Article

Deep Learning Method for Selecting Effective Models and Feature Groups in Emotion Recognition Using an Asian Multimodal Database

Jun-Ho Maeng, Dong-Hyun Kang and Deok-Hwan Kim *

Department of Electronic Engineering, Inha University, Incheon 22212, Korea; or2926@gmail.com (J.-H.M.); kawow123@gmail.com (D.-H.K.)

* Correspondence: deokhwan@inha.ac.kr

Received: 20 October 2020; Accepted: 17 November 2020; Published: 24 November 2020

Abstract: Emotional awareness is vital for advanced interactions between humans and computer systems. This paper introduces a new multimodal dataset called MERTI-Apps based on Asian physiological signals and proposes a genetic algorithm (GA)—long short-term memory (LSTM) deep learning model to derive the active feature groups for emotion recognition. This study developed an annotation labeling program for observers to tag the emotions of subjects by their arousal and valence during dataset creation. In the learning phase, a GA was used to select effective LSTM model parameters and determine the active feature group from 37 features and 25 brain lateralization features extracted from the electroencephalogram (EEG) time, frequency, and time–frequency domains. The proposed model achieved a root-mean-square error (RMSE) of 0.0156 in terms of the valence regression performance in the MAHNOB-HCI dataset, and RMSE performances of 0.0579 and 0.0287 in terms of valence and arousal regression performance, and 65.7% and 88.3% in terms of valence and arousal accuracy in the in-house MERTI-Apps dataset, which uses Asian-population-specific 12-channel EEG data and adds an additional brain lateralization (BL) feature. The results revealed 91.3% and 94.8% accuracy in the valence and arousal domain in the DEAP dataset owing to the effective model selection of a GA.

Keywords: dataset; emotion analysis; feature evaluation and selection; signal analysis; machine learning

1. Introduction

The technology of recognizing human emotion is essential to developing a human computing interface (HCI) for human–robot interactions. Furthermore, emotion recognition is attracting attention in other fields, including artificial intelligence. In recent studies, emotion recognition can be divided into two areas according to the type of data used. One area is external emotion recognition based on voice or image. Because the current voice information is sparse, there is a fundamental limitation in extracting continuous emotions. Research on emotion recognition has recently been conducted using dialogue context and voice tone [1,2]. On the other hand, emotion recognition research based on image information has shown the best performance. Emotions are categorized by recognizing changes in facial expressions based mainly on the features obtained from facial images [3,4]. Recently, mechanisms for classifying emotions end-to-end using a convolution neural network without an additional feature extraction process have been developed and have performed with high accuracy [5,6].

In contrast, internal emotion recognition uses the changes in bio-signals to recognize emotions without considering external changes. Emotion recognition is conducted using an electroencephalogram (EEG) produced from electrical signals generated in the human brain [7,8]. Frequency-domain features,

such as power spectral density (PSD), are extracted, and emotions are recognized using typical learning algorithms. Recent studies have improved the accuracy of emotion recognition by extracting the features of asymmetry between the channels of the EEG signal and applying deep learning algorithms [2,9,10]. On the other hand, emotion recognition using bio-signals presents difficulties in acquiring data. Public datasets, such as DEAP [11] and MAHNOB-HCI [12], are often used. The type of data strongly influences the results of bio-signal based emotion recognition studies. Previous studies used a Western-based dataset and various EEG features [13,14].

This study aimed to improve the accuracy of emotion classification by searching the models, features, and channels used for emotion recognition using a Korean dataset instead of a public dataset to perform emotion recognition based on the EEG signals. The features were extracted from the time, frequency, and time–frequency domains for comparison. A genetic algorithm (GA) was used to select valid features, and a learning algorithm was trained based on long short-term memory (LSTM). From the LSTM model results, all features were weighted using repeated feature evaluations by the GA. To evaluate all features consistently, the learning used the maximum number of repetitions, mixed randomly, to prevent the features from being removed by the GA. After completing the proposed GA–LSTM process, the models, channels, and features of the EEG signal that were valid for emotion recognition could be distinguished by weighting the model parameters and features. Experiments using the MAHNOB-HCI dataset with continuous-time annotation applied to the valence label and other experiments using the DEAP dataset showed that the proposed method outperformed the continuous emotion recognition method using the representative PSD feature [15] and the recent emotion classification methods [16–18]. To appraise the Asian-oriented dataset, the multimodal emotion for the robot interface was constructed using the Asian physiological signals (MERTI-Apps) dataset, which is similar to the MAHNOB-HCI dataset. By applying the model to the MERTI-Apps dataset, this study verified that the proposed technique could provide effective model selection and feature group selection.

The following are the key contributions of this paper:

- A GA for a deep-learning network was proposed to select effective network models and search for valid EEG channels and features. GA was combined with LSTM to select the effective LSTM models and remove the features that interfere with emotion recognition judgment. The proposed GA-LSTM architecture outperforms previous methods [15–18] in terms of the regression performance (RMSE) and classification performance (accuracy).
- An Asian physiological signals (MERTI-Apps) dataset that provides valence and arousal annotation labeling and various feature sets (time, frequency, time–frequency domain, and brain lateralization feature) was established to perform emotion recognition based on the EEG signals.
- The brain lateralization feature further enhances the outcome of emotional recognition.

This paper is organized as follows. Section 2 reviews the related work, and Section 3 describes the MAHNOB-HCI and DEAP public datasets and the MERTI-Apps dataset. Section 4 details the proposed method, and Section 5 presents the experimental results and analysis of further experiments on brain deflection. Finally, the conclusions are presented in Section 6.

2. Related Works

This section introduces the HCI and previous studies using public datasets for biological signals and emotion recognition.

2.1. Emotion Recognition for Human Computing Interaction

Human emotions have long been equated to facial expressions. Therefore, methods based on facial images have been developed intensively for emotion recognition. Tong et al. detected the area where movement is activated in a facial image and determined the action unit (AU) according to the facial position and shape [19]. Facial expressions were classified according to the AU type and

combination. On the other hand, as the performance of algorithms for extracting landmarks has been improved dramatically, new methods for determining facial expressions have emerged [20]. In this context, landmarks are defined as the location information of crucial points in facial images, and facial expressions are characterized by the location of each landmark and the distance between markers. In an actual environment, human emotions may differ from those in the datasets. For example, the same emotions may have different means of expression, or emotions may change without being revealed by facial expressions. Therefore, it is desirable to use information, such as bio-signals and facial images, for accurate emotion recognition in an actual environment.

An EEG records the electrical signals generated by the human brain, which can be used for internal emotion recognition. The state of emotion can be estimated by analyzing the position and waveform characteristics of the EEG. The DEAP dataset and the MAHNOB-HCI dataset, which measure the EEG in the context of human emotion, were released in 2012. Soleymani et al. [15] and Kim et al. [21] developed representative EEG-based emotion recognition techniques. Soleymani et al. proposed a method for the regression of positive/negative valence values in real time using the MAHNOB-HCI dataset annotated in a continuous-time domain. Kim et al. reported that the channel activating the EEG signals varies with the state of emotion. Moreover, the features extracted using the connectivity between the EEG channels were significant. Convolutional LSTM was applied to the extracted features to classify the emotional states in the valence and arousal domains [7,22]. A paper surveying recent research [14] showed that classifier research is dominant. Classifier research [16–18,23] was still active, and emotion recognition in the order of image, voice, and brain waves showed good accuracy. On the other hand, the accuracy of EEG was still lower compared to that of the video and audio fields. This can be seen as the need for regression and fusion with other biological signals. In addition to emotion recognition, the field of brain–machine interface systems analyzes EEG signals to improve performance. The technology to understand the driver’s condition and apply it to future automotive technology is very important. Gao et al. [24] measured a driver’s fatigue and applied features, including PSD, to the model to improve performance. Because an analysis of EEG is difficult, there is active research on finding effective features and channels for emotional recognition [18,25,26]. Hao Chao et al. [25] reported an improvement in performance using the maximal information coefficient (MIC) feature. Li et al. [26] adopted a multiple feature fusion approach to combine the compensative activation and connection information for emotion recognition. They selected the optimal feature subsets among the feature combinations after determining the combination of the feature sets. Wang et al. [18] selected effective channels using an EEG spectrogram and normalized mutual information. Becker et al. [23] provided a database of High-Resolution (HR) EEG recordings to analyze emotions elicited by film excerpts and other simultaneously recorded physiological measurements: galvanic skin response (GSR), electrocardiogram (ECG), respiration, peripheral oxygen saturation, and pulse rate. The classification accuracy using brain source reconstruction was above 70% or 75%. They analyzed the frequency of feature selection for each channel/brain region, which was averaged over all subjects. On the other hand, they did not compare the classification performance with other databases, nor did they consider the feature selection technique. Halim et al. [27] presented a method to identify dynamically the drivers’ emotions and applied an ensemble of Support Vector Machine (SVM), Random Forest (RF), and Neural Network (NN) classifiers. Among them, the classification accuracy using SVM was 88.75%. They also used feature selection using principal component analysis (PCA), mutual information (MI), and random subset feature selection (RSFS). Krishna et al. [28] presented an emotion classification using a generalized mixture model and achieved 89% classification accuracy using the confusion matrix for happy, sad, bored, and neutral in terms of valence. On the other hand, they only collected the EEG dataset using six mentally impaired subjects and used features, such as Mel-frequency Cepstrum Coefficients, but they did not apply feature or channel selection.

This study did not determine the initial models and feature groups. All models, feature groups, and channels were evaluated by the GA to explore the optimal models and feature subset. Through the

GA, the models, features, and channels that interfere with emotion recognition were removed to show a better effect.

2.2. Biosignals and Dataset

With the rapid development of deep learning technology in emotion recognition, research is actively being conducted to classify emotions by combining biosignals other than EEG. As shown in Figure 1, the central nervous system (CNS) signal refers to the EEG signal generated from the brain. The peripheral nervous system (PNS) signals can be observed using electrooculogram (EOG), electromyogram (EMG), photoplethysmogram (PPG), and galvanic skin response (GSR), which indicate the electrical signals generated through muscle movements and eye movements. The EEG is the most critical signal for internal emotion recognition. Nevertheless, severe artifacts can result from data acquisition, noise, temperature, and individual differences because EEG signals are very susceptible. The PNS signals can be generalized differently than the EEG signal. Research often uses PNS signals as an aid because it is challenging to classify emotions using only EEG signals. One study [29] used the heart rate information obtained from an electrocardiogram (ECG) to recognize emotion by assessing the facial muscle movement through an EMG signal. Another survey [30] measured facial expressions through the tail muscles to read the changes in emotion. In those studies, the EEG signals use the PSD features. Few studies have defined useful features other than PSD for emotion recognition.

Figure 1. Bio-signal type and position for central nervous system (CNS) and peripheral nervous system (PNS).

Bio-signal data, including EEG signals, require specialized equipment, but the experimental conditions where emotions are felt are difficult to reproduce. Therefore, most studies use an open dataset. Table 1 lists the datasets according to the research purpose. On the other hand, it is unclear whether the experimental results will apply equally to an Asian population because most datasets are Western-based, including the MAHNOB-HCI.

Table 1. Comparison of the physiological datasets for emotion recognition.

Database	No. Part.	Induced or Natural	Audio	Visual	Peripheral Physio.	EEG	Annotation Labeling
MIT [31]	17	Natural	No	No	Yes	No	No
VAM [32]	19	Natural	Yes	Yes	No	No	No
SEMAINE [33]	20	Induced	Yes	Yes	Yes	No	No
DEAP [11]	32	Induced	No	Yes	Yes	Yes	No
MAHNOB-HCI [12]	28	Induced	Yes	Yes (for 22)	Yes	Yes	Valence
MERTI-Apps	62	Induced	Yes	Yes	Yes	Yes	Valence, Arousal

3. Multimodal Datasets

Three datasets were used in this study: the MAHNOB-HCI dataset [12], which is an open dataset with observer tagging for valence for continuous emotion recognition; the DEAP dataset [11], which is a multimodal dataset for analyzing people's discrete emotions; and the MERTI-Apps dataset produced for this study for use with Asian populations. This chapter describes three datasets and introduces the annotation-labeling program used in MERTI-Apps.

3.1. MAHNOB-HCI and DEAP Datasets

Twenty videos were used in the MANHOB-HCI dataset [12] (hereafter, MANHOB dataset) to induce continuous emotions. In the preliminary study, the participants assisted in video selection by reporting their feelings through their self-assessment. The excerpt's 20 source videos include emotions, such as disgust, amusement, joy, fear, sadness, and neutral. The length of each video clip was approximately 34 to 117 s. The participants comprised 27 healthy subjects: 13 males and 16 females. The EEG signals were obtained using 32 activation electrodes positioned according to the criteria of the Biosemi Active II system [34] and the 10–20 International System [35]. The faces of the participants were filmed in a 720 × 580 video at 60 Hz and synchronized with the EEG signal. Table 2 provides a detailed description of the dataset production. Two hundred and thirty-nine records were produced, each with corresponding label information. In addition, five educated commentators provided continuous comments on the participants' facial expressions, and the valence of the frontal facial expressions was determined using FEELTRACE [36] and joysticks.

Table 2. Summary of MAHNOB-HCI dataset.

Participants and Modalities	
Nr. of Participants	27, 11 male and 16 female
Recorded signals	32-channel EEG (256 Hz), Peripheral physiological signals (256 Hz), Face and body video using 6 cameras (60 f/s), Eye gaze (60 Hz) and Audio (44.1 kHz)
Self-report	Emotional keyword, arousal, valence, dominance, predictability
Rating values	discrete scale of 1–9
Nr. of videos	20
Selection method	Subset of online annotated videos

The DEAP dataset [11] is a multimodal dataset for analyzing people's discrete emotions. The dataset is the extraction of EEG and peripheral physiological signals of 32 participants. In the extraction process, 40 music videos were viewed at 1 minute intervals. As shown in Table 3, the participants were 32 subjects: 16 males and 16 females. The participants rated the labels of each video as arousal, valence, like/dislike, dominance, and familiarity. The data collected were EEG, EMG, EOG, GSR, Blood Volume Pulse (BVP), temperature, and breathing data. The length of the video was extracted in 1 min segments through highlighted detection. The EEG data were collected from 32 electrodes at a sampling rate of 512 Hz, and PNS bio-signals were collected from 13 electrodes with four EOG, four EMG, two GSR, one BVP, temperature, and breathing signals.

Table 3. Summary of DEAP dataset.

Participants and Modalities	
Nr. of Participants	27, 11 male and 16 female
Recorded signals	32-channel EEG (256 Hz), Peripheral physiological signals (256 Hz), Face and body video using 6 cameras (60 f/s), Eye gaze (60 Hz) and Audio (44.1 kHz)
Self-report	Emotional keyword, arousal, valence, dominance, predictability
Rating values	discrete scale of 1–9
Nr. of videos	20
Selection method	Subset of online annotated videos

3.2. MERTI-Apps Dataset

The MERTI-Apps dataset collected the records of multimodal recordings of participants in their responses to fifteen emotional videos. Initially, the classification system with eight emotions was established based on the arousal and valence domains: happiness, excited, sad, bored, disgust, anger, calm, and comfortable, to build an experimental video set. The videos to be presented to induce emotions were searched randomly and collected on YouTube using the emotional vocabulary keywords. Five to seven research assistants reviewed the collected videos, and videos judged to induce emotions were classified by search keywords. To select videos optimized for emotion induction, the content validity of the emotion inducement suitability (emotion type, intensity, etc.) was then checked through a field test. As a result, the final emotion collection videos were selected except for images with a Content Validity Index (CVI) of 2.5 or lower. The final 32 videos were selected from the four valence–arousal bases (high arousal positive valence: HAPV, high arousal negative valence: HANV, low arousal positive valence: LAPV, low arousal negative valence: LANV) and neutral domains. Among them, 15 were selected for the participant’s concentration time. Each video spanned from 60 to 206 s, an average length of 114.6 s. Sad emotions were difficult to induce quickly compared to other public databases, so the length of sad images was longer than that of other databases. The recruitment process, consent, laboratory environment, emotion-inducing videos, participant treatment, follow-up management, and other measurement items showed a CVI value of 0.92 between experts. Expert CVI measurements secured the validity, and a pilot test was conducted. This video selection process was conducted by a research team at Inha University Department of Social Welfare. Table 4 lists the three experiments. Experiment 1 was performed using EMG and EOG, excluding EEG to analyze the peripheral nervous signals other than EEG. Experiment 2 considered only emotional recognition using the CNS, such as the EEG signal, and measured the EEG, EMG, and EOG signals together to remove the noise caused by movements of the head and facial muscles using EMG and EOG signals. In this case, the problem of the EMG electrode covering the face was identified. In experiment 3, the PPG and GSR signals, which are from the peripheral nervous system, were synchronized with the EEG signal, excluding the EMG signal, and different nerve signals, such as CNS and PNS, were measured in a multimodal form.

Table 4. Summary of MERTI-Apps dataset.

Participants and Modalities		
Nr. of Participants	62, 28 male and 34 female	
Recorded signals	12-channel EEG (1 kHz), 2-channel EMG (1 kHz), 1-channel EOG (1 kHz), PPG (1 kHz), GSR (1 kHz)	
Selection method	Content Validity Index(CVI) score	
Self-report	arousal, valence	
Rating values	9 kind of Emotion	
Experiment 1 (Emotion response to videos)	Nr. of videos	5 (Induced: sad 1, happy 1, angry 2 scared 1)
	Use to signals	EMG, EOG, PPG, GSR
Experiment 2 (Emotion response to videos)	Nr. of videos	5 (Induced: sad 1, happy 1, angry 1 scared 1, neutral 1)
	Use to signals	EEG, EMG, EOG
Experiment 3 (Emotion response to videos & tagging)	Nr. of videos	5 (Induced: sad 1, happy 2, angry 1 scared 1)
	Use to signals	EEG, EOG, PPG, GSR

The participants consisted of 62 healthy subjects: 28 males and 34 females aged from 21 to 29 years. Figure 2 shows the data collecting procedure. The EEG signals were obtained with 12 active electrodes located in accordance with the 10–20 International System using a BIOPAC MP150 instrument. Along with the EEG signal, a video of the participant's face was shot in 1080 p at 30 Hz. Because the video runtime was short, an EOG channel was used to remove the artifacts caused by eye blinks. The participant's self-emotional questionnaire and data inconsistent with the evaluation were excluded from the records. The valid records in the MERTI-Apps dataset were 283, 312, and 236 in experiments 1, 2, and 3, respectively. Initially, in each experiment, 320 records were produced using 62 participants and five videos. The records causing serious artifacts from the induced emotion were excluded. In experiment 3, 236 records were used for annotation labeling, which was designed without discomfort in facial movement. Each record contained the corresponding valence and arousal label information. As with the MAHNOB dataset, trained commentators commented continuously on the participants' facial expressions, and the program evaluated the valence and arousal of frontal facial expressions. The measurement program related to annotation labeling was recorded, as shown in Figure 3.

Figure 2. MERTI-Apps data collecting procedure.

Figure 3. MERTI-Apps data acquisition example, (a) video, (b) face video, (c) bio-signals, (d) valence annotation labeling, (e) arousal annotation labeling.

3.3. Annotation Labeling

Annotation labeling was performed by the observer to evaluate the emotion of the bio-signals in the MERTI-Apps dataset. Only one emotion was induced in the video, and it was confirmed that the induced emotion and the participant's emotion matched. The participant's valence and arousal were evaluated using the recorded facial video. After viewing the video, the participant produced a self-assessment label, which displayed the emotions they felt. The participant's self-emotional questionnaire and data inconsistent with the evaluation were excluded from the experiment. Arousal is a measure of the level of the excitement of emotions. A smaller arousal value indicates calmer emotions, which corresponds to boredom, comfort, and sleepiness. On the other hand, high-arousal emotions include excitement, anger, and fear. The valence represents a positive or negative degree of emotion. Fear has a very negative valence. Boredom or excitement has a moderate valence. Happiness or comfort has a positive valence. As it is difficult to subdivide emotions further using the valence alone, this study evaluated arousal, which refers to feelings, from high to low. Both valence and arousal were evaluated in the range of -100 to 100 , and data collection occurred at 0.25 s intervals in accordance with the bio-signal data.

The most important aspect when evaluating annotation labeling is that observers should make the most objective evaluation possible. Therefore, an in-house annotation-labeling program was used. The observer consisted of five males and females aged 22–25 years old. After training and coordinating the observers' opinions of the evaluation of the participants' emotions through the pilot experiment video, the labeling work of the observers began in earnest. In case of disagreement, two additional observers excluded the abnormal values and used the average valence or arousal of the remaining observers. The video and labeling data were matched by displaying the bio-signal data and the point at the start and end of the face video of the participant in the experiment. The observers watched the facial video and evaluated the valence and arousal through a scroll bar on the right side of the program at 0.25 s intervals. The five observers recorded the labeling data (see Figure 3d), which were used as the target data. To find data with the same emotion, this study checked whether the emotion-inducing

video and the actual emotion felt by the participant were the same. Only data whose labeling value matched were used in the experiment.

4. Proposed Method

The features of the EEG signal were extracted, as shown in Table 5, to derive an effective group of features for emotion recognition through deep learning. The proposed deep learning model was combined with the GA–LSTM, as shown in Figure 4. The initialization step was performed randomly to transmit effective features in the GA continuously.

Table 5. EEG feature list for genetic algorithm—long short-term memory (GA-LSTM).

Time domain	
mean, max, min, 1st difference, Normalized 1st difference	
Frequency domain	
Power spectrum density(PSD) from Slow alpha, Alpha, Beta, Gamma	mean, max, integral
Time-frequency domain	
Power of Discrete Wavelet Transform(DWT) from Slow alpha, Alpha, Beta, Gamma	mean, max, absolute
Log of Power of Discrete Wavelet Transform(DWT) from Slow alpha, Alpha, Beta, Gamma	
Abs(Log) of Power of Discrete Wavelet Transform(DWT) from Slow alpha, Alpha, Beta, Gamma	
Brain Lateralization feature	
1×5 (C3: F8 C4 T4 T6 O2)	5×1 (F7 C3 T3 T5, O1: C4)
3×3 (F7 T3 O1: F8 T4 O2)	5×5 (F7 C3 T3 T5, O1: F8 C4 T4 T6 O2)

Figure 4. Proposed GA-LSTM model architecture.

4.1. Feature Extraction

The active features for emotion recognition were selected by extracting the EEG features from three domains (time, frequency, and time–frequency), as summarized in Table 5. The proposed method can distinguish which channels and features of the EEG signals are valid for emotion recognition through the weighting of features. This is why the previous work [15] used only PSD features, whereas various

feature sets were used in the present study. Thirty-seven EEG features were extracted per channel, some of which were used in addition to the PSD in other studies [9,37]. A variety of factors, such as the small movement of the participants, sweat, body temperature, and tension, act as noise in the EEG signal. Therefore, a notch filter on Biopac's M150 equipment was used to eliminate very-low-frequency bands and >50 Hz frequency bands. The eyeblink pattern was also removed via the EOG signal. A fast Fourier transform (FFT) was used in the frequency domain to divide the EEG signal into slow alpha, alpha, beta, and gamma waves. The features were extracted and converted to a one-dimensional vector and used as the input data. The final feature dimension was calculated as (number of channels) $\times$ (amount of features). For the MAHNOB dataset, the feature dimension was $32 \times 37 = 1184$. For MERTI-Apps, the feature dimension was $12 \times 37 = 444$. In addition, a brain deflection feature consisting of 1:5, 5:1, 3:3, and 5:5 was used. A feature was extracted from the previous 2 s data stream in the proposed technique because the data corresponded to a continuous-time annotation.

Time domain feature. A characteristic of the time domain can be represented as the change in the EEG signal over time. In the time domain, the mean, minimum, maximum, 1st difference, and normalized 1st difference were used to recognize changes in emotions over time [33].

Frequency domain feature. In previous studies, the components of the frequency domain were used because of the excellent spatial resolution of the EEG. Outputs appearing in different frequency bands are good factors to detect different emotional states. In this study, the PSD features were divided into four areas, Slow Alpha (8–10 Hz), Alpha (8–12.9 Hz), Beta (13–29.9 Hz), and Gamma (30–50 Hz), and extracted. The mean, max, and integral values were compared in the PSD features.

Time–frequency domain feature. The time–frequency domain was divided into four frequency domains, and five characteristics were selected. A discrete wavelet transform (DWT) decomposes a signal into bits based on time. This feature was used in the speech field [38] and in the field of emotion recognition [39]. The mean, maximum, and absolute values were used in each frequency range of DWT; the log and absolute (log) values were also used.

Brain lateralization feature. The values of both electrodes are expressed in the power spectrum. The deflection value of two electrodes is in the range of 0 to 1. A deflection value close to 1 means that the two electrodes are strongly connected, and a deflection value close to 0 means that the correlation between the two electrodes is small.

In this study, the brain lateralization features were extracted in four ways, as depicted in Figure 5.

Figure 5. Brain lateralization feature extraction for EEG signal (a) 1:5, (b) 5:1, (c) 3:3, (d) 5:5.

- 1 $\times$ 5: when lateralization change from left-brain center to the right brain occurs, ensure the lateralization between one left-center electrode and five right-brain electrodes
- 5 $\times$ 1: when lateralization change from right-brain center to the left brain occurs, measure the lateralization between five right-brain electrodes and one left-brain center electrode
- 3 $\times$ 3, 5 $\times$ 5: when lateralization change between the left brain and the right brain occurs, measure the lateralization between the same number of electrodes in the left and right brains

Figure 5 shows the 10–20 system, and the locations of the channels used are the same as in Figure 6. Figure 5 is simplified because it is difficult to see when all electrodes are displayed.

Figure 6. Effective channel comparison between (a) MAHNOB and (b) MERTI-Apps datasets.

4.2. Genetic Algorithm

GA is a heuristic search algorithm that mimics genetic biological evolution as a problem-solving strategy. Algorithm 1 lists the detailed pseudo code of the genetic algorithm. In the first stage, a genetic algorithm (GA) was used to select an effective LSTM model suitable for the given data, as shown in Figure 4. Forty percent of the population were chosen randomly to produce parent objects. After learning, it selected the top 20% of the parent objects and moved them to the next-generation model. In addition, the child objects were produced using selection, mutation, and crossover. This process was repeated until the 10th next-generation model was produced, or the RMSE of the current model was no longer improved. The probability of a mutation was set to 10%, and the mutation process only updated the coefficients of the epoch, LSTM_cell, and drop_out of the models. It was assumed that the epoch, LSTM_cell, drop_out, activation, and optimizer parameters of the models were set randomly at the initial time.

Algorithm 1. The pseudo-code of the genetic algorithm.

Genetic Algorithm

1. **Input data:** *EEG_data*
 2. **Output data:** *model_result, features_result;*
 - 3.
 4. **set** parameters *NoOfPopulations, NoOfGenerations, threshold;*
 - 5.
 6. *GA-Model_Selection:*
 7. Generate randomly *model_populations* **using** *Epochs, lstm_cells, Activation, Dropout, Optimize* and repeat this step until *NoOfPopulations;*
 8. Learn *model_populations* using input data in *lstm* and then save only 20% of top results for next generation, and repeat this step until *NoOfGenerations*
 9. **if** (*result_RMSE > threshold*) {
 10. *Crossover();* // create child objects by mixing parent objects
 11. *Mutate();* // re-determines *Epochs, lstm_cell, and Dropout* within probability 10%
 12. repeat *GA-Model_Selection;*
 13. }
 14. **else**
-

```

15.         save model_result;
16.
17. GA-Feature_Selection:
18.     Generate randomly feature_populations using Input_data;
19.     Learn feature_populations using input data and model_result and save only 20% of top results for next
    generation, and repeat this step until NoOfGenerations;
20.     if (result_RMSE > threshold) {
21.         Crossover(); // create child objects by mixing parent objects
22.         Mutate(); // re-determine feature_list using input_data within probability 10%
23.     }
24.     else
25.         save features_result;

```

In the second stage, the dominant feature set was selected using GA again and the predetermined LSTM model. This stage selects 25–50% of the features in the whole records randomly. Invalid features may confuse the training model if all features are used for LSTM training. As shown in Figure 6, approximately 20% of all the features in MAHNOB-HCI performed well. If all the features with good performance are in the set, it is difficult to add new features and evaluate them. Therefore, 25–50% of features were used as the dominant features. To produce the next-generation features, the GA consisted of three main tasks: selection, crossover, and mutation. The feature set was an array of integers, where each integer represented a weight. Selection was to choose features with high weights from the dominant features to enter the next dominant features. The crossover swaps the features remaining after being selected from the unused features to the high-weight features. Mutations are selected randomly from dominant features and mutated to new features to prevent the cases where there are fewer random selections.

After learning, the model selects the top 20% of parent objects and moves to the next-generation feature group. The crossover ratio of the features fetched when crossover selects features and produces child objects for the next generation was 8:2. The probability of a mutation was set to 10%, which prevent the features that were not selected for genetic diversity. This process was repeated until the 10th generation feature group was formed, or the RMSE of the current model was no longer improved. Choosing the right fitness function is important for the effectiveness and efficiency of a GA.

$$V_{\text{new}} = \beta p_1 + (1 + \beta)p_2, \quad (1)$$

where β represents a random number from -0.25 to 1.25 , p_1 and p_2 represent the parent gene values, and V represents the child gene value. When the two parental genes were similar, multiple similar traits were evaluated each time in repeated inheritance. Randomness was given when a new offspring gene was produced. According to Equation (1), each parent gene was selected for a new child gene, and the characteristics were reduced and selected randomly when crossing, and the child gene was determined using the GA.

4.3. GA-LSTM

The feature extraction step extracts three domain features as one-dimensional features from each EEG signal. Regression was performed through GA-LSTM. The LSTM-fully connected (FC) model applied to GA-LSTM consists of three LSTM layers and two FC layers, as illustrated in Figures 6 and 7. The input layer consisted of GA outputs (for example, eight timestamps, 142 features, 0.02 drop_out, LSTM_cells, and activation function, optimizer). The output layer produced a value between -0.1 and 0.1 using the active function of the FC layer with one neuron in the last layer to predict one valence value. As LSTM input data, the EEG data purified by feature extraction were input at 0.25 s intervals in 2 s increments. The model was the same as the general LSTM model but included a circulatory neural network capable of maintaining conditions. The state learned in a single batch was transmitted

to the next batch. The LSTM was stacked in three layers, allowing a deeper inference than one layer. A dropout layer was added between the LSTM layers to prevent overfitting.

Figure 7. LSTM-fully connected (FC) model architecture.

Based on the one-dimensional feature, the initial feature group was selected randomly through the GA to be used as the input of the LSTM. One valence value was regressed with a one-dimensional feature extracted from each image in a sequence. Before being processed by the LSTM, the feature groups were adjusted by the GA. The hidden state vector output from the LSTM allowed new feature groups to be recruited before the next LSTM. The final product of the GA–LSTM was the valence. Both valence and arousal experiments were performed using the MERTI-Apps dataset.

5. Experimental Results

This study used 239 records from the MAHNOB dataset validated in the previous study and 236 records from experiment 3 from the MERTI-Apps dataset. The sequence of each record was annotated with the valence value at 4 Hz intervals. In this study, when the emotional state was recognized, the emotional state was estimated using the data of the previous 2 s, and a scenario suitable for the actual HCI situation was assumed. The learning information of the proposed algorithm was as follows. To compare the recognition performance using the PSD features and all additional features, the LSTM-FC model with the reimplemented model of the previous study was first examined. The performance of each of the three domain and BL features of the GA–LSTM model was confirmed. The maximum epoch was set to 100 times. The feature extraction step was implemented using MATLAB, and the deep learning algorithm was implemented using PyTorch [40]. The learning environment used Intel i5–7 generation RTX 2080.

5.1. MAHNOB-HCI Dataset

For the MAHNOB-HCI dataset, all data except for the continuous-time annotation information were disclosed. The analysis was performed after receiving the continuous-time annotation from the author to ensure that the experiment was conducted in the same environment and dataset as those by Soleymani et al. [15]. Two hundred and thirty-nine records were used. The test accuracy was measured using a 10-fold validation to compare with that by Soleymani et al. For EEG, 2 s of 256 Hz sampling

and 32 channels of data were used as the unit data. For the EEG feature, information extracted from feature data of a 2 s interval was used as input to the LSTM as a one-dimensional feature.

Table 6 summarizes the results of the valence experiments using only the PSD feature in the previous study with the MAHNOB dataset and using all extracted features proposed in this study. “N/A” denotes the not applicable cells. The GA-LSTM model took approximately six times longer to evaluate the ultimate features through the GA than the existing learning model [15]. Using only the PSD feature, the LSTM-FC model shows a 3% performance improvement over the previous study [15]. A root-mean-square error (RMSE) of 0.0156 with the GA-LSTM model yielded a 24% improvement over [15] when the brain lateralization (BL) feature was added. This suggests that the feature selection using the GA and many features is useful for emotion recognition. Figure 8 gives an example of the test result of annotation labeling using the MANHOB-HCI dataset. The solid line indicates the change in target valence over time, and the valence predicted by GA-LSTM is indicated by the dotted line. The thresholds for annotation labeling ranged from -0.1 to 0.1 . Figure 8a shows positive valence changes over time, and Figure 8b shows the negative valence changes over time. The target and GA-LSTM results tend to be similar. The peculiar thing is that, after the positive or negative valence expression, the GA-LSTM valence tends to return to the neutral value faster than the target valence. The reason is the imbalance of the annotation data. Because the number of positive annotations is greater than the number of negative annotations, it affects the model more, and negative valence or arousal may not be regressed accurately. Furthermore, positive emotions are almost good for facial expressions, while negative emotions are difficult to label by facial expressions because of the severe individual differences. Hence, the labeling method by facial expression is vulnerable to negative emotional expression.

Figure 8. Example of the MAHNOB-HCI test result. (a) Positive valence, (b) negative valence.

Table 6. Experimental results of discrete and continuous emotion recognition according to the model architectures and datasets.

Authors	Dataset	Features	Features Selection	Model Selection	Model	Emotion	Results	
							RMSE	Accuracy
Alhagry et al. [16]	DEAP	raw EEG signals	N/A	N/A	LSTM-RNN	Valence, Arousal	N/A	85.45%/85.65%
Salama et al. [17]	DEAP	raw EEG signals	N/A	N/A	3D-CNN	Valence, Arousal	N/A	87.44%/88.49%
Wang et al. [18]	DEAP	Spectrogram, Normalized mutual information-	Channel selection	N/A	SVM	Valence, Arousal	N/A	74.41%/73.64%
Becker et al. [23]	HR-EEG recordings	Band power, Connectivity, Spectral Crest Factor	Channel/Brain region	N/A	SVM	Valence	N/A	75%
Halim et al. [27]	drivers' EEG dataset	Time and frequency domain set	PCA, MI, RSFS	N/A	ensemble of SVM, RF, and NN	Valence	N/A	88.75%
Krishna et al. [28]	Dataset of mentally impaired subjects	Mel-frequency Cepstrum Coefficients	N/A	N/A	generalized mixture model	Valence	N/A	89%
Soleymani et al. [15] Our Works	MAHNOB-HCI	PSD	N/A	N/A	LSTM	Valence	0.0530 ± 0.0290	N/A
	MAHNOB-HCI	PSD	N/A	N/A	LSTM-FC	Valence	0.0515 ± 0.0150	N/A
	MAHNOB-HCI	3 domain set	Genetic Algorithm	N/A	GA-LSTM	Valence	0.0485 ± 0.0130	N/A
	MAHNOB-HCI	3 domain set, BL	Genetic Algorithm	Genetic Algorithm	GA-LSTM	Valence	0.0156 ± 0.0110	96.2%
	MERTI-Apps	PSD	N/A	N/A	LSTM-FC	Valence	0.0768 ± 0.0150	N/A
	MERTI-Apps	3 domain set	Genetic Algorithm	N/A	GA-LSTM	Valence, Arousal	0.0679 ± 0.0150/0.0752 ± 0.0350	N/A
	MERTI-Apps	3 domain set, BL	Genetic Algorithm	Genetic Algorithm	GA-LSTM	Valence, Arousal	0.0579 ± 0.019/0.0287 ± 0.0151	65.7%/88.3%
	DEAP	3 domain set, BL	Genetic Algorithm	Genetic Algorithm	GA-LSTM	Valence, Arousal	0.0290 ± 0.012/0.0249 ± 0.027	91.3%/94.8%

5.2. MERTI-Apps Dataset

The performance of the proposed GA-LSTM model was also verified using the MERTI-Apps dataset obtained from Asian participants. Like the MAHNOB-HCI dataset, 60% and 40% of the records except for the test records were used as training and validation, respectively. The test data were measured using a 10-fold validation. The EEG signals were collected through 1024 Hz sampling and a 12-channel electrode and used as the LSTM input as a one-dimensional feature using data from a 2 s interval.

Table 6 lists the performance using the MERTI-Apps dataset. The RMSE 0.0579 of GA-LSTM when a three-domain set and the BL feature were added represents a 33% improvement compared to the RMSE 0.0768 of LSTM-FC in the valence domain. With the MERTI-Apps dataset, the GA applying the model selection and the feature selection for various features was useful for emotion recognition. Figure 9 presents the estimated values of the annotation labeling (target) and GA-LSTM over time for positive valence and negative valence using the MERTI-Apps dataset. Figure 9a suggests that the estimated value of the positive valence at the time of positive expression accurately determines the target value. Figure 9b shows that the estimated value of the negative valence is different from the target value at the time of negative expression. On the other hand, the direction and slope of valence are similar. Table 6 also lists the results of arousal experiments using the MERTI-Apps dataset. The regression performance (RMSE = 0.0287) using the arousal data was better than that using the valence data (RMSE = 0.0579), even though human expression was absent. Hence, continuous emotion regression using arousal data is robust to inner emotion recognition.

Figure 9. Example of MERTI-Apps test result. (a) Positive valence, (b) negative valence.

5.3. DEAP Dataset

In the field of emotion recognition using biosignals, there are more state-of-the-art studies on the classification model than on the regression model. Therefore, by converting the proposed regression model into a classification model, as shown in Table 6, the classification performance of the proposed GA-LSTM in the DEAP dataset was measured and compared with the existing state-of-the-art studies [16–18,23,27,28]. The last column shows the classification accuracy in terms of the valence and arousal perspective. Alhagry et al. [16] and Salama et al. [17] examined the classification performance using the raw EEG signals in the DEAP dataset. Salama et al. reported the highest value among the classification accuracies in Table 6, and the accuracies with respect to valence and arousal were 87.44% and 88.49%, respectively, when emotion was divided into fear, sad, happy, and satisfied. The accuracies reported by Wang et al. [18] with respect to arousal and valence were 74.41% and 73.64%, respectively, when channel selection was applied. They conducted discrete emotion recognition experiments by classifying emotions into high valence, neutral, and low valence.

Owing to the high regression power of continuous emotion recognition, the accuracies of emotion classification using the proposed GA-LSTM were 91.3% and 94.8% with respect to valence and arousal, respectively. The proposed GA-LSTM showed good performance in discrete emotion recognition even when compared to state-of-the-art studies. Therefore, the GA-LSTM model is effective in regression and classification.

6. Discussion and Conclusions

This paper proposed the MERTI-Apps dataset for Asian emotion recognition and developed a method of searching for effective models and feature groups required for emotional classification with EEG signals. To collect data for dataset creation, short and robust stimulation videos were prepared through a content validity inspection in the video collection. Training was conducted to enable an objective evaluation through an in-house production program for annotation labeling. In addition, valence and arousal were evaluated according to complex human emotions. To minimize inconvenience to the users, the data were collected in three experiments to prevent excessive sensor attachment. The models and feature groups that were effective and ineffective for emotion classification were determined by identifying the useful models, features, and channels for emotion recognition. Figure 10 shows the characteristics and channel weights according to the GA-LSTM learning results when the MAHNOB dataset was applied. A darker color means that the features are used more commonly in the GA-LSTM, and the result can be seen as a valid feature. In the MANHOB-HCI dataset, the experimental result showed that the classification accuracy of GA-LSTM was 96.2%. Therefore, the effect of the GA increases with more data.

Figure 10. EEG feature weights, channel weights of GA-LSTM in the MANHOB-HCI dataset.

The regression performance of GA-LSTM was improved by 24% compared to Soleymani et al. [15], owing to effective model selection and active feature group selection using a GA. In the MERTI-Apps dataset based on Asian participants, similar trends were found in the regression performance and classification accuracy for emotion recognition. Although similar results were obtained, the experiment showed that the RMSE in MERTI-Apps dataset was 0.0579, whereas the RMSE in the MAHNOB dataset was 0.0156 in the valence domain. The MERTI-Apps dataset involved a multi-step verification for the accuracy of annotations. For the MAHNOB-HCI dataset, however, the model can be assumed to have performed more rigorously because more experiments were conducted with more cases. In addition, annotating the regression value is more difficult than annotating the binary classification of positive and negative because the assessment of emotion is subjective. To compensate for the discrepancy, additional research on annotation and quantitative measurement methods that exclude subjective factors will be needed.

Based on Figure 6, valid channels were searched and expressed, as shown in Figure 11. In Figure 6a, the MAHNOB dataset results indicated that the electrodes of the right brain, parietal lobes, and occipital lobe were helpful for emotion recognition. Similarly, in Figure 6b, the experimental results of MERTI-Apps also confirm that the electrodes at similar positions are valid. In other words, there is no significant difference between the MAHNOB-HCI dataset based on Westerners and the MERTI-Apps dataset based on Asians. In the experiment using the MERT-Apps dataset, the arousal performance was better than the valence performance when human expression was absent. Moreover, additional bio-signals, such as EMG, EOG, and PPG, can compensate for the characteristics not found by EEG signals.

Figure 11. Training loss and validation loss of GA-LSTM Model using (a) MERT-Apps Arousal, (b) MERTI-Apps Arousal, (c) MAHNOB, and (d) DEAP Arousal, (e) DEAP Valence datasets.

The RMSE value of the proposed GA-LSTM showed a 33% performance improvement over LSTM-FC owing to the weight evolution effects of the GA for selecting models, features, and channels. On the other hand, learning was slow because the emotion classification was conducted while exploring valid models, features, and channels through GA-LSTM. Therefore, it may be difficult to apply directly to real-time emotion recognition.

The experimental result using the MERTI-Apps dataset still showed a higher RMSE than MAHNOB-HCI. The number of data used was the same, but the MERTI-Apps dataset may have more low-quality data than MANHOB-HCI because it used fewer videos than MANHOB-HCI. Furthermore,

EEG is a very sensitive and difficult signal to analyze. Additional bio-signals, such as PPG, GSR, EMG, and EOG, are needed to compensate for this. In experiment 1 of the MERTI-Apps dataset, PNS signals, their valence, and arousal information were collected and studied to determine how to use them efficiently. Methods using observers, such as participants' evaluation of emotion annotation and annotation labeling, have been used. This is because it is unclear what emotions the signals represent. Moreover, the labeling method by facial expression is vulnerable to negative emotional expression, as shown in Figures 8 and 9. Therefore, it is important to consider annotating methods that can be classified more accurately. In addition, the fusion of biometric signals with voice and image data will improve emotion recognition patterns. Figure 11 shows the training loss and validation loss of the GA-LSTM model in MERTI-Apps, MAHNOB, and DEAP datasets. The stability of the model was improved considerably, owing to the evolutionary effect of a genetic algorithm. The present study used a continuous-time annotation for labeling the MAHNOB-HCI and MERTI-Apps databases, but used only self-assessment for labeling the DEAP database. Because it is difficult to apply the label method using only self-assessment to the training of the proposed model, it cannot reduce the loss rate quickly in the case of the DEAP database.

Author Contributions: Conceptualization, J.-H.M., D.-H.K. (Dong-Hyun Kang) and D.-H.K. (Deok-Hwan Kim); methodology, J.-H.M.; software, J.-H.M. and D.-H.K. (Dong-Hyun Kang); validation, J.-H.M., D.-H.K. (Dong-Hyun Kang) and D.-H.K. (Deok-Hwan Kim); formal analysis, D.-H.K. (Deok-Hwan Kim); investigation, J.-H.M. and D.-H.K. (Dong-Hyun Kang); resources, J.-H.M.; data curation, J.-H.M. and D.-H.K. (Dong-Hyun Kang); writing—original draft preparation, J.-H.M.; writing—review and editing, D.-H.K. (Deok-Hwan Kim); visualization, J.-H.M.; supervision, D.-H.K. (Deok-Hwan Kim); project administration, D.-H.K. (Deok-Hwan Kim). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This work was supported by the Industrial Technology Innovation Program funded by the Ministry of Trade, Industry, and Energy (MI, Korea) (10073154, Development of human-friendly human–robot interaction technologies using human internal emotional states) and in part by Inha University Research Grant.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sinith, M.S.; Aswathi, E.; Deepa, T.M.; Shameema, C.P.; Rajan, S. Emotion recognition from audio signals using Support Vector Machine. In *Intelligent Computational Systems (RAICS)*; IEEE Recent Advances in 2005; IEEE: Piscataway, NJ, USA, 2015; pp. 139–144.
2. Busso, C.; Deng, Z.; Yildirim, S.; Bulut, M.; Lee, C.M.; Kazemzadeh, A.; Narayanan, S. Analysis of emotion recognition using facial expressions, speech, and multimodal information. In *Proceedings of the 6th International Conference on Multimodal Interfaces*, State College, PA, USA, 13–15 October 2004; pp. 205–211.
3. Yang, H.; Ciftci, U.; Yin, L. Facial Expression Recognition by De-Expression Residue Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2168–2177.
4. Jung, H.; Lee, S.; Yim, J.; Park, S.; Kim, J. Joint fine-tuning in deep neural networks for facial expression recognition. In *Proceedings of the IEEE International Conference on Computer Vision 2015*, Santiago, Chile, 7–13 December 2015; pp. 2983–2991.
5. Ranganathan, H.; Chakraborty, S.; Panchanathan, S. Multimodal emotion recognition using deep learning architectures. In *Proceedings of the 2016 IEEE Winter conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, 7–10 March 2016.
6. Valstar, M.; Pantic, M. Induced disgust, happiness, and surprise: An addition to the mmi facial expression database. In *Proceedings of the 3rd International Workshop on EMOTION (Satellite of LREC): Corpora for Research on Emotion and Affect*, Caserta, Italy, 15–19 March 2010; pp. 65–70.
7. Wang, X.-H.; Zhang, T.; Xu, X.-M.; Chen, L.; Xing, X.-F.; Chen, C.P. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks and Broad Learning System. In *Proceedings of the IEEE Conference on Bioinformatics and Biomedicine (BIBM)*, Madrid, Spain, 3–6 December 2018; pp. 1240–1244.

8. Mavratzakis, A.; Herbert, C.; Walla, P. Emotional facial expressions evoke faster-orienting responses, but weaker emotional responses at neural and behavioral levels compared to scenes: A simultaneous EEG and facial EMG study. *Neuroimage* **2016**, *124*, 931–946. [\[CrossRef\]](#)
9. Jenke, R.; Peer, A.; Buss, M. Feature extraction and selection for emotion recognition from EEG. *IEEE Trans. Affect. Comput.* **2014**, *5*, 327–339. [\[CrossRef\]](#)
10. Ding, Y.; Hu, X.; Xia, Z.; Liu, Y.J.; Zhang, D. Inter-brain EEG feature Extraction and Analysis for Continuous Implicit Emotion Tagging during Video Watching. *IEEE Trans. Affect. Comput.* **2018**. [\[CrossRef\]](#)
11. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Patras, I. Deep: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. [\[CrossRef\]](#)
12. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A multimodal database for affect recognition and implicit tagging. *IEEE Trans. Affect. Comput.* **2012**, *3*, 42–55. [\[CrossRef\]](#)
13. Soraia, M.A.; Manuel, F.J. Emotions Recognition Using EEG Signals: A Survey. *IEEE Trans. Affect. Comput.* **2017**, *10*, 374–393.
14. Maria, E.; Matthias, L.; Sten, H. Emotion Recognition from Physiological Signal Analysis: A Review. *Electron. Notes Theor. Comput. Sci.* **2019**, *343*, 35–55.
15. Soleymani, M.; Asghari-Esfeden, S.; Fu, Y.; Pantic, M. Analysis of EEG signals and facial expressions for continuous emotion detection. *IEEE Trans. Affect. Comput.* **2016**, *1*, 17–28. [\[CrossRef\]](#)
16. Alhagry, S.; Fahmy, A.A.; El-Khoribi, R.A. Emotion recognition based on EEG using LSTM recurrent neural network. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 8–11. [\[CrossRef\]](#)
17. Salama, E.S.; El-Khoribi, R.A.; Shoman, M.E.; Shalaby, M.A.W. EEG-based emotion recognition using 3D convolutional neural networks. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 329–337. [\[CrossRef\]](#)
18. Wang, Z.M.; Hu, S.Y.; Song, H. Channel selection method for eeg emotion recognition using normalized mutual information. *IEEE Access* **2019**, *7*, 143303–143311. [\[CrossRef\]](#)
19. Lucey, P.; Cohn, J.F.; Kanade, T.; Saragih, J.; Ambadar, Z.; Matthews, I. The extended Cohn-Kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) 2010, San Francisco, CA, USA, 13–18 June 2010; pp. 94–101.
20. Granados, L.S.; Organero, M.M.; Gonzalez, G.R.; Abdulhay, E.; Arunkumar, N. Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset(AMIGOS). *IEEE Access* **2018**, *7*, 57–67. [\[CrossRef\]](#)
21. Kim, B.H.; Jo, S. Deep Physiological Affect Network for the Recognition of Human Emotions. *IEEE Trans. Affect. Comput.* **2020**, *11*, 230–243. [\[CrossRef\]](#)
22. Song, T.; Zheng, W.; Song, P.; Cui, Z. EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks. *IEEE Trans. Affect. Comput.* **2020**, *11*, 532–541. [\[CrossRef\]](#)
23. Becker, H.; Fleureau, J.; Guillote, P.; Wendling, F.; Merlet, I.; Albera, L. Emotion recognition based on high-resolution EEG recordings and reconstructed brain sources. *IEEE Trans. Affect. Comput.* **2020**, *11*, 244–257. [\[CrossRef\]](#)
24. Gao, Z.; Wang, X.; Yang, Y.; Mu, C.; Cai, Q.; Dang, W.; Zuo, S. EEG-Based Spatio-Temporal Convolutional Neural Network for Driver Fatigue Evaluation. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2755–2763. [\[CrossRef\]](#)
25. Chao, H.; Dong, L.; Liu, Y.; Lu, B. Improved Deep Feature Learning by Synchronization Measurements for Multi-Channel EEG Emotion Recognition. *Hindawi Complex.* **2020**, *2020*, 6816502. [\[CrossRef\]](#)
26. Li, P.; Liu, H.; Si, Y.; Li, C.; Li, F.; Zhu, X.; Huang, X.; Zeng, Y.; Yao, D.; Zhang, Y.; et al. EEG Based Emotion Recognition by Combining Functional Connectivity Network and Local Activations. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 2869–2881. [\[CrossRef\]](#)
27. Halim, Z.; Rehan, M. On identification of driving-induced stress using electroencephalogram signals: A framework based on wearable safety-critical scheme and machine learning. *Inf. Fusion* **2020**, *53*, 66–79. [\[CrossRef\]](#)
28. Krishna, N.M.; Sekaran, K.; Vamsi, A.V.N.; Ghantasala, G.P.; Chandana, P.; Kadry, S.; Blažauskas, T.; Damaševičius, R. An efficient mixture model approach in brain-machine interface systems for extracting the psychological status of mentally impaired persons using EEG signals. *IEEE Access* **2019**, *7*, 77905–77914. [\[CrossRef\]](#)

29. Katsigiannis, S.; Ramzan, N. DREAMER: A Database for Emotion Recognition through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 98–107. [[CrossRef](#)] [[PubMed](#)]
30. Mithbavkar, S.A.; Shah, M.S. Recognition of Emotion through Facial Expressions Using EMG Signal. In Proceedings of the 2019 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 4–5 January 2019.
31. Healey, J.A.; Picard, R.W. Detecting Stress during Real-World Driving Tasks Using Physiological Sensors. *IEEE Trans. Intell. Transp. Syst.* **2005**, *6*, 156–166. [[CrossRef](#)]
32. Grimm, M.; Kroschel, K.; Narayanan, S. The Vera am Mittag German Audio-Visual Emotional Speech Database. In Proceedings of the IEEE International Conference Multimedia and Expo, Hannover, Germany, 23–26 June 2008; pp. 865–868.
33. Feradov, F.N.; Ganchev, T.D. Ranking of EEG Time-domain Features on the Negative Emotions Recognition Task. *Annu. J. Electron.* **2015**, *9*, 26–29.
34. Biosemi Active II System. Available online: <https://www.biosemi.com/products.htm> (accessed on 3 November 2020).
35. Herwing, U.; Satrapi, P.; Ronfeldt-Lecuona, C. Using the International 10–20 EEG system for Positioning of Transcranial Magnetic stimulation. *Brain Topogr.* **2003**, *16*, 95–99. [[CrossRef](#)] [[PubMed](#)]
36. Kumfor, F.; Landin-Romero, R.; Devenney, E.; Hutchings, R.; Grasso, R.; Hodges, J.R.; Piguet, O. On the right side? A longitudinal study of left- versus right-lateralized semantic dementia. *Brain* **2016**, *139*, 986–998. [[CrossRef](#)] [[PubMed](#)]
37. Yin, Z.; Zhao, M.; Wang, Y.; Yang, J.; Zhang, J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Comput. Methods Prog. Biomed.* **2017**, *140*, 93–110. [[CrossRef](#)] [[PubMed](#)]
38. Ali, H.; Ahmad, N.; Zhou, X.; Iqbal, K.; Ali, S.M. DWT features performance analysis for automatic speech recognition of Urdu. *SpringerPlus* **2014**, *3*, 204. [[CrossRef](#)]
39. Candra, H.; Yuwono, M.; Chai, R.; Nguyen, H.T.; Su, S. EEG emotion recognition using reduced channel wavelet entropy and average wavelet coefficient features with normal Mutual Information method. In Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017.
40. Ketkar, N. Introduction to PyTorch. In *Deep Learning with Python*; Apress: Berkeley, CA, USA, 2017; pp. 195–208.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).