Time series-dependent feature of EEG signals for improved visually evoked emotion classification using EmotionCapsNet

Nandini Kumari[1] · **Shamama Anwar[1]** · **Vandana Bhattacharjee[1]**

Abstract

In recent studies, machine learning and deep learning strategies have been explored in many EEG-based application for best performance. More specifically, convolutional neural networks (CNNs) have demonstrated incredible capacity in electroencephalograph (EEG)-evoked emotion classification tasks. In preexisting case, CNN-based emotion classification techniques using EEG signals mostly involve a moderately intricate phase of feature extrication before any network model implementation. The CNNs are not able to well describe the natural interrelation among the various EEG channels, which basically provides essential data for the classification of different emotion states. In this paper, an efficacious and advanced version of CNN called Emotion-based Capsule Network (EmotionCapsNet) for multi-channel EEG-based emotion classification to achieve better classification accuracy is presented. EmotionCapsNet has been applied to the raw EEG signals as well as 2D image representation generated from EEG signals which can extricate descriptive and complex features from the EEG signals and decide the different emotional states. The proposed system is then compared with the other conventional machine learning and deep learning-based CNN model. Our strategy accomplishes an average accuracy of 77.50%, 78.44% and 79.38% for valence, arousal and dominance on the DEAP, 79.06%, 78.90% and 79.69% on AMIGOS and attains an average accuracy of 80.34%, 83.04% and 82.50% for valence, arousal and dominance on the DREAMER, respectively. These outcomes demonstrate that adapted strategy yields comparable precision on raw EEG signal and it also provides better classification results on spatiotemporal feature of EEG signal for emotion classification task.

Keywords Time series data · Spatiotemporal · Spectrogram · Electroencephalograph (EEG) · EmotionCapsNet

1 Introduction

Emotions play a significant role in human social interaction and are acquired by studying the psychological condition of an individual which includes the cognition, behavior and decision-making capability [1]. Human emotional states can be unobtrusively impacted by multiple external and mental factors and are a combination of time, space, experience, environment and cultural background. This diversity gives rise to the main challenges faced by researches for emotion recognition (ER). Emotion is categorized by two essential models, i.e., the dimensional and discrete models [2]. Happy, anger, sad, fear, disgust and surprise are six essential emotions that are indicated by discrete model to portray emotion [3]. Howbeit, the most popular dimensional hypotheses is the 3D space: valence, arousal and dominance (VAD). Valence, in particular, alludes to the level of a person's joy from optimist to pessimist whereas, arousal maps the degree of energy from inactive to dynamic and dominance goes from a powerless and feeble inclination to an enabled feeling.

Hence, EEG signals are broadly utilized for ER model by exploring data from the different frequency bands, electrode position, temporal resolution and accomplished agreeable outcomes. However, some existing work used combination of physiological signals which provided prominent accuracy for emotion classification. In this

✉ Nandini Kumari
phdcs10009.18@bitmesra.ac.in

Shamama Anwar
shamama@bitmesra.ac.in

Vandana Bhattacharjee
vbhattacharya@bitmesra.ac.in

[1] Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, Ranchi 835215, India

paper, we ought to achieve comparable classification accuracy using EEG signal only (single modality) and compare this with the outcomes of the techniques using all the other signals like GSR, ECG, EOG, etc. (multiple modality). EEG diagnostic tests essentially assist clinicians to identify medical complications and determine the most effective therapy to reduce long-term repercussions. There are other applications that track viewers' emotions based on their emotional responses to videos clip [4]. EEG-based ER model can also be used to help children with autism spectrum disorder (ASD) with their social skills [5]. Now it is possible to capture brainwave patterns and evaluate a person's mental state while wearing a wireless EEG headset. The emotions of a person are identified and shown on his/her avatar in real time, giving human computer interfaces an additional "emotion dimension" [6]. An EEG-enabled music therapy facility has also been developed that plays music to the patients to help them cope with things like pain and sadness. A web-enabled music player with an EEG-based system that displays the music based on the user's current emotional states was also developed and implemented [7].

The previously adapted models for EEG-based emotion classification incorporates convolutional neural networks (CNNs) [8, 9], deep belief networks (DBNs) [10], RNN and long short-term memory (LSTM) [11], capsule network (CapsNet) [12] and so on. These methods have been also implemented to computer vision, remote sensing, semantic segmentation, robotics-related task and other applications as in [13, 14]. Among the diverse deep learning models, CNN shows the most precise and promising outcomes for classification tasks. The existing work [15, 16] on CNN-based visual evoked stimuli evoked EEG signal classification and its prominent accuracy has encouraged us to attempt an emotion classification task using CNN. Therefore, we have implemented CNN architecture on raw EEG signals to achieve better accuracy than conventional machine learning methods.

Although, we figured out two challenges; firstly, major consideration is given on deciding how to acquire the remarkable data related to emotion states from the temporal and frequential domain, and secondly, the time–frequency qualities of EEG signals. A few investigations have dissected the spatial space attributes of multi-channel EEG, that consist of notable data. Further, a couple of spatial attributes which are restricted to the deviation between the electrode pair sets [17]. In this manner, deciding how to incorporate and represent the regular attributes of the EEG signal by virtue of the spatial qualities is a crucial issue. Moreover, numerous deep learning techniques have less affect-ability to spatial features while handling 2D objects, for example, the Stacked AutoEncoder (SAE) and Deep Belief Network (DBN). Despite the fact that CNNs can

deal with 2D objects, they do not have the capacity to depict the general connection between various local features and the complete object, that can deliver applicable data for classification. This paper also proposes an EEG-based ER framework to deal with the challenges mentioned above that converts 1D EEG signal to 2D representation using STFT algorithm which represents spatiotemporal information of EEG signal. Further, a capsule network (CapsNet) is implemented which can deal excellently with the spatiotemporal features of EEG signals. As compared to CNNs, the CapsNet has a lot of potential for recognizing the spatial relationship between confined features in the spatial space, which can help to improve the overall accuracy of system. The overall proposed capsule-based emotion detection architecture EmotionCapsNet is shown in Fig. 1 and implemented using three different emotion datasets namely DEAP (dataset for emotion analysis using physiological and video signals), DREAMER and AMIGOS to accomplish our task.

The layout of the paper is as follows: Sect. 2 presents the related work in traditional and deep learning techniques applied in EEG-based ER. Section 3 presents the brief description of dataset, experimentation and architecture of proposed model on ER. Section 4 reports the exploratory accuracy results assessed on three popular emotion datasets and provides a discussion on comparative analysis of proposed model with conventional methods. Finally, Sect. 5 provides an overall conclusion of the work presented in this paper.

2 Related work

Emotion recognition is an exceptionally vital research field, with numerous novel approaches being proposed and researched over the past decade. With the availability of EEG dataset and great computational power, deep learning strategies are quickly turning out to be efficacious. EEG signal has been one of the most favored physiological signal structures in ER tasks, which provides a good chance for exploration in this field. The common features essentially include the three classes, namely: time-domain features, frequency domain features and time–frequency domain features. Time-domain features extract and represents the temporal features through EEG signals, for example, differential entropy (DE), power spectral density (PSD), the rational asymmetry (RASM) and so on. Different machine learning-based strategies have been utilized as classifiers for EEG-based ER model with acceptable model's performance, for example, linear discriminant analysis (LDA) [18], support vector machine (SVM) [19], random forest(RF), k-nearest neighbors (k-NN) [20], naive bayes (NB), neural networks [21, 22], etc. For comparison
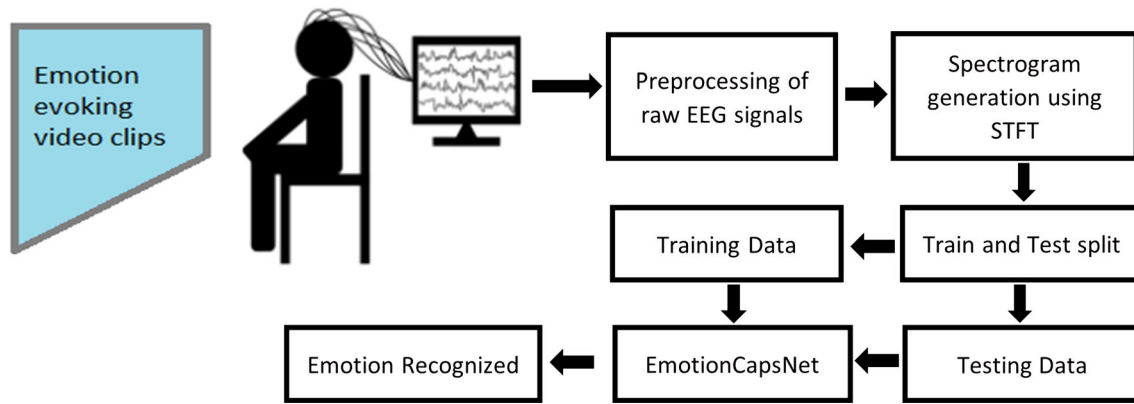
Fig. 1 The overall workflow of proposed system for EEG-based ER

purpose, we implemented these traditional machine learning techniques on raw EEG data and calculated the classification accuracy which motivated us for further improvement on results. Various deep learning (DL)-based models have been applied in classification of EEG signals to solve different types of classification problems.

CNN architecture has accomplished unrivaled achievement in computer vision tasks, for example, classification, object recognition and semantic division. Howbeit, CNN actually has a deficiency induced by pooling activity. CNN cannot encode the orientation and position of learned features from images. In the pooling layer, only the most effective neurons are selected to be shifted to the next layer resulting in a significant number of information lost. To beat the weaknesses of CNN, an advance DL-based model called the CapsNet has been proposed [23].

Considering the previously acknowledged preferences, the CapsNet has been implemented in numerous expanse in the previous years, for example, natural language processing [24], clinical image segmentation and speech recognition [25], and has outperformed the traditional methods. The emotion-based model has been proposed using CapsNet, which is equipped for performing viewpoint location and emotion classification at the same time [12]. Some researchers used the CapsNet to detect the brain tumor using Magnetic Resonance Imaging (MRI). In [26], CapsNet-based model was proposed with three convolutional layers to classify the spectral-based images. Turan et al. converted sound signals to spectrogram images and afterward utilized the CapsNet to perceive a newborn child's cry [25]. CNN-based EEG emotional recognition algorithms often require a very sophisticated pre-extraction stage. The integral relationship between multiple channels is a significant piece of evidence for characterizing emotional states in multi-channel EEG-based ER. However, CNN models may overlook certain inextricable relationship between multiple channels of EEG signals and cannot accurately differentiate the spatial relationship between

various features. Chao et al. [12] proposed a multiband feature matrix (MFM) and a CapsNet-based deep learning framework for ER to deal with intrinsic relationship of emotion states. The experimental results show that the three MFM characteristics were complementing, and that the capsule network was superior, but performance was not up to mark. Liu et al. [27] proposed a multi-level emotion guided capsule networks (MLF-CapsNet) to increase accuracy percentage in recognition tasks. The MLF-CapsNet framework can also be used at the same time for feature extraction from raw EEG signals and then emotional states are determined with competently classification accuracy. Ali et al. [28] have implemented capsule network for ER and achieved good accuracy for each subject-dependent EEG data. Howbeit, these approaches are applied only on subject-dependent ER and do not deal with subject-independent ER tasks. Subject-dependent technique to predicting emotion from a specific subject (which collects and only develops from that subject's emotional input), is ideal. It is required to create a subject-independent model that can apply prior knowledge from a wide variety of subjects to a new one. Almost all extracted attributes have a beneficial impact on the new subject's capacity to recognize emotions. The tremendous performance of CapsNet encouraged the researchers to implement this emerging architecture to solve various problems. The aim of this paper is to propose a novel CapsNet-based architecture, named emotion-based capsule network (EmotionCapsNet), for subject-independent EEG-based ER. The following are two main contribution of this paper:

1. This paper exquisitely deals with the issues that shallow networks have frail ability to manage, where 2D convolution requires more data of temporal dependency and they could not deal with spatial information of EEG signals. To tackle this problem STFT (short-term Fourier transform) algorithm is used to transform raw 1D EEG signals to 2D spectrogram

images which retains the spatiotemporal information of EEG signals of each electrode position.

2. An enhanced version of CNN and a modified version of the original capsule network namely EmotionCapsNet model is proposed to manage relevant data and spatial connections for ER, which not only classifies the temporal and frequential information but also classifies the spatial features of EEG signals. This technique ensures better accuracy compared with the other techniques for EEG-based ER on the DEAP, DREAMER and AMIGOS datasets.

3 Methodology

3.1 Dataset description

We implemented and evaluated our structure on three openly accessible multi-modular datasets in particular DEAP [29], AMIGOS [30] and DREAMER [31]. Table 1 represents the three datasets, with an emphasis on the modalities that we utilized in this work. The trials in the DEAP and AMIGOS have been labeled on a scale of 1–9 for VAD by test subjects. For DREAMER, the VAD has been labeled on a discrete scale utilizing numbers from 1 to 5, separately. For each dataset, the corresponding labels were reported by each subject individually after visualizing each trial video clip as visual stimuli. The emotion has been classified into 2 classes for High-Valence and Low-Valence, High-Arousal and Low-Arousal, High-Dominance and Low-Dominance as we considered a rating point that is greater than 4 for DEAP and AMIGOS and greater than 3 for DREAMER dataset has been set for high valance, arousal and dominance (labeled as 1) otherwise will be considered for low valance, arousal and dominance (labeled as 0).

Previous work on EEG-based ER using single-subject techniques has long been criticized for their lack of global recognition. The heterogeneity in specific neural activities within the same emotion response is the first obstacle in emotion detection utilizing brain signals. To put it another way, directly transmitting knowledge about the neurological system from one subject to another results in incompatibility or poor generalization. For example, if they had 1000 subjects, they would have 1000 particular models in their system, resulting in enormous processing overhead which may be suboptimal for practical systems. Therefore, in this research experimentation, a subject-independent strategy has been chosen for generalization of proposed EmotionCapsNet that applies prior knowledge from a wide variety of subjects to a new one. Almost all raw EEG data samples of different electrode positions w.r.t different subjects have been merged trail-wise that have a beneficial impact on recognizing new subject's emotion states. In this manner, the total of $1280 \times 258,048$, i.e., $1280 = 40 \times 32$ (40 trials of each of 32 subjects) and $258,048 = 32 \times 8064$ (8064 samples of each of 32 electrode channels) raw EEG data points for DEAP has been created in subject-independent manner. Similarly, a total of 640×7168 and 414×7168 for AMIGOS and DREAMER dataset, respectively, were re-sampled where 640 in AMIGOS represents 16 trial $\times$ 40 subjects and 414 in Dreamer represents 18 trails $\times$ 23 subjects with 7168 data points, i.e., 14 channels $\times$ 512 samples of each channel. These raw EEG signals are taken as input for the proposed system.

3.2 Preprocessing

These EEG signals can be decomposed into delta (1–4 Hz), theta (4–7 Hz), alpha (8–15 Hz), beta (16–31 Hz) and gamma (> 32 Hz) bands, as per their frequency band powers. The range of EEG frequency band to be used peculiarly depends on the motivation of research. As per

Table 1 Dataset description based on the single modality (EEG)

DEAP dataset	AMIGOS dataset	DREAMER dataset
32 subjects	40 subjects	23 subjects
$40 \times 32 \times 32$ (trial / video$\times$channel$\times$subject)	$16 \times 14 \times 40$ (trial/video $\times$channel$\times$ subject)	$18 \times 14 \times 23$ (trial / video$\times$channel$\times$subject)
32-channel EEG system (Two different raw and preprocessed data available EEG systems used. Channel locations: Fp1, AF3, F7, F3, FC1, FC5, T7, C3, CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, P8, CP6, CP2, C4, T8, FC6, FC2, F4, F8, AF4, Fp2, Fz, Cz)	Raw and preprocessed data available 14-channel EEG system (A single EEG system used for all subjects. Channel locations: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4)	Raw data available 14-channel EEG system (A single EEG system used for all subjects. Channel locations: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4)
Sampling rate 128 Hz	Sampling rate 128 Hz	Sampling rate EEG/ECG: 128/256 Hz
Valence/Arousal/Dominance rated using a continuous scale between 1 and 9	Valence/Arousal/Dominance rated using a continuous scale between 1 and 9	Valence/Arousal/Dominance rated using a discrete scale of integers from 1 to 5

our literature survey, many previous researches have utilized 4–45 Hz band power for detection of emotions and they perceived good accuracy with these frequency band of EEG signals [32]. Therefore, in this paper, the recorded raw EEG signals from all datasets were preprocessed with a band-pass filter of 4–45 Hz and a notch filter of 50 Hz to eliminate physiological artifacts, power band noises and for interference noise cancellation [33, 34].

3.3 Spatial encoding of EEG signals into spectrogram

In this study, the STFT algorithm, which interprets temporal signals into spatiotemporal signals, was used to transform EEG signals into an image-like representation called spectrogram. The EEG information is initially segmented into shorter pieces of equal length and then Fourier transform is computed on each shorter segment. The various other research applicability of Fourier transform has been experimented [35, 36]. The STFT calculation is described in Eq. 1:

$$\text{STFT}\{s(t)\}(m,f) \equiv S(m,f) = \sum s(t)w(t-m)e^{-ift}\text{d}t \tag{1}$$

where $s(t)$ is the signal to be transformed using $w(t)$ window function, here Hamming window function has been used. $S(m, f)$ is basically the Fourier Transform of $s(t)(m, f)$, a complex function representing the phase and magnitude of the signal over time (m) and frequency f. Hence, to maintain both time and spatial persistence of the acquired EEG signal, the STFT algorithm is applied to the filtered EEG signals, for converting 1D EEG signals of every EEG electrode position E to 2D spectrogram images. For instance, in the experiment, the EEG signals in DEAP dataset were acquired from 32 electrode points, therefore, 32 2D spectrogram images were generated for each of the 40 emotion evoking video clips. The data has been collected from 32 subjects while viewing 40 videos therefore, $32 \times 40 \times 32 = 40,960$ spectrogram images with 64×64 image size is generated. Whereas, AMIGOS and DREAMER dataset consists of 14 EEG channels, which were used to create 14 2D spectrogram images for each 40 and 23 subjects, respectively. Further, each subject's EEG signals are recorded for each of the 16 and 18 emotion evoking video clips, therefore, for each subject, $14 \times 16 \times 40 = 8960$ and $14 \times 18 \times 23 = 5796$ spectrogram images are produced. These generated spectrograms are then labeled with the 2 classes names low/high valence, low/high arousal and low/high dominance. The generated data was split into 80:20 ratio, i.e., 80% data is used for training and 20% is used for testing of the proposed model. The detailed structure of the datasets is described in Table 1.

3.4 EEG-based emotion classification with deep learning-based CNN using raw EEG signal

Apart from the advanced deep learning-based proposed methodology, some traditional machine learning techniques such as support vector machine (SVM), random forest (RF), decision tree (DT), K-nearest neighbor (KNN) and Gaussian Naive Bayes (GNB) has been implemented on raw EEG signals without any feature extraction. Further, convolutional neural network (CNN) and advance version of CNN architecture called capsule network have been implemented using raw EEG signals and an image-like representation, i.e., 2D spectrogram images generated from EEG signals to classify emotional states. In the current work, the proposed CNN architecture comprises 3 Conv1D modules, a Batch Normalization and MaxPool1-D layer along with 2 Fully connected (FC) layers. A nonlinear activation function LeakyRelu (leaky rectified linear units) is applied to all convolutional layers to convert the outcomes between 0.01 and 1 and a dropout layer prior to output layer has been added. The first Conv1D layer acquires the raw EEG signals of dimension $1280 \times 258{,}048$; 640×7168 and 414×7168 as input for DEAP, AMIGOS and DREAMER dataset, respectively, where 1,280; 640 and 414 represent the total number of trials and 258,048 and 7,168 represents the data size which has been flattened to get a 1D array (as described in the earlier Sect. 3.1). A brief explanation with the number of layers, input and output size, different operations along with the filter size in the proposed CNN model for EEG based emotion classification is shown in Table 2. This architecture has been tested using all the three datasets and provides a slight improvement from the accuracy achieved through the Machine learning techniques.

Table 2 Architecture of proposed convolutional neural network

Layer	Operation	Filter size/parameters
1	Conv1D + LeakyRelu	256×3
2	Batch Normalization	–
3	MaxPool1D	2
4	Conv1D + LeakyRelu	128×3
5	Conv1D + LeakyRelu	64×3
6	Dropout	0.2
7	Flatten	–
8	Dense + LeakyRelu	64
9	Dense + Softmax	2

3.5 Emotion recognition with proposed emotion capsule network (EmotionCapsNet) using raw EEG signal

The proposed EEG-based ER model based on CapsNet has four modules, namely, convolutional layer with rectified linear unit (ConvReLU), primary capsule (PrimaryCaps), emotion capsule (EmotionCaps) and fully connected layer. The overall structural details are illustrated in Table 3a. A 1D convolutional layer, which has 128 convolutional kernels and 8 kernel size with a stride of 1 and ReLU activation has been used to classify 1D raw EEG signals. The three datasets described in Sect. 3.1 is used as input to 1D convolutional layer that generates sample point values to feature map, which can then be used as input variables to the PrimaryCaps which consist a Conv1D layer with 256×8 filters with 32 channels of convolutional 8D capsules (in other words each emotion capsule contains 8 convolutional units) followed by EmotionCaps layer which has 32D capsules that correlate to two components of emotional states since this framework perform binary classification tasks, such as low/high valence, low/high arousal and low/high dominance. The routing-by-agreement process, implemented between PrimaryCaps and EmotionCaps, which integrates the existing EmotionCaps module with the preceding PrimaryCaps layer. Figure 2 depicts the structural details of routing-by-agreement process. Finally, the two fully connected layers with 256 and 2 dense sizes are implemented for classification of emotional states. The overall architecture description of EmotionCapsNet is described in Table 3 and the classification accuracy result has been described in Sect. 4.
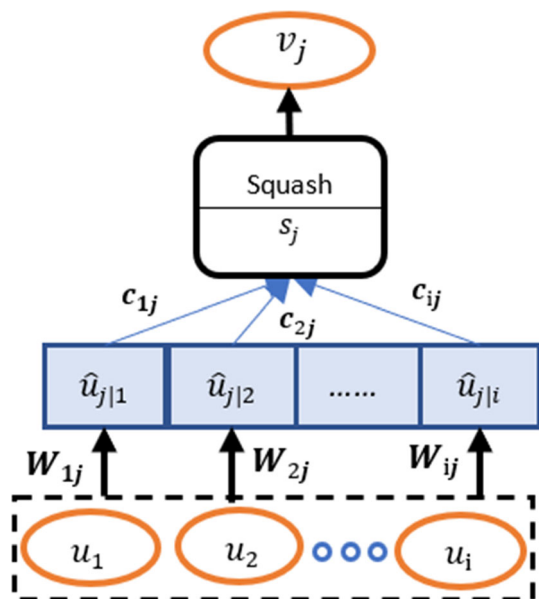

Fig. 2 Routing-by-agreement process

3.6 Emotion recognition using EEG spectrogram images with proposed emotion capsule network (EmotionCapsNet)

The proposed EmotionCapsNet uses spectrogram images generated from emotion-based EEG signals, which converts 1D EEG signals to 2D Spectrogram images with size 64×64 using STFT algorithm as input data. Here, three different datasets are used for this experiment (as described in the earlier Sect. 3.1). The preprocessing and spectrogram formulation is also described in Sects. 3.2 and 3.3 which act as input data in current scenario. The generated spectrogram images are split into training and testing data. The training data is then fed to the proposed model to train the EmotionCapsNet and the performance of the proposed system can be tested using the testing data. The proposed EmotionCapsNet model comprises of four sections: the initial segment consists of a 2D Conv+ReLU and convolution activities are performed on each input image to extract the significant and decisive features. PrimaryCaps is the subsequent part, that comprises of a convolution procedure and transforms data to the capsules. The last capsule layer is EmotionCaps, which incorporates the dynamic routing process among capsules and is utilized for emotion-based spectrogram image classification. At the final stage, decoder attempts to reconstruct the information from the final capsule output and classify namely low/high valence, low/high arousal, low/high dominance, the two emotion states for DEAP, AMIGOS and DREAMER dataset. The proposed model's overall accuracy shows the capability and the probability of proposed system to distinguish between three emotion states from the evoked EEG signals of a subject while viewing a video clip. The workflow and the architecture of emotion capsule network are same for all the datasets which is described in this section. The layers and module description of EmotionCapNet for classification of emotion states using raw EEG signal and spectrogram images formulated from EEG signal is given in Table 3a, and 3b, respectively. Figure 3 shows the design of the proposed EmotionCapsNet-based methodology.

A brief description of the emotion capsule network architecture on DEAP, AMIGOS and DREAMER dataset is:

1. The EEG-based spectrogram images with 64×64 size is taken as input.
2. The first Conv2D layer consists of 64 kernels with 7×7 2D convolution kernel block, no padding and a stride of 1. This Conv2D layer yields 64 feature maps with size 58×58.

Table 3 Architecture of proposed emotion capsule network (EmotionCapsNet) for (a) raw EEG signal and (b) spectrogram images formulated from EEG signal

Modules	Operation	Parameters	Shape/value
(a)			
Input	Input	–	DEAP: 32 × 8064 AMIGOS:14 × 512 Dreamer: 14 × 512
ConvReLU	Conv1D	Kernel	128 × 8
Primary capsule	Conv1D	Kernel	256 × 8
	Reshape	size	256 × 1
Emotion capsule	Dynamic routing	W_{ij}	8 × 32
Dense	Fully connected layer	256	–
Dense	Fully connected layer	2	–
Loss	–	m^+	0.9
	–	m^-	0.1
(b)			
Input	Input	–	DEAP: 64 × 64 × 3 AMIGOS:64 × 64 × 3 Dreamer: 64 × 64 × 3
ConvReLU	Conv1D	Kernel	64 × 7 × 7
Primary capsule	Conv1D	Kernel	128 × 9 × 9
	Reshape	Size	128 × 1
Emotion capsule	Dynamic routing	W_{ij}	4 × 32
Dense	Fully connected layer	128	–
Dense	Fully connected layer	2	–
Loss	–	m^+	0.9
	—	m^-	0.1

3. PrimaryCaps comprises of 32 filter channels with 4D capsule blocks (i.e., a primary capsule layer has 128 filters; 32 channels × 4 capsules; of size 9 × 9 with stride of 1). After that 128, 50 × 50 output of primary capsule is reshaped to yield u_i of 50×50 32×4 (each output is an 4D vector) as input shape for the upcoming emotion capsule layers. For these inputs, a capsule prediction vector $\hat{u}_{ij}$ is determined by multiplying an input vector u_i of capsule i and weighting metrics W_{ij} as in Eq. 2.

$$\hat{u}_{j|i} = W_{ij}u_i. \tag{2}$$

The weighted sum over all prediction vectors $\hat{u}_{j|i}$ is calculated as s_j in Eq. 3:

$$s_j = \sum_i c_{ij}\hat{u}_{j|i} \tag{3}$$

whereby c_{ij} is the "coupling coefficient" between ith primary capsule and jth emotional capsule that is calculated using the dynamic routing algorithm demonstrated in ([23]). The output of jth emotional capsule, v_j is then determined (as shown in Fig. 2) using a nonlinear squash function as in Eq. 4, which "squashes" vectors near zero when input capsule s_j (Eq. 3), is short and near 1 in case when vector length is long.

$$v_j = \frac{||s_j||^2}{1 + ||s_j||^2}\frac{s_j}{||s_j||}. \tag{4}$$

4. The EmotionCaps has 32Demotional capsules that pertain to two states of emotion as the proposed framework is implemented to perform binary classification tasks, such as low/high valence, low/high arousal and low/high dominance. In this regard, the classification loss is calculated using the distance and orientation of each capsule vector which indicate the presence of an emotional class in the EmotionCaps layer. In this phase, the routing-by-agreement process is used to connect EmotionCaps and PrimaryCaps layer to improve the learning process in comparison with conventional pooling procedure. Furthermore, this model employs a unique margin loss for each emotion capsule. For a capsule representing class c, the margin loss L_c is as follows:

$$L_c = P_c \max\left(0, m^+||v_c||\right)^2 + \lambda(1 - P_c)\max\left(0, ||v_c|| - m^-\right)^2. \tag{5}$$

In the loss function formula, the correct label determines the value of P_c, where P_c is 1 if the correct label c matches with the image of the particular EmotionCapule and 0 otherwise. The hyper-parameters m^+
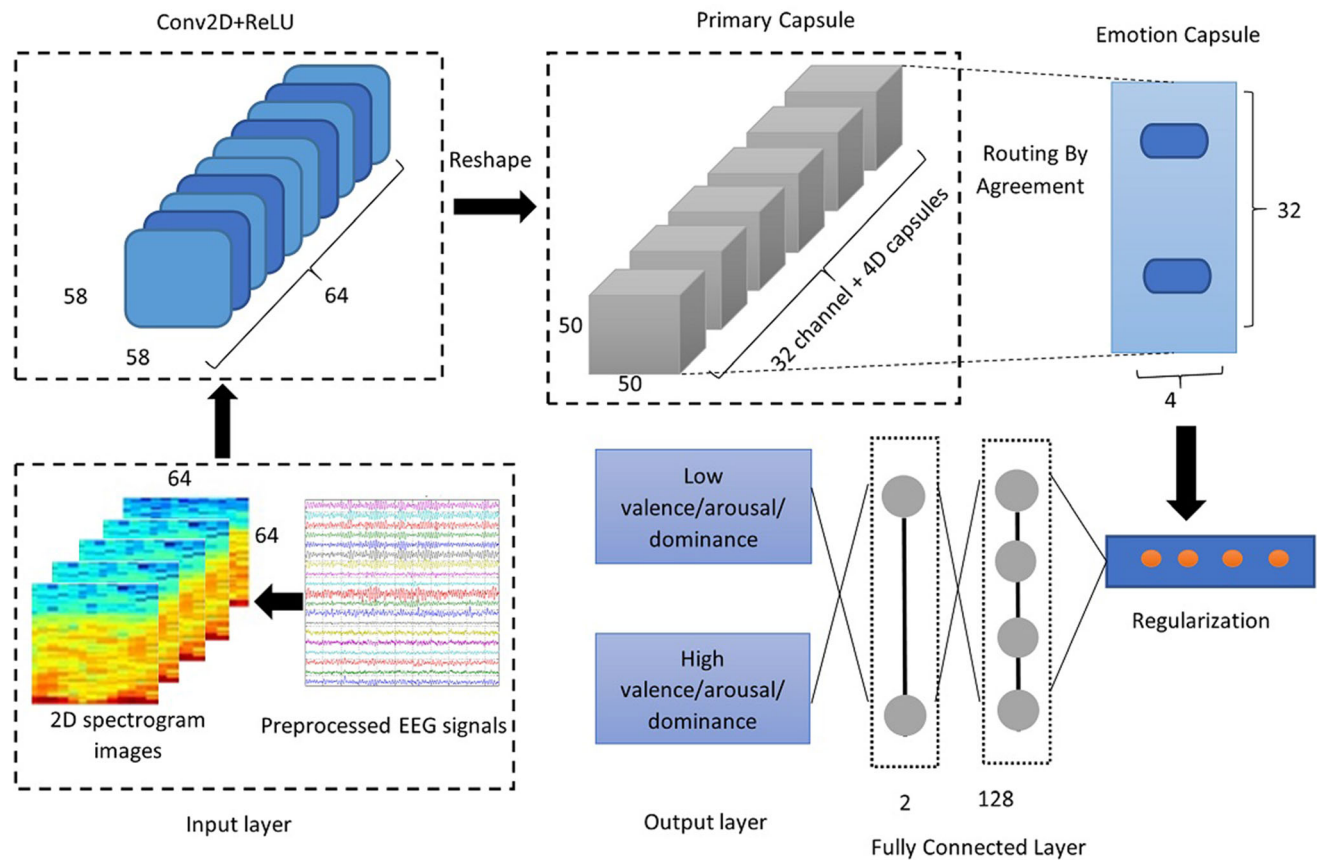
Fig. 3 The proposed architecture of EmotionCapsNet on DEAP, AMIGOS and DREAMER dataset

is set to 0.9 meaning if an object of class c is present $||v_c||$ should not be less than 0.9 and m^- 0.1 means if an object class is not present then $||v_c||$ should not more than 0.1. The λ parameter weakens the impact of the labels that mismatch with the correct class label is set to 0.5 for numerical stability. Though traditional CNN avoids overfitting by applying dropout layer, Capsule systems are regularized with a reconstruction autoencoder. The reconstruction loss is obtained from the Euclidean distance of recreated and original data.

5. The decoder includes two fully connected layers along with 128 and 2 neurons, respectively. The number of neurons in the last fully connected layer is the same as the number of emotion classes identified.

4 Results and discussion

4.1 Parameter optimization and classification accuracy of proposed EmotionCapsNet

To begin with, different designs with different parameters have been evaluated for EmotionCapsNet, utilizing the above referenced datasets, to decide on the ideal network environment. As per the description of the DEAP, AMIGOS and DREAMER dataset, a classifier was prepared and tried for all subject. As clarified in the above sections, the raw EEG signals are converted into spectrogram images using STFT, which then serve as an input data for the proposed CapsNet-based technique. Additionally, this methodology was built and trained in 32 mini-batches for all three datasets. The stochastic gradient descent (SGD) optimizer was used for 100 epochs. In routing-by-agreement algorithm, the agreement parameter is set as 3 and the margin loss function is used while compilation of the model. These hyper-parameters are optimized after many trials. The hyper-parameters of the CapsNet to be streamlined include: The iteration number in Dynamic Routing Process (1, 2, 3), number of filters used in convolutional layers (32, 64, 128), kernel size used in convolutional layers (3, 4, 7, 9), number of the capsules and dimensions in the PrimaryCaps layer (64, 128, 256) and the ImageCaps capsules (32, 16, 8). During the parameter improvement stage, the number and type of trainable parameters influence the overall performance and classification accuracy of CapsNet model. To find the ideal combination of model parameters, various models were tested on the

datasets. These model parameters were chosen by alluding to references [23, 37]. The complexity of the system is reduced by applying Reshaping, L2-norm Regularization and SGD optimizer in EmotionCapsNet model. Extremely complicated models are more difficult to comprehend, have a higher risk of overfitting and are likely to be computationally costly. In this paper, Regularization and dimensionality reduction strategies are applied to reduce model complexity. In capsule network pooling operation has been eliminated and reshaping has been done for dimensionality reduction between convolutional layer and primary capsule layer which compresses the multi-dimensional feature vector to 1D vector without losing of prominent information for further processing in capsule layer which helps to build feature-wise capsule for higher layers. Regularization essentially preserves all features but reducing (or penalizing) their impact on the model's projected values. The reduced effect comes from shrinking the magnitude, and therefore the effect, of some of the model's term's coefficients. L2 regularization is applied to reduce overfitting issue and hence uses a penalty factor to the loss function of "squared magnitude" of the coefficient. The penalty increases with the magnitude of the term's coefficient, which simply indicates that the optimization factor encourages the coefficient to be near to 0. If the dataset size is huge, training will be slow and computationally expensive [38]. In this case, gradient descent optimization is the preferred way to optimize machine and deep learning algorithms which updates the weight and bias using the entire training dataset but causes high computational complexity. To overcome the gradient descent issue, a stochastic gradient descent (SGD) optimization technique is introduced that updates parameters only using a single record. In this work, the learning rate for SGD was set to 0.0001 and momentum to 0.7 is used. In SGD with momentum, a momentum in a gradient function is added which means the present gradient is dependent on its previous gradient and so on that accelerates the converge of SGD.

4.2 Comparison with the other implemented conventional methods

As discussed in Related Work section, many researchers have achieved good accuracy results for ER task using these signals as well as combination of multi-modality signals. In this work, three objective has been accomplished, first to achieve comparable accuracy result using EEG signals as single modality for emotion classification. Second, to bypass the feature extraction step without degrading the accuracy. Third, an Advance DL-based EmotionCapsNet has been proposed to achieve better accuracy results using raw EEG signals as well as from 2D

spectrogram images generated from raw EEG signals. Table 4 represents the accuracy along with F1 score using single modality EEG-based emotion classification on DEAP, AMIGOS and DREAMER dataset where F1 score is metric that produces a comparison between precision and recall which describe a probability relevant predictions are selected by the model [39]. As it can be seen that the reported accuracy achieved using raw EEG signal without going through any feature extraction step performs equally well. The achieved accuracy results of deep learning-based proposed CNN architecture described in Sect. 3.4 has significantly improved the recognition accuracy as compared to other machine learning methods which has been implemented on raw EEG signals only. However, the classification accuracy obtained from proposed CNN architecture is not that much good, to achieve better accuracy results we tried to implement an Advanced DL model.

4.3 Comparison between the classification accuracy of proposed EmotionCapsnet on raw EEG signals and 2D spectrogram images generated from EEG signals

Table 5 represents the training and testing accuracy along with F1 score for three emotion states on three datasets. The classification accuracy and F1 score of the proposed EmotionCapsNet architecture has been implemented using raw EEG signal and using 2D spectrogram images generated from raw EEG signals for three datasets is compared in Table 5. The proposed architecture and implementation of EmotionCapsNet model for raw EEG-based emotion classification with some modification on general architecture of CapsNet has been already described in Sect. 3.5. It can be observed from Table 5 that accuracy results of EmotionCapsNet on raw EEG signal have been slightly improved but it cannot be considered as good accuracy results and encouraged us to work further to achieve better accuracy results. As per the literature review, the original CapsNet performed well for the handwritten digital images classification in the MNIST dataset, therefore we have converted the raw EEG signals into 2D spectrogram images using STFT algorithm as described in Sect. 3.3 to evaluate the proposed EmotionCapsNet. Table 5 shows the average training and testing accuracy along with F1 score of proposed EmotionCapsNet on raw EEG data and EEG generated spectrogram data (given in bold). It can be observed from the accuracy results described in Table 5, that the proposed EmotionCapsNet architecture has significantly improved the classification accuracy using spectrogram data (Sect. 3.6), as compared to the raw EEG-based EmotionCapsNet and other conventional methods. In these traditional methods, the most significant and

Table 4 Average accuracy (%) along with F1 score (%) of different machine learning-based methods on three datasets for valence, arousal and dominance emotional states classification tasks

Dataset	Emotion states	SVM		RF		KNN		GNB		DT		CNN	
		Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
DEAP	Valence	60.41	62.00	61.97	75.00	60.00	68.00	56.26	68.00	56.77	67.00	64.07	73.00
	Arousal	58.07	68.00	63.02	76.00	62.25	75.00	45.31	45.00	53.12	65.00	69.92	75.00
	Dominance	63.28	74.00	69.27	81.00	69.45	80.00	42.70	43.00	64.58	75.00	71.88	76.00
AMIGOS	Valence	55.72	68.00	65.05	67.00	52.33	68.00	46.87	50.00	52.08	63.00	67.97	71.00
	Arousal	54.68	66.00	51.04	65.00	44.56	50.00	52.08	57.00	52.08	58.00	64.69	69.00
	Dominance	48.43	54.00	52.08	63.00	52.20	64.00	50.52	62.00	52.60	47.00	70.00	73.00
DREAMER	Valence	54.48	66.00	55.37	68.00	52.33	64.00	55.68	63.00	56.88	66.00	69.09	71.00
	Arousal	58.63	56.00	58.76	63.00	57.61	66.00	48.80	64.00	47.16	48.00	65.01	68.00
	Dominance	51.45	45.00	29.60	55.00	50.00	54.00	42.40	56.00	44.48	44.00	63.40	65.00

Table 5 Average training and testing accuracy (%) along with F1 score (%) of proposed EmotionCapsNet on three datasets

Dataset	Input	Emotion states								
		Valence			Arousal			Dominance		
		Train Acc.	Test Acc.	F1	Train Acc.	Test Acc.	F1	Train Acc.	Test Acc.	F1
DEAP	Raw EEG signal	63.17	63.02	0.62	62.17	66.93	0.69	60.38	66.15	0.67
	Spectrogram	**73.12**	**77.50**	**0.71**	**82.81**	**78.44**	**0.78**	**81.56**	**79.38**	**0.77**
AMIGOS	Raw EEG signal	66.92	61.04	0.63	66.96	66.15	0.67	59.92	55.73	0.56
	Spectrogram	**81.88**	**79.06**	**0.76**	**80.07**	**78.90**	**0.79**	**77.87**	**79.69**	**0.79**
DREAMER	Raw EEG signal	60.55	60.80	0.60	67.09	61.20	0.63	60.87	65.00	0.62
	Spectrogram	**78.12**	**80.34**	**0.79**	**79.06**	**83.04**	**0.81**	**78.28**	**82.50**	**0.80**

challenging issues is overfitting, as the recorded EEG signal are generally noisy, inconsistent and unstable in nature. The evoked EEG signals can vary across the same subject as well as situation or stimuli. EEG signals have more intricate intramural representations of emotion states which require deep learning to learn its complex feature layer-wise. EmotionCapsNet learns the temporal, frequential as well as spatial features of encoded EEG signal and identify the significant low level features and forward it to high level feature to provide more descriptive and premised features sets for classification tasks.

Table 6 presents the comparison studies between proposed model and other existing models which uses single modality as EEG signals. It can be seen that the proposed model has achieved better accuracy on raw EEG as compared to other traditional classifier's accuracy (given in Table 5). The proposed EmotionCapsNet using spectrogram images formulated from EEG signals has also overridden the accuracy achieved from CapsNet implemented by [12] as well as other conventional methods.

Further, considering the privacy aspect of the model, the DL life span can be divided into two parts: training and testing/inference. The privacy devastation caused by DL is classified into two types: model extraction threats and model inversion threats. Differential privacy, homomorphic encryption, secure multi-party computing and trusted execution environment are the four basic defenses against them. There are two types of DL security threats: adversarial attacks and poisoning attacks. Adversarial defenses are being developed in three primary directions: preprocessing, malware detection and model robustness improvement [46]. As this paper works on subject-independent ER task, secure multi-party computing is applied to overcome the problem of collaborative computing that preserves each subject's EEG data privacy in a legion of the numerous subject's data. Fundamentally, all participant's EEG signals are merged for collaborative computation of proposed model. Preprocessing step has been also incorporated in this work by conducting band-pass and notch filter to remove noise and artifacts from raw EEG signals. This defense technique is applied before the input and the first layer of the model. Training techniques and regularization, applied to improve the model's robustness seek to enhance the model's capacity to resist adversarial data. For this purpose, Batch normalization used in proposed CNN model and L2-norm regularization has been applied on EmotionCapsNet model [47, 48].

Table 6 Comparison of proposed EmotionCapsNet with other existing work

Studies	Input	Classifier	Dataset	Accuracy(%)		
				Valence	Arousal	Dominance
[29]	Power spectral density (PSD)	GNB	DEAP	62.00	0.5760	–
[40]	DT-CWPT(dual-tree complex wavelet packet transform)	SVM	DEAP	66.90	65.30	69.10
[41]	DBN features	SVM	DEAP	64.20	58.40	65.80
[42]	PCA features	SAE	DEAP	49.52	46.03	–
[43]	Statistical features	CNN	DEAP	73	81	–
[12]	Multiband feature matrix (MFM)	Capsule network	DEAP	68.28	66.73	67.25
[44]	Statistical features	VAE	AMIGOS	67	68.80	–
[31]	PSD	SVM	DREAMER	62.49	62.17	61.84
[45]	PSD-based images	VGG-16 network	DREAMER	78.99	79.23	–
Proposed work	Raw EEG signals	EmotionCapsNet	DEAP	63.02	66.93	66.15
			AMIGOS	61.04	66.15	55.73
			DREAMER	60.80	61.20	65.00
Proposed work	Spectrogram images	EmotionCapsNet	DEAP	77.50	78.44	79.38
			AMIGOS	79.06	80.07	79.69
			DREAMER	80.34	83.04	82.50

5 Conclusion

In the current research work, advance deep learning-based EmotionCapsNet model has been implemented for multi-channel EEG-based ER model. The proposed framework can determine the inherent interaction encompassed by various EEG channels well. The raw EEG data is combined together w.r.t. EEG channels only to build a subject-independent ER model. Then the prominent features extracted from different convolution layers to form primary capsules which encapsulates the most prominent extricated features from various convolution operations into a capsule-like structure. An EmotionCaps layer is then used to select the most descriptive feature capsule for classification. Finally, experiments are conducted on DEAP, AMIGOS and DREAMER. The proposed method achieves average accuracy of 77.50%, 78.44% and 79.38% for VAD on the DEAP dataset, 79.06%, 78.90% and 79.69% on AMIGOS dataset and accomplishes average accuracy of 80.34%, 83.04% and 82.50% for VAD on the DREAMER, respectively. The experimental outcomes demonstrate that the proposed EmotionCapsNet-based methodology produces higher accuracy than some classical machine learning-based classifiers, such as the SVM, RF, KNN, GNB, DT and CNN methods. The advancement of this work has been presented in two primary perspectives. The first one is, a promising feature representation technique STFT is used to convert the visually evoked 1D EEG signals into 2D spectrogram images for acquiring improvements in classification results. The accuracy results represent that the proposed technique accomplished acceptable outcomes with three datasets. This shows that a spectrogram which contain spatial as well as time-frequency domain information are useful to recognize the various emotion states (valence, arousal and dominance), and the model dependent on capsule network could effectively utilize this data for classifying VAD emotion states. In the future, the proposed EmotionCapsNet-based model will be implemented for subject-independent EEG-based ER and can work to reduce the complexity of the system through enhancing layers and sharing parameters between capsule layers. The advantages and findings of our proposed multilayer capsule network approach are as follows: The need for sophisticated feature extraction is eliminated. It also outperforms other methods that are subject independent. The improved accuracy on subject-independent-based recognition performance can be attributed to the fact that capsules are able to record the fundamental spatial relationship from the spectrogram images, between the components of an object of seen images, while EEG recording. Therefore, this approach offers a practical means of describing the intrinsic interactions between the varied EEG channels to be specific; the main capsules encapsule the neural circuits and the transition matrices reflect the relationship between distinct parts of the brain. The advance CNN model-based capsule network is useful for getting to the highest levels of discrimination in an ER model. Moreover, the huge dimension of diversified

EEG signal is often considerable, that increases the length and size of parameters and the network's complexity. The suggested framework tends to decrease the amount of data and complexity using subject-independent EEG channels. Dimension reduction and regularization techniques are applied here to reduce overall complexity of the proposed methodology.

Declaration

Conflict of interest The authors declare no conflict of interest.

References

1. Dolan RJ (2002) Emotion, cognition, and behavior. Science 298(5596):1191–1194
2. Al-Nafjan A, Hosny M, Al-Ohali Y, Al-Wabil A (2017) Review and classification of emotion recognition based on EEG brain-computer interface system research: a systematic review. Appl Sci 7(12):1239
3. Van den Broek EL (2013) Ubiquitous emotion-aware computing. Pers Ubiquitous Comput 17(1):53–67
4. Malandrakis N, Potamianos A, Evangelopoulos G, Zlatintsi A (2011) A supervised approach to movie emotion tracking. In: 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 2376–2379
5. Aslam AR, Altaf MAB (2019) An 8 channel patient specific neuromorphic processor for the early screening of autistic children through emotion detection. In: 2019 IEEE international symposium on circuits and systems (ISCAS). IEEE, pp 1–5
6. Liu Y, Sourina O, Nguyen MK (2011) Real-time EEG-based emotion recognition and its applications. In: Transactions on computational science XII. Springer, pp 256–277
7. Lin YP, Wang CH, Jung TP, Wu TL, Jeng SK, Duann JR, Chen JH (2010) EEG-based emotion recognition in music listening. IEEE Trans Biomed Eng 57(7):1798–1806
8. Yang Y, Wu Q, Fu Y, Chen X (2018) Continuous convolutional neural network with 3d input for EEG-based emotion recognition. In: International conference on neural information processing. Springer, pp 433–443
9. Wu W, Yin Y, Wang X, Xu D (2018) Face detection with different scales based on faster r-CNN. IEEE Trans Cybern 49(11):4017–4028
10. Zheng WL, Zhu JY, Peng Y, Lu BL (2014) EEG-based emotion classification using deep belief networks. In: 2014 IEEE international conference on multimedia and expo (ICME). IEEE, pp 1–6
11. Sun B, Wei Q, Li L, Xu Q, He J, Yu L (2016) LSTM for dynamic emotion and group emotion recognition in the wild. In: Proceedings of the 18th ACM international conference on multimodal interaction, pp 451–457
12. Chao H, Dong L, Liu Y, Lu B (2019) Emotion recognition from multiband EEG signals using CapsNet. Sensors 19(9):2212
13. Zhou W, Liu J, Lei J, Yu L, Hwang JN (2021) Gmnet: graded-feature multilabel-learning network for RGB-thermal urban scene semantic segmentation. IEEE Trans Image Process 30:7790–7802
14. Ding L, Huang L, Li S, Gao H, Deng H, Li Y, Liu G (2020) Definition and application of variable resistance coefficient for wheeled mobile robots on deformable terrain. IEEE Trans Robot 36(3):894–909
15. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
16. Yang H, Han J, Min K (2019) A multi-column CNN model for emotion recognition from EEG signals. Sensors 19(21):4736
17. Thammasan N, Moriyama K, Ki Fukui, Numao M (2016) Continuous music-emotion recognition based on electroencephalogram. IEICE Trans Inf Syst 99(4):1234–1241
18. Estepp JR, Christensen JC (2015) Electrode replacement does not affect classification accuracy in dual-session use of a passive brain-computer interface for assessing cognitive workload. Front Neurosci 9:54
19. Zheng WL, Lu BL (2015) Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans Auton Mental Dev 7(3):162–175
20. Naji M, Firoozabadi M, Azadfallah P (2015) Emotion classification during music listening from forehead biosignals. Signal Image Video Process 9(6):1365–1375
21. Liu F, Zhang G, Lu J (2020) Multi-source heterogeneous unsupervised domain adaptation via fuzzy-relation neural networks. IEEE Trans Fuzzy Syst
22. Dong J, Cong Y, Sun G, Fang Z, Ding Z (2021) Where and how to transfer: knowledge aggregation-induced transferability perception for unsupervised domain adaptation. IEEE Trans Pattern Anal Mach Intell
23. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. In: Advances in neural information processing systems, pp 3856–3866
24. Wang Y, Sun A, Huang M, Zhu X (2019) Aspect-level sentiment analysis using as-capsules. In: The world wide web conference, pp 2033–2044
25. Turan MAT, Erzin E (2018) Monitoring infant's emotional cry in domestic environments using the capsule network architecture. In: Interspeech, pp 132–136
26. Yin J, Li S, Zhu H, Luo X (2019) Hyperspectral image classification using CAPSNET with well-initialized shallow layers. IEEE Geosci Remote Sens Lett 16(7):1095–1099
27. Liu Y, Ding Y, Li C, Cheng J, Song R, Wan F, Chen X (2020) Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network. Comput Biol Med 123(103):927
28. Ali U, Li H, Yao R, Wang Q, Hussain W, ud Duja SB, Amjad M, Ahmed B (2020) EEG emotion signal of artificial neural network by using capsule network. Int J Adv Comput Sci Appl 11(1):434–443
29. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2011) Deap: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31
30. Correa JAM, Abadi MK, Sebe N, Patras I (2018) Amigos: a dataset for affect, personality and mood research on individuals and groups. IEEE Trans Affect Comput
31. Katsigiannis S, Ramzan N (2017) Dreamer: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE J Biomed Health Inform 22(1):98–107
32. Rahman MA, Anjum A, Milu MMH, Khanam F, Uddin MS, Mollah MN (2021) Emotion recognition from EEG-based relative power spectral topography using convolutional neural network. Array, p 100072
33. Feng Y, Zhang B, Liu Y, Niu Z, Dai B, Fan Y, Chen X (2021) A 200–225-GHZ manifold-coupled multiplexer utilizing metal waveguides. IEEE Trans Microw Theory Tech
34. Jiang Y, Li X (2021) Broadband cancellation method in an adaptive co-site interference cancellation system. Int J Electron (just-accepted)

35. Yan Y, Feng L, Shi M, Cui C, Liu Y (2020) Effect of plasma-activated water on the structure and in vitro digestibility of waxy and normal maize starches during heat-moisture treatment. Food Chem 306(125):589

36. Shi M, Wang F, Lan P, Zhang Y, Zhang M, Yan Y, Liu Y (2021) Effect of ultrasonic intensity on structure and properties of wheat starch-monoglyceride complex and its influence on quality of norther-style Chinese steamed bread. LWT 138(110):677

37. Ha KW, Jeong JW (2019) Motor imagery EEG classification using capsule networks. Sensors 19(13):2854

38. Che H, Wang J (2020) A two-timescale duplex neurodynamic approach to mixed-integer optimization. IEEE Trans Neural Netw Learn Syst 32(1):36–48

39. Martínez-Tejada LA, Yoshimura N, Koike Y (2020) Classifier comparison using EEG features for emotion recognition process. In: 2020 IEEE 18th world symposium on applied machine intelligence and informatics (SAMI). IEEE, pp 225–230

40. Daimi SN, Saha G (2014) Classification of emotions induced by music videos and correlation with participants' rating. Expert Syst Appl 41(13):6057–6065

41. Li X, Zhang P, Song D, Yu G, Hou Y, Hu B (2015) EEG based emotion identification using unsupervised deep feature learning

42. Jirayucharoensak S, Pan-Ngum S, Israsena P (2014) EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. Sci World J

43. Tripathi S, Acharya S, Sharma RD, Mittal S, Bhattacharya S (2017) Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In: Proceedings of the thirty-first AAAI conference on artificial intelligence, pp 4746–4752

44. Yang HC, Lee CC (2019) An attribute-invariant variational learning for emotion recognition using physiology. In: ICASSP 2019–2019 IEEE international conference on acoustics. Speech and signal processing (ICASSP). IEEE, pp 1184–1188

45. Siddharth S, Jung TP, Sejnowski TJ (2019) Utilizing deep learning towards multi-modal bio-sensing and vision-based affective computing. IEEE Trans Affect Comput 1–1

46. Liu X, Xie L, Wang Y, Zou J, Xiong J, Ying Z, Vasilakos AV (2020) Privacy and security issues in deep learning: a survey. IEEE Access 9:4566–4593

47. Debie E, Moustafa N, Vasilakos A (2021) Session invariant EEG signatures using elicitation protocol fusion and convolutional neural network. IEEE Trans Dependable Secure Comput

48. Shen Z, Luo J, Zimmermann R, Vasilakos AV (2011) Peer-to-peer media streaming: insights and new developments. Proc IEEE 99(12):2089–2109

49. Afshar P, Mohammadi A, Plataniotis KN (2018) Brain tumor type classification via capsule networks. In: 25th IEEE international conference on image processing (ICIP) (pp 3129-3133). IEEE