold of 90% was used, the average accuracy was 78% and the average number of features extracted was 2.4. However, when the energy thresholds of 85% and 80% were used, the average number of features extracted was 2.4, with similar average accuracy. These results can interpret that both energy thresholds of 80% and 85% include the importance of information of similar amounts. Similarly, in the DREAMER dataset, as shown in Fig. 4(c), the average number of features extracted was reduced by approximately 30% when an energy threshold of 90% was used rather than 95%; however, the classification accuracy was approximately the same. Therefore, in this article, we set the energy threshold to 90% in all the experiments to achieve a high accuracy rate with a lower demand for computer processing time and memory resources.

In this article, we extracted features according to each emotional state by applying ET-MCSP to the data segmented using the sliding-window segmentation method. Subsequently, the LSTM model was built using the extracted features as the input data. We compared our ET-MCSP method to the multivariate EMD (MEMD) [48] and improved VMD (IVMD) [49] methods, which were used to extract the features of the EEG data comprising multiple classes. The ET-MCSP method proposed for extracting features in our new framework automatically determines the number of features corresponding to 90% of the energy threshold of the entire dataset. Table III presents the results of the classification of the SEED data.

As presented in Table III, as a result of extracting features using MEMD and measuring the classification accuracy rate with the extracted features as the input data for the LSTM model, the average classification accuracy was 64.4 (±15.78)% when three IMFs were used as feature values, whereas the average classification accuracy when six IMFs was used as input values was 67.1 (±14.23)%. Similarly, the IVMD method showed an average classification accuracy of 65.6 (±14.06)% when three IMFs were used and 69.3 (±11.76)% when six IMFs were used. In contrast, the ET-MCSP used in our proposed framework automatically

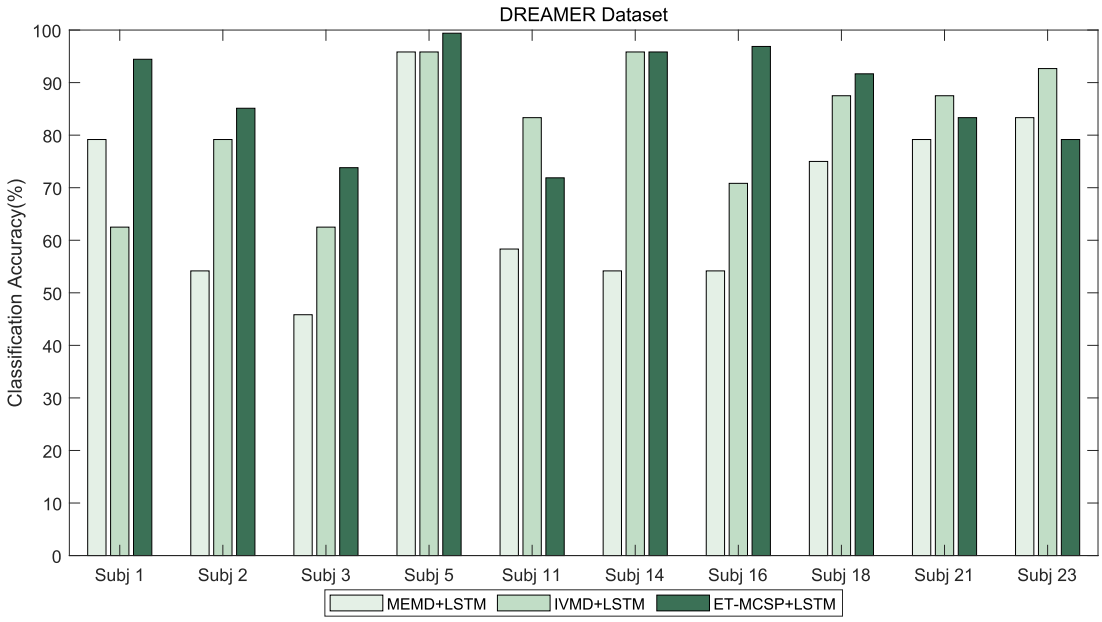

Fig. 5. Classification accuracy using methods for extracting features and LSTM on DREAMER dataset.

extracted an average of 12.5 features as the input data for the LSTM model, with an average classification accuracy of 81.1 ($\pm$13.35)%. In other words, the classification accuracy based on the feature extraction using ET-MCSP on the eRAD-Fe was improved by approximately 14% and 11% compared to the existing methods.

Table IV shows the classification accuracy when using the DEAP data. We segmented the DEAP data by applying sliding-window segmentation method with an overlapping of 30% and extracted the features using existing methods and the proposed method. Subsequently, we measured the classification accuracy with an LSTM model. An experiment was conducted using five subjects who met the conditions listed in Table I. The existing IVMD method showed a higher classification accuracy than the MEMD method for all the subjects, except for Subject 28 with six IMFs. In other words, the classification accuracy of the IVMD method showed an average improvement of approximately 12% over the MEMD method. However, overall, it had a lower classification accuracy than the ET-MCSP method. The classification accuracy of the ET-MCSP method showed improvements of 20% and 8% in terms of classification accuracy over the MEMD and IVMD existing methods, respectively. Fig. 5 shows the classification accuracy when using the DREAMER dataset. Ten subjects who met the conditions listed in Table I were used in the experiment. As a result, our applied methods showed the enhanced results of 19% and 5% compared to those of MEMD and IVMD with six IMFs, respectively. Thus, the ET-MCSP on the proposed framework was considered a more suitable method for extracting features for emotion recognition than the existing methods.

3) Emotional States Classification: We applied a deep learning model used in recent recognition research, LSTM, for the classification phase of eRAD-Fe. This method requires various configuration parameters. One of the most important considerations for the entire neural network is determining the number of neurons in the hidden layers. Therefore, we experimentally determined the number of neurons in the hidden layers of the LSTM model for emotional states classification. Fig. 6 shows the experimental results for the three EEG datasets. In particular, we randomly selected one subject dataset from each dataset and measured the classification accuracy by tenfold cross validation, during which the number of nodes in hidden layers was altered. In Fig. 6, the x-axis indicates the number of neurons in the first hidden layer (or the number of neurons in the second hidden layer) and the y-axis indicates the average classification accuracy according to the neuron number.

In this experiment, the number of neurons in the second hidden layer was set to 80% of the number of neurons in the first hidden layer. For example, if the number of neurons in the first hidden layer is 100, the number of neurons in the second hidden layer is set to 80. As shown in Fig. 6, all the three datasets exhibited relatively high accuracy with the number of neurons between 75 and 150 in the first hidden layer. Among them, SEED and DREAMER datasets had the highest accuracy, with 125 neurons in the first hidden layer and 100 neurons in the second hidden layer. The DEAP dataset showed the highest accuracy when the numbers of neurons in the first and second hidden layers were 100 and 80, respectively. Furthermore, it also showed high accuracy when the numbers of neurons in the first and second hidden layers were 125 and 100, respectively. Therefore, we set the number of neurons in the first and second hidden layers as 125 and 100, respectively, because they performed well on all the three datasets based on the experimental results.

In this article, to verify the suitability of the LSTM model used in the classification phase of the proposed new framework, we conducted a comparative analysis with bidirectional LSTM (Bi-LSTM) [50], which can improve model

TABLE IV
DEAP DATA CLASSIFICATION ACCURACY USING FEATURES EXTRACTION METHODS AND LSTM

No.Subject	MEMD				IVMD				ET-MCSP	
	Num. IMFs	ACC(%)	Num. IMFs	ACC(%)	Num. IMFs	ACC(%)	Num. IMFs	ACC(%)	Num. Feature	ACC(%)
Subject 1	3	45.8	6	54.2	3	74.5	6	77.5	2	87.5
Subject 2	3	45.6	6	41.7	3	70.3	6	63.3	2	69.2
Subject 7	3	58.3	6	79.2	3	75.8	6	85.8	2	87.5
Subject 14	3	41.7	6	54.5	3	65.8	6	75.8	4	88.9
Subject 28	3	75.0	6	83.3	3	61.6	6	71.7	3	80.6
Average	3	53.3(±13.6)	6	62.6(±17.9)	3	69.6(±5.9)	6	74.8(±8.2)	2.6	82.7(±8.2)

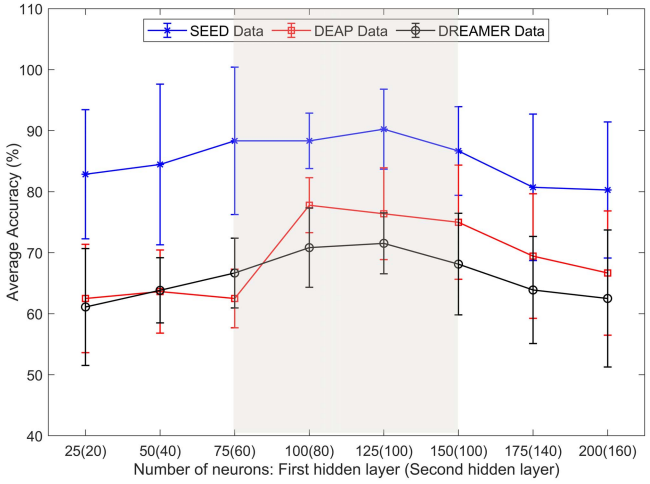

Fig. 6. Classification accuracy according to the number of neurons and hidden layers in the LSTM model.

performance on sequence classification problems; SVM [27], which exhibits the best performance among the existing machine learning methods for classification; and DCNN model [34], which is occasionally used in time-series classification. The classification accuracy was measured using tenfold cross validation for each subject. Fig. 7 shows the classification accuracy results using Bi-LSTM, SVM, DCNN, and LSTM classifiers with the three datasets. Fig. 7(a) shows the results obtained using the SEED dataset, in which the average accuracy rate was 81.33% when the Bi-LSTM classifier was used, 73.01% when SVM was used, 81.30% when DCNN was used, and 82.86% when LSTM was used. The DCNN, Bi-LSTM, and LSTM classifiers had improved accuracies of approximately 8%, 8%, and 9.8%, respectively, over the SVM method, with the LSTM classifier having the highest accuracy among the four methods. Fig. 7(b) shows the classification results obtained using the DEAP dataset. The average accuracy of LSTM was 72.22%, which was 2%, 9%, and 6% higher than those of Bi-LSTM, SVM, and DCNN, respectively. As shown in Fig. 7(c), using the DREAMER dataset, the average accuracy was 81.51% when the LSTM classifier was used, which is an improvement of 2%, 9%, and 10% over those of Bi-LSTM, SVM, and DCNN, respectively. In other words, as shown in Fig. 7, the performance was enhanced when the LSTM model was applied to classify emotional states in the classification stage of the new emotion recognition-assisted framework rather than the Bi-KSTM,

TABLE V
CLASSIFICATION ACCURACY FOR EPILEPSY EEG, EEG MOTOR ACTIVITY, AND EMG PHYSICAL ACTION DATASET

Dataset	Num.Feautes	Accuracy(%)
Epilepsy EEG [51]	67	81.67
Motor imagery EEG [52]	3	81.18
EMG physical action [53]	5	72.22

SVM, or DCNN methods. Therefore, the LSTM classifier is more suited to the novel framework than the Bi-LSTM, SVM, and DCNN classifiers.

In this article, we measured the execution complexity to verify the scalability of the proposed framework. Fig. 8 shows the computational cost of eRAD-Fe for varying numbers of samples. Fig. 8(a) shows the measurement results of the time required to process the SEED data using eRAD-Fe for 1200 samples. Fig. 8(b) shows the computational cost of the DEAP data. Fig. 8(c) shows the computational cost for different numbers of samples on the DREAMER data. The result showed that the processing time of eRAD-Fe linearly increased with the number of samples in the data. The computational complexity of eRAD-Fe is $O(i + k * h)$, where i denotes the number of classes, k denotes the number of output, and h denotes the number of cells in the hidden layer. Therefore, we can utilize our framework on other real-time data composed of multiple classes. In addition, to verify the applicability of eRAD-Fe, we evaluated the classification accuracies of three time-series datasets, namely epilepsy EEG dataset with three classes (ictal, inter-ictal, and pre-ictal) [51], EEG motor imagery dataset with four classes (left-hand, right-hand, foot, and tongue) [52], and EMG physical action dataset with six classes (clapping, jumping, seating, hammering, headering, and kneeing) [53]. Table V lists the classification accuracies, which were measured by tenfold cross validation on these three datasets. As a result, the epilepsy data had 81.67% classification accuracy, whereas the motor imagery data had 81.18% classification accuracy. Finally, the EMG data had a 72.22% accuracy. Through these experimental results, our proposed framework demonstrates its applicability on data with temporal characteristics.

V. DISCUSSION

Recently, developments in emotion recognition technology have trended toward developing to fusion the sensing technology with deep learning and database technology [3]. Emotions

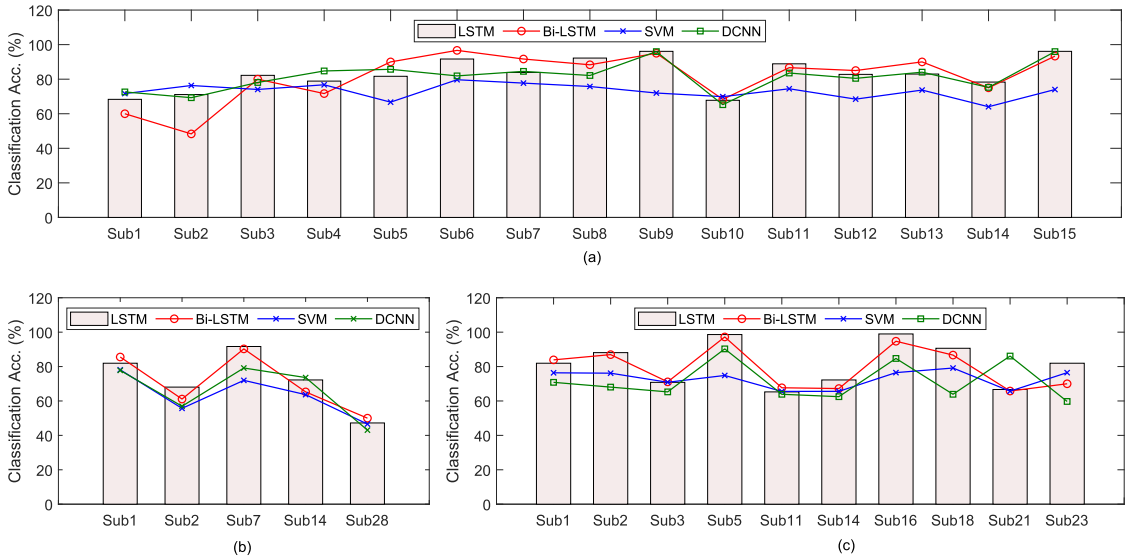

Fig. 7. Classification accuracies of Bi-LSTM, SVM, DCNN, and LSTM model. (a) SEED data. (b) DEAP data. (c) DREAMER data.

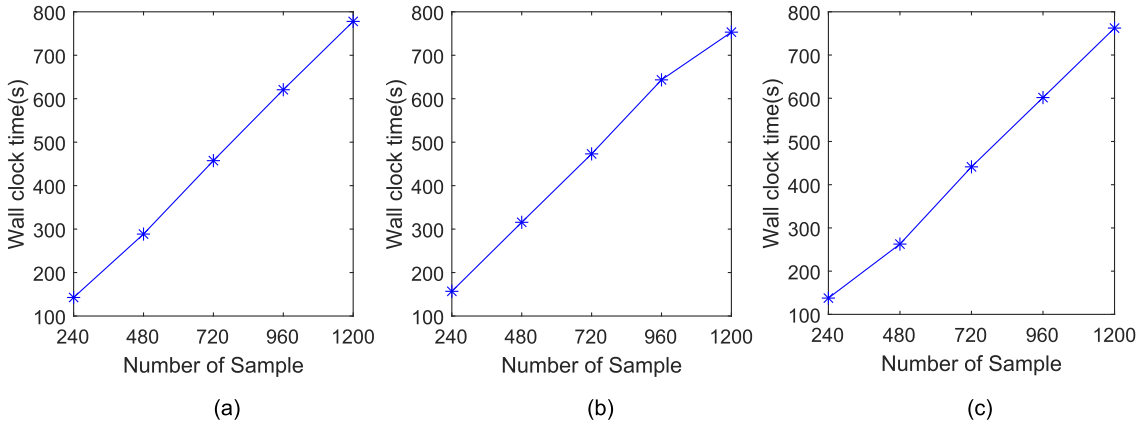

Fig. 8. eRAD-Fe scales linearly: wall clock time versus number of sample. (a) SEED. (b) DEAP. (c) DREAMER.

can be expressed verbally or nonverbally through the tone of voice, facial expression, and physiological changes in the nervous system. However, voice and facial expressions are not reliable indicators of emotion because they can be intentionally manipulated by the individual [54]. In comparison, physiological signals are more accurate than other data used in emotion recognition because they cannot be artificially generated [6]. In particular, because internal human emotional states can be recognized in emotion recognition research using EEG, this type of physiological signal is typically used in emotion recognition research.

An EEG, which measures small electrical signals in the brain, is very sensitive to external noise. Accordingly, in the EEG signal analysis, preprocessing is required, which removes unnecessary high-frequency and low-frequency components as well as artifacts caused by movements [55]. In this article, our emotion recognition-assisted framework does not place significant weight on these preprocessing phases needed in EEG analysis because the SEED and DEAP datasets used in our experiments provide preprocessed data in which noise and artifacts have been removed. Thus, we used the datasets without additional preprocessing. Meanwhile, the DREAMER dataset provided raw data that were not preprocessed. To verify the efficiency and scalability of the proposed framework, we used the DREAMER data recorded in a wireless environment. However, as previously mentioned, we filtered the noise and artifacts from the DREAMER data using EEGLAB [17], a widely known tool for EEG analysis, because eRAD-Fe does not significantly consider the preprocessing phases. Our emotion recognition-assisted framework, eRAD-Fe, comprises stages for sliding-window segmentation for data increment, automatic feature extraction of the emotional states, and deep learning for classification.

In this article, we applied sliding-window segmentation, which is used for forecasting and classifying time-series data, as it can substantially solve the time complexity problem [56]. Because it is useful when the amount of data is limited, it is also used to create the additional sample data needed for deep learning [56]. The sliding-window segmentation method can be used with or without an overlap. Meanwhile, EEG data have

temporal and spatial characteristics, and time-series data such as EEGs mostly use the sliding-window segmentation method with overlapping. We used sliding-window segmentation with a 10-s window length and a 30% overlap. Meanwhile, a fixed-size sliding window may not always be an effective method because the length of the transition activities in the data can be different. A small window size can separate one activity, whereas a large window size can include multiple activities. Therefore, in both cases, they can elicit incorrect information for classification [36]. We used a sliding window with a fixed length proposed in [57] and applied the sliding-window segmentation approach to reduce the computational complexity and generate additional training data. However, additional research is required to determine the optimal window length for the best performance.

Recently, MEMD and IVMD methods have been proposed to correctly extract the features of EEG signals. EMD is a powerful multiresolution signal decomposition technology that induces basis function from the signals [58]. It generates IMFs, including the band related to each emotion, and extracts statistical features such as mean absolute deviation, arithmetic mean, standard deviation, and root mean square using fast FT. However, the method uses a recursive approach that does not allow backward error correction for the time–frequency decomposition approach, and noise cannot be handled. To overcome these problems, a VMD method was proposed. VMD decomposes into IMFs through a concurrent approach rather than a recursive approach [59]. VMD chooses the decomposed IMFs and extracts features using PSD. It is known to be less sensitive to noise than EMD [59]. These methods were expanded to extract the features of the time-series data with good performance. Therefore, we used the expanded MEMD and IVMD adapted for multidimensional time-series data for our comparative experiments.

In the comparative experiments, we selected the IMFs in both MEMD and IVMD considering the computational complexity and memory resources. In other words, we considered three and six IMFs to minimize the number of features to reduce the computational complexity and memory consumption and found that the classification accuracy was higher when six IMFs were used as input data for the learning phase. We compared the proposed framework with two other methods. ET-MCSP on eRAD-Fe automatically extracted the optimal features using an energy threshold of 90%, as discussed in Section IV, and the classification accuracy was measured using the extracted features as the input data for deep learning. As a result, eRAD-Fe using ET-MCSP afforded a higher classification accuracy with fewer extracted features than the existing methods. Therefore, we can assert that the ET-MCSP in the proposed framework extracts only the optimal features according to the emotional state and that our method can reduce complexity and save memory compared to the MEMD and IVMD methods.

In this article, we used the RNN-based LSTM deep learning method in the classification stage of the new emotion recognition-assisted framework. An LSTM model is used to classify time-series data such as EEGs with better performance than DCNN model [60], as shown in the results in Section IV. However, it is necessary to consider factors, such as time and memory for learning, overfitting, and sensitivity of the weight initialization. Thus, we composed an LSTM layer with two hidden layers, as shown in Fig. 2. To prevent overfitting problems in which the classification algorithm only adapts to the training data, we applied a dropout rate of 50% to the fully connected layer [61]. In addition, we considered the computational complexity and memory resources of the LSTM model by applying the sliding-window segmentation approach.

VI. Conclusion

Recently, numerous studies have been conducted on human's emotion detection and recognition to improve communication between humans and machines. To improve the efficiency and accuracy of the emotion recognition system, it is necessary to identify and understand the user's emotional state. The main contribution of this article is to propose a novel framework, eRAD-Fe, which provides an effective and reliable emotion recognition-assisted framework based on EEG in terms of data generation, feature extraction, and classification. In particular, the proposed framework has stages comprising sliding-window segmentation by applying an overlap for deep learning, automatic feature extraction using ET-MCSP, and an RNN-based LSTM model. According to the experimental results, the proposed framework with ET-MCSP had an accuracy rate that was 5%–20% better than those of existing methods for extracting features, and the LSTM model exhibited results that were, on average, 1%–10% better than that of the DCNN method for all considered datasets. In future research, we will seek a new solution that can improve the recognition rate from more complex emotion data based on our emotion recognition technology. In addition, we will expand the research on emotion recognition exploiting multimodal data structure as well as the research that is able to improve the recognition performance by removing noise and artifacts included for measuring the data.

References

[1] S. Gu, F. Wang, N. P. Patel, J. A. Bourgeois, and J. H. Huang, "A model for basic emotions using observations of behavior in drosophila," *Frontiers Psychol.*, vol. 10, pp. 1–13, Apr. 2019.

[2] P. Li *et al.*, "EEG based emotion recognition by combining functional connectivity network and local activations," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, Oct. 2019.

[3] T. Thanapattheerakul, K. Mao, J. Amoranto, and J. H. Chan, "Emotion in a century: A review of emotion recognition," in *Proc. 10th Int. Conf. Adv. Inf. Technol. (IAIT)*, Bangkok, Thailand, 2018, pp. 1–8.

[4] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[5] B. T. Atmaja and M. Akagi, "Speech emotion recognition based on speech segment using LSTM with attention model," in *Proc. IEEE Int. Conf. Signals Syst. (ICSigSys)*, Bandung, Indonesia, Jul. 2019, pp. 40–44.

[6] Z. Gao, X. Wang, Y. Yang, Y. Li, K. Ma, and G. Chen, "A channel-fused dense convolutional network for EEG-based emotion recognition," *IEEE Trans. Cogn. Devel. Syst.*, early access, Feb. 25, 2020, doi: 10.1109/TCDS.2020.2976112.

[7] C. I. Hooker, S. C. Verosky, L. T. Germine, R. T. Knight, and M. D'Esposito, "Mentalizing about emotion and its relationship to empathy," *Social Cognit. Affect. Neurosci.*, vol. 3, no. 3, pp. 204–217, Sep. 2008.

[8] R. M. Mehmood, H.-J. Yang, and S.-H. Kim, "Children emotion regulation: Development of neural marker by investigating human brain signals," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[9] T.-E. Kam, H.-I. Suk, and S.-W. Lee, "Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification," *Neurocomputing*, vol. 108, pp. 58–68, May 2013.

[10] D. L. Schomer and F. H. L. da Silva, *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 7th ed. Oxford, U.K.: Oxford Univ. Press, 2017.

[11] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 374–393, Jul./Sep. 2019.

[12] G. K. P. Veeramallu, Y. Anupalli, S. K. Jilumudi, and A. Bhattacharyya, "EEG based automatic emotion recognition using EMD and random forest classifier," in *Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Kanpur, India, Jul. 2019, pp. 1–6.

[13] P. Pandey and K. R. Seeja, "Subject independent emotion recognition from EEG using VMD and deep learning," *J. King Saud Univ.-Comput. Inf. Sci.*, Nov. 2019, doi: 10.1016/j.jksuci.2019.11.003.

[14] A. Al-Nafjan, M. Hosny, Y. Al-Ohali, and A. Al-Wabil, "Review and classification of emotion recognition based on EEG brain computer interface system research: A systematic review," *Appl. Sci.*, vol. 7, no. 12, pp. 1–34, 2017.

[15] P. S. Ghare and A. N. Paithane, "Human emotion recognition using non linear and non stationary EEG signal," in *Proc. Int. Conf. Automat. Control Dyn. Optim. Techn. (ICACDOT)*, Pune, India, Sep. 2016, pp. 1013–1016.

[16] C. Y. Sai, N. Mokhtar, H. Arof, P. Cumming, and M. Iwahashi, "Automated classification and removal of EEG artifacts with SVM and wavelet-ICA," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 3, pp. 664–670, May 2018.

[17] A. Delorme and S. Makeig, "EEGLAB: An open-source toolbox for analysis of single-trial EEG dynamics," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004. Accessed: Sep. 28, 2021. [Online]. Available: https://sccn.ucsd.edu/eeglab/index.php

[18] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, San Diego, CA, USA, Nov. 2013, pp. 81–84.

[19] M. Ali, A. H. Mosa, F. Al Machot, and K. Kyamakya, "EEG-based emotion recognition approach for e-healthcare applications," in *Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Vienna, Austria, Jul. 2016, pp. 946–950.

[20] A. Q.-X. Ang, Y. Q. Yeong, and W. Wee, "Emotion classification from EEG signals using time-frequency-DWT features and ANN," *J. Comput. Commun.*, vol. 5, no. 3, pp. 75–79, 2017.

[21] P. Ackermann, C. Kohlschein, J. A. Bitsch, K. Wehrle, and S. Jeschke, "EEG-based automatic emotion recognition: Feature extraction, selection and classification methods," in *Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Munich, Germany, Sep. 2016, pp. 1–6.

[22] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 186–197, Mar. 2010.

[23] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Trans. Affective Comput.*, vol. 1, no. 2, pp. 81–97, Jul. 2010.

[24] M. D. Basar, A. D. Duru, and A. Akan, "Emotional state detection based on common spatial patterns of EEG," *Signal, Image Video Process.*, vol. 14, no. 3, pp. 473–481, Apr. 2020.

[25] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Phys. Rev. Lett.*, vol. 89, Jul. 2002, Art. no. 068102.

[26] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy: A tool for complexity analysis of multichannel data," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 84, no. 6, Dec. 2011, Art. no. 061918.

[27] O. Bazgir, Z. Mohammadi, and S. A. H. Habibi, "Emotion recognition with machine learning using EEG signals," in *Proc. 25th Nat. 3rd Int. Iranian Conf. Biomed. Eng. (ICBME)*, Qom, Iran, Nov. 2018, pp. 1–5.

[28] J. Huang, X. Xu, and T. Zhang, "Emotion classification using deep neural networks and emotional patches," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Kansas City, MO, USA, Nov. 2017, pp. 958–962.

[29] P. Wang, A. Jiang, X. Liu, J. Shang, and L. Zhang, "LSTM-based EEG classification in motor imagery tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 11, pp. 2086–2095, Nov. 2018.

[30] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.

[31] X. Li, D. Song, P. Zhang, G. Yu, Y. Hou, and B. Hu, "Emotion recognition from multi-channel EEG data through convolutional recurrent neural network," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Shenzhen, China, Dec. 2016, pp. 352–359.

[32] Y. Li, J. Huang, H. Zhou, and N. Zhong, "Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks," *Appl. Sci.*, vol. 7, no. 10, p. 1060, Oct. 2017.

[33] H. Tang, W. Liu, W. L. Zheng, and B. L. Lu, "Multimodal emotion recognition using deep neural networks," in *Proc. Int. Conf. Neural Inf. Process. (ICONIP)*, Guangzhou, China, Nov. 2017, pp. 811–819.

[34] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44317–44328, 2019.

[35] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS ONE*, vol. 12, no. 2, pp. 1–20, 2017.

[36] M. H. M. Noor, Z. Salcic, and K. I.-K. Wang, "Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer," *Pervasive Mobile Comput.*, vol. 38, pp. 41–59, Jul. 2017.

[37] S.-H. Kim, H.-J. Yang, N. A. T. Nguyen, and S.-W. Lee, "AsEmo: Automatic approach for EEG-based multiple emotional state identification," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 5, pp. 1508–1518, May 2021.

[38] T. Yan, T. Jingtian, and G. Andong, "Multi-class EEG classification for brain computer interface based on CSP," in *Proc. Int. Conf. Biomed. Eng. Informat.*, Sanya, China, May 2008, pp. 469–472.

[39] W. Wu, X. Gao, and S. Gao, "One-versus-the-rest (OVR) algorithm: An extension of common spatial patterns (CSP) algorithm to multi-class case," in *Proc. IEEE Eng. Med. Biol. 27th Annu. Conf.*, Shanghai, China, 2005, pp. 2387–2390.

[40] T. T. A. Hoang, "Multivariate features for multi-class brain computer interface systems," Ph.D. dissertation, Dept. Inf. Sci. Eng., Univ. Canberra, Canberra, NSW, Australia, Jan. 2014. [Online]. Available: https://researchprofiles.canberra.edu.au/en/studentTheses/xmultivariate-features-for-multi-class-brain-computer-interface-sy

[41] I. T. Jollie, *Principal Component Analysis*, 2nd ed. New York, NY USA: Springer, 2002. [Online]. Available: http://www.stats.org.uk/pca/pca.pdf

[42] K. S. Ng, H.-J. Yang, S.-H. Kim, and S.-H. Kim, "Incremental non-Gaussian analysis on multivariate EEG signal data," *IEICE Trans. Inf. Syst.*, vol. E95.D, no. 12, pp. 3010–3016, 2012.

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

[44] L. Fraiwan and M. Alkhodari, "Classification of focal and non-focal epileptic patients using single channel EEG and long short-term memory learning system," *IEEE Access*, vol. 8, pp. 77255–77262, 2020.

[45] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[46] S. Koelstra *et al.*, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan./Mar. 2012.

[47] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.

[48] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," *Expert Syst. Appl.*, vol. 95, pp. 201–211, Apr. 2018.

[49] D. Xie, H. Sun, and J. Qi, "A new feature extraction method based on improved variational mode decomposition, normalized maximal information coefficient and permutation entropy for ship-radiated noise," *Entropy*, vol. 22, no. 6, pp. 1–19, 2020.

[50] B. Jang, M. Kim, G. Harerimana, S. Kang, and J. W. Kim, "Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism," *Appl. Sci.*, vol. 10, no. 17, pp. 5841–5855, 2020.

[51] B. Schelter *et al.*, "Do false predictions of seizures depend on the state of vigilance? A report from two seizure-prediction methods and proposed remedies?" *Epilepsia*, vol. 47, no. 12, pp. 2058–2070, 2006. [Online]. Available: http://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database
[52] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl, and G. Pfurtscheller. (Accessed: Sep. 28, 2021). *BCI Competition 2008—Graz Data Set A*. [Online]. Available: http://www.bbci.de/competition/
[53] T. Theodoridis, A. Agapitos, H. Hu, and S. M. Lucas, "Ubiquitous robotics in physical human action recognition: A comparison between dynamic ANNs and GP," in *Proc. IEEE Int. Conf. Robot. Automat.*, Pasadena, CA, USA, May 2008, pp. 19–23. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/EMG+Physical+Action+Data+Set/
[54] X. Zhang, C. Xu, W. Xue, J. Hu, Y. He, and M. Gao, "Emotion recognition based on multichannel physiological signals with comprehensive nonlinear processing," *Sensors*, vol. 18, pp. 1–16, Nov. 2018.
[55] S. Phadikar, N. Sinha, and R. Ghosh, "Automatic eyeblink artifact removal from EEG signal using wavelet transform with heuristically optimized threshold," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 475–484, Feb. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9095264
[56] Q. Ni, T. Patterson, I. Cleland, and C. Nugent, "Dynamic detection of window starting positions and its implementation within an activity recognition framework," *J. Biomed. Inform.*, vol. 62, pp. 171–180, Aug. 2016.
[57] B. Nakisa, M. N. Rastgoo, A. Rakotonirainy, F. Maire, and V. Chandran, "Automatic emotion recognition using temporal multimodal deep learning," *IEEE Access*, vol. 8, pp. 225463–225474, 2020.
[58] Y. Zhou, B. W.-K. Ling, X. Mo, Y. Guo, and Z. Tian, "Empirical mode decomposition-based hierarchical multiresolution analysis for suppressing noise," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1833–1845, Apr. 2020.
[59] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[60] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, "Deep learning with long short-term memory for time series prediction," *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 114–119, Jun. 2019.
[61] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*. [Online]. Available: http://arxiv.org/abs/1207.0580

Sun-Hee Kim received the B.S. degree in multimedia from the Korean Educational Development Institute, Jincheon, South Korea, in 2004, the M.S. degree in computer science from Dongguk University, Seoul, South Korea, in 2006, and the Ph.D. degree in computer science from Chonnam National University, Gwangju, South Korea, in 2011.

From 2011 to 2013, she was visited as a Post-Doctoral Fellow at Carnegie Mellon University, Pittsburgh, PA, USA, where she worked on data mining problems in biosignals at the Database Group. She is currently a Research Professor at the Department of Artificial Intelligence, Korea University, Seoul. Her current research interests are data mining, machine learning, bioinformatics, and their applications.

Ngoc Anh Thi Nguyen received the B.S. degree from the Faculty of Mathematics and Informatics, Danang Education University, Da Nang, Vietnam, in 2006, the M.S. degree from the Department of Electronics and Computer Engineering, Chonnam National University, Gwangju, South Korea, in 2011, and the Ph.D. degree in electronics and computer engineering from Chonnam National University, in 2016.

She is currently a Lecturer at the Faculty of Information Technology, The University of Danang—University of Science and Education, Da Nang. Her research interests are data mining, machine learning, data analysis, signal processing, pattern recognition, and mathematical modeling. She is mainly working on time series mining, with applications in bioinformatics and neuroscience.

Hyung-Jeong Yang received the B.S., M.S., and Ph.D. degrees from Chonbuk National University, Jeonju-si, South Korea, in 1991, 1993, and 1998, respectively.

She is currently a Professor at Chonnam National University, Gwangju, South Korea. Her main research interests include multimedia data mining, pattern recognition, artificial intelligence, e-learning, and e-design.

Seong-Whan Lee (Fellow, IEEE) received the B.S. degree in computer science and statistics from Seoul National University, Seoul, South Korea, in 1984, and the M.S. and Ph.D. degrees in computer science from Korea Advanced Institute of Science and Technology, Seoul, in 1986 and 1989, respectively.

He is currently the Hyundai-Kia Motor Chair Professor and the Head of the Department of Artificial Intelligence and the Department of Brain and Cognitive Engineering, Korea University, Seoul. His research interests include pattern recognition, artificial intelligence, and brain engineering.

Dr. Lee is a fellow of the International Association on Pattern Recognition (IAPR) and Korea Academy of Science and Technology.