



EEG Channel Correlation Based Model for Emotion Recognition



Md. Rabiul Islam^{a,b}, Md. Milon Islam^c, Md. Mustafizur Rahman^d, Chayan Mondal^b, Suvojit Kumar Singha^b, Mohiuddin Ahmad^b, Abdul Awal^e, Md. Saiful Islam^f, Mohammad Ali Moni^{g,*}

^a Electrical and Electronic Engineering, Bangladesh Army University of Engineering & Technology, Natore, 6431, Bangladesh

^b Electrical and Electronic Engineering, Khulna University of Engineering & Technology, Khulna, 9203, Bangladesh

^c Computer Science and Engineering, Khulna University of Engineering & Technology, Khulna, 9203, Bangladesh

^d Electrical and Electronic Engineering, Jashore University of Science and Technology, Jashore, 7408, Bangladesh

^e Electronics and Communication Engineering, Khulna University, Khulna, 9208, Bangladesh

^f School of Information and Communication Technology, Griffith University, Gold Coast, Australia

^g School of Health and Rehabilitation Sciences, The University of Queensland, St Lucia, QLD, 4072, Australia

ARTICLE INFO

Keywords:

Emotion
Convolutional neural network
Feature extraction
EEG
Pearson's correlation coefficient
Complexity

ABSTRACT

Emotion recognition using Artificial Intelligence (AI) is a fundamental prerequisite to improve Human-Computer Interaction (HCI). Recognizing emotion from Electroencephalogram (EEG) has been globally accepted in many applications such as intelligent thinking, decision-making, social communication, feeling detection, affective computing, etc. Nevertheless, due to having too low amplitude variation related to time on EEG signal, the proper recognition of emotion from this signal has become too challenging. Usually, considerable effort is required to identify the proper feature or feature set for an effective feature-based emotion recognition system. To extenuate the manual human effort of feature extraction, we proposed a deep machine-learning-based model with Convolutional Neural Network (CNN). At first, the one-dimensional EEG data were converted to Pearson's Correlation Coefficient (PCC) featured images of channel correlation of EEG sub-bands. Then the images were fed into the CNN model to recognize emotion. Two protocols were conducted, namely, protocol-1 to identify two levels and protocol-2 to recognize three levels of valence and arousal that demonstrate emotion. We investigated that only the upper triangular portion of the PCC featured images reduced the computational complexity and size of memory without hampering the model accuracy. The maximum accuracy of 78.22% on valence and 74.92% on arousal were obtained using the internationally authorized DEAP dataset.

1. Introduction

Emotion is a feeling that implies how we act for a particular instance. Now in the era of technological development, the research of emotion recognition from EEG has become very popular due to its non-invasive feature. The remarkable development has occurred on emotion-based non-invasive Brain-Computer Interface (BCI) and signal processing technology. Therefore, the analysis of EEG-based emotion recognition has become a very attractive issue in affective computing. Various information and important data can easily be extracted from the human brain and analyzed for knowing the inner truth of humans.

Nowadays machine especially robot has been utilized in many

industries, hospitals and even in household applications. People are setting higher expectations for robots as they become more prevalent in many parts of daily life. For better human-machine interaction, it is hoped the super ability of decision making, self-thinking, emotion sensing. The assurance of emotion recognition is an inevitable need to make a robot more practical for real-life applications. The patient's affective information, including emotional state, is considered a key factor of his/her mental and physical status. A patient's emotional state has a significant effect on the treatment management process [1]. Therefore clinics, hospitals, and other healthcare providers should keep a proper system for emotional state assessment to strengthen healthcare. Since human behavior is mainly dependent on emotion and psychologist

* Corresponding author.

E-mail addresses: rabiul.kuet.bd@gmail.com (Md.R. Islam), milonislam@cse.kuet.ac.bd (Md.M. Islam), mustafizur.170710@gmail.com (Md.M. Rahman), chayan.eee.92@gmail.com (C. Mondal), singha10.suvojit@gmail.com (S.K. Singha), ahmad@eee.kuet.ac.bd (M. Ahmad), m.awal@ece.ku.ac.bd (A. Awal), saiful.islam@griffith.edu.au (Md.S. Islam), m.moni@uq.edu.au (M.A. Moni).

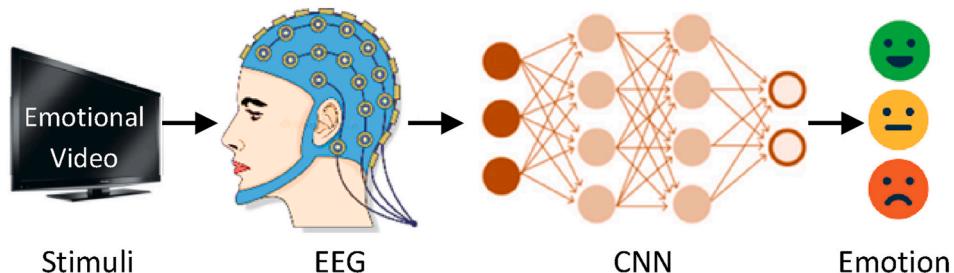


Fig. 1. Emotion recognition from EEG signal using CNN.

wants to analyze the condition of the human mind; recognition of emotion has also become a hot topic in psychology. Many psychological experiments demonstrated the relation between emotions with EEG signals [2,3]. Even emotion is now treated as a governor of music and video recommendation. For instance, in a sad mood, it recommends a funny event for making a person happy. In gassing, the level may become harder in positive emotional states. In court, during the criminal's statement, emotion can also be recorded to judge his authenticity and reliability. Accordingly, emotion recognition has become an indispensable part of our daily life. Consequently, computer scientists, AI specialists, physiologists, and biologists are continuously generating enormous studies using computer technology to understand human emotions [4].

The research on emotion recognition demands the emotionally aroused signal of the human being. Humans remain emotionally excited in their real-life activities. However, it is too tough to accumulate the aroused data on that situation. For research, emotion-relevant EEG data is usually collected by listening music, watching videos, and playing emotion-related games. The audiovisual stimuli create the proper excitement on human brain signals among the different types of stimuli as audiovisual stimuli affect the human mind by both audio and video context [5]. To recognize human emotion, many scholars used various types of raw signals [6]. Many used EEG signals [2,7–9] and facial expressions [10,11]; and few used gesture, speech signals [12], autonomous nervous signals [13]. Subjects need to express emotion explicitly when facial expressions and speech signals are being used. Moreover, from the facial expression, only these types of emotion can be classified whose effect will change the human face structure. However, subjects who feel happy internally but do not express it by their face cannot be classified. So facial expression is not considered as a good emotion recognizer raw signal. For speech signals, the subject's emotion can be classified with the help of the voice's intensity. For that reason, this type of emotion recognition method is not applicable to autistic persons. Moreover, the people who are not able to speak remain out of this type of emotion recognition method. Some researchers tried to classify emotion from gesture [14] and body movement [15]. People with physically handicapped are not able to present body language and gestures to express emotion. Furthermore, the autonomous nervous system is not too applicable because of its complex acquisition technique. Interestingly, subjects cannot control the spontaneously generated EEG signal. When we extract emotion from EEG, then the emotion of the people who cannot speak or unable to express themselves by gesture and posture can also be recognized. Besides, the EEG signal acquisition approach is not as tricky as multimodal signal acquisition. Moreover, the low cost, wireless flexible and portable acquisition medium makes it more popular. Consequently, the best and peerless signal for emotion recognition is EEG for any kind of person at any time in any place.

Different researchers have applied multiple methods and approaches in recognizing and classifying emotion [16–18]. Recently, the wavelet transform has become a prevalent analysis method due to its good performance both in the time and frequency domain [8]. EEG signal is a non-linear, non-stationary, and temporal asymmetry type signal on the microvolt range. The computation and analysis of this type of signal are

too challenging. Some researches on emotion recognition have been performed by selecting proper features and shallow machine learning algorithms like Support Vector Machine (SVM) [19–23], k Nearest Neighbor (kNN) [19], Decision Tree (DT), or Multi-Layer Perceptron (MLP) etc. Soleymani et al. developed an emotion recognition method based on user-independent analysis. They achieved a good result in terms of accuracy, like 68.5% for valence and 76.4% for arousal upon the classification of 3 classes [24]. Zheng et al. investigated stable EEG patterns using machine learning and systematically evaluated the performance of different feature extraction methods, selection, and smoothing [25]. Atkinson and Campos suggested a valence arousal-based method combined with mutual information-based feature selection methods and kernel classifiers [26]. They proposed an EEG-based Brain-Computer Interface to explore a set of emotional types and incorporate additional features relevant for signal preprocessing and classification. In the conventional feature extraction methods, several features or feature set are being considered to train the recognition system. In contrast, deep learning eliminates the difficulties of selecting the useful and significant features. It allows the machine to learn the feature from the input data set automatically and transfer its learning to the classifier.

Längkvist et al. proposed a Deep Belief Network (DBN) architecture to reduce the complexity and necessity of multimodal sleep data [27]. They also recovered the fact of time consumption in the classification stage. Li et al. used differential entropy feature with a novel DBN and achieved 11.5% and 24.4% improvements on the task of affective state recognition [28]. Martinez et al. classified four different emotional states using CNN, considering the skin conductance and blood volume pulse signal [29]. The researcher Wen et al. [30] proposed a method using channel correlation, but they did not use the emotion-relevant sub-band data. Another researcher used deep and convolutional neural networks on mean, median, standard deviation, and many other time-domain features [31], but time-domain features are not too helpful. The author of [32] calculates the PCC of multichannel EEG data considering four sub-bands.

This paper proposed an emotion recognition method of lower complexity from audio-visual stimuli-based EEG signals using CNN. The architectural flow diagram is shown in the following Fig. 1. Emotion is highly related to beta and gamma sub-bands, moderately related to alpha sub-band, and very poorly connected to theta sub-band. Therefore, we have considered only alpha, beta gamma sub-bands despite considering four sub-bands. That makes our information more significant and reduces computational complexity. Again, we proved that only the upper triangular portion of the PCC featured images was substantial to recognize emotion. Our work used a deep architecture-based Convolution Neural Network as it could automatically extract internal features. Thus, the main contributions of the proposed research are as follows:

- 1) The development of an improved emotion recognition method with lower computational complexity, lower memory requirement, lower time consumption for processing; without hampering the performance.

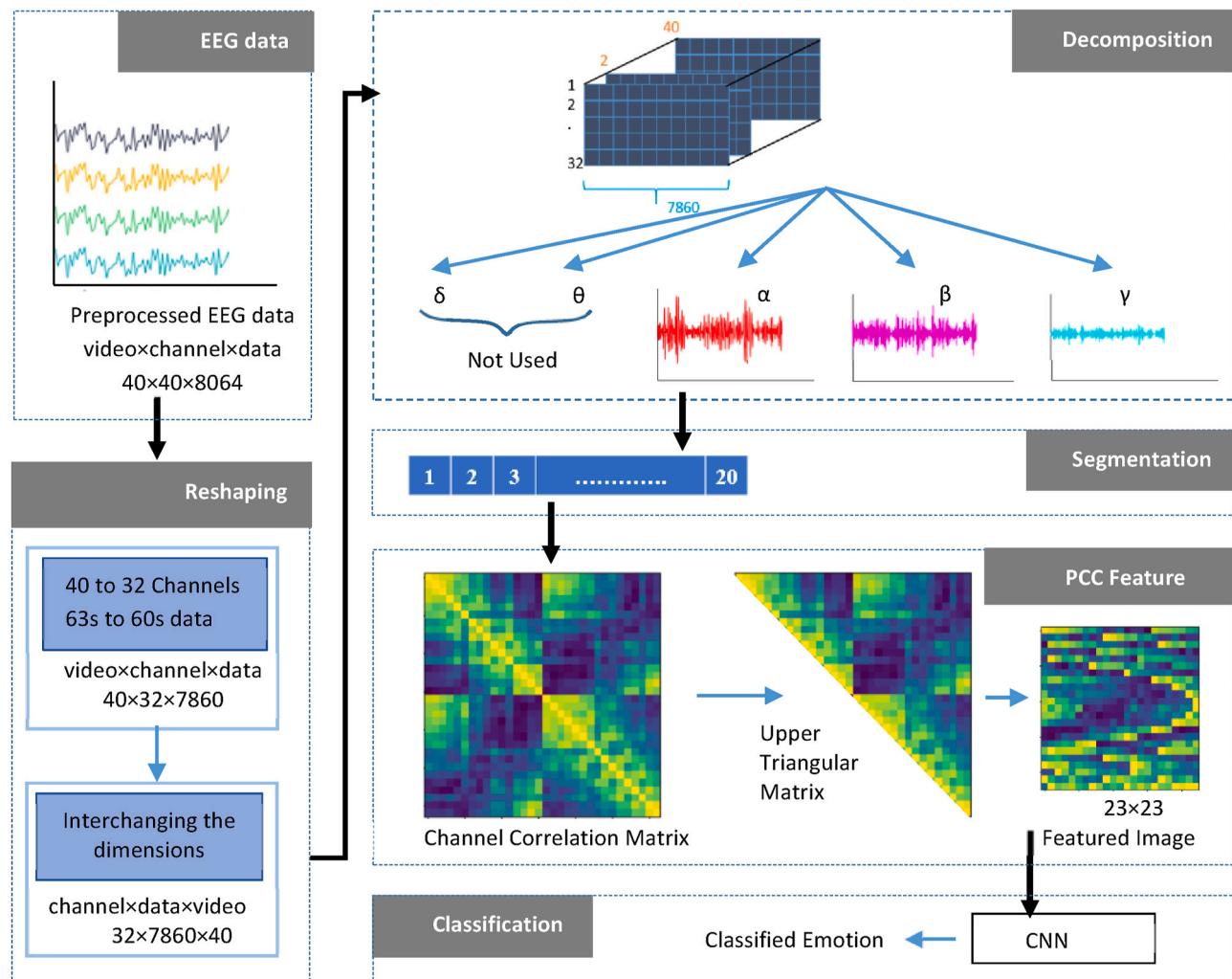


Fig. 2. The successive process of our proposed method includes reshaping, decomposition, segmentation, PCC featured image extraction, and classification to recognize emotion from raw EEG data.

- 2) Developing a novel procedure to construct a matrix of significant two-dimensional data based on channel correlation from the one-dimensional EEG signals is proposed.
- 3) The development of a CNN-based modified model that can recognize emotion using a single feature Pearson's Correlation Coefficient of EEG sub-bands.

This work is the updated version of our previous work [33]. Here, the computation complexity is significantly reduced by reducing input images' size considering only the upper triangular matrix of PCC featured images. Moreover, the convolution neural network model is modified according to the size of our newly generated PCC featured images that applied as input.

The rest of the paper is organized as follows. Section II demonstrates the necessary dataset collection and description with data preprocessing including some preprocessing techniques. In addition, the development of network architecture is also described in this section. The experimental results for two different protocols with the DEAP dataset are illustrated in section III. The discussions along with the comparative study with the current works are depicted in section IV. Finally, the conclusion with some significant future works is discussed in section V.

2. Materials and methods

The proposed system takes EEG data as input and generates classified

emotion as output. The system diagram of our experiment is illustrated in Fig. 2. Later the individual portion of the system is described separately. Firstly, the raw EEG data were preprocessed. Secondly, just one feature named Pearson's Correlation Coefficient between every possible combination of two channels among 32 channels was determined. Thus, we formulated the PCC featured images of channel correlation. Afterward, the PCC featured images containing two-dimensional channel correlation data were taken as input into our designed CNN classification algorithm to recognize emotion.

2.1. Dataset

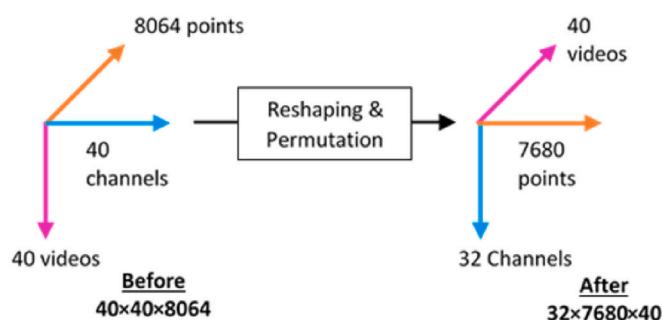
We have considered the internationally accredited 'DEAP' dataset [34] in our study to perform the experiment and measured our model's performance. There remain sixteen different emotional class-based labeled EEG data of 32 participants. 'DEAP' stands for Dataset for Emotion Analysis using Physiological signals. It is a commonly useable open dataset for emotion analysis that contains EEG, physiological and video signals.

Researchers used the dataset to evaluate the performance of their own designed system of emotion recognition. The 16 male and 16 female participants (age ranges from 19 to 37 years) were watched forty different emotional music videos and the corresponding physiological recordings and participant ratings were stored. Each video had a duration of 1 min long. They recorded EEG signals using the International

Table 1

Information about ‘DEAP’ dataset.

Attributes	Details Information
EEG acquisition system	Biosemi ActiveTwo
Channels of recorded signals	32 EEG channel (512 Hz) 12 Peripherals 1 Status channel 3 Unused channels Total of 48 channels
EEG electrodes	AF3, AF4, C3, C4, Cz, CP1, CP2, CP5, CP6, F3, F4, F7, F8, Fz, FC1, FC2, FC5, FC6, Fp1, Fp2, O1, O2, Oz, P3, P4, P7, P8, Pz, PO3, PO4, T7 and T8
Rated parameter	Valence, arousal, dominance, liking, and familiarity
Range of rating	1–9, except familiarity (1–5)
Available data format	video × channel × data = 40 × 40 × 8760

**Fig. 3.** Dimension 3D of data before and after reshaping and permutation.

10/20 electrode placement system. The relevant information about the ‘DEAP’ dataset is shown in **Table 1**.

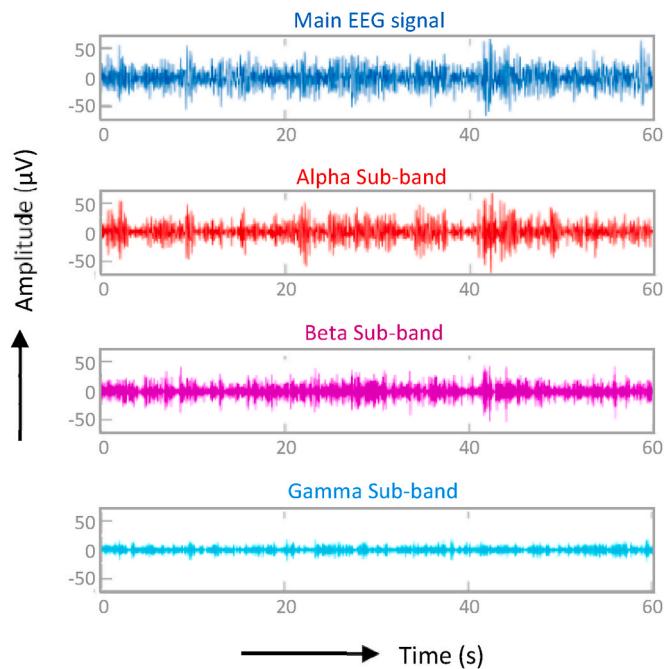
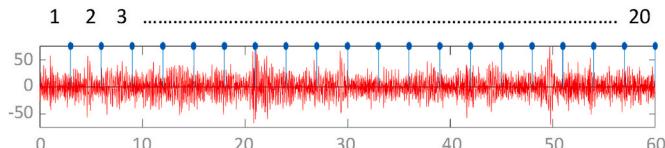
As the EEG signal is very low amplitude signals, the extraction of information from this type of signal is too complicated. Moreover, an Electrooculogram (EOG) and another noise signal may hamper the original signal’s effectiveness if it is not preprocessed. Firstly, the raw data were downsampled to 128 Hz from 512 Hz to reduce the volume. Later, EOG artifacts were removed. Finally, it was passed through a bandpass filter of frequency (4–45) Hz. We collected the preprocessed data from the official website of ‘DEAP’.

2.2. Reshaping and Permutation

In the emotion recognition task, the downloaded data shape format was like that video × channel × data = 40 × 40 × 8064. It illustrated 40 different videos, 40 channels, and 8064 data points (63s data, sampling frequency 128 Hz, data points = 128 × 63). Among them, only the first 32 electrodes (channels) were used for the EEG signals recording. Therefore, we extracted these 32 channel data only. Also, the last 60s data were original. We reshaped the data as video × channel × data = 40 × 32 × 7680. Afterward, the reshaped data were permuted to arrange the data as following format channel × data × video = 32 × 7680 × 40. The data structure before and after reshaping and permutation has demonstrated in **Fig. 3**.

2.3. Decomposition

The activity of human work affects fully on brain waves. As a result, the mental states and conditions can be extracted from the EEG signals. Generally human EEG is a composite type signal that consists of five different types of brain waves of different frequency namely delta ($1 \text{ Hz} < f < 4 \text{ Hz}$), theta ($4 \text{ Hz} < f < 8 \text{ Hz}$), alpha ($8 \text{ Hz} < f < 12 \text{ Hz}$), beta ($12 \text{ Hz} < f < 30 \text{ Hz}$) and gamma ($30 \text{ Hz} < f < 60 \text{ Hz}$). The different sub-bands are incorporated with different mental states and activities. For an instance, the delta sub-band is related to relaxing or calm

**Fig. 4.** Decomposition of main preprocessed EEG signal (of p1, v1, Fp1) into alpha, beta, and gamma sub-band signals.**Fig. 5.** Every decomposed signal of the 60s is segmented into 20 portions.

activities like deep sleep and unconsciousness. The theta and alpha sub-bands are related to low-level excitement like drowsiness, imagination, closing the eye etc. The medium-level excitements including thinking, anxiety and stressed are incorporated with the beta sub-band signal. Lastly, the gamma sub-band signals affect hyperactivity, such as alertness, agitation, object matching and sensory processing. In this part of our work, the EEG signals were decomposed into the five sub-bands using Discrete Wavelet Transform. Later, as delta and theta sub-band associated with relaxing to low-level brain activity, we ignored it for emotion recognition. The original EEG signal and our decomposed alpha, beta, and gamma sub-band signals of channel Fp1, of the first participant (p1) for the first video (v1), are shown in **Fig. 4**.

2.4. Segmentation

As CNN requires many training data for an exclusive performance, the EEG data have to be segmented. One EEG signal’s length is the 60s, which means the number of total data points is 7680. Before calculating the PCC of different channels, we need to segment the data. Here the EEG signals were segmented into 20 parts in which every segment contains 384 data points (data of 3s). The segmented part of a single channel EEG is shown in **Fig. 5**. Accordingly, for a single participant, the total segmented partition number will be 20 times 32 times 40, i.e., 25,600.

2.5. PCC featured image Formation

To recognize emotion from EEG signal, many researchers used various types of the method such as Short Time Fourier transform

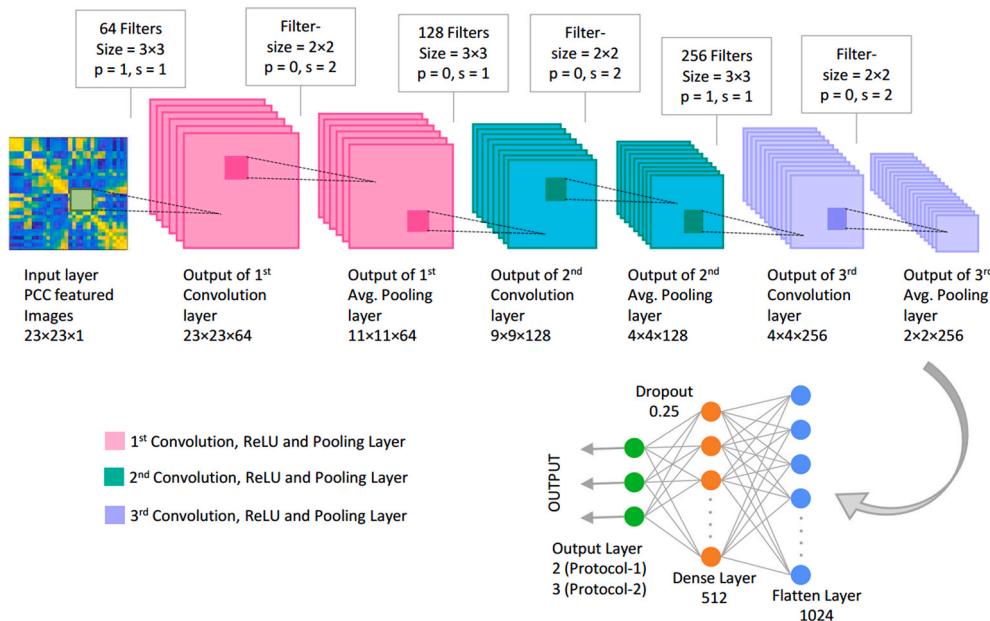


Fig. 6. The proposed CNN model for emotion recognition consists of three sets of convolution, Rectified Linear Unit (ReLU), and pooling layers in addition to one flatten and one dense layer. The size of images, number of filters, padding, and stride, etc., on each level, are shown.

(STFT), Common Spatial Pattern (CSP), Discrete Wavelet Transform, Statistical feature, Higuchi fractal dimension (HFD), Magnitude Squared Coherence Estimate (MSCE), Power Spectral Density (PSD), Fast Fourier Transform (FFT), Higher-Order component (HOC), Differential Entropy (DE), Differential Asymmetry (DASM), Rational Asymmetry (RASM), Affective Signal Processing (ASP), etc. We used Pearson's Correlation Coefficient to generate the PCC-featured effective images for emotion recognition.

Pearson's correlation coefficient is the scale of indication that represents the statistical relationship between two continuous variables. It is based on covariance; it is the best method of measuring the relationship between two variables. The value of Pearson's Correlation Coefficient ranges from +1 to -1, where +1 indicates a perfect positive correlation and -1 is the opposite of this. The calculation of PCC for two series of the dataset a and b is

$$\rho_{ab} = \frac{\text{cov}(a, b)}{\sigma_a \sigma_b} \quad (1)$$

where $a = (a_1, a_2, a_3, \dots, a_n)$ and $b = (b_1, b_2, b_3, \dots, b_n)$. Here $\text{cov}(a, b)$ indicates the covariance between a and b , and σ_a, σ_b implies the standard deviation of the dataset a and b , respectively. In our work, we calculated the PCC directly by using a user-defined function in MATLAB by (2).

$$\rho_{ab} = \frac{n \sum_{i=1}^n (a_i b_i) - \sum_{i=1}^n a_i \sum_{i=1}^n b_i}{\sqrt{n \sum_{i=1}^n a_i^2 - (\sum_{i=1}^n a_i)^2} \sqrt{n \sum_{i=1}^n b_i^2 - (\sum_{i=1}^n b_i)^2}} \quad (2)$$

2.6. Convolutional neural network model

Convolutional Neural Network is a branch of Deep Neural Networks (DNN), which has proven outstanding image classification and computer vision performance. The main structure of CNN is very similar to the connectivity pattern of the neuron of the human brain. One of the main advantages of CNN is that it requires a little bit of preprocessing or sometimes needs no preprocessing like other traditional classification algorithms. It can automatically learn many features from the training dataset and use them to predict the test data. The first operation of a CNN is making convolution with the images and a filter. In the neural network discussion, the filter and kernel are the same terms. The convolutional layer is the primary layer of any CNN network. This layer

performs matrix multiplication between a particular portion of input images and a specific shape of the kernel.

The kernel moves through an image and makes a new convolved image. Our proposed CNN architecture from the PCC-based images to the classified emotion is shown in Fig. 6. In the architecture, there remain three sets of of

Convolution, ReLU, and Pooling layers. Afterward, we used a flatten layer. Lastly, the dense layer was connected. Here we used a dropout of 0.25 for reducing the network complexity.

After a single convolution layer, the dimension of the output images will be $n_H \times n_W \times n_C$, considering the size of input images, filters, weight, and bias are as similar as expressed in (3), (4), (5), and (6), respectively.

$$\text{size_of_input} = n_H^{[l-1]} \times n_W^{[l-1]} \times n_C^{[l-1]} \quad (3)$$

$$\text{size_of_filter} = f_H^{[l]} \times f_W^{[l]} \times n_C^{[l-1]} \quad (4)$$

$$\text{size_of_weight} = (f_H^{[l]} \times f_W^{[l]} \times n_C^{[l-1]}) \times n_C^{[l]} \quad (5)$$

$$\text{size_of_bias} = 1 \times 1 \times 1 \times n_C^{[l]} \quad (6)$$

The output layer at level l can be calculated as follows.

$$n_H = \frac{n_H^{[l-1]} + 2p^{[l]} - f^{[l]}}{S^{[l]}} + 1 \quad (7)$$

$$\text{size_of_output} = n_H^{[l]} \times n_W^{[l]} \times n_C^{[l]} \quad (8)$$

The calculation of n_w is precisely similar to n_H like (7). Here, $p[l]$ and $S[l]$ represent the padding and stride values, respectively at a level l . The filters of size 3×3 were used in the 1st, 2nd, and 3rd convolution layers. The number of filters is shown in Fig. 6. Since we did not want to lose the information from the pixels of a corner of the input PCC-based image, we used the same convolution (padding layer, $p = 1$) for the 1st and 3rd convolution layers. A valid convolution (i.e. no padding, $p = 0$) method was used for the 2nd convolution layer to reduce computational complexity. The number of padding (p) and stride (s) in every layer is indicated clearly in Fig. 6.

In the classification task for normal images, the most popular type of pooling is max pooling. Since the PCC-featured images are texture-type images, we used average pooling despite using max pooling. It may be

Table 2

The value or types of the major hyper-parameters of our trained CNN model.

Model Hyper-parameter	Value or Type
Learning rate	0.001
Momentum	0.80
Number of epochs	50
Batch size	512
Dropout	0.25
Pooling method	Average pooling
Activation function	ReLU and sigmoid (for protocol 1) ReLU and softmax (for protocol 2)
Optimization algorithm	Adam

noted that a texture-type image does not contain any particular pattern like an ordinary or natural image. The 2×2 sized filters with stride 2 were used in every pooling layer. The major hyper-parameters and their corresponding information, such as the value or the type of them of our trained CNN model, are shown in [Table 2](#).

2.7. Activation functions

In our model, the ReLU activation function is used that best fits CNN architecture and faster. The ReLU has a positive feature compared to ‘sigmoid’ and ‘tanh’ that is never saturated with a significant value of x . Besides, it is more reliable and accelerates the convergence by six times. The mathematical expression of ReLU is given in [\(9\)](#).

$$f(x) = \max(0, x) \quad (9)$$

As our input data consists of values from -1 to $+1$, normalization is not necessary here. We used $L2$ type kernel regularizer to remove the overfitting in dense layers where the regularization parameter is 0.001.

In protocol 1, we made a binary classification. For doing this, we used the ‘sigmoid’ activation function in the last layer. The ‘sigmoid’ is a smooth non-linear activation function that returns the probabilities of a class, and since the probability ranges from 0 to 1, its range is also the same. The equation of the ‘sigmoid’ activation function is given in [\(10\)](#). Protocol 2 is not a binary classification task because we classified valence and arousal into three classes. Therefore, ‘sigmoid’ is not suitable for this purpose. For multiclass classification ‘softmax’ activation function is perfect. The ‘softmax’ activation function returns the probability of every class, and lastly, it targets the class that belongs to the most considerable probability. The ‘softmax’ function is defined in [\(11\)](#).

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (10)$$

$$h_{\theta}(x) = \frac{1}{1 + \exp(-\theta^T x)} \quad (11)$$

Here, h_{θ} is the scalar output of ‘softmax’ in the range of $h_{\theta}(x) \in R$ and $0 < h_{\theta}(x) < 1$. The θ and x are the vectors of weights and input values, respectively.

2.8. Optimization algorithm and cost function

In a neural network, updating the model parameter like weight and bias values needs an optimization algorithm. The optimization algorithm runs a procedure of finding the optimum or satisfactory solution. An optimization algorithm’s primary purpose is to minimize the loss or error of a neural network model. Some popular examples of the optimization algorithm are Stochastic Gradient Descent (SGD), Batch Gradient Descent (BGD), Nadam (Nesterov Adam optimizer), Adagrad (Adaptive Gradient), Adadelta (Adaptive Delta), Adam (Adaptive Moment Estimation), Adamax, RMSprop, etc. The Adaptive Moment Estimation (Adam) gradient descent algorithm was being used to optimize our neural network. The main target of a perfect neural network is not to increase the accuracy but to reduce the loss. The optimization

Table 3

Cost function of the binary and multi-class classification problem.

Type of Problem	Configuration of the Output Layer	Cost Function
Binary Classification	One node with a ‘sigmoid’ activation function.	Binary Cross-Entropy
Multi-Class Classification	One node for each class using the ‘softmax’ activation function.	Categorical Cross-Entropy

Table 4

Shape and parameter values of our CNN model in protocol 1.

Layer (Type)	Output Shape	Parameter
Input	23, 23, 1	–
Conv_1 (Conv2D)	23, 23, 64	640
AvgPool_1 (AveragePooling)	11, 11, 64	0
Conv_2 (Conv2D)	9, 9, 128	73856
AvgPool_2 (AveragePooling)	4, 4, 128	0
Conv_3 (Conv2D)	4, 4, 256	295168
AvgPool_3 (AveragePooling)	2, 2, 256	0
Flatten_1 (Flatten)	1024	0
Dense_1 (Dense)	512	524800
Dropout_1 (Dropout)	512	0
Dense_2 (Dense)	2	1026
Total parameter: 895,490		
Trainable parameter: 895,490		
Non-trainable parameter: 0		

Table 5

Shape and parameter values of our CNN model in protocol 2.

Layer (Type)	Output Shape	Parameter
Input	23, 23, 1	–
Conv_1 (Conv2D)	23, 23, 64	640
AvgPool_1 (AveragePooling)	11, 11, 64	0
Conv_2 (Conv2D)	9, 9, 128	73856
AvgPool_2 (AveragePooling)	4, 4, 128	0
Conv_3 (Conv2D)	4, 4, 256	295168
AvgPool_3 (AveragePooling)	2, 2, 256	0
Flatten_1 (Flatten)	1024	0
Dense_1 (Dense)	512	524800
Dropout_1 (Dropout)	512	0
Dense_2 (Dense)	3	1539
Total parameter: 896,003		
Trainable parameter: 896,003		
Non-trainable parameter: 0		

algorithm always updates the values of weights and biases to reduce the value of the loss.

In protocol 1, where the number of output classes $M = 2$, the cross-entropy was calculated by using [\(12\)](#). Similarly, in the 2nd protocol, as the number of output classes is three, the categorical cross-entropy was calculated by [\(13\)](#). The output layer configuration and relevant cost function are represented in [Table 3](#).

$$\text{cost}_{p=1} = -[y \log(p) + (1-y)\log(1-p)] \quad (12)$$

$$\text{cost}_{p=2} = - \sum_{c=1}^M y_{o,c} \log(P_{o,c}) \quad (13)$$

where ‘ M ’ indicates the number of output classes, ‘ y ’ is the binary indicator (0,1) that defines the correct or incorrect classification, and ‘ P_{oc} ’ implies the predicted probability on observation ‘ o ’ of class ‘ c ’.

We used the ‘Google-Colab’ user interface to execute the code. The summary of the proposed emotion recognition model is shown in [Table 4](#) and [Table 5](#), respectively. It may be noted that the two architectures were almost the same except for the output shape of the last dense layer. As we classified 2 and 3 different classes in protocols 1 and 2, respectively, the last dense layer’s shape was 2 and 3.

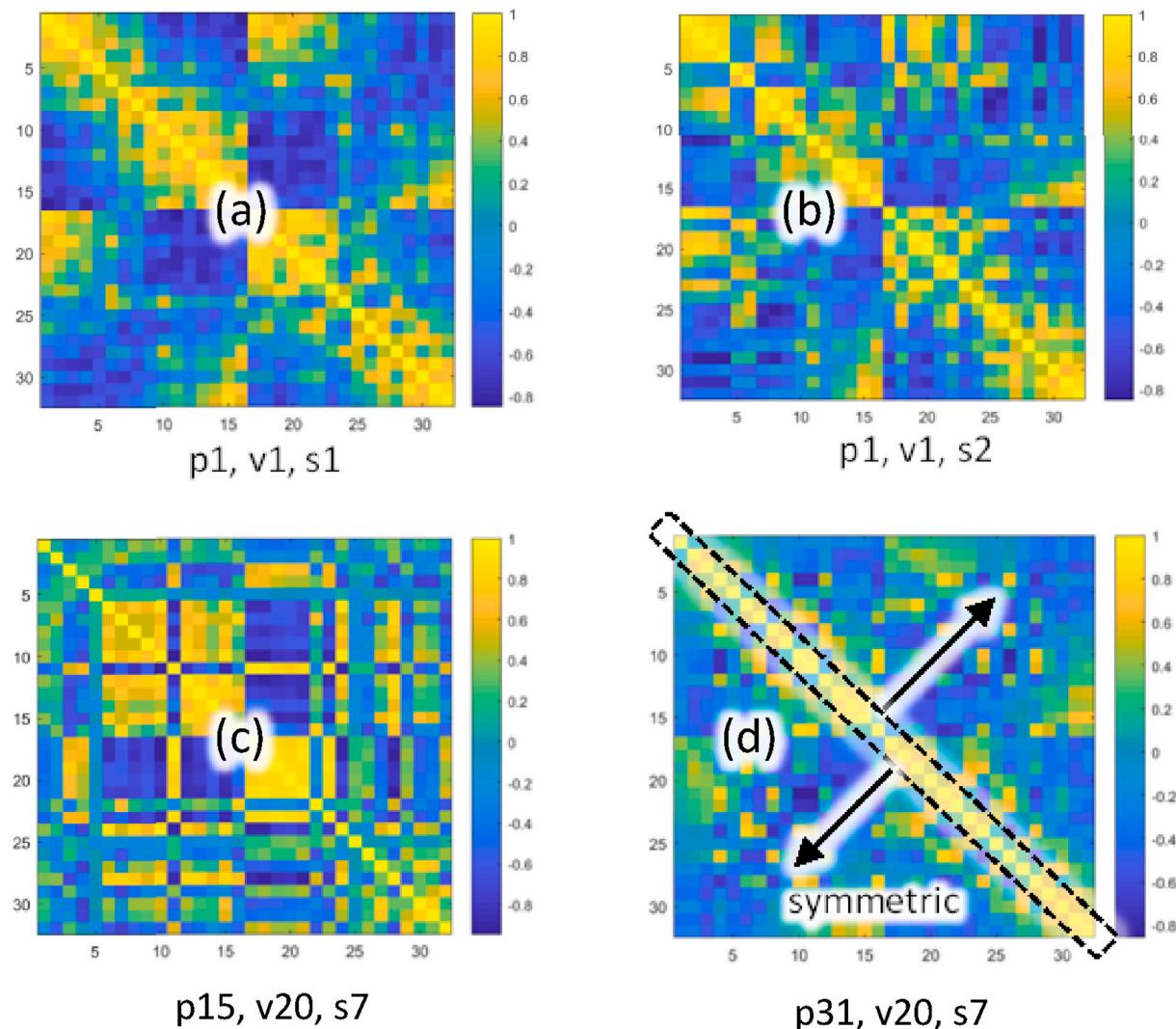


Fig. 7. Random samples of PCC featured images in which every pixel indicates the PCC of the relevant two channels data. (a) Image for participant number 1, video number 1, segmentation number 1 (b) Image for participant number 1, video number 1, segmentation number 2 (c) Image for participant number 15, video number 20, segmentation number 7 (d) Image for participant number 31, video number 20, segmentation number 7.

3. Result

In our experiment, we used the ‘DEAP’ dataset of the EEG signals to classify emotion. The emotion-related EEG signals were firstly converted into PCC featured images. Here, we calculated the correlations of different channels data for the same emotional video and the same segmentation. This channel correlation matrix contains the PCC between the combinations of every two channels for similar segmentation. Thus, the matrix of sized 32×32 could be found for every segmentation. As in every 60s the EEG signals were segmented into 20 segments, and a participant showed 40 different emotional videos; the number of the total square matrices will be $20 \times 40 = 800$. Thus for 32 participants, 800×32 , i.e., 25600 square matrices of channel correlation were generated for a single sub-band. The channel correlation-based images were formulated with the highest correlation coefficient 1 as yellow color and the lowest correlation coefficient -1 as blue color. Some samples of PCC-based images are shown in Fig. 7, where ‘p’ indicates the certain participant, ‘v’ is for video, and ‘s’ represents the segment number. Afterward, these images were fed into a CNN-based classification algorithm. In protocol-1, we distinguish low and high levels of valence and arousal. Here, the levels are distinguished by following the scale as low: $0 \leq \text{value} \leq 4.5$ and high: $4.5 < \text{value} \leq 9$. In protocol-2 low, medium, and high levels of valence and arousal are differentiated by

following the range of low: $0 \leq \text{value} \leq 3$, medium: $3 < \text{value} \leq 6$ and high: $6 < \text{value} \leq 9$.

Alarcao and Fonseca [7] and Zheng and Lu [9] published their investigation that emotion was closely related to the beta and gamma band and moderately associated with the alpha band. Furthermore, the theta band had a slight activity depending on the emotion. On the contrary, as delta band was correlated with dreamless and deep sleep, unconscious mind, and very.

poorly related to emotion. Therefore, we considered the three sub-bands except the delta, theta and main EEG data to extract emotion using CNN.

Since we used a set of 25600 PCC featured images of the main frequency and three sub-band frequencies, there remains a total of $25600 \times 4 = 102,400$ labeled PCC featured images. The data were divided into training (90%), validation (5%), and testing (5%) data. During the training period, the model validated the accuracy of the data selected in the validation portion. The accuracy and loss of 2 classes and 3 classes classification of valence and arousal are shown in Fig. 8 and Fig. 9. From these graphs, it is clear that the minimum loss occurs between the 45 to 50 epochs for both of the two protocols. As in the machine learning approach, the minimization of loss is more desirable than the maximization of accuracy, we used to train the system up to 50 epochs.

The confusion matrices of our model are illustrated in Fig. 10. The

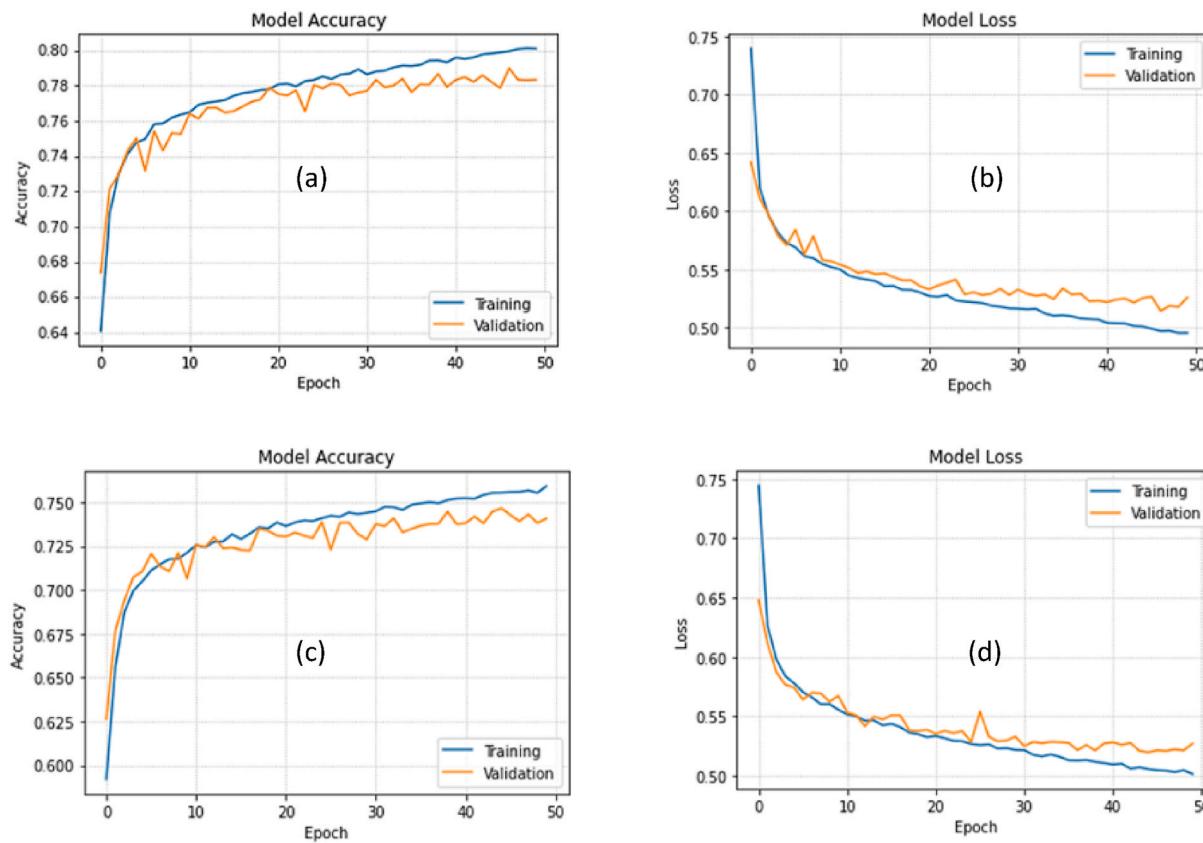


Fig. 8. Model accuracy and loss curves for two-class classification: (a) accuracy in valence, (b) loss in valence, (c) accuracy in arousal, (d) loss in arousal.

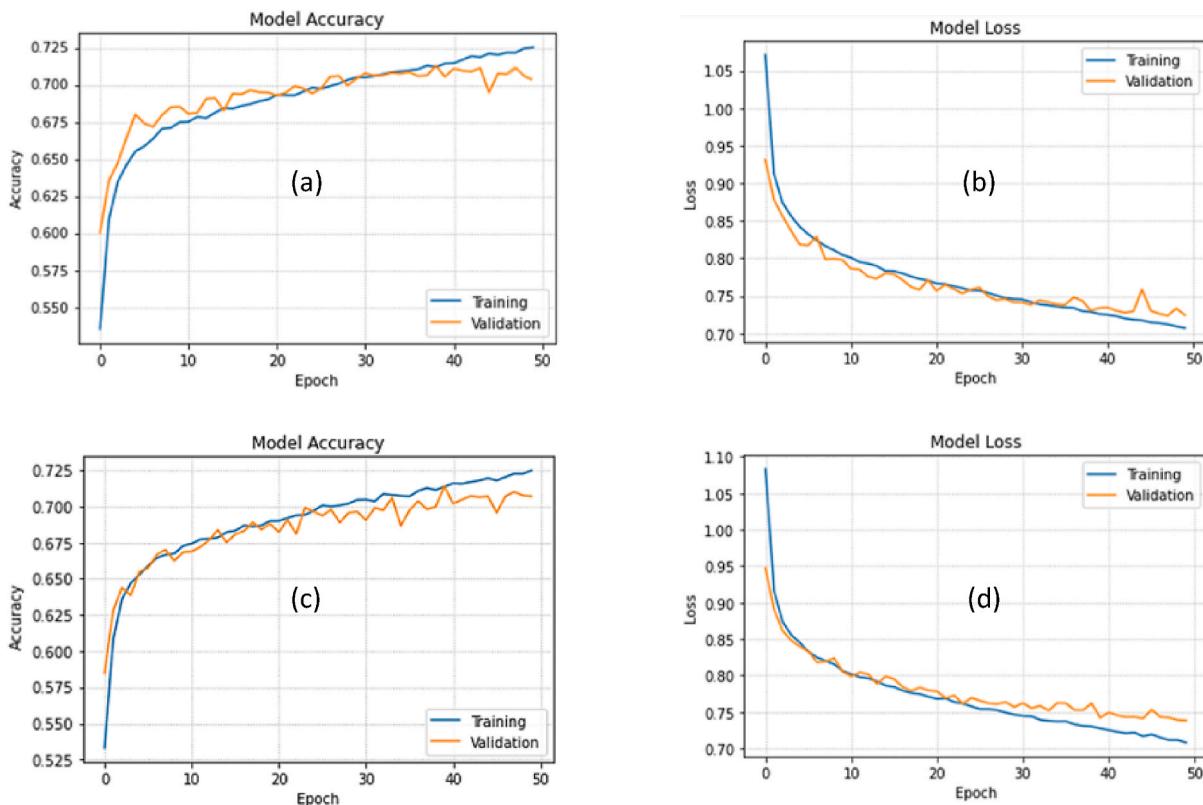


Fig. 9. Model accuracy and loss curves for three-class classification: (a) accuracy in valence, (b) loss in valence, (c) accuracy in arousal, (d) loss in arousal.

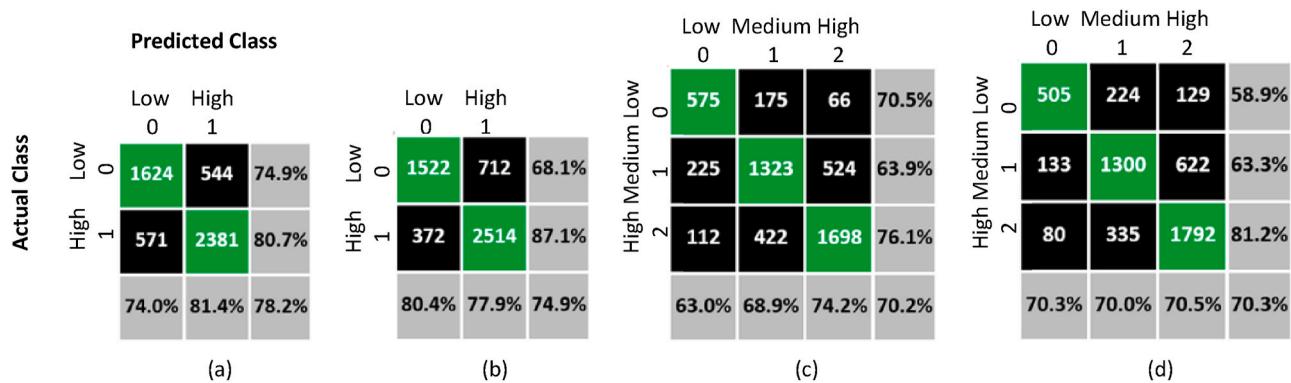


Fig. 10. Confusion Matrices (a) two-class valence (b) two-class arousal (c) three class valence, (d) three class arousal.

Table 6
Classification outcomes of our model.

Valence/Arousal	Class	Precision	Recall	F1 Score	Support
Protocol 1: 0 = Low, 1 = High					
Valence	0	0.74	0.75	0.74	2168
	1	0.81	0.81	0.81	2952
Arousal	0	0.80	0.68	0.74	2234
	1	0.78	0.87	0.82	2886
Protocol 2: 0 = Low, 1 = Medium, 2 = High					
Valence	0	0.63	0.70	0.66	816
	1	0.69	0.64	0.66	2072
	2	0.74	0.76	0.75	2232
Arousal	0	0.70	0.59	0.64	858
	1	0.70	0.63	0.66	2055
	2	0.70	0.81	0.75	2207

four separate confusion matrices are generated for the two-class valence, two-class arousal, three-class valence, and three-class arousal classification task. In every confusion matrix, the last row represents the value of precision, and the last column represents the value of recall. Besides, the last row and the last column data indicates the percentage accuracy of the specific task.

The overall classification report containing the value of precision, recall, F1 score, and accuracy of every class of two different protocols is given in Table 6. The high value of the F1 score indicates that the importance of precision and recall were balanced and satisfactory for both of the two protocols.

Here, only the inter-channel correlations were considered. The inter-channel correlations were more meaningful as the human brain remains aroused by any single emotional stimuli in only a few seconds (1s–4s). On the other hand, the inter-sample correlations may be another option for future work. It was important as it allowed one to consider the variation of temporal characteristics of EEG signals for different emotional states. In addition, the inter-sample correlation technique will enlarge the volume of data that promotes the deep machine learning algorithm performance for large-sized data.

4. Discussion

To recognize emotion firstly, the channel correlation-based PCC featured images were generated, as stated earlier. The PCC-featured images of size 32×32 are consisted by the diagonal symmetrical matrices that were showed clearly in Fig. 7. Since the diagonal elements represented the correlation between the same channels, the value will always be 1 for all of the PCC matrices. As a result, in every image, the color of the diagonal element was yellow.

Since the upper or lower triangular matrix data of a single PCC-based image were the same, it was unnecessary to use the same data in double time. Therefore, before sending the PCC matrices to our CNN model, the

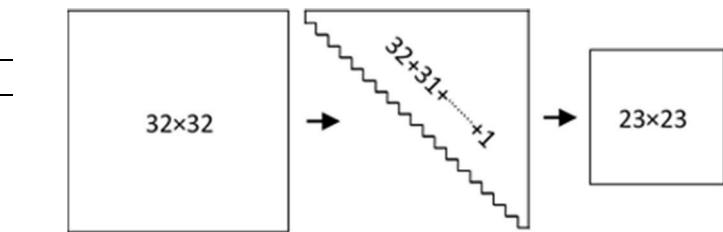


Fig. 11. Complexity reduction considering only the upper triangular matrix.

Table 7

Computational Complexity Reduction based on Input Size, Parameters, Execution Time in New Model.

Factors for Reducing Complexity	Input Size	Number of Total Trainable Parameters	Execution Time (for 50 Epochs)
Protocol-1			
Old Model	$32 \times 32 = 1024$	1,550,850	351s
New Model	$23 \times 23 = 529$	895,490	207s
Complexity Reduction (%)	48.34%	42.59%	41.03%
Protocol-2			
Old Model	$32 \times 32 = 1024$	1,551,363	302s
New Model	$23 \times 23 = 529$	896,003	101s
Complexity Reduction (%)	48.34%	42.24%	66.56%

lower triangular data were removed. Thus, the data points were reduced from 1024 (32×32) to 528 ($1 + 2 + \dots + 32$) as explained in Fig. 11. An extra 0 data was added with this 528 data points to formulate a new shape PCC featured image of size 23×23 . Eventually, only 51.1% of data points of the square matrix of PCC-based images were fed to the CNN model because these points consist of significant information for emotion recognition. Using this technique, the size of memory and the computational complexity had been reduced without degrading the emotion recognition accuracy.

The data points of input images were reduced from 1024 to 529, indicating a 48.34% memory reduction only in this step of execution. The reduction of complexity in protocols 1 and 2 has been given in Table 7 considering some specific factors. In protocol-1, the total trainable parameter was reduced to 895,490 from 1,550,850, which tends to a 42.59% reduction in computation for the whole program. We used the GPU of ‘Google Colab’ with a 5 Mbps internet connection through the laptop of configuration Intel(R) Core(TM) i5-8250U CPU, 1.60 GHz, 7.88 GB DDR4 RAM, 256 GB SSD, 64-bit operating system for

Table 8

Comparison of different emotion recognition methods.

Author	Extracted Features	Classifier	Number of Classified emotion	Accuracy (%)	Dataset
Koelstra et al. (2011) [34]	Multimedia Content Analysis (MCA)	Gaussian Naive Bayes classifier	4 (HVHA, HVLA, LVHA, LVLA)	Valence = 57.6 Arousal = 62.0	DEAP
Jirayucharoensak et al. (2014) [35]	Power Spectral Density (PSD)	Deep learning network with SAE	3 class	Valence = 49.5 Arousal = 46.0	DEAP
Ackermann et al. (2016) [4]	Statistical feature	SVM and Random Forest	3 (anger, surprise, etc.)	Average Accuracy = 55	DEAP
Tripathi et al. (2017) [31]	Statistical features in time and frequency domain	Convolutional Neural Networks (CNN)	2 (low, high) 3 (low, medium high)	Valence = 81.4 Arousal = 73.4 Valence = 66.7 Arousal = 57.6	DEAP
Song et al. (2018) [36]	PSD with graph	Dynamical Graph Convolutional Neural Networks (DGCNN)	2 (low, high)	Valence= (86.2 ± 12.3) Arousal= (84.5 ± 10.8) Dominance= (85 ± 10.3)	DREAMER
Cheng et al. (2019) [37]	2D frame sequence considering spatial position	Deep Forest	2 (low, high)	Valence= (97.69 ± 1.22) Arousal= (97.53 ± 1.52) Valence= (89.03 ± 5.56) Arousal= (90.41 ± 5.33) Dominance= (89.89 ± 6.19)	DREAMER
Fang et al. (2020) [38]	Power Spectral Density (PSD) and Differential Entropy (DE)	Multi-Feature Deep Forest (MFDF)	5 (angry, happy, sad, pleasant, and neutral)	Overall accuracy = 71.05	DEAP
OurMethod,P1	Pearson's Correlation Coefficient (PCC)	Convolutional Neural Networks (CNN)	2 (low, high)	Valence = 78.22 Arousal = 74.92	DEAP
OurMethod,P2			3 (low, medium high)	Valence = 70.23 Arousal = 70.25	

executing our programming code. The new model also needed less time for execution by using the same processor and same internet speed. The total time for 50 epochs was reduced to 207s from 351s. Similarly, in protocol-2, (see Table 7); the new model shows the 48.34%, 42.24%, 66.56% reduction of complexity in the input size, number of total parameters and execution time respectively.

The overall accuracy of our proposed model is shown in Table 8. In Table 8, it is represented that the accuracy for two-class valence and arousal recognition tasks is 78.22% and 74.92%, respectively. In the three-class classification, the accuracy is 70.23% and 70.25% for valence and arousal recognition tasks. Another important fact that the percentage accuracy of protocol-1 is comparatively higher than the accuracy.

of protocol-2. The cause is that whenever the number of classes to be categorized increases, then the loss will also increase. As a result, the accuracy will decrease. From the comparison table of work like ours, the accuracy of our proposed method is indeed satisfactory. The accuracy is not the highest, then how can it be acceptable? Our target is not to design a model of maximum accuracy, the target was to develop a model of lower computational complexity and lower memory and time required. For instance, the model of Tripathi et al. [31] works with lots of features, whereas we used only one feature named PCC. Besides, the accuracy in just 2 class valence is only higher; on the contrary, the remaining accuracy is lower than our method. The model of Song et al. [36] is very effective, but as the dataset is not the same as ours, the proper comparison is not too meaningful here. Recently, some deep learning-based models, such as the deep forest-based model [37,38] and rhythm-specific deep learning-based model [39] perform very well in the context of emotion recognition. Cheng et al. [37] used a deep forest algorithm on the 2D frame sequence considering the spatial position across channels. They applied it on DEAP and DREAMER datasets and achieved top-ranked results in terms of accuracy stated in Table 8. Fang et al. [38] used Multi-Feature Deep Forest as a classifier on PSD and DE

features and classified emotion into five classes. The model was tested using the DEAP dataset and achieved 71.05% overall accuracy, which is the highest for many classes classification. However, we focused on developing a model of lower computational complexity with a satisfactory level of classification accuracy. Our work's novelty is to use the technique of using only the upper triangular matrix data that reduces the size of input PCC-featured images from 32×32 to 23×23 . It drastically reduces the computational complexity and time of training and operation. Moreover, we used only the PCC featured images with the CNN model; except for this feature, no additional effort was required to find many features. It reduces the trouble to extract features manually.

5. Conclusion

In this paper, we have used the Convolutional Neural Network model to recognize emotion from EEG signals. Despite using raw EEG data a systematically developed PCC featured images were considered for lower computation complexity and short operation time. The whole process of emotion recognition was completed by following the rule of classification using logistic regression. Here, two different protocols were used; in both protocols, we observed the accuracy, model loss, and classification report. We achieved 78.22% and 70.23% accuracy in the valence classification task and 74.92% and 70.25% accuracy in the arousal classification task for two and three classes classification respectively. More accuracy and real-time operation compatibility are the fundamental prerequisites to enjoy the advantages of emotion recognition in the practical field. Consequently, more research and studies are essential in channel reduction, significant feature extraction, and deep network optimization in the future.

Declaration of competing interest

The authors have no conflict of interest.

References

- [1] P. Romaniszyn-Kania, et al., Affective state during physiotherapy and its analysis using machine learning methods, *Sensors* 21 (14) (2021) 4853, <https://doi.org/10.3390/s21144853>.
- [2] D. Matherl, L.M. Williams, P.J. Hopkinson, A.H. Kemp, Investigating models of affect: relationships among EEG alpha asymmetry, depression, and anxiety, *Emotion* 8 (4) (2008) 560.
- [3] D. Sammler, M. Grigutsch, T. Fritz, S. Koelsch, Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music, *Psychophysiology* 44 (2) (2007) 293–304.
- [4] P. Ackermann, C. Kohlschein, J.A. Bitsch, K. Wehrle, S. Jeschke, EEG-based automatic emotion recognition: feature extraction, selection and classification methods, in: 2016 IEEE 18th International Conference on E-Health Networking, Applications and Services, Healthcom, 2016, pp. 1–6, <https://doi.org/10.1109/HealthCom.2016.7749447>, 2016.
- [5] M.R. Islam, et al., Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques, *IEEE Access* 9 (2021) 94601–94624, <https://doi.org/10.1109/access.2021.3091487>.
- [6] M.M. Rahman, et al., Recognition of human emotions using EEG signals: a review, *Comput. Biol. Med.* (2021) 104696, <https://doi.org/10.1016/j.combiomed.2021.104696>.
- [7] S.M. Alarcão, M.J. Fonseca, Emotions recognition using EEG signals: a survey, *IEEE Trans. Affect. Comput.* 10 (3) (2019) 374–393, <https://doi.org/10.1109/TAFFC.2017.2714671>.
- [8] M.R. Islam, M. Ahmad, Wavelet Analysis Based Classification of Emotion from EEG Signal, 2019, <https://doi.org/10.1109/ECACE.2019.8679156>.
- [9] W.L. Zheng, B.L. Lu, Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks, *IEEE Trans. Auton. Ment. Dev.* 7 (3) (2015) 162–175, <https://doi.org/10.1109/TAMD.2015.2431497>.
- [10] E. Barsoum, C. Zhang, C.C. Ferrer, Z. Zhang, Training deep networks for facial expression recognition with crowd-sourced label distribution, in: Proceedings of the 18th ACM International Conference on Multimodal Interaction, 2016, pp. 279–283.
- [11] M. Soleimani, S. Asghari-Esfeden, Y. Fu, M. Pantic, Analysis of EEG signals and facial expressions for continuous emotion detection, *IEEE Trans. Affect. Comput.* 7 (1) (2016) 17–28, <https://doi.org/10.1109/TAFFC.2015.2436926>.
- [12] S.R. Kadiri, P. Gangamohan, S.V. Gangashetty, B. Yegnanarayana, Analysis of excitation source features of speech for emotion recognition, in: Proceedings of the Annual Conference of the International Speech Communication Association, 2015–January, INTERSPEECH, 2015, pp. 1324–1328.
- [13] M.T. Valderas, J. Bolea, P. Laguna, M. Vallverdú, R. Bailón, Human emotion recognition using heart rate variability analysis with spectral bands based on respiration, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2015, pp. 6134–6137.
- [14] S. Piana, A. Stagliano, F. Odone, A. Camurri, Adaptive body gesture representation for automatic emotion recognition, *ACM Trans. Interact. Intell. Syst.* 6 (1) (2016) 1–31, <https://doi.org/10.1145/2818740>.
- [15] A. Vaskinn, K. Sundet, T. Østefjells, K. Nymo, I. Melle, T. Ueland, Reading emotions from body movement: a generalized impairment in schizophrenia, *Front. Psychol.* 6 (JAN) (2016) 2058, <https://doi.org/10.3389/fpsyg.2015.02058>.
- [16] J. Shukla, M. Barreda-Angeles, J. Oliver, G.C. Nandi, D. Puig, Feature extraction and selection for emotion recognition from electrodermal activity, *IEEE Trans. Affect. Comput.* 5 (3) (2019) 327–339, <https://doi.org/10.1109/TAFFC.2019.2901673>.
- [17] K. Schaff, T. Schultz, Towards emotion recognition from electroencephalographic signals, in: Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII, 2009, pp. 1–6, <https://doi.org/10.1109/ACII.2009.5349316>, 2009.
- [18] Z. Mohammadi, J. Frounchi, M. Amiri, Wavelet-based emotion recognition system using EEG signal, *Neural Comput. Appl.* 28 (8) (2017) 1985–1990, <https://doi.org/10.1007/s00521-015-2149-8>.
- [19] S.K. Hadjidimitriou, L.J. Hadjileontiadis, EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings, *IEEE Trans. Affect. Comput.* 4 (2) (2013) 161–172.
- [20] M. Li, B.-L. Lu, Emotion classification based on gamma-band EEG, in: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 1223–1226.
- [21] L. He, D. Hu, M. Wan, Y. Wen, K.M. Von Deneen, M.C. Zhou, Common Bayesian network for classification of EEG-based multiclass motor imagery BCI, *IEEE Trans. Syst. Man, Cybern. Syst.* 46 (6) (2016) 843–854, <https://doi.org/10.1109/TSMC.2015.2450680>.
- [22] Y. Velchev, S. Radeva, S. Sokolov, D. Radev, Automated estimation of human emotion from EEG using statistical features and SVM, in: 2016 Digital Media Industry & Academic Forum, DMIAF, 2016, pp. 40–42.
- [23] M. Chen, J. Han, L. Guo, J. Wang, I. Patras, Identifying valence and arousal levels via connectivity between EEG channels, in: 2015 International Conference on Affective Computing and Intelligent Interaction, ACII, 2015, pp. 63–69, <https://doi.org/10.1109/ACII.2015.7344552>, 2015.
- [24] M. Soleimani, M. Pantic, T. Pun, Multimodal emotion recognition in response to videos (Extended abstract), in: 2015 Int. Conf. Affect. Comput. Intell. Interact. vol. 3, ACII, 2015, pp. 491–497, <https://doi.org/10.1109/ACII.2015.7344615>, 2015.
- [25] W.L. Zheng, J.Y. Zhu, B.L. Lu, Identifying stable patterns over time for emotion recognition from eeg, *IEEE Trans. Affect. Comput.* 10 (2019) 417–429, <https://doi.org/10.1109/TAFFC.2017.2712143>, 3.
- [26] J. Atkinson, D. Campos, Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers, *Expert Syst. Appl.* 47 (2016) 35–41, <https://doi.org/10.1016/j.eswa.2015.10.049>.
- [27] M. Långkvist, L. Karlsson, A. Loutfi, Sleep stage classification using unsupervised feature learning, *Adv. Artif. Neural Syst.* 2012 (2012).
- [28] K. Li, X. Li, Y. Zhang, A. Zhang, Affective state recognition from EEG with deep belief networks, in: 2013 IEEE International Conference on Bioinformatics and Biomedicine, 2013, pp. 305–310.
- [29] H.P. Martinez, Y. Bengio, G. Yannakakis, Learning deep physiological models of affect, *IEEE Comput. Intell. Mag.* 8 (2) (2013) 20–33, <https://doi.org/10.1109/MCI.2013.2247823>.
- [30] Z. Wen, R. Xu, J. Du, “A Novel Convolutional Neural Networks for Emotion Recognition Based on EEG Signal,” in 2017 International Conference on Security, Pattern Analysis, and Cybernetics, SPAC, 2017, pp. 672–677.
- [31] S. Tripathi, S. Acharya, R.D. Sharma, S. Mittal, S. Bhattacharya, Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 4746–4752 [Online]. Available: <https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/15007/13731>.
- [32] H. Mei, X. Xu, EEG-based emotion classification using convolutional neural network, in: 2017 International Conference on Security, Pattern Analysis, and Cybernetics, SPAC, 2017, pp. 130–135.
- [33] M.R. Islam, M. Ahmad, Virtual Image from EEG to Recognize Appropriate Emotion Using Convolutional Neural Network, 2019, <https://doi.org/10.1109/ICASERT.2019.8934760>.
- [34] S. Koelstra, et al., DEAP: a database for emotion analysis; Using physiological signals, *IEEE Trans. Affect. Comput.* 3 (1) (2012) 18–31, <https://doi.org/10.1109/T-AFFC.2011.15>.
- [35] S. Jiraycharoensak, S. Pan-Ngum, P. Israsena, EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation, *Sci. World J.* 2014 (2014), <https://doi.org/10.1155/2014/627892>.
- [36] T. Song, W. Zheng, P. Song, Z. Cui, EEG emotion recognition using dynamical graph convolutional neural networks, *IEEE Trans. Affect. Comput.* 11 (3) (2018) 532–541.
- [37] J. Cheng, et al., Emotion recognition from multi-channel EEG via deep forest, *IEEE J. Biomed. Heal. Informatics* 25 (2) (2021) 453–464, <https://doi.org/10.1109/JBHI.2020.2995767>.
- [38] Y. Fang, H. Yang, X. Zhang, H. Liu, B. Tao, Multi-feature input deep forest for EEG-based emotion recognition, *Front. Neurorob.* 14 (2021) 617531, <https://doi.org/10.3389/fnbot.2020.617531>.
- [39] D. Maheshwari, S.K. Ghosh, R.K. Tripathy, M. Sharma, U.R. Acharya, Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel EEG signals, *Comput. Biol. Med.* 134 (2021) 104428.