

Received 10 March 2024, accepted 24 March 2024, date of publication 2 April 2024, date of current version 15 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3384303

RESEARCH ARTICLE

Developing an EEG-Based Emotion Recognition Using Ensemble Deep Learning Methods and Fusion of Brain Effective Connectivity Maps

SARA BAGHERZADEH¹, AHMAD SHALBAF², AFSHIN SHOEIBI³, MAHBOOBEH JAFARI⁴, RU-SAN TAN^{5,6}, AND U. RAJENDRA ACHARYA^{4,7}

¹Department of Biomedical Engineering, Islamic Azad University Science and Research Branch, Tehran 1477893855, Iran

²Department of Biomedical Engineering and Medical Physics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran 1516745811, Iran

³Data Science and Computational Intelligence Institute, University of Granada, 18071 Granada, Spain

⁴School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia

⁵National Heart Centre Singapore, Singapore 169609

⁶Duke-NUS Medical School, Singapore 169857

⁷Centre for Health Research, University of Southern Queensland, Toowoomba, QLD 4350, Australia

Corresponding author: Ahmad Shalbf (shalbf@sbm.ac.ir)

This work was supported by the Shahid Beheshti University of Medical Sciences under Grant 43004477.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the ethics committee of Shahid Beheshti University of Medical Sciences under Application No. IR.SBMU.MSP.REC.1401.652.

ABSTRACT The objective of this paper is to develop a novel emotion recognition system from electroencephalogram (EEG) signals using effective connectivity and deep learning methods. Emotion recognition is an important task for various applications such as human-computer interaction and, mental health diagnosis. The paper aims to improve the accuracy and robustness of emotion recognition by combining different effective connectivity (EC) methods and pre-trained convolutional neural networks (CNNs), as well as long short-term memory (LSTM). EC methods measure information flow in the brain during emotional states using EEG signals. We used three EC methods: transfer entropy (TE), partial directed coherence (PDC), and direct directed transfer function (dDTF). We estimated a fused image from these methods for each five-second window of 32-channel EEG signals. Then, we applied six pre-trained CNNs to classify the images into four emotion classes based on the two-dimensional valence-arousal model. We used the leave-one-subject-out cross-validation strategy to evaluate the classification results. We also used an ensemble model to select the best results from the best pre-trained CNNs using the majority voting approach. Moreover, we combined the CNNs with LSTM to improve recognition performance. We achieved the average accuracy and F-score of 98.76%, 98.86%, 98.66 and 98.88% for classifying emotions using DEAP and MAHNOB-HCI datasets, respectively. Our results show that fused images can increase the accuracy and that an ensemble and combination of pre-trained CNNs and LSTM can achieve high accuracy for automated emotion recognition. Our model outperformed other state-of-the-art systems using the same datasets for four-class emotion classification.

INDEX TERMS Effective connectivity, electroencephalography, emotion recognition, long short-term memory, transfer learning.

I. INTRODUCTION

Emotions are mental states that are either evoked by internal stimuli like past memory or induced by external stimuli like

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kafiul Islam¹.

video content [1], [2]. The emotional state can influence important brain functions such as perception and decision-making, which carry implications for social interactions and societal well-being. Recently, emotion recognition has been applied in real-world applications such as meditation, relaxation, and mode detection using mobile devices and smart

watches. Many applications are developed and available on mobile devices based on the recognition of emotions. For example, researchers are using artificial intelligence (AI) algorithms to detect the amount of stress using the brain signals recorded from headsets during daily life. These applications are aimed at improving lifestyle. For example, by detecting anger or fear, or any other emotion of the user on his/her watch or mobile device, feedback can be emitted to alert to the user. Such systems can be used in the electronic learning of students during online courses to improve the quality of education. Indeed, recognizing emotional states has increasingly garnered interest among researchers in neuroscience [3], [4] and mental disorders [5], [6], [7], [8]. The electroencephalography (EEG) is a suitable technique used for emotions recognition [9], [10], [11], [12]. This is because EEG has portability advantages, lower cost, better tolerability, and higher signal temporal resolution than functional magnetic resonance imaging. It is also an established modality that has been applied in the clinic for diagnosing mental disorders or neurological conditions [9], [13], [14], [15], [16].

Many researchers have developed machine learning methods to classify emotions using EEG. An emotion recognition system comprises three steps: feature extraction, feature selection/reduction, and classification [17], [18]. In some studies, handcrafted features in the frequency, time, and time-frequency domains are extracted [17], [18], [19]. Functional brain connectivity constitutes another feature that can be extracted, e.g., using coherence, correlation, and mutual information to estimate the statistical dependencies between EEG signals from different channels [20]. For example, Liu et al. in [21] extracted mutual information features from EEG signals in the Database for Emotional Analysis using Physiological signals (DEAP) dataset [1] and classified them using support vector machine (SVM) and random forest (RF). They attained 75.75% average accuracy for the binary classification tasks of discriminating high arousal versus low arousal and negative valence versus positive valence states. Principal component analysis (PCA) [17], [21], [22], [23], [24] is the most popular feature selection/reduction methods in emotion recognition systems. In the classification step, SVM and k-nearest neighbor are commonly used to classify two- or multi-class emotions [17], [23], [24].

Recently, deep learning methods like deep autoencoders [25], and convolutional neural networks (CNNs) [25], [26], [27], [28], [29], [30], [31] have been applied to emotion recognition studies. Among these, CNNs have garnered the most success using EEG signal-based emotion recognition. CNNs are more generalizable, versatile and can overcome issues like non-stationary, low signal-to-noise, and inter-subject variances associated with EEG signals [29], [30]. Khare and Bajaj in [31] used smoothed pseudo-Wigner-Ville distribution to convert self-recorded 1D EEG signals into time-frequency images to feed to a configurable CNN. The latter attained the highest classification accuracy of

93.01% for emotion classes versus other pre-trained networks. In some studies, a combination of these methods was used to recognize emotions. For example, CNN was combined with a sparse autoencoder and deep neural network to achieve the accuracy of 89.49% and 92.86% to discriminate valence and arousal classes, respectively [25]. Some studies [15], [16], [30] showed the effectiveness of adding time dependency by combining CNNs and long short-term memory (LSTM). This combination has increased the classification accuracy of EEG signals in neuroscience. Yin et al. [29] used a network that was developed based on graph CNN combined with long short-term memory from 6-second time windows of EEG signals in the DEAP dataset. The model attained 90.62% accuracy for binary classification. Also, some studies demonstrated performance improvement in the ensemble approach using the majority voting method with deep learning techniques [30].

One of the methods of the converting one-dimensional multichannel EEGs signals to two-dimensional images for using deep neural networks is brain connectivity. Brain connectivity has three types structural, functional, and effective. The structural determines information flow between adjacent regions, the functional estimates the strengths of coincidence of two separate regions from the point of coherence, correlation or similar measures [20]. Effective connectivity (EC) estimates the dependencies of distinct brain regions. This method describes the causal information flow of different brain regions based on EEG channels. EEG is one of suitable techniques for mapping brain effective connectivity [10]. Directed transfer functions (DTF), partial directed coherence (PDC) and transfer entropy (TE) are three well established methods of brain EC. These well-known methods are used to convert one-dimensional multichannel EEGs to two-dimensional images to represent emotional brain interaction between distinct pairs of electrodes in the point of flow of information. The DTF determines the causal information flows between brain regions, while direct DTF (dDTF, extension of DTF) considers only direct flows. The PDC estimates direct and indirect relations and provides information about the direction of brain regions [9]. TE is based on the information theory and estimates the nonlinear relationships of transfer of information between pairs of EEG channels. In other words, TE is a model-free, non-parametric and nonlinear effective connectivity measure that estimate causality between pairs of two brain regions according to the conditional entropy [16]. Whereas PDC and dDTF are linear measures and, consequently, more restricted. Also, PDC and TE discard in-directed connections. These dDTF and PDC were used to classify emotional states from EEG using [32], [33] and achieved accuracies above 95% on the both DEAP and MAHNOB-HCI databases. The contribution of this study is as follows:

- Generating a novel fused 2D image that combines three effective connectivity measures and different overlapping time windows obtained from 32 channels EEG signals. This

new fused image adds time information and utilizes the advantages of each measure, resulting in a more precise representation of brain connectivity during emotional states. This novel image improved our previous model [32], [33].

-Extracting novel deep features from fused connectivity images. Our aim is to improve emotion recognition performance by providing relevant and different information about brain flow during desired emotional classes. We have fed these new fused connectivity images into a more recent and powerful set of pre-trained CNNs models and LSTM structure for classification into four emotion classes via transfer learning for DEAP and MAHNOB-HCI databases.

-Fusion of pre-trained CNNs models and LSTM models to improve the accuracy. We have combined CNN models and LSTM to utilize the advantages of each deep learning method, with LSTMs adding time dependency of sequences to the recognition of emotion.

-Taking the final decision using ensemble CNN-LSTM models using the majority voting method. To further improve classification accuracy, we have presented an ensemble of these state-of-the-art models based on the majority voting method. This approach is particularly useful in dealing with non-linear behavior in multi-channel recordings of EEG signals during emotion function and the high variance in performance of CNN-LSTM models due to the high number of parameters. By combining different models, we can construct a good model with the best achievable performance. Each basic model may have a skill in some part of data, and the combination of these models using majority voting yields better results. Overall, our study presents a promising approach for improving emotion recognition using EEG signals and deep learning techniques.

II. RELATED STUDIES

Yao et al. [34] proposed a method to recognize emotions from multi-channel EEG data using multi-feature fusion. They extracted 1D power values from each frequency band of the EEG signals and exploited them to construct 2D spatial image features based on the electrode locations. Then, they combined the spatial image features for each frequency band to create 3D multispectral images that capture the spatial and spectral information of the EEG data. They designed a CNN framework that uses feature fusion modules and dilated bottleneck-based convolutional neural networks (DBCN) to process the spatio-spectral image features and classify the emotions.

In [35], a new method for emotion recognition from EEG signals using holographic features and CNN was introduced. The proposed method employed computer-generated holography (CGH) to create 2D maps of EEG signal features and fed them to a CNN for feature extraction. Authors also applied ReliefF and Neighborhood Component Analysis (NCA) to select the optimal electrodes for each gender, which reduced the dimensionality and complexity of the EEG data.

Gao et al. [36] integrated time-domain features and frequency-domain features for the purpose of EEG emotion

analysis. The objective of this approach was to leverage the information present between EEG channels and the contextual information of EEG signals to enhance emotion recognition. The effectiveness of this method was evaluated through emotion recognition experiments conducted on the DEAP dataset, which resulted in high accuracy in valence and arousal classification. Subsequently, they used a GoogleNet-based CNN with inception modules to learn the spatial and contextual information of EEG electrodes and used SVM to classify the fused features.

In [37], a novel deep learning framework is presented that is capable of recognizing emotions from subject-independent EEG signals. The framework consists of two components: the first component utilizes LSTM and channel-attention autoencoder (AE) to create a subject-invariant latent vector space from each individual's EEG data. The second component employs CNN and attention structure to classify emotions from latent space representations obtained from LSTM and channel-attention AE. In this paper, all the experiments have been performed on three available datasets including DEAP, SEED and CHB-MIT.

Jana et al. [38] proposed a novel approach to integrate spatial and temporal information from EEG signals in a sparse spatio-temporal framework. The authors introduced a data splitting technique that enables the model to learn from the least familiar information in the problem of EEG-based emotion recognition. The CapsNet architecture was trained on the DEAP dataset to perform binary classification, and Bayesian optimization was leveraged to fine-tune hyperparameters. The proposed method achieved recognition accuracy that is comparable to state-of-the-art models.

Ngai et al. [39] proposed a method to enhance the performance of emotion classification by utilizing multimodality features and making decisions based on the combined features. The authors employed multi-source and heterogeneous data, including EEG signals from two channels, eye data, and face data, and designed a multi-branch deep convolutional neural network. To evaluate the effectiveness of the proposed method, extensive experiments were conducted on modality and emotion data, which yielded promising results.

In [40], the authors introduced a novel approach that uses spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition. The proposed method employs a spatio-temporal attention mechanism to automatically select the crucial temporal and spatial components of EEG signals. Additionally, a self-adaptive brain network adjacency matrix is utilized to measure the connection strength between channels, which can reveal distinct patterns of brain activation associated with different emotions.

Samavat et al. [41] propose a novel deep learning methodology that combines CNN and Bi-LSTM for emotion recognition from raw EEG signals. The proposed approach employs a two-layer CNN with varying filter sizes to extract time and frequency features and integrates them with differential entropy feature. Furthermore, adaptive regularization

is utilized to account for the spatial information of EEG electrodes. The efficacy of the proposed method is evaluated on two datasets, namely SEED and DEAP.

Zhang et al. [42] propose a two-stage framework that leverages spatial and temporal information for emotion recognition from EEG signals. The authors employ a hierarchical self-attention network to model local and global temporal information, select the most relevant segments, and reduce noise at the temporal level. Additionally, squeeze-and-excitation (SE) and channel correlation loss (CC-loss) modules are utilized to select the most important channels and extract distinctive features at the spatial level. The proposed method is evaluated on three datasets, namely SEED, DEAP, and MAHNOB-HCI.

Pandey et al. [43] investigated emotional responses to EEG signal stimuli using the DEAP dataset and two feature extraction techniques, namely variational mode decomposition (VMD) and empirical mode decomposition (EMD). The authors subsequently utilized a DNN model to classify emotions. The results of the study indicate that the DNN classifier outperforms SVM classifiers for emotion recognition, and the VMD-based feature method is more effective than the EMD-based method, reducing signal complexity.

Reference [44] proposes a DL model based on meta-transfer learning that can automatically differentiate and classify generated EEG signals into various emotional states. By utilizing meta-learning, the model can quickly adapt to new subjects with minimal data, which is highly advantageous in practical applications. Moreover, this approach enhances model performance and mitigates the issue of low accuracy after transfer. The proposed model is evaluated on two datasets, SEED and SEED-IV, and outperforms existing methods in terms of data accuracy and efficiency.

An improved graph convolution model with dynamic channel selection by Lin et al. [45] has been suggested. The proposed model combines the advantages of 1D Conv and graph convolution to capture intra- and inter-channel EEG features. Additionally, they employed functional connectivity in the graph structure to help further simulate the relationship between brain regions. Various subject-dependent and subject-independent experiments were performed on two datasets, DEAP and SEED, to evaluate the effectiveness of the model and optimization method. The experimental results demonstrate the efficacy of the proposed approach.

In [46], Li et al. propose a fusion graph convolutional network (FGCN) architecture for extracting and combining various relationships in EEG data to obtain a more comprehensive representation for emotion recognition. FGCN initially identifies brain connectivity features based on topology, causality, and function. Subsequently, a local fusion strategy is introduced to merge these three graphs, enabling the utilization of valuable channels with strong topological, causal, and functional relationships. Finally, a graph convolutional neural network is employed to enhance the representation of EEG data for improved emotion recognition. Experimental results on SEED and SEED-IV datasets

indicate the effectiveness of combining different graphs in enhancing emotion recognition capabilities.

III. MATERIAL AND METHODS

A. EEG SIGNALS

We downloaded EEG signals from the well-known open access DEAP [1] and MAHNOB-HCI [2] datasets, which have been widely used for emotion recognition research. Both datasets comprise simultaneous EEG and physiological signals recorded while healthy subjects were shown video clips designed to induce various emotional reactions from all quarters of the 2D valence-arousal model [47].

In the DEAP dataset, 32-channel EEG signals were recorded from healthy subjects (16 male, 16 female; age range 17 to 37 years) while they were shown short music video clips [1]. Forty music video clips thought to induce emotions encompassing all quarters of the 2D valence-arousal model had been selected from 120 clips by volunteer online rating. Each music video lasted 60 seconds, and a length deemed sufficient to induce an emotional state [48], [49]. After watching the music video clips, subjects filled the self-reported self-assessment manikin form, which quantified emotional responses in valence and arousal parameters and scored using float values between 1 (lowest) and 9 (highest). EEGs were recorded (sampling frequency 512 Hz) using the BioSemi system based on the 10-20 international recording system. In the preprocessing step, EEG signals were down-sampled to 128 Hz, and line noise, blinks, and other artifacts were removed.

In the MAHNOB-HCI dataset, 32-channel EEG signals were recorded from healthy subjects (11 male, 13 female; age range 19 to 40 years) who were shown 17 emotional video clips of lengths ranging from 34.9 to 117 seconds [2], [50]. The subjects spoke different languages and had diverse cultural and educational backgrounds. The video clips were chosen to trigger emotions in all quarters of the 2D valence-arousal model: joy, fear, disgust, sadness, amusement, and neutral. Like the DEAP dataset, each participant filled out self-reported questionnaires that rated valence and arousal measures using integer values from 1 to 9. The EEG signals were recorded (sampling frequency 256 Hz) using the BioSemi active II system based on the 10-20 international recording system. In preprocessing step, line noise, blinks, and other artifacts were removed from EEG signals. In brief, the reference of the EEG signal was first changed using the averaging method in the EEGLAB toolbox, MATLAB 2021b. The finite impulse response (FIR) low- and high-pass filters were used to remove noise below 0.5 Hz and higher than 45 Hz, respectively. Then, the signal was passed through an FIR notch filter to remove 49-51 Hz power line noise. The remaining artifacts, like blinking, were manually removed using the EEGLAB above toolbox. The EEG signals from four subjects (3 male, and 1 female) were excluded from analysis due to excessive artifacts. To standardize comparison with the DEAP dataset, the length of each EEG signal was truncated at 60 seconds from the start of stimulation.

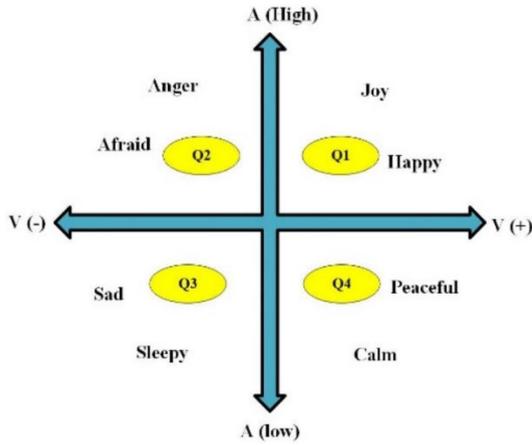


FIGURE 1. 2D valence-arousal model. Various emotions are depicted in the four quarters (Q1, Q2, Q3, and Q4) corresponding to positive and negative valence (V) as well as low and high arousal (A).

B. EMOTION CLASSES

Russell first introduced the 2D valence-arousal model to characterize the emotions [47]. Valence embodies the pleasantness of the induced feeling, which can range from positive to negative. For example, sadness is negative, whereas happiness and entertainment are positive. Arousal denotes the level of evoked emotion, ranging from low to high. For example, when a person becomes angry, the arousal level is higher than sad or bored. In the DEAP dataset, each subject's emotional class when watching each video clip was stratified into one of four quarters (Q) of the 2D valence-arousal model using self-assessed valence and arousal scores: Q1, valence, and arousal ≥ 4.5 ; Q2, valence < 4.5 and arousal ≥ 4.5 ; Q3, valence, and arousal < 4.5 ; and Q4, valence ≥ 4.5 and arousal < 4.5 (FIGURE 1).

In the MAHNOB-HCI dataset, each subject's emotional class when watching each video clip was stratified into one of four quarters (Q) of the 2D valence-arousal model using self-assessed valence and arousal scores: Q1, valence, and arousal ≥ 5 ; Q2, valence ≤ 4 and arousal ≥ 5 ; Q3, valence and arousal ≤ 4 ; and Q4, valence ≥ 5 and arousal ≤ 4 (FIGURE 1).

C. EFFECTIVE CONNECTIVITY METHODS

1) TE

TE is a nonlinear EC measure that estimates the causal interactions between two signals or anatomical brain regions based on conditional entropy. It has the advantage of not requiring a priori assumption of connectivity patterns. TE measures have been applied to detect schizophrenia [21], diagnose Alzheimer's dementia and mild cognitive impairment [51], detect drowsiness [52], as well as recognize emotions [24].

If we have two signals of $x(t)$ and $y(t)$ from a Markov process, where $x_t^m = (x_t, x_{t+1}, \dots, x_{t-m+1})$ and $y_t^n = (y_t, y_{t+1}, \dots, y_{t-n+1})$ are m and n memory (or embedding dimension) of Markov processes in x and y , respectively. The

directed information flow, TE relation, is calculated from the time series x_t to y_t as Eq. (1) [53]:

$$TE_{X \rightarrow Y} = \sum_{y_{t+1}, y_t^{d_y}, x_t^{d_x}} p(y_{t+1} | y_t^{d_y}, x_t^{d_x}) \log \left(\frac{p(y_{t+u} | y_t^{d_y}, x_t^{d_x})}{p(y_{t+u} | y_t^{d_y})} \right) \quad (1)$$

where t is the time-index (discrete value); and u the prediction time (discrete value). $y_t^{d_y}$ and $x_t^{d_x}$ denote d_y - and d_x -dimensional delay vectors are defined as in Eq. (2) and (3), wherein τ represents the time delay and is autocorrelation time of the signal.

$$x_t^{d_x} = (x(t), x(t - \tau), \dots, x(t - d_x - 1)\tau) \quad (2)$$

$$y_t^{d_y} = (y(t), y(t - \tau), \dots, y(t - d_y - 1)\tau) \quad (3)$$

TE is estimated in MATLAB 2021b using the HERMES connectivity toolbox, version 2020. The number of neighbors (note: TE is calculated through the K-Nearest Neighbor estimator), the embedding dimension, and the time delay to estimate TE was set at 4, 3, and 10, respectively.

2) DDTF

dDTF is a linear parametric EC measure that estimates common frequency components between two EEG channels [54], [55]. dDTF from j th channel to i th channel is estimated using Eq. (4):

$$dDTF_{ij} = \frac{|H_{ij}(f)|^2}{\sum_f \sum_{k=1}^M |H_{ik}(f)|^2} \times \frac{\hat{S}_{ij}(f)}{\sqrt{\hat{S}_{ii}(f) \hat{S}_{jj}(f)}} \quad (4)$$

where $H(f)$ is the transfer matrix of the system; and $S(f)$; spectral density matrix of a multi-variable auto-regressive model that is estimated from $x(t)$, i.e., $S(f) = X(f)X(f)^* = H(f)\Sigma H(f)^*$. dDTF was estimated in MATLAB 2021b using the source information flow toolbox, version 0.1a. The model order was set at 10.

3) PDC

PDC is a linear parametric EC measure in the frequency domain that estimates the causal influences between two EEG signals. PDC of i th EEG channel from j th channel at frequency component f is estimated using Eq. (5) [55], [56], [57]:

$$\pi_{ij}(f) = \frac{\bar{A}_{ij}(f)}{\sqrt{\sum_{m=1}^N \bar{A}_{mj}(f) \bar{A}_{mj}^*(f)}} \quad (5)$$

where $\bar{A}_{ij}(f)$, is the frequency component of the autoregressive coefficients of $A(a_{ij})$; and is computed using Eq. (6):

$$\bar{A}_{ij}(f) = \delta_{ij} - \sum_{r=1}^p a_{ij}(r) e^{-j2\pi fr} \quad (6)$$

where p is model order. PDC is estimated in MATLAB 2021b using the HERMES connectivity toolbox, version 2020. The model order was set at 10.

D. PRE-TRAINED CONVOLUTIONAL NEURAL NETWORKS

CNN is an effective deep learning method with an integrated structure for extracting, selecting, and classifying features that are known to yield high performance among contemporary machine learning methods [26]. In a CNN, several convolution operations extract low- and high-level features in the convolutional layer to create feature maps. Rectified linear unit is the activation function after each convolutional layer. In the pooling layer, maximum and average operations reduce feature map size using filter masks. After the fully connected layer, the softmax function calculates the probability of each sample for each class, while cross-entropy computes the loss function. Batch normalization and dropout techniques are used to regularize values and prevent overfitting. In this work, pre-trained CNNs were used as transfer learning models. These pre-trained CNNs had been trained on the large public ImageNet database, comprising more than one million images of animals, objects, etc., with 1,000 classes [58]. Transfer learning is done by considering a powerful deep network to extract comprehensive features along with its pre-trained weights on large datasets and then adapt the network weights on the target task that has limited training data. Transfer learning with pre-trained CNNs overcomes the need a lot of data for good training and prevent overfitting. This is because the pre-trained CNN model has already learned to extract informative features from natural images, which can then be a starting point to learning new study-specific tasks using a smaller dataset of the training set. Accordingly, transfer learning with pre-trained CNN networks is better than training a custom CNN model with random weights. The pre-trained CNNs were previously used for the detection of schizophrenia [16], etc.

In this work, six powerful pre-trained CNNs—ResNet-50, Inception-v3, Xception, DenseNet-201, EfficientNetB0, and NasNet-Mobile—were incorporated as transfer learning models to classify four emotion classes. We selected these pre-trained CNNs as they are designed in various architectures and were used in the previous studies [7], [16], [30], [32], [33] on EEG signals and demonstrated high performance.

In CNNs, the classification layers, which each consisted of 1,000 neurons to classify the 1,000 classes of the ImageNet database, were modified to four neurons each to classify four emotion classes. Loss function was the cross-entropy, and an adaptive moment estimation optimizer algorithm was deployed in the optimization phase.

1) RESNET-50

ResNet-50 belongs to the residual network family of CNNs, which won the ImageNet Large Scale Visual Recognition Challenge 2015 [59]. It consists of one convolutional layer with several 7×7 filters and 16 residual units that each contains two or three stacked convolutional layers (several 3×3 filters) with shortcut from the beginning to end of them. It has one max pool layer at the start and one fully connected

TABLE 1. Attributes of the pre-trained CNNs.

Net	Depth	Parameters (millions)	Input size
NasNet-Mobile	*	5.3	$224 \times 224 \times 3$
DenseNet-201	201	20	$224 \times 224 \times 3$
EfficientNetB0	82	5.3	$224 \times 224 \times 3$
Xception	71	22.9	$299 \times 299 \times 3$
ResNet-50	50	25.6	$224 \times 224 \times 3$
Inception-v3	48	23.9	$299 \times 299 \times 3$

layer at the end. This network inputs $224 \times 224 \times 3$ images and outputs 25.6 million weighted parameters after training on the ImageNet database (TABLE 1).

2) DENSENET-201

DenseNet contains a sequence of 5, 12, 48, and 32 stacked residual blocks that are separated by batch normalization, pooling, and one simple convolutional layer, respectively [60]. These residual blocks are constituted using several 3×3 and 1×1 filters. This network inputs $224 \times 224 \times 3$ images and outputs 20 million parameters after training on the ImageNet database (TABLE 1).

3) EFFICIENTNETB0

EfficientNetB0 was initially designed for light devices like mobile phones [61]. EfficientNetB0 starts with a stem unit that comprises the input, rescaling, normalization, zero-padding, convolutional, batch normalization, and activation layers. To this are added seven mobile inverted bottleneck blocks of convolutional layers with 3×3 and 5×5 kernels. This network inputs $224 \times 224 \times 3$ images and outputs 5.3 million weighted parameters after training on the ImageNet database (TABLE 1).

4) INCEPTION-V3

Inception-v3 contains five convolutional layers with several 3×3 filters and two max-pooling layers [62]. To this are added eleven Inception units, which are multiple parallel convolutional layers with various filter sizes, and batch normalization layer to reduce overfitting. This hierarchical and compact structure contributes to the efficiency of Inception-v3, which was awarded runner-up in the ImageNet Large Scale Visual Recognition Challenge 2015. This network inputs $299 \times 299 \times 3$ images and outputs 23.9 million weighted parameters after training on the ImageNet database (TABLE 1).

5) XCEPTION

Xception is an advanced version of Inception that was designed by Google [63]. It contains two simple convolutional layers with multiple 3×3 filters and twelve separable

convolutional units (instead of an Inception unit with shortcuts like in residual networks). Separable convolutional units contain stacked groups and simple convolutional layers with various filter sizes, and each unit has a residual shortcut. Xception inputs $299 \times 299 \times 3$ images and outputs 22.9 million weighted parameters after training on the ImageNet database (TABLE 1).

6) NASNET-MOBILE

NasNet-Mobile, designed by Google, possesses the largest and deepest architecture among the pre-trained CNN models [64]. NasNet-Mobile has stem units that contain many separable groups of convolutional layers with nested 11×11 , 5×5 , 7×7 , and 3×3 filters. This network uses a search strategy to evolve the best architecture. It inputs $224 \times 224 \times 3$ and outputs 20 million weighted parameters after training on the ImageNet database (TABLE 1).

E. PROPOSED EMOTION RECOGNITION SYSTEM BASED ON COMBINED PRE-TRAINED CNN AND LSTM MODELS

EEG signals from the DEAP and MAHNOB-HCI datasets were input into the model. TE (nonlinear nonparametric EC measure in the time domain) and PDC and dDTF (linear parametric EC measures in the time and frequency domains) were used to estimate the causal influences between two EEG signals. We chose these EC measures based on their successful individual performances in extracting features from multichannel EEG signals [55], [56], [57]. In the feature extraction phase of our model, TE, PDC, and dDTF EC measures were first estimated from 5-second time windows of 1D EEG signals from every one of the 32 channels and used to create 2D images. To add temporal information, the 2D images derived from EC measures estimated from three consecutive overlapping 5-second time windows were arranged horizontally in series. The series of time-resolved 2D image representations for the dDTF, PDC, and TE functions were then stacked vertically (dDTF, PDC, and TE from top to bottom) to form a novel three-dimensional connectivity image that effectively fuses the serial three-phase time windows of all three EC measures (FIGURE 2). The input signal length was 60 seconds. There was 80% overlap between pairs of consecutive time windows. Accordingly, 56-time frames and $32 \times 32 \times 56$ arrays for each EC measure and 18 (= 56-time frames/3) fused connectivity images with size 96×96 were obtained from the simultaneous 32-channel EEG signals during each viewed video clip. These 18 fused connectivity images effectively capture the spatial and time-resolved directionalities of all three EC measures estimated from multichannel 1D EEG signals acquired in response to each experimentally induced emotional state. EC measures were computed in MATLAB 2021b programming environment—TE and PDC using HERMES connectivity toolbox, version 2020; and dDTF, source information flow toolbox plugin to EEGLAB toolbox—on a laptop with Intel (R) Core (TM) i7-6500U and @2.50 GHz 2.60 GHz CPU.

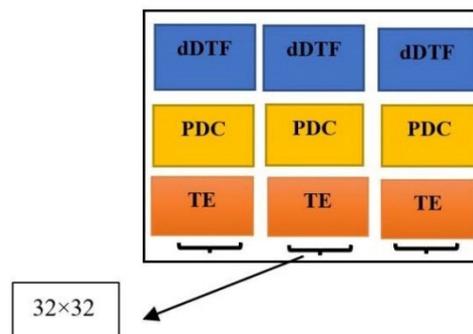


FIGURE 2. Creation of a new fused connectivity image of size 96×96 based on the fusion of a series of three-time windows of 2D image represents converted from 1D 32-channel EEG signals using three different EC measures (PDC, dDTF, and TE). Each window is an EC measure with a size of 32×32 . Three consecutive 5-second windows from each EC measure were arranged horizontally, and PDC, dDTF, and TE were arranged vertically to construct an input fused EC image.

Six pre-trained CNNs—ResNet-50, Inception-v3, Xception, DenseNet-201, EfficientNetB0, and NasNet-Mobile—were deployed as transfer learning models to classify fused EC images derived from multichannel EEG signals in the DEAP and MAHNOB-HCI datasets into four emotion classes of the 2D valence-arousal model using leave-one-subject-out (LOSO) cross-validation. The fused EC images were input to each pre-trained CNN model separately to fine-tune its parameters. As the size of each fused EC image was 96×96 and not matched to the pre-trained CNN model's input, these images were expanded to the required input image sizes of the individual pre-trained CNNs (Table 1). A simple ensemble model based on a majority voting algorithm [30], [50] of results was deployed to select the best results by pre-trained CNN networks.

As we know, LSTM is a RNN that learns the dependency of time series to apply classification [28], [29]. This method achieved good performance in schizophrenia detection [15] and emotion recognition [28], [29] studies. Hence, it is a good choice to improve recognition of four emotional classes. Combining pre-trained CNN and LSTM can increase decision ability by using deep extracted features from CNN and considering the dependency of time samples from the LSTM model. In this study, every five sequential fused images (size = $224 \times 224 \times 3$) were used as input to the CNN-LSTM model. Then, the Time Distributed layer of CNN models was used to handle all five images across the temporal dimension. 128 features were extracted from each sample input. Bidirectional LSTM layers used in this study exploited the most advantageous features from each input sample by analyzing it in both directions of the temporal dimension. 32 features were extracted from the first and second bidirectional LSTM layers. Then, the sigmoid activation function classified in to four classes (Q1, Q2, Q3, and Q4). After that, another ensemble approaches using the same method vote decides the final class. The training was run on a GPU with 1.08 GB/12.68 GB RAM from Google Colaboratory on Python programming. Python programs are available

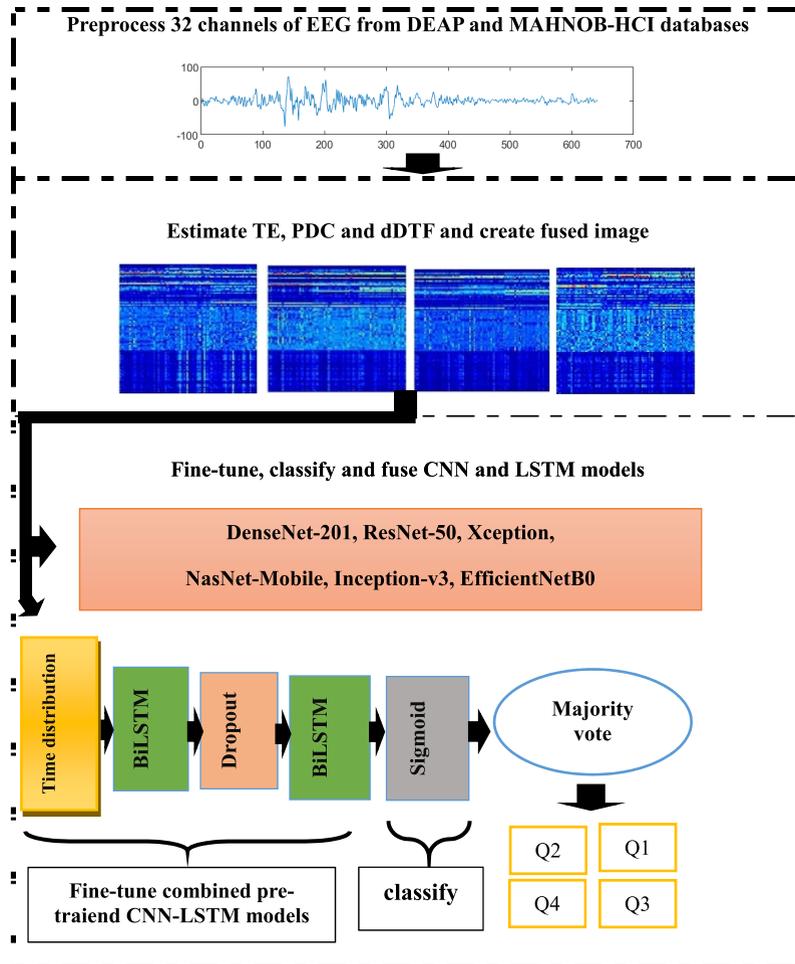


FIGURE 3. Block diagram of the proposed emotion recognition system based on fused connectivity images and ensemble of combined pre-trained CNN-LSTM models.

on GitHub link (<https://github.com/SARAHbagh65/fused-effective-connectivity-and-pre-trained-cnns>). FIGURE 3 depicts the workflow of the proposed emotion recognition system.

F. STATISTICAL METRICS

LOSO cross-validation was used to evaluate the classification performance of the four classes of emotions, i.e., Q1, Q2, Q3, and Q4. At each repetition, data from all subjects except for one subject were used to train the model; and data from the latter to test the model. This process was performed 32 and 20 times for the DEAP and MAHNOB-HCI datasets, respectively, corresponding to the number of analyzable subjects. For the DEAP dataset, 22,320 images (18 fused EC images per each of 40 videos in 31 subjects) were used as the training set for each model per repetition; and 720 images (18 fused EC images per each of 40 videos in one subject) were used for testing. For the MAHNOB-HCI dataset, 5,814 images (18 fused EC images per each of 17 videos in 19 subjects) were used as the training set for each model per repetition; and 306 images (18 fused EC images per each of 17 videos in one subject) were used for testing.

TABLE 2. Details of DEAP and MAHNOB-HCI databases.

Name	# Train	# Test	Total
DEAP	22320	720	23040
MAHNOB-HCI	5814	306	6120

The means and standard deviations of standard performance metrics—accuracy, sensitivity, precision, and F-score [66]—were calculated to evaluate the model. TABLE 2 summarizes the two final datasets used in the manuscript, including the total number of 2D fused images generated and used to train and evaluate the performance of the model.

IV. RESULTS

FIGURE 4 shows examples of fused EC images of the four emotion classes from one subject each in the DEAP and MAHNOB-HCI datasets. The four-class emotion recognition classification performance of the six pre-trained CNN models based on fused connectivity images and, by comparison, the individual TE, PDC, and dDTF measures separately, for the DEAP and MAHNOB-HCI datasets are summarized in TABLES 3 and 4, respectively. For the DEAP dataset, the

TABLE 3. Performance metrics of individual pre-trained CNN models and the ensemble of best model results, stratified by use of fused connectivity images or individual TE, PDC, dDTF measures, for the classification of four emotion classes in the DEAP dataset using LOSO Cross-Validation. Values are presented as means \pm standard deviations.

Image	CNN	Accuracy	Precision	Sensitivity	F-score
Fused	NasNet-Mobile	96.25 \pm 0.54	96.21 \pm 0.62	96.32 \pm 0.53	96.26 \pm 0.59
	EC				
EC	EfficientNetB0	95.55 \pm 0.62	95.33 \pm 0.54	95.66 \pm 0.56	95.49 \pm 0.63
	ResNet-50	95.18 \pm 0.62	95.25 \pm 0.65	95.04 \pm 0.67	95.15 \pm 0.65
	DenseNet-201	94.78 \pm 0.66	94.85 \pm 0.62	94.70 \pm 0.54	94.76 \pm 0.62
	Xception	94.25 \pm 0.58	94.14 \pm 0.59	94.36 \pm 0.70	94.25 \pm 0.64
	Inception-v3	92.93 \pm 0.62	92.78 \pm 0.55	92.98 \pm 0.63	92.85 \pm 0.59
	Ensemble	97.91 \pm 0.55	98.03 \pm 0.62	97.97 \pm 0.57	97.99 \pm 0.58
TE	NasNet-Mobile	92.24 \pm 0.65	92.39 \pm 0.75	92.48 \pm 0.60	92.41 \pm 0.64
	EfficientNetB0	89.97 \pm 0.75	89.82 \pm 0.64	90.10 \pm 0.65	89.92 \pm 0.65
	ResNet-50	91.33 \pm 0.71	91.42 \pm 0.63	91.27 \pm 0.66	91.35 \pm 0.70
	DenseNet-201	90.34 \pm 0.65	90.54 \pm 0.59	90.25 \pm 0.63	90.36 \pm 0.68
	Xception	89.43 \pm 0.66	89.35 \pm 0.72	89.52 \pm 0.64	89.44 \pm 0.65
	Inception-v3	89.22 \pm 0.64	98.34 \pm 0.70	89.16 \pm 0.61	89.27 \pm 0.66
	Ensemble	94.25 \pm 0.62	94.35 \pm 0.65	94.17 \pm 0.65	94.23 \pm 0.63
	Ddtf				
Ddtf	NasNet-Mobile	90.45 \pm 0.76	90.62 \pm 0.72	90.37 \pm 0.71	90.51 \pm 0.75
	EfficientNetB0	88.21 \pm 0.75	88.38 \pm 0.70	88.26 \pm 0.74	88.32 \pm 0.71
	ResNet-50	89.66 \pm 0.72	89.55 \pm 0.74	89.76 \pm 0.75	89.62 \pm 0.67
	DenseNet-201	88.15 \pm 0.74	88.37 \pm 0.73	88.24 \pm 0.72	88.29 \pm 0.76
	Xception	88.45 \pm 0.73	88.24 \pm 0.70	88.50 \pm 0.73	88.35 \pm 0.70
	Inception-v3	87.48 \pm 0.70	87.39 \pm 0.74	87.52 \pm 0.75	87.45 \pm 0.73
	Ensemble	93.79 \pm 0.78	93.64 \pm 0.72	93.80 \pm 0.69	93.74 \pm 0.75
PDC	NasNet-Mobile	85.50 \pm 0.87	85.62 \pm 0.85	85.39 \pm 0.82	85.55 \pm 0.79
	EfficientNetB0	84.25 \pm 0.82	84.19 \pm 0.85	84.39 \pm 0.83	84.26 \pm 0.80
	ResNet-50	83.02 \pm 0.85	82.89 \pm 0.83	83.12 \pm 0.84	82.95 \pm 0.85
	DenseNet-201	82.75 \pm 0.80	82.70 \pm 0.88	82.83 \pm 0.85	82.79 \pm 0.84
	Xception	82.10 \pm 0.88	82.25 \pm 0.82	81.92 \pm 0.79	82.14 \pm 0.80
	Inception-v3	81.84 \pm 0.87	81.75 \pm 0.80	81.92 \pm 0.76	81.89 \pm 0.86
	Ensemble	87.56 \pm 0.84	87.63 \pm 0.85	87.48 \pm 0.87	87.60 \pm 0.85

ensemble of NasNet-Mobile, EfficientNetB0, and ResNet-50 yielded the best results for fused connectivity images, attaining accuracy and F-score of 97.91% and 97.99%, respectively. For the MAHNOB-HCI database, the ensemble of NasNet-Mobile, EfficientNetB0, and ResNet-50 yielded the best results for fused connectivity images, attaining accuracy and F-score of 98.03% and 98.03%, respectively.

Pre-trained on the ImageNet database, the fully connected layers of the used CNN models contained 1,000 neurons necessary for classifying the 1,000 classes in ImageNet. These were changed to four neurons to perform the classification of four emotion classes in this work, and the classification layers

were replaced by new ones to match the new fully connected layers. In addition, weight learns rate factor and bias learn rate factor values were increased from 1 to 20 to accelerate the fine-tuning process in both layers compared with the transferred (unchanged) layers. The training curves—average accuracy for LOSO cross-validations with increasing numbers of training epochs—for all pre-trained CNN models on the DEAP (FIGURE 5) and MAHNOB-HCI (FIGURE 6) datasets show exponential behavior before plateauing to reaching the highest accuracy rates, which implies stability of all models. The initial learning rate was set at 0.0004 for the five pre-trained CNN models (at single approach and

TABLE 4. Performance metrics of individual pre-trained CNN models and the ensemble of best model results, stratified by use of fused connectivity images or individual TE, PDC, dDTF measures, for the classification of four emotion classes in the MAHNOB-HCI dataset using LOSO Cross-Validation. Values are presented as means \pm standard deviations.

Image	CNN	Accuracy	Precision	Sensitivity	F-score	
Fused	NasNet-Mobile	96.40 \pm 0.63	96.41 \pm 0.62	96.40 \pm 0.64	96.40 \pm 0.54	
	EC	EfficientNetB0	96.04 \pm 0.55	95.94 \pm 0.72	96.13 \pm 0.59	96.05 \pm 0.59
		ResNet-50	95.79 \pm 0.56	95.60 \pm 0.60	95.84 \pm 0.67	95.74 \pm 0.54
		DenseNet-201	95.26 \pm 0.62	95.15 \pm 0.55	95.42 \pm 0.55	95.30 \pm 0.59
		Xception	94.82 \pm 0.60	94.95 \pm 0.59	94.72 \pm 0.64	94.64 \pm 0.60
		Inception-v3	94.25 \pm 0.54	94.12 \pm 0.61	94.34 \pm 0.60	94.20 \pm 0.53
		Ensemble	98.03 \pm 0.54	98.03 \pm 0.53	98.03 \pm 0.55	98.03 \pm 0.57
		TE	NasNet-Mobile	94.42 \pm 0.72	94.65 \pm 0.67	94.35 \pm 0.76
EfficientNetB0	93.20 \pm 0.67		93.05 \pm 0.76	93.36 \pm 0.56	93.24 \pm 0.75	
ResNet-50	92.89 \pm 0.76		92.78 \pm 0.73	92.90 \pm 0.76	92.85 \pm 0.72	
DenseNet-201	92.64 \pm 0.73		92.59 \pm 0.68	92.65 \pm 0.72	92.63 \pm 0.65	
Xception	92.39 \pm 0.72		92.45 \pm 0.65	92.35 \pm 0.74	92.40 \pm 0.72	
Inception-v3	91.67 \pm 0.75		91.72 \pm 0.72	91.57 \pm 0.75	91.68 \pm 0.65	
Ensemble	95.79 \pm 0.67		95.86 \pm 0.75	95.54 \pm 0.65	95.63 \pm 0.65	
dDTF	NasNet-Mobile		92.35 \pm 0.57	92.26 \pm 0.78	92.45 \pm 0.65	92.34 \pm 0.65
	EfficientNetB0	91.52 \pm 0.64	91.63 \pm 0.64	91.44 \pm 0.62	91.49 \pm 0.63	
	ResNet-50	90.89 \pm 0.68	90.92 \pm 0.62	90.80 \pm 0.72	90.87 \pm 0.59	
	DenseNet-201	90.55 \pm 0.67	90.46 \pm 0.76	90.62 \pm 0.74	90.57 \pm 0.65	
	Xception	89.93 \pm 0.65	89.89 \pm 0.72	89.96 \pm 0.65	89.87 \pm 0.64	
	Inception-v3	89.75 \pm 0.72	89.82 \pm 0.74	89.69 \pm 0.76	89.73 \pm 0.71	
	Ensemble	93.82 \pm 0.76	93.75 \pm 0.69	93.89 \pm 0.65	93.83 \pm 0.76	
	PDC	NasNet-Mobile	91.25 \pm 0.64	91.35 \pm 0.65	91.16 \pm 0.66	91.30 \pm 0.64
EfficientNetB0		90.34 \pm 0.65	90.27 \pm 0.60	90.45 \pm 0.68	90.33 \pm 0.68	
ResNet-50		90.04 \pm 0.70	90.12 \pm 0.73	89.95 \pm 0.65	90.06 \pm 0.73	
DenseNet-201		88.94 \pm 0.73	89.21 \pm 0.63	88.82 \pm 0.64	88.98 \pm 0.74	
Xception		88.55 \pm 0.67	88.42 \pm 0.75	88.64 \pm 0.59	88.53 \pm 0.76	
Inception-v3		87.92 \pm 0.66	87.76 \pm 0.71	87.97 \pm 0.71	87.85 \pm 0.71	
Ensemble		91.74 \pm 0.64	91.60 \pm 0.68	91.46 \pm 0.64	91.53 \pm 0.62	

combination approach), and squared gradient decay factor, max epochs, and mini-batch size were set at 0.99, 40, and 32, respectively.

Then, the fusion of pre-trained CNN models and LSTM was performed to improve the accuracy of the best results of the previous phase, i.e., on fused images obtained from three connectivity measures of TE, PDC, dDTF. All the CNN-based models were trained under same condition. TABLES 5 and 6 show the results obtained for various CNN-based models using DEAP and MAHNOB-HCI databases, respectively. It can be observed from both tables

that the highest accuracy of 96.83% and 97.12% was obtained using a combination of NasNet-Mobile-LSTM for DEAP and MAHNOB-HCI databases, respectively. Also, using ensemble model the accuracy increased to 98.76% and 98.86% for DEAP and MAHNOB-HCI, respectively.

V. DISCUSSION

In this research, a new fused connectivity image constructed from multichannel EEG signals using nonlinear TE and linear PDC and dDTF methods were combined with transfer learning and applied for the automatic classification of four

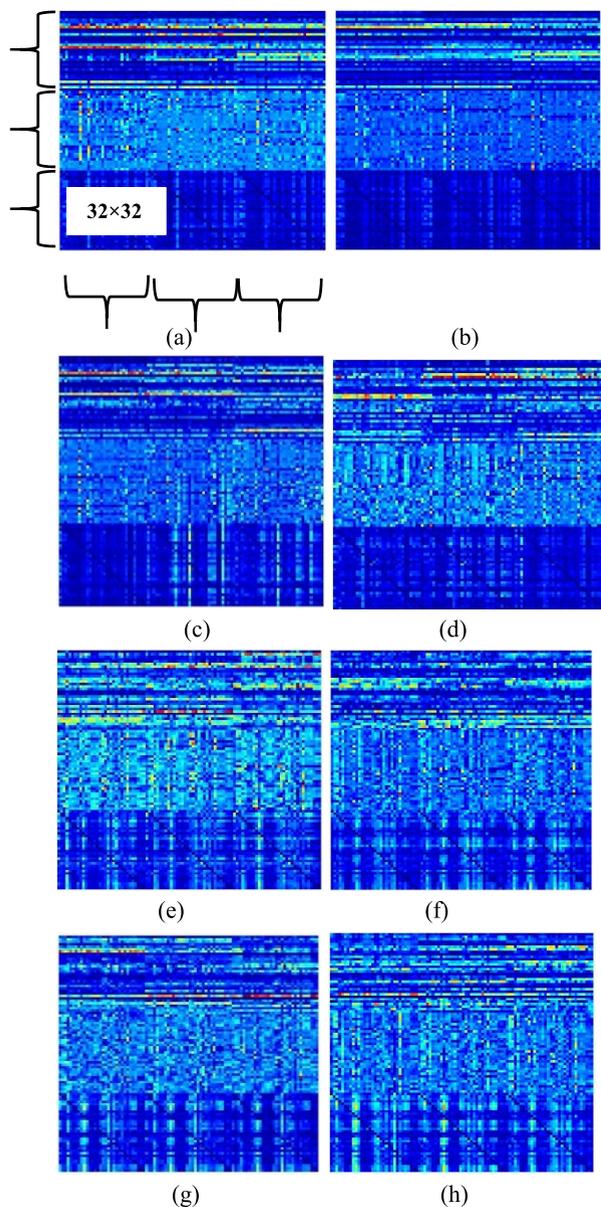


FIGURE 4. Example fused connectivity images in the emotion classes Q1 (a, e), Q2 (b, f), Q3 (c, g), and Q4 (d, h) from one subject each in the DEAP (a to d) and MAHNOB-HCI (e to h) datasets. EC measures were extracted from 5-second 1D EEG signals to create 2D images corresponding to the time windows. Each individual time window occupying one-third was an EC measure with a size of 32×32 . EC values assigned to the individual elements represented the causal influences between all pair combinations of the 32 simultaneous EEG signal channels at any one time window. The fused connectivity images were created from the fusion of three consecutive time windows of all three EC measures (PDC, dDTF, and TE) and therefore have a size of 96×96 .

emotional classes. Six powerful CNN models—ResNet-50, Inception-v3, Xception, DenseNet-201, EfficientNetB0, and NasNet-Mobile—pre-trained on the huge ImageNet database were deployed to extract deep features from the novel fused connectivity images for downstream classification. To enhance the model’s accuracy and stability, the predictions of these pre-trained CNN models were merged, and the best ensemble results were subsequently selected by majority

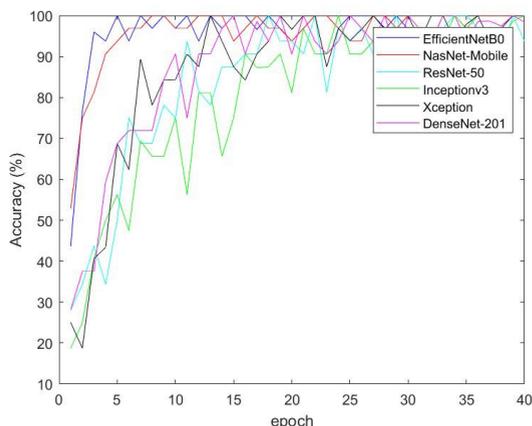


FIGURE 5. The training process for EfficientNetB0, NasNet-Mobile, ResNet-50, Xception, Inception-v3 and DenseNet-201 on the DEAP dataset. The vertical axis depicts accuracy values, and the horizontal axis shows epochs for every pre-trained CNN model during the training process.

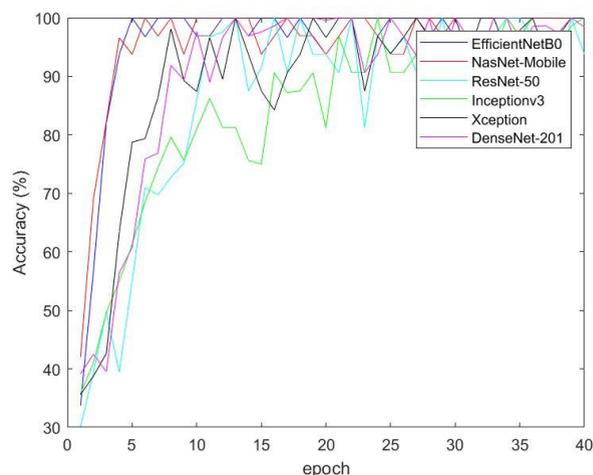


FIGURE 6. The training process for EfficientNetB0, NasNet-Mobile, ResNet-50, Xception, Inception-v3 and DenseNet-201 on the MAHNOB-HCI dataset. The vertical axis depicts accuracy values, and the horizontal axis shows epochs for every pre-trained CNN model during the training process.

voting. As a result, the ensemble of EfficientNetB0, NasNet-Mobile, and ResNet-50 yielded superior performance with fused connectivity images, attaining average accuracy rates of 97.91% and 98.03% on the DEAP and MAHNOB-HCI datasets, respectively. Also, the combination of CNNs and LSTM improved the final accuracy and achieved 98.76% for the DEAP database and 98.86% for the MAHNOB-HCI database. It means the time dependency that LSTM added to good features of CNNs improved the performance of our proposed emotion recognition system.

Without fused connective image, among the three individual EC methods, TE measures attained the best average four-class classification accuracy in terms of ensemble results and individual pre-trained network performance for both datasets (TABLES 3 and 4). A potential explanation for this superior performance compared with PDC and dDTF is that

TABLE 5. Performance metrics (means \pm standard deviations) obtained for various combinations of CNN-based models using DEAP database with LOSO cross-validation strategy.

Net	Accuracy (%)	Precision (%)	Sensitivity (%)	F-score (%)
NasNet-Mobile-LSTM	96.83 \pm 0.55	96.75 \pm 0.63	96.89 \pm 0.57	96.86 \pm 0.54
EfficientNetB0-LSTM	96.12 \pm 0.57	95.85 \pm 0.55	96.20 \pm 0.61	95.93 \pm 0.53
ResNet-50-LSTM	95.75 \pm 0.64	95.82 \pm 0.63	95.63 \pm 0.71	95.67 \pm 0.65
DenseNet-201-LSTM	95.27 \pm 0.62	95.35 \pm 0.66	95.13 \pm 0.58	95.30 \pm 0.62
Xception-LSTM	94.64 \pm 0.55	94.42 \pm 0.67	94.72 \pm 0.71	94.58 \pm 0.59
Inception-v3-LSTM	93.51 \pm 0.57	93.38 \pm 0.64	93.60 \pm 0.62	93.46 \pm 0.63
Ensemble	98.76\pm0.53	98.57\pm0.61	98.79\pm0.58	98.66\pm0.62

TABLE 6. Performance metrics (means \pm standard deviations) obtained for various combinations of CNN-based models using MAHNOB-HCI database with LOSO cross-validation strategy.

Net	Accuracy (%)	Precision (%)	Sensitivity (%)	F-score (%)
NasNet-Mobile-LSTM	97.12 \pm 0.59	97.25 \pm 0.65	96.89 \pm 0.61	97.15 \pm 0.61
EfficientNetB0-LSTM	96.63 \pm 0.63	96.54 \pm 0.71	96.73 \pm 0.65	96.62 \pm 0.64
ResNet-50-LSTM	96.24 \pm 0.55	96.17 \pm 0.63	96.35 \pm 0.58	96.26 \pm 0.61
DenseNet-201-LSTM	95.85 \pm 0.63	95.64 \pm 0.59	95.89 \pm 0.66	95.74 \pm 0.62
Xception-LSTM	95.45 \pm 0.66	95.63 \pm 0.54	95.57 \pm 0.65	95.59 \pm 0.57
Inception-v3-LSTM	95.03 \pm 0.72	94.69 \pm 0.68	95.12 \pm 0.68	94.78 \pm 0.65
Ensemble	98.86\pm0.57	98.83\pm0.59	98.79\pm0.61	98.88\pm0.64

TE is based on the information theory and estimates the nonlinear relationships of transfer of information between pairs of EEG channels. In contrast, PDC and dDTF are linear measures and unable to extract hidden nonlinear information from the data. In the absence of ensemble majority voting, the NasNet-Mobile network yielded the best performance among individual pre-trained CNN models, attaining accuracy rates of 96.25 and 96.40% on fused connectivity images in the DEAP and MAHNOB-HCI datasets, respectively (Tables 3 and 4). This pre-trained CNN model uses reinforcement learning to seek out the best architecture of cells and is comparatively lightweight computationally, despite its deep architecture.

The main contribution of this work is a novel method that fuses three established brain EC measures—dDTF, PDC, and TE—to represent spatially and temporally resolved multichannel 1D EEG signals as compact images that can be input to pre-trained CNN models for downstream classification using transfer learning. By way of comparison, time-spectral methods [67], [68] such as wavelet transform, is the classical time-frequency method used to convert 1D signals into 2D image representations of the derived time-frequency image to feed to pre-trained CNN transfer learning models. The EC methods employed in this study have been successfully

applied to process EEG signals for the detection of various mental conditions [55], [56], [57]. In the current work, these measures are distinguished by their ability to provide linear and nonlinear high-level contents related to emotional state-induced time-resolved information flow between spatially distributed signal channels located at separate brain regions. The excellent classification performance of the proposed model conclusively demonstrates its effectiveness for emotion recognition.

We performed a non-systemic literature review to compare our model performance with recent related studies on EEG signal-based emotion recognition using the same datasets. The results are summarized in TABLE 7. Our proposed model clearly outperformed related studies in the literature. Our models spent more computational time and load in the training phase in comparison of traditional machine learning methods like SVM or decision tree, but they are quick in front of unknown samples (test set). Therefore, improvements around 1% and 3% on accuracy rates have its worth to be applied in real-World scenario. In summary, the excellent performance provides support for (1) the ability of our proposed fusion EC method to construct discriminative images from multichannel EEG signals that accurately represent the comprehensive spatially and temporally resolved brain

TABLE 7. Recent studies (in chronological order) of EEG-based emotion recognition on the DEAP and MAHNOB-HCI datasets.

Works	Dataset	method	# Classes	Evaluation method	Accuracy (%)
[34]	DEAP	3D spatial-spectral features, DBCN	2 ¹	LOSO CV	83.98 (valence), 79.45 (arousal)
[36]	DEAP	DE, SampEn, Hjorth parameter, PSD, CNN+inception structure, SVM	2	LOSO CV	80.52 (valence), 75.22 (arousal)
[37]	DEAP	Attention based CNN+ LSTM	2	LOSO CV	65.9 (valence), 69.5 (arousal)
[38]	DEAP	SpTemp features, 2D matrix mapping, CapsNet	2	LOSO CV	58.52 (arousal), 48.21 (valence)
[40]	DEAP	Different Feature Extraction Methods, EEG-GCN	2	10-fold CV	81.77 (valence), 81.95 (arousal)
[42]	MAHNOB-HCI, DEAP	Short-term continuity modeling, Hierarchical self-attention network	4 (Q1-Q4)	LOSO CV	59.03 (MAHNOB-HCI), 51.48 (DEAP)
[25]	DEAP	PCA, DNN-CNN	2	Hold out CV	89.49 (valence), 92.86 (arousal)
[27]	DEAP	Multi-column CNN	2	Hold out CV	90.01 (valence), 90.65 (arousal)
[28]	DEAP	Stacked AE, PSD ² , Pearson correlation coefficients, LSTM	2	k-fold CV	81.10 (valence), 74.38 (arousal)
[29]	DEAP	DE, Graph CNN, LSTM	2	LOSO CV	84.81 (valence), 85.27 (arousal)
[32]	DEAP, MAHNOB-HCI	PDC, pre-trained CNNs	5 (Q1-Q4 and neutral)	LOSO CV	94.27 (DEAP), 95.25 (MAHNOB-HCI)
[33]	DEAP, MAHNOB-HCI	dDTF, pre-trained CNNs	5 (Q1-Q4 and neutral)	k-fold CV	97.78
[43]	DEAP	Variational Mode Decomposition, empirical mode decomposition, Feature Extraction, CNN	2	Hold out	62.50 (valence), 61.25 (arousal)
[69]	DEAP	Coordinate Attention, AP-CapsNet ³ , MobileNet	2	k-fold CV	93.89 (valence), 95.04 (arousal)
[70]	DEAP	Z-score normalization, LP-1D-CNN ⁴	2	k-fold CV	98.43 (arousal), 97.65 (arousal)
[71]	DEAP	ID sequences to 2D frames, ICA ⁵ , Temporal relative encoding	5	k-fold on each subject	93.99
[72]	DEAP	Few-shot model, 3D CNN-Bi-LSTM	2	k-fold CV	69.92 (valence), 68.89 (arousal)
This work	DEAP, MAHNOB-HCI	Fused EC image, pre-trained CNNs-LSTMs, majority voting	5 (Q1-Q4 and neutral)	LOSO CV	98.76 (DEAP), 98.86 (MAHNOB-HCI)

¹ 2 classes: low valence/high valence; low arousal/high arousal, ²PSD= power spectral density, ³attention mechanism and the pre-trained convolutional CapsNet network, ⁴LP-1D-CNN= lightweight pyramidal 1D-CNN, ⁵ICA= independent component analysis

connectivity during induced emotional states; (2) the use of pre-trained transfer learning models; (3) the use of majority voting algorithm to enhance model accuracy and stability; and (4) the combined pre-trained transfer learning models and LSTM is used to improve model accuracy.

The main limitation of the study is the small sample size, which can render the developed deep learning model vulnerable to bias and less generalizable. In this study, we mitigated this problem using transfer learning, in which the CNNs were pre-trained on the large ImageNet database. Another limitation of our proposed model is the high computational cost associated with the fine-tuning and evaluation steps involved in LOSO cross-validation. While the ensemble results on the fused connectivity images were the best, almost similar results were obtained with the comparatively computationally lightweight NasNet-Mobile network, which may be more

suitable for real-world implementation. Lastly, the ground truth of the four emotion classes in the 2D valence-arousal model for both DEAP and MAHNOB-HCI datasets was established through self-reported questionnaires and not by experts. However, this practice has been long-established. The final drawback of this study is that our emotion recognition systems do not recognize individual emotions per se (e.g., we cannot state that our system predicts with 98% accuracy that the subject is angry). These models classify signals according to differences in their level of valence/arousal, and those levels are associated with different emotional states.

A. LIMITATIONS OF STUDY

In this paper, pre-trained models for emotion recognition using EEG signals are presented. The study uses two datasets, DEAP and MAHNOB-HCI, which have a limited number

of subjects for emotion recognition. Due to this limitation, advanced DL models like self-supervised learning techniques cannot be employed. In some medical research, it has been shown that emotion recognition is crucial in the diagnosis of some brain disorders such as schizophrenia (SZ), although currently available EEG datasets for diagnosing such disorders have not been provided based on emotion recognition, which is another limitation. In the processing step of this work, three effective connectivity techniques including TE, PDC and dDTF have been exploited and fused with each other. Although these methods perform better than functional connectivity techniques, they have a higher computational cost. As mentioned in the previous sections, the fusion of TE, PDC and dDTF images is the first novelty of this work, making this section more complex than using functional connectivity methods. In the following, the latest pre-trained architectures are used, which is another novelty of this paper. These networks have achieved satisfactory results with low input data, but their initial training takes some time, and they cannot be used for 1D EEG signals. These architectures are trained on ImageNet dataset, but if they are trained on medical data, their results can be more reliable.

VI. CONCLUSION

In this study, we proposed a new method to represent multichannel 1D EEG signals as images using fusion of three well-known brain EC measures (dDTF, PDC, and TE). Then, the fused images were fed to pre-trained CNN models to classify four emotional classes via transfer learning. NasNet-Mobile outperformed other base pre-trained models, attaining accuracy rates of 96.25 and 96.40% using fused connectivity images of EEG signals from the DEAP and MAHNOB-HCI datasets, respectively. Then, an ensemble model based on the best pre-trained CNN models (EfficientNetB0, NasNet-Mobile, and ResNet-50) on fused connectivity images yielded the best performance, attaining average accuracy rates of 97.91% and 98.03% on the DEPA and MAHNOB-HCI datasets, respectively. Finally, the individual pre-trained CNNs are combined with LSTM models to improve the accuracy of recognition of emotional classes. ensemble of the outputs of the three best models (EfficientNetB0-LSTM, NasNet-Mobile-LSTM, and ResNet-50-LSTM), increased the accuracy to 98.76% and 98.86% for DEAP and MAHNOB-HCI databases, respectively. The excellent performance achieved in this study demonstrates the discriminative utility of the model compared to other deep and machine learning studies, and lends support to its real-world implementation for automated emotion recognition.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Sara Bagherzadeh: Conceptualization, Methodology, Software, Investigation, Validation, Writing – original draft, Writing – review & editing. **Ahmad Shalbaf:** Conceptualization, Methodology, Investigation, Validation, Supervision,

Writing – original draft. **Afshin Shoeibi:** Writing – original draft. **Mahboobeh Jafari:** Writing – original draft. **Ru-San Tan:** Investigation, Validation, Writing – original draft. **U. Rajendra Acharya:** Investigation, Validation, Writing – original draft.

STATEMENTS AND DECLARATIONS

COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

APPENDIX

Algorithm Pseudocode of the proposed method based on the fusion of effective connectivity measures and pre-trained CNN models to classify emotional states.

Input: 32 EEG channels from DEAP and MAHNOB-HCI databases

Output: four emotional classes

-
- 1- Preprocess 32 EEG channels
 - 2- *for* $i = 1$ to 32 *do*
 - 3- *for* $j =$ to 40 video clips *do* (this number pertains to DEAP; the corresponding number is 17 for MAHNOB-HCI)
 - 4- Estimate TE, PDC and dDTF from five-second windows
 - 5- Build images based on fusion of the three estimation methods from step 4
 - 6- *end for* j
 - 7- Fine-tune six pre-trained CNNs of DenseNet-201, ResNet-50, NasNet-Mobile, Inception-v3, Xception, EfficientNetB0 using LOSO cross-validation separately,
%% [In LOSO cross-validation, CNNs are trained on images of 31 subjects (this number pertains to DEAP; the corresponding number is 19 for MAHNOB-HCI) and tested on images of the remaining one at each repetition]
 - 8- Evaluate six pre-trained CNNs on images of the one unseen participant
 - 9- Calculate final means and standard deviations of accuracy, precision, sensitivity, and F-score
 - 10- *end for* i
 - 11- Vote emotion classes based on majority from outputs from three or five pre-trained CNNs toward increasing final decision
 - 12- Calculate final means and standard deviations of accuracy, precision, sensitivity and F-score
 - 13- Combine pre-trained CNNs with a designed LSTM model
 - 14- Calculate final means and standard deviations of accuracy, precision, sensitivity and F-score
 - 15- Vote emotion classes based on majority from outputs from three or five pre-trained CNN-LSTMs toward increasing final decision
 - 16- Calculate final means and standard deviations of accuracy, precision, sensitivity, and F-score
-

REFERENCES

- [1] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.
- [2] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 42–55, Jan. 2012.
- [3] R. Yuvaraj, M. Murugappan, U. R. Acharya, H. Adeli, N. M. Ibrahim, and E. Mesquita, "Brain functional connectivity patterns for emotional state classification in Parkinson's disease patients without dementia," *Behavioural Brain Res.*, vol. 298, pp. 248–260, Feb. 2016.
- [4] J. De Lope and M. Graña, "A hybrid time-distributed deep neural architecture for speech emotion recognition," *Int. J. Neural Syst.*, vol. 32, no. 6, Jun. 2022, Art. no. 2250024.
- [5] S. Akdemir Akar, S. Kara, S. Agambayev, and V. Bilgiç, "Nonlinear analysis of EEGs of patients with major depression during different emotional states," *Comput. Biol. Med.*, vol. 67, pp. 49–60, Dec. 2015.
- [6] M. Murugappan, W. Alshuaib, A. K. Bourisly, S. K. Khare, S. Sruthi, and V. Bajaj, "Tunable Q wavelet transform based emotion classification in Parkinson's disease using electroencephalography," *PLoS One*, vol. 15, no. 11, Nov. 2020, Art. no. e0242014.
- [7] M. S. Shahabi, A. Shalhaf, and A. Maghsoudi, "Prediction of drug response in major depressive disorder using ensemble of transfer learning with convolutional neural network based on EEG," *Biocybern. Biomed. Eng.*, vol. 41, no. 3, pp. 946–959, Jul. 2021.
- [8] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG," *Comput. Biol. Med.*, vol. 41, no. 12, pp. 1110–1117, Dec. 2011.
- [9] S. S. Panicker and P. Gayathri, "A survey of machine learning techniques in physiology based mental stress detection systems," *Biocybern. Biomed. Eng.*, vol. 39, no. 2, pp. 444–469, Apr. 2019.
- [10] H. Shahabi and S. Moghimi, "Toward automatic detection of brain responses to emotional music through analysis of EEG effective connectivity," *Comput. Hum. Behav.*, vol. 58, pp. 231–239, May 2016.
- [11] A. Subasi, T. Tuncer, S. Dogan, D. Tanko, and U. Sakoglu, "EEG-based emotion recognition using tunable q wavelet transform and rotation forest ensemble classifier," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102648.
- [12] W. X. Cheng, R. Gao, P. N. Suganthan, and K. F. Yuen, "EEG-based emotion recognition using random convolutional neural networks," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105349.
- [13] F. Afshani, A. Shalhaf, R. Shalhaf, and J. Sleigh, "Frontal-temporal functional connectivity of EEG signal by standardized permutation mutual information during anesthesia," *Cognit. Neurodynamics*, vol. 13, no. 6, pp. 531–540, Dec. 2019.
- [14] A. M. Roy, "Adaptive transfer learning-based multiscale feature fused deep convolutional neural network for EEG MI multiclassification in brain-computer interface," *Eng. Appl. Artif. Intell.*, vol. 116, Nov. 2022, Art. no. 105347.
- [15] A. Shoeibi, D. Sadeghi, P. Moridian, N. Ghassemi, J. Heras, R. Alizadehsani, A. Khadem, Y. Kong, S. Nahavandi, Y.-D. Zhang, and J. M. Gorriz, "Automatic diagnosis of schizophrenia in EEG signals using CNN-LSTM models," *Frontiers Neuroinform.*, vol. 15, pp. 1–16, Nov. 2021.
- [16] S. Bagherzadeh, M. S. Shahabi, and A. Shalhaf, "Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal," *Comput. Biol. Med.*, vol. 146, Jul. 2022, Art. no. 105570.
- [17] M. R. Islam, M. A. Moni, M. M. Islam, M. Rashed-Al-Mahfuz, M. S. Islam, M. K. Hasan, M. S. Hossain, M. Ahmad, S. Uddin, A. Azad, S. A. Alyami, M. A. R. Ahad, and P. Lió, "Emotion recognition from EEG signal focusing on deep learning and shallow learning techniques," *IEEE Access*, vol. 9, pp. 94601–94624, 2021.
- [18] M. M. Rahman, A. K. Sarkar, M. A. Hossain, M. S. Hossain, M. R. Islam, M. B. Hossain, J. M. W. Quinn, and M. A. Moni, "Recognition of human emotions using EEG signals: A review," *Comput. Biol. Med.*, vol. 136, Sep. 2021, Art. no. 104696.
- [19] T. Tuncer, S. Dogan, M. Baygin, and U. Rajendra Acharya, "Tetromino pattern based accurate EEG emotion classification model," *Artif. Intell. Med.*, vol. 123, Jan. 2022, Art. no. 102210.
- [20] T. Mullen, "Source information flow toolbox (SIFT)," Swartz Center Comput. Neurosci. Inst. Neural Comput., UC San Diego, Qusp Labs, pp. 1–69, 2010.
- [21] J. Liu, H. Meng, M. Li, F. Zhang, R. Qin, and A. K. Nandi, "Emotion detection from EEG recordings based on supervised and unsupervised dimension reduction," *Concurrency Comput., Pract. Exper.*, vol. 30, no. 23, p. 4446, Dec. 2018.
- [22] M. A. Rahman, M. F. Hossain, M. Hossain, and R. Ahmmed, "Employing PCA and *t*-statistical approach for feature extraction and classification of emotion from multichannel EEG signal," *Egyptian Informat. J.*, vol. 21, no. 1, pp. 23–35, Mar. 2020.
- [23] R. Nawaz, K. H. Cheah, H. Nisar, and V. V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 910–926, Jul. 2020.
- [24] Q. Gao, C.-H. Wang, Z. Wang, X.-L. Song, E.-Z. Dong, and Y. Song, "EEG based emotion recognition using fusion feature extraction method," *Multimedia Tools Appl.*, vol. 79, nos. 37–38, pp. 27057–27074, Oct. 2020.
- [25] J. Liu, G. Wu, Y. Luo, S. Qiu, S. Yang, W. Li, and Y. Bi, "EEG-based emotion classification using a deep neural network and sparse autoencoder," *Frontiers Syst. Neurosci.*, vol. 14, p. 43, Sep. 2020.
- [26] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.
- [27] H. Yang, J. Han, and K. Min, "A multi-column CNN model for emotion recognition from EEG signals," *Sensors*, vol. 19, no. 21, p. 4736, Oct. 2019.
- [28] X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," *Frontiers Neurobot.*, vol. 13, p. 37, Jun. 2019.
- [29] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106954.
- [30] M. S. Shahabi and A. Shalhaf, "Prediction of treatment outcome in major depressive disorder using ensemble of hybrid transfer learning and long short term memory based on EEG signal," *IEEE Trans. Cognit. Develop. Syst.*, vol. 15, no. 3, pp. 1279–1288, Sep. 2023.
- [31] S. K. Khare and V. Bajaj, "Time-frequency representation and convolutional neural network-based emotion recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 2901–2909, Jul. 2021.
- [32] S. Bagherzadeh, K. Maghooli, A. Shalhaf, and A. Maghsoudi, "Recognition of emotional states using frequency effective connectivity maps through transfer learning approach from electroencephalogram signals," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103544.
- [33] S. Bagherzadeh, K. Maghooli, A. Shalhaf, and A. Maghsoudi, "Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals," *Cognit. Neurodyn.*, vol. 16, no. 5, pp. 1087–1106, Oct. 2022.
- [34] Q. Yao, H. Gu, S. Wang, and X. Li, "A feature-fused convolutional neural network for emotion recognition from multichannel EEG signals," *IEEE Sensors J.*, vol. 22, no. 12, pp. 11954–11964, Jun. 2022.
- [35] A. Topic, M. Russo, M. Stella, and M. Saric, "Emotion recognition using a reduced set of EEG channels based on holographic feature maps," *Sensors*, vol. 22, no. 9, p. 3248, Apr. 2022.
- [36] Q. Gao, Y. Yang, Q. Kang, Z. Tian, and Y. Song, "EEG-based emotion recognition with feature fusion networks," *Int. J. Mach. Learn. Cybern.*, vol. 13, no. 2, pp. 421–429, Feb. 2022.
- [37] Arjun, A. S. Rajpoot, and M. R. Panicker, "Subject independent emotion recognition using EEG signals employing attention driven neural networks," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103547.
- [38] G. C. Jana, A. Sabath, and A. Agrawal, "Capsule neural networks on spatio-temporal EEG frames for cross-subject emotion recognition," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103361.
- [39] W. K. Ngai, H. Xie, D. Zou, and K.-L. Chou, "Emotion recognition based on convolutional neural networks and heterogeneous bio-signal data sources," *Inf. Fusion*, vol. 77, pp. 107–117, Jan. 2022.
- [40] Y. Gao, X. Fu, T. Ouyang, and Y. Wang, "EEG-GCN: Spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition," *IEEE Signal Process. Lett.*, vol. 29, pp. 1574–1578, 2022.
- [41] A. Samavat, E. Khalili, B. Ayati, and M. Ayati, "Deep learning model with adaptive regularization for EEG-based emotion recognition using temporal and frequency features," *IEEE Access*, vol. 10, pp. 24520–24527, 2022.

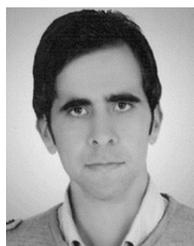
- [42] Y. Zhang, H. Liu, D. Zhang, X. Chen, T. Qin, and Q. Zheng, "EEG-based emotion recognition with emotion localization via hierarchical self-attention," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2458–2469, Jul./Sep. 2023.
- [43] P. Pandey and K. R. Seeja, "Subject independent emotion recognition from EEG using VMD and deep learning," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1730–1738, May 2022.
- [44] H. Tang, G. Jiang, and Q. Wang, "Deep neural network for emotion recognition based on meta-transfer learning," *IEEE Access*, vol. 10, pp. 78114–78122, 2022.
- [45] X. Lin, J. Chen, W. Ma, W. Tang, and Y. Wang, "EEG emotion recognition using improved graph neural network with channel selection," *Comput. Methods Programs Biomed.*, vol. 231, Apr. 2023, Art. no. 107380.
- [46] M. Li, M. Qiu, W. Kong, L. Zhu, and Y. Ding, "Fusion graph representation of EEG for emotion recognition," *Sensors*, vol. 23, no. 3, p. 1404, Jan. 2023.
- [47] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, Dec. 1980.
- [48] X. Rottenberg, R. D. Ray, and J. J. Gross, "Emotion elicitation using films," in *Handbook of Emotion Elicitation and Assessment* (Series in Affective Science). London, U.K.: Oxford Univ. Press, 2007, pp. 9–28.
- [49] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition Emotion*, vol. 24, no. 7, pp. 1153–1172, Nov. 2010.
- [50] S. Koelstra and I. Patras, "Fusion of facial expressions and EEG for implicit affective tagging," *Image Vis. Comput.*, vol. 31, no. 2, pp. 164–174, Feb. 2013.
- [51] J. McBride, X. Zhao, N. Munro, G. Jicha, C. Smith, and Y. Jiang, "Discrimination of mild cognitive impairment and Alzheimer's disease using transfer entropy measures of scalp EEG," *J. Healthcare Eng.*, vol. 6, no. 1, pp. 55–70, Mar. 2015.
- [52] C.-S. Huang, N. R. Pal, C.-H. Chuang, and C.-T. Lin, "Identifying changes in EEG information transfer during drowsy driving by transfer entropy," *Frontiers Human Neurosci.*, vol. 9, p. 570, Oct. 2015.
- [53] G. Niso, R. Bruña, E. Pereda, R. Gutiérrez, R. Bajo, F. Maestú, and F. del-Pozo, "HERMES: Towards an integrated toolbox to characterize functional and effective brain connectivity," *Neuroinformatics*, vol. 11, no. 4, pp. 405–434, Oct. 2013.
- [54] A. Korzeniewska, M. Mańczak, M. Kamiński, K. J. Blinowska, and S. Kasicki, "Determination of information flow direction among brain structures by a modified directed transfer function (dDTF) method," *J. Neurosci. Methods*, vol. 125, nos. 1–2, pp. 195–207, May 2003.
- [55] T. F. Tafreshi, M. R. Daliri, and M. Ghodousi, "Functional and effective connectivity based features of EEG signals for object recognition," *Cognit. Neurodyn.*, vol. 13, no. 6, pp. 555–566, Dec. 2019.
- [56] L. Astolfi, F. Cincotti, D. Mattia, M. G. Marciani, L. A. Baccala, F. de Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J. C. Edgar, G. A. Miller, B. He, and F. Babiloni, "Comparison of different cortical connectivity estimators for high-resolution EEG recordings," *Human Brain Mapping*, vol. 28, no. 2, pp. 143–157, Feb. 2007.
- [57] L. Astolfi, F. De Vico Fallani, F. Cincotti, D. Mattia, M. G. Marciani, S. Salinari, J. Sweeney, G. A. Miller, B. He, and F. Babiloni, "Estimation of effective and functional cortical connectivity from neuroelectric and hemodynamic recordings," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 3, pp. 224–233, Jun. 2009.
- [58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [59] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [60] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [61] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.
- [62] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.
- [63] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2016, *arXiv:1610.02357*.
- [64] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.
- [65] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Hoboken, NJ, USA: Wiley, 2014.
- [66] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.
- [67] P. Thangavel, J. Thomas, W. Y. Peh, J. Jing, R. Yuvaraj, S. S. Cash, R. Chaudhari, S. Karia, R. Rathakrishnan, V. Saini, N. Shah, R. Srivastava, Y.-L. Tan, B. Westover, and J. Dauwels, "Time-frequency decomposition of scalp electroencephalograms improves deep learning-based epilepsy diagnosis," *Int. J. Neural Syst.*, vol. 31, no. 8, Aug. 2021, Art. no. 2150032.
- [68] P. Peng, L. Xie, and H. Wei, "A deep Fourier neural network for seizure prediction using convolutional neural network and ratios of spectral power," *Int. J. Neural Syst.*, vol. 31, no. 8, Aug. 2021, Art. no. 2150022.
- [69] S. Liu, Z. Wang, Y. An, J. Zhao, Y. Zhao, and Y.-D. Zhang, "EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network," *Knowl.-Based Syst.*, vol. 265, Apr. 2023, Art. no. 110372.
- [70] M. Hussain, E.-U.-H. Qazi, H. A. AboAlSamh, and I. Ullah, "Emotion recognition system based on two-level ensemble of deep-convolutional neural network models," *IEEE Access*, vol. 11, pp. 16875–16895, 2023.
- [71] G. Peng, K. Zhao, H. Zhang, D. Xu, and X. Kong, "Temporal relative transformer encoding cooperating with channel attention for EEG emotion analysis," *Comput. Biol. Med.*, vol. 154, Mar. 2023, Art. no. 106537.
- [72] S. Bhosale, R. Chakraborty, and S. K. Koppurapu, "Calibration free meta learning based approach for subject independent EEG emotion recognition," *Biomed. Signal Process. Control*, vol. 72, Feb. 2022, Art. no. 103289.



SARA BAGHERZADEH received the Ph.D. degree in biomedical engineering from Islamic Azad University Science and Research Branch, in 2021. She is currently a Neuroscience Researcher, with a focus on EEG signal processing techniques for healthcare. Her research interests include recognition of emotional states, detection of mental disorders from EEG signal, brain-computer interface (BCI) systems, and deep learning methods based on imaging techniques.



AHMAD SHALBF received the M.S. degree in biomedical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2006, and the Ph.D. degree in biomedical engineering from Iran University of Science and Technology, in 2013. He is currently an Associate Professor of biomedical engineering with the Shahid Beheshti University of Medical Sciences, Tehran. His research interests include the biomedical signal processing and cognitive neuroscience.



AFSHIN SHOEIBI received the M.Sc. degree in telecommunication and digital signal processing engineering from the Electrical Engineering Faculty, Islamic Azad University, Gonabad Branch, in 2016. He has been a Research Intern with the Biomedical Machine Learning Laboratory, Graduate School of Biomedical Engineering, UNSW Sydney, Australia, since 2022. Additionally, he is a Research Assistant with the University of Granada, Spain, and Macquarie Group,

Australia. His research interests include biomedical signal processing, fuzzy systems, computational neuroscience, deep learning, and VLSI for machine learning.



MAHBOOBEH JAFARI received the M.Sc. degree in electrical engineering from the Electrical and Computer Engineering Faculty, Semnan University, Semnan, Iran, in 2014. She is currently a Research Assistant and a Research Intern with the University of Granada and UNSW, respectively. Her current research interests include deep learning, advanced biomedical signal processing, and neuroimaging.



RU-SAN TAN is currently a Senior Consultant with the Department of Cardiology, National Heart Centre Singapore. His specialization is in non-invasive diagnostic cardiac imaging: cardiovascular magnetic resonance imaging, echocardiography, and nuclear cardiology. His research interests include advanced cardiac imaging, cardiac biomechanics, and computational modelling. He is also a site PI and a member of the steering committees of multinational clinical trials of novel cardiology drugs and notably novel anticoagulants.



U. RAJENDRA ACHARYA received the Ph.D., D.Eng., and D.Sc. degrees. He is currently a Professor of artificial intelligence in healthcare with the School of Mathematics, Physics and Computing, University of Southern Queensland, Australia; a Distinguished Professor with the International Research Organization for Advanced Science and Technology, Kumamoto University, Japan; an Adjunct Professor with the University of Malaya, Malaysia; and an Adjunct Professor with Asia University, Taiwan. He has authored over 800 publications, including 750 in refereed international journals, 42 in international conference proceedings, and 17 books. His research interests include biomedical imaging and signal processing, data mining, and visualization, as well as applications of biophysics for better healthcare design and delivery. His funded research has accrued cumulative grants exceeding six million Singapore dollars. He received more than 75,000 citations on Google Scholar (with an H-index of 137). He has been ranked in the top 1% of the highly cited researchers for the last seven consecutive years (2016–2022) in computer science, according to the Essential Science Indicators of Thomson. He is on the editorial boards of many journals and has served as the guest editor for several AI-related issues.

• • •