IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 23, 2020, accepted January 12, 2021, date of publication January 26, 2021, date of current version February 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3054670

EEG Classification by Factoring in Sensor Spatial Configuration

LUBNA SHIBLY MOKATREN[1], (Student Member, IEEE), RASHID ANSARI[1], (Life Fellow, IEEE), AHMET ENIS CETIN[1], (Fellow, IEEE), ALEX D. LEOW[2], (Member, IEEE), OLUSOLA A. AJILORE[2], HEIDE KLUMPP[2], (Member, IEEE), AND FATOS T. YARMAN VURAL[3], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607, USA
[2]Department of Psychiatry, University of Illinois at Chicago, Chicago, IL 60607, USA
[3]Department of Computer Engineering, Middle East Technical University, 06800 Ankara, Turkey

Corresponding author: Lubna Shibly Mokatren (lshibl2@uic.edu)

This work was supported in part by NIMH K23 MH093679 (HK) and the Center for Clinical and Translational Research (CCTS) UL1RR029879.

ABSTRACT Electroencephalography (EEG) serves as an effective diagnostic tool for mental disorders and neurological abnormalities. Enhanced analysis and classification of EEG signals can help improve performance in classifying the disorders and abnormalities. A new approach is examined here for enhancing EEG classification performance using a novel model of data representation that leverages knowledge of spatial layout of EEG sensors. An investigation of the performance of the proposed data representation model provides evidence of consistently higher classification accuracy of the proposed model compared with a model that ignores the sensor layout. The performance is assessed for models that represent the information content of the EEG signals in two different ways: a one-dimensional concatenation of the channels of the frequency bands and a proposed image-like two-dimensional representation of the EEG channel locations. The models are used in conjunction with different machine learning techniques. Performance of these models is examined on two tasks: social anxiety disorder classification, and emotion recognition using a dataset, DEAP, for emotion analysis using physiological signals. We hypothesize that the proposed two-dimensional model will significantly outperform the one-dimensional model and this is validated in our results as this model consistently yields 5–8% higher accuracy in all machine learning algorithms investigated. Among the algorithms investigated, Convolutional Neural Networks provide the best performance, far exceeding that of Support Vector Machine and k-Nearest Neighbors algorithms.

INDEX TERMS Machine learning, EEG, CNN, spatio-temporal features, emotion recognition, SAD.

I. INTRODUCTION[1]

Electroencephalography (EEG) is a widely used mechanism for diagnosing mental states and brain disorders. EEG records electrical patterns resulting from brain activity using electrodes wired onto the scalp. EEG is among a variety of brain sensing methods, used for diagnosis in research and clinical practice, which include positron emission tomography (PET), computed tomography (CT), and functional magnetic resonance imaging (fMRI). EEG has key distinguishing attributes of excellent temporal resolution and cost-effectiveness compared with other methods [2]. The waveform of each EEG sensor is divided into five main frequency bands [3], labeled as Delta, Theta, Alpha, Beta, and Gamma ($\delta, \theta, \alpha, \beta, \gamma$) waves. EEG is widely used for noninvasive monitoring and it serves an important role as a diagnostic tool in brain-computer interface (BCI) applications [4]. EEG data are used to evaluate various mental disorders, such as Alzheimer's disease, strokes, migraine, sleep disorders, and Parkinson's disease [5]. The modeling of the data and subsequent analysis may not, however, always yield the best results as the data are complex and degraded by noise and artifacts. It is beneficial to examine new models that may lead to improved performance of the EEG analysis. In the past, many classification algorithms were devised for analyzing EEG data [6], using machine learning techniques. These include linear discriminant analysis, neural networks, support vector machines (SVM), nonlinear Bayesian classifiers, k Nearest-Neighbor (kNN) classification, hidden Markov models, a combination of classifiers, and other techniques for EEG-based BCI

The associate editor coordinating the review of this manuscript and approving it for publication was Mouloud Denai.

[1]A preliminary version of this manuscript is presented in IEEE EMBS Conference on Neural Engineering [1]

applications [7]. Of the many past studies on EEG classification, only a few have paid attention to spatial locations and configuration of the EEG channel sensors for the purpose of creating models that may achieve improved performance in the analysis tasks. Furthermore, they were primarily focused on the asymmetry between electrodes in the left and right hemispheres. Since EEG signals are highly correlated in the spatial, temporal, and spectral domains, we propose a model that is designed to preserve the information extracted from all three domains. Factoring in the sensor topology is a key driving factor in our research. We investigate two data representation models for classification - a 1-Dimensional (1D) Concatenated Data Representation (1D-CDR) model and a 2-Dimensional (2D) Interpolated Data Representation (2D-IDR) model. Note that the model labels refer only to the spatial dimension and the time dimension is not factored into the labeling. The 1D-CDR model ignores the sensor layout while the 2D-IDR model factors it in with different interpolation methods. We hypothesize that 2D-IDR model would outperform 1D-CDR model. This hypothesis is validated in our results as the performance of 2D-IDR model surpassed that of 1D-CDR model by yielding 5–8% higher accuracy in all machine learning algorithms investigated. Among the machine learning methods examined, Convolutional Neural Networks (CNN) provided the best performance far exceeding that of Support Vector Machine (SVM) and k-Nearest Neighbors (kNN) algorithms.

To assess the efficacy of the 2D-IDR model, its performance on two important classification tasks is examined. The first task focuses on Social Anxiety Disorder (SAD) detection. SAD, the world's third-largest mental health problem, affects about 7% of the population [8]. It is characterized by the fear of negative evaluations and avoidance of social interactions [9]. The process of diagnosing SAD was officially recognized in 1980 by Diagnosis and Statistical Manual for Mental Disorders (version DSM-III). Over the years, the criteria evolved and are now described in the fifth edition of the manual (DSM-5) [10]. The effectiveness and reliability of the process of DSM-5 diagnosis are critical in accurate assessment of the underlying disorder [11]. Analysis of EEG data using machine learning in SAD diagnosis has seen only limited study. Identifying SAD patients by visual detection of differences in the EEG signals is impractical. Automated methods of detecting SAD such as those based on machine learning algorithms are therefore adopted to pave the way for potentially highly accurate diagnosis, better connectivity analysis, and improved understanding of treatment responses in SAD [12].

The second classification task we examined for assessing the efficacy of the 2D-IDR model is emotion recognition using the DEAP dataset [13]. The study of EEG-based emotion recognition is very widely used in many fields such as psychology and neuroscience. Emotions are a very important factor in correct interpretation of actions and play a crucial role in everyday communication. EEG-based emotion recognition task can be subject-dependent or subject-independent [14]. In this paper, both subject-dependent and subject-independent approaches are investigated. A preliminary version of a part of this study was presented at IEEE EMBS Conference on Neural Engineering [1]. The work reported here goes well beyond the initial study as summarized in Section V. One question that we sought to address is whether our proposed method was applicable in analyzing tasks other than SAD diagnosis. We wanted to use publicly available datasets to demonstrate the applicability of our method. Toward that end, we decided to apply our method to the DEAP dataset, which is publicly available, and it is recorded with stimuli present, unlike SAD. Several changes were made in our algorithms and these are described in Section V. All these changes resulted in a more robust performance and better accuracy than the previous model, as can be seen in the 92.19% SAD accuracy reported in this work which exceeded the 87% accuracy reported in our previous paper.

The main contributions of this work are summarized as follows:

- We introduce a new model 2D-IDR for EEG analysis that integrates the spatial configuration of the EEG sensors in the analysis of the data. This integration allows us to utilize the channel spatial configuration information and build a 2D image-like representation of the EEG samples.
- We provide a novel robust framework that can be applied to various recognition and classification tasks. The experiments conducted using our approach demonstrated significant improvements over other methods that disregarded the spatial topology of the sensors. The model's ability to capture spatial information can be utilized in other state-of-the-art classifiers.
- To the best of our knowledge, no EEG studies to date utilize machine learning techniques for SAD detection.

II. RELATED WORK

Over the years, several studies have been conducted to determine the underlying estimators and treatment outcomes in SAD patients. For instance, Harrewijn *et al.* [15] found that delta-beta correlation can be used as an EEG measure of anxiety for subjects carrying out a social performance task. In a different study, Miskovic *et al.* [16] showed that a cognitive behavioral treatment for SAD was associated with reductions at delta-beta frequency coupling. Finally, a recent review of the most frequently studied EEG predictive biomarkers for SAD was presented by AL-Ezzi *et al.* [17], which included frontal alpha asymmetry and delta-beta coupling as estimators. While many studies focused on SAD analysis, to the best of our knowledge, no EEG studies to date have investigated SAD detection using machine learning methods.

On the other hand, EEG-based emotion recognition has been widely studied over the years. Since the DEAP dataset is publicly available, it allows us to have a better perspective on the differences between the methods and on the variability in their performances. Piho and Tjahjadi [18] investigated

TABLE 1. Survey of recent studies in emotion recognition using DEAP dataset.

Study	Classification Method	Spatial configuration used	Valence Accuracy	Arousal Accuracy
Piho and Tjahjadi [18]	SVM/kNN/NB	No	89.61%	89.84%
Chen et al. [19]	combined Features Using CNN	No	88.76%	86.98%
Chao et al. [20]	DBN-GCs	No	76.83%	75.92%
Ullah et al. [21]	SDEL	No	82.81%	74.53%
Li et al. [22]	hybrid CNN and LSTM	Yes	$\leq 82\%$	$\leq 82\%$
Hao et al. [23]	CapsNet	Yes	66.73%	68.28%
Wang et al. [24]	EmotioNet	Yes	72.1%	73.3%
Proposed method	CNN	Yes	91.85%	91.06%

reduced EEG data of emotions using mutual information-based adaptive windowing and achieved an average accuracy of 89.61% and 89.84% for valence and arousal, respectively. Chao *et al.* [20] integrated deep belief networks with glia chains learning framework using multichannel EEG data and achieved average accuracy of 76.83% and 75.92% for valence and arousal states classification, respectively. In their study, Ullah *et al.* [21] proposed a sparse discriminative ensemble Learning (SDEL) algorithm for computing the most discriminative subset of EEG channels, and obtained 82.81% and 74.53% accuracy for valence and arousal, respectively. Another recent study on DEAP dataset was conducted by Chen *et al.* [19] using deep CNNs with temporal features, frequential features, and their combinations to achieve an average accuracy of 88.76% and 86.98% for valence and arousal, respectively.

The studies mentioned earlier followed procedures that did not exploit the knowledge of EEG sensor configuration. In contrast, Li *et al.* [25] proposed a method called R2G-STNN to learn discriminative spatial-temporal EEG features using a region-attention layer that adjusts its weights to reflect the contributions of specific brain regions. In another study, Jin *et al.* [26] designed a novel spatial feature selection method based on an improved objective function for the common spatial pattern algorithm using L1-norm and Dempster-Shafer theory, which resulted in an optimal spatial feature selection. In particular, for DEAP dataset, Li *et al.* [22] took the spatial configuration into account and proposed an emotion recognition method using EEG multidimensional feature images and hybrid deep neural networks, where the highest accuracy achieved for valence and arousal is 82%. Another study that utilizes spatial topology is proposed by Chao *et al.* [23]. Here, the authors suggested a deep learning framework (CapsNet) based on a multiband feature matrix of the EEG signal. Finally, a 3D convolutional neural network for EEG-based emotion recognition is presented by Wang *et al.* [24], where the electrode topology and time domain information are considered in constructing the input. A summary of these emotion recognition research studies is listed in Table 1, for valence and arousal.

While several studies considered the spatial configuration in the construction process of their models, the work presented in this paper proposes a significantly different method, which is described in Section III. The models differ from each other with respect to input construction and classification techniques. This includes, but is not limited to, input dimensions, input representation, and interpolation method (if used) for the locations of the missing electrodes.

III. METHOD

A. EEG DATASETS USED IN THIS STUDY

The two datasets investigated in this study are the SAD and DEAP datasets that are described next.

1) SAD DATASET

In this study, one of the EEG datasets we used was the SAD dataset which was obtained from the Department of Psychiatry at the University of Illinois at Chicago (UIC). The multi-channel EEG data were collected from two subject groups. The first consists of 32 SAD patients, and the second of 32 healthy subjects. The brain activity was analyzed in resting state, that is, in the absence of any stimulus. The duration of the EEG recordings was 4 minutes on average. Data were collected from 34 electrodes positioned on the left and right mastoids, using a BioSemi (Amsterdam, Netherlands) 34-channel cap (32 channels based on 10-20 system plus FCz and Iz), with electrooculogram recorded from four facial electrodes. The signals were digitized at 24-bit resolution with a Least Significant Bit (LSB) value of 31.25 nV and a sampling frequency of 1024 Hz.

2) DEAP DATASET

The DEAP dataset is used for emotion analysis. Specifically, in this research, valence and arousal were assessed. DEAP is a publicly available EEG dataset [13] that contains signals from 32 participants. Each participant watched 40 one-minute long videos and evaluated themselves on the basis of four emotional states: arousal, valence, liking, and dominance on a scale of 1-9. The data are recorded over 32 channels on the scalp. For our analysis, a classification task of two output labels is considered. A rating value greater than or equal to 5 is mapped to 1 (aroused; pleasant), otherwise, it is mapped to 0 (relaxed; unpleasant).

B. DATA PREPROCESSING

EEG data are complex with high temporal resolution. The signals are easily contaminated with various artifacts and undesired noise, such as electromyogram artifacts, residual

eye movements, and other muscle activities. Hence, preprocessing the data which includes artifacts removal, is very critical for proper analysis.

1) SAD DATASET PREPROCESSING

The preprocessing of the SAD dataset is implemented offline using Fieldtrip and Brain Vision Analyzer (Brain Products, Gilching Germany) software. The data were converted to a linked mastoid reference and band-pass filtered. Eye movements and ocular artifact corrections were performed [27], and semi-automated rejection of epochs containing artifacts with voltage step higher than 50 μV between samples was applied. Additional artifacts were removed by visual inspection. The frequencies of interest are in the range 0-48 Hz, covering the five different frequency bands (δ,θ,α,β,γ).

2) DEAP DATASET PREPROCESSING

The EEG data in the DEAP dataset are first downsampled to 128 Hz to adequately capture the data content in the 0-48 Hz range. Next, the electrooculogram and electromyogram artifacts are removed. As delta waves usually correspond to deeper sleep [28], useful and informative data for emotion analysis are known to lie in the frequency range 4-45 Hz. Hence, a bandpass filter was applied to take the content in the first band, δ [0-4 Hz] out of the analysis process. Eye artifacts were removed using a blind source separation technique, and the data were averaged to the common average reference (CAR) where the common average of the entire electrode is subtracted from a specific channel of interest, resulting in a zero-mean signal distribution [29]. The data are partitioned into 60-second segments for each recording, and a 3-second pre-trial baseline is removed.

C. DATA ANALYSIS AND FEATURE EXTRACTION

The constant switching of meta-stable states of neurons assembling during brain activity is manifested as nonstationary phenomena in the EEG data, which consequently limits the reliability of the conventional analysis. However, the data is quasistationary over short time periods. Therefore, in our data analysis, each channel signal is normalized and divided into N windows in which the data are assumed to be stationary. For each window, a time-domain wavelet packet decomposition (WPD) is applied to extract the content of the five frequency bands, from which the energy and entropy of these bands are extracted as features. Based on the sampling frequency and statistical properties of the signals, multiple sizes of window segments were examined and a segment of size N=5 seconds and N=4 seconds were found to provide better results for SAD and DEAP datasets, respectively. The subjects' cognition and emotional states can be assessed by analyzing the EEG signal content in all frequency bands.

The energy and entropy of the content of each windowed segment are computed for the k frequency bands (k = 5 for SAD, k = 4 for DEAP) separately as described later in Eqs. (1) to (4). The analysis is based on the energy and entropy content of these signals represented in two different ways: (i) concatenation of the channels of the k frequency bands and (ii) an image-like 2D representation of the EEG channel locations. The latter method is further discussed in Section III-E.

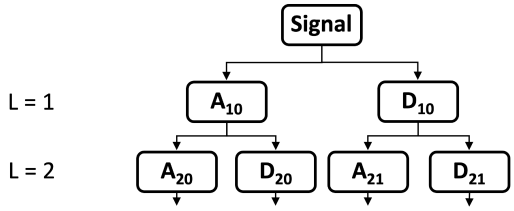


FIGURE 1. Packet wavelet transform decomposition tree.

D. WAVELET DECOMPOSITION

After segmenting the data into multiple windows, wavelet packet decomposition (WPD) is applied to the data in the time domain to extract the EEG features. A wavelet packet transform decomposition tree with 2 levels is shown in Fig.1. Here A_{mn} and D_{mn} denote the approximation and detail coefficients, respectively, at level m with n denoting the number $(0, 1, 2, \ldots, 2^{m-1} - 1)$ of the parent unit labeled from left to right at level $m - 1$.

Since EEG signals are non-stationary, Fourier methods are not suitable for the time-frequency analysis of such signals. However, wavelet transforms can capture the local behavior of the signal and can adequately capture both frequency and time information of transient non-stationary signals. Hence, they are more appropriate and preferable to use for EEG analysis and decomposition [30], [31]. For both datasets, WPD was used to extract the EEG frequency bands. In WPD, both the detail and the approximation coefficients obtained at each level are decomposed at subsequent levels to create a full binary tree. In each level of WPD, the signals are processed with both low-pass and high-pass filters followed by downsampling by a factor 2. Each output at every level before computing the final set of outputs is decomposed into (i) a detail signal obtained using the branch containing the high-pass filter, and (ii) an approximation signal obtained using the branch containing the low-pass filter. For SAD dataset, the data are filtered and downsampled to 128 Hz and decomposed using a 4-level WPD and Daubechies 4 (*db4*) mother wavelet. All five frequency bands are used for the analysis of SAD. Table 2 contains the ideal subbands chosen for each frequency and the relevant decomposition coefficients. In the case of the DEAP dataset, for each data segment, a 4-level WPD was applied to the input similarly to SAD dataset. However, unlike in the case of SAD analysis, only 4 frequency bands, Gamma, Beta, Alpha, and Theta are used for the emotion recognition task.

It should be noted that *db4* wavelet is chosen due to its orthogonality and smoothing features, which are used for optimal detection of changes in the EEG signal [32].

In both datasets, the energy and entropy content are both extracted as features. The mean wavelet energy E_j of wavelet

TABLE 2. Wavelet packet decomposition for SAD.

Frequency band	Frequency range(Hz)	Decomposition level
Gamma	$32-48$ Hz	A_{21}
Beta	$12-32$ Hz	D_{41}, D_{20}
Alpha	$8-12$ Hz	A_{41}
Theta	$4-8$ Hz	D_{40}
Delta	$0-4$ Hz	A_{40}

coefficients at resolution level j is defined as:

$$\bar{E}_j = \frac{\sum_k |C_{j,k}|^2}{N_j} \tag{1}$$

where $C_{j,k}$ is the k-th wavelet coefficient at resolution level j and N_j is the number of wavelet coefficients at that level. For each frequency band, C_j wavelet coefficients at a resolution level j correspond to the C_{mn} coefficients detailed in Table 2.

The total energy is defined as

$$E_{tot} = \sum_{j=1}^{N} \bar{E}_j \tag{2}$$

and the relative wavelet energy is calculated as follows

$$q_j = \frac{\bar{E}_j}{E_{tot}} \tag{3}$$

The wavelet entropy is defined as

$$w_j = -q_j log q_j \tag{4}$$

E. IMAGE REPRESENTATION OF THE EEG DATA

Data acquisition is performed by positioning M electrodes over five areas on the scalp: Frontal (F), Central (C) Temporal (T), Parietal (P) and, Occipital (O), where M is 34 for SAD data and 32 for DEAP data. It is hypothesized that knowledge of the location of the channels can provide improved detection accuracy in the analysis of the data. To investigate this hypothesis, two main data models are examined using M channels and B extracted features. Suitable choices of features considered in this study are the energy of the frequency bands or a combination of energy and entropy. $B \in \{5, 10\}$ for SAD, where the choice B=5 features corresponds to the case when only the energy of the 5 frequency bands is used for the analysis, and B=10 corresponds to the case when both energy and entropy are used. In the analysis of the DEAP dataset, $B \in \{4, 8\}$ as only 4 frequency bands were used. In the first model, i.e. 1D-CDR, the M channels of the B features are concatenated by creating a $M\times B$ feature matrix over each window, without accounting for the location of the channel electrodes ($34\times B$ for SAD, and $32\times B$ for DEAP). Therefore, an input sample to the network is a $M\times B$ matrix, in which the i-th row consists of B feature values corresponding to the i-th of the M electrodes. For the second model, i.e. 2D-IDR, a 3D array of size $K\times K\times B$ is constructed consisting of a stack of B two-dimensional $K\times K$ arrays each corresponding to one of B features estimated from the electrode signals over a uniform $K\times K$ grid. As explained later, we choose K=15.

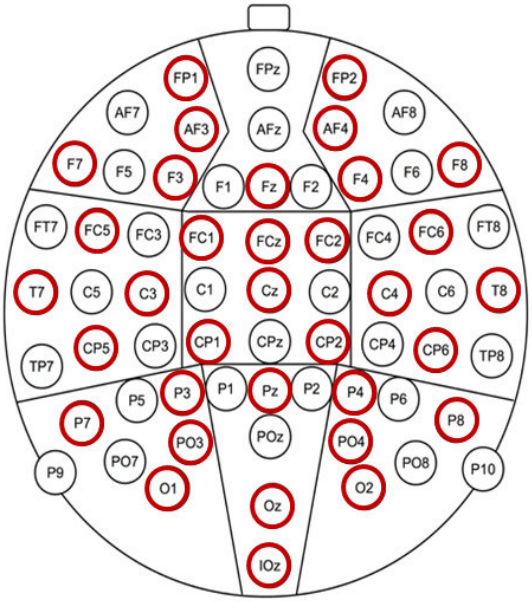


FIGURE 2. Layout of 34 electrodes on scalp.

In this model, the image pixel locations may not coincide with any of the M electrode locations and the B feature values over the $K\times K$ grid are estimated by applying various interpolation techniques. Hence, an input sample for this model is a $K\times K\times B$ array in which each $K\times K$ grid represents an image-like spatially interpolated configuration of the electrodes over the scalp for each of the B features.

For both datasets, an image of size 15×15 was computed to construct an image-like representation of the channels layout. The M channels are mapped to specific pixels in the image based on their locations. A layout of the channels' locations for SAD data is shown in Fig. 2. The red circles represent the electrode locations used in recording the data. Both datasets follow the international 10-20 System that allows EEG electrode placement to be standardized. A 2D mapping array of the electrodes' location is shown in Fig. 3 to illustrate a possible mapping of size 15×15.

To fill in the missing pixel feature values, the Inverse Distance Weighting (IDW) interpolation method was used [33]: for an interpolated value e at point x, only the samples $u_i = u(x_i)$ with sensor locations x_i, $i = 1, \ldots n_i$, that are within a radius d_{max} from the grid point x, are used in interpolating the value with the weighted average:

$$u(x) = \frac{\sum_{i=1}^{n_i} w_i(x) u_i}{\sum_{i=1}^{n_i} w_i(x)} \tag{5}$$

If $d(x, x_i) \leq d_{max}$ is the euclidean distance between points x and x_i, then $w_i(x) = \frac{1}{d(x,x_i)}$. At locations x where no sensors lie within a distance of d_{max} i.e. "Border Points" (BP), the estimate is made with the nearest sensor value. The value of d_{max} is experimentally determined by investigating different choices of d_{max} in the $K\times K$ grid.

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & FP_1 & 0 & 0 & 0 & FP_2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & AF3 & 0 & 0 & 0 & AF4 & 0 & 0 & 0 & 0 & 0 \\
0 & F7 & 0 & 0 & F3 & 0 & 0 & F2 & 0 & 0 & F4 & 0 & 0 & F8 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & FC5 & 0 & 0 & FC1 & 0 & FCz & 0 & FC2 & 0 & 0 & FC6 & 0 & 0 \\
T7 & 0 & 0 & 0 & C3 & 0 & 0 & C1 & 0 & 0 & C4 & 0 & 0 & 0 & T8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & CP5 & 0 & 0 & CP1 & 0 & 0 & 0 & CP2 & 0 & 0 & CP6 & 0 & 0 \\
0 & 0 & 0 & 0 & P3 & 0 & 0 & Pz & 0 & 0 & P4 & 0 & 0 & 0 & 0 \\
0 & P7 & 0 & 0 & 0 & PO3 & 0 & 0 & 0 & PO4 & 0 & 0 & 0 & P8 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & O1 & 0 & Oz & 0 & O2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & IOz & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

FIGURE 3. A mapping matrix of the EEG channels.

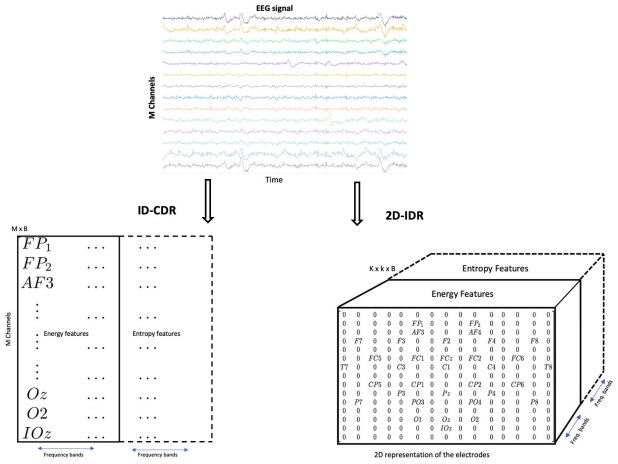


FIGURE 4. Input data organization diagram of 1D-CDR (left) and 2D-IDR (right) models.

Other interpolation methods were also considered, such as padding all locations with no sensors with zeros, IDW with zero values at border points (IDW with 0 BP), nearest-neighbor interpolation, bilinear interpolation, and cubic b-spline interpolation [34]. However, they were all found to be inferior to the method mentioned above. This is further discussed in Section V, where a summary of average performances of different interpolation methods in SAD, tested on the main CNN, is provided.

The value of K was selected by investigating performance accuracy after constructing images of different sizes $K\times K$. The value $K = 15$ was found to give higher accuracies when compared with other sizes tried on both datasets. In this investigation, images of size $K\times K$ were considered for the analysis where $K = \{10, 15, 20, 25\}$. For SAD dataset, $K = 15$ yielded the highest accuracy. For DEAP dataset, both $K = 15$ and $K = 20$ gave significantly good results. However, $K = 15$ was chosen as the location is adequately captured without making the image size too large thereby increasing the computational load.

Fig. 4 shows the organization of the sensor data in the two models. The energy and entropy features are calculated in each time segment. In the 2D-IDR model, the points with zero values correspond to locations with no sensor value. The values of such points are filled out using interpolation techniques such as the one indicated in (5) for IDW interpolation.

IV. EXPERIMENTS

After the data are collected and preprocessed, two models for each dataset are used as previously discussed.

SAD Dataset Experiments: In the first model, 1D-CDR, the energy and entropy for the five frequency bands are calculated separately in each window of 5 seconds. Hence, for each window, a feature matrix of dimensions $34\times B$ is constructed. There are 34 channels or electrodes and 2 values of energy and entropy are extracted from each of the 5 frequency bands, giving a choice $B \in \{5, 10\}$ as discussed earlier. Each row of the $M = 34$ rows in the matrix represents one channel. The second model, 2D-IDR, adopts the image representation technique of the EEG data described earlier in Section III-E. A 3D array of $15 \times 15\times B$ is constructed for each segment, where the third dimension comprises of the energy and/or entropy values of all the frequency bands.

DEAP Dataset Experiments: The models are again built in a manner similar to SAD models. However, the dimensions are different. For the first model, 1D-DCR, the energy and entropy for only 4 frequency bands are calculated separately in each window of 4 seconds. Hence, for each window an energy matrix of dimensions $32\times B$ is constructed, as 32 channels were considered, $B \in \{4, 8\}$. In the second model, 2D-IDR, a 3D energy array of $15 \times 15\times B$ is built for each window. For both models in each dataset, each matrix is considered as a single sample for the training or testing dataset.

A. ACQUISITION OF TRAINING AND TESTING EEG DATA

1) TRAINING AND TESTING USING SAD DATASET

To train the network, a stratified 8-fold cross-validation procedure is applied. where the classifier is trained 8 times using a different fold for testing in each run. A total of 7 folds (56 subjects) are used for training and validation (20% validation and 80% training) and the remaining fold (8 subjects) is used in the testing stage. We employed Bayesian optimization using Keras API for hyper-parameter selection, where the validation set was used for hyper-parameter tuning and to prevent overfitting. The hyper-parameters included the optimizer, activation functions, learning rate, dropout rate, number of layers, and the filter size used to extract the features. In both models, for every time window of size N seconds, a feature matrix is constructed and considered as a single sample, where $N = 5$ seconds. For each subject, multiple samples are gathered by sliding a moving window of size N with no overlap of windows over all channels. The samples collected for the training, validation, and testing sets do not overlap, that is, different samples collected from a specific subject cannot be used for training and testing at the same trial. Samples are labeled 1 if they belong to SAD patients, and 0 otherwise. For every trial, each testing subject is evaluated as follows:

$$prediction = \begin{cases} patient & \text{if } \frac{p_i}{t_i} \geq Th \\ control & \text{else} \end{cases} \quad (6)$$

where $Th = 0.5$, p_i is number of samples classified as 1 for subject i, t_i is total number of samples for subject i. The

performance variation using a set of different threshold values was investigated. In both datasets, for $Th \in [0.4, 0.65]$, the performance of each model remained almost unchanged. If we increase the threshold to $Th = 0.7$, the accuracies drop. Therefore, $Th = 0.5$ was kept as the chosen threshold value.

2) TRAINING AND TESTING USING DEAP DATASET

A very important aspect of the emotion recognition task is whether the task is subject-dependent or independent. A subject-dependent task entails partitioning the training and testing data from same participant. However, different samples taken from the same video cannot be used in both the training and testing data. A Subject-independent task entails testing on a group of participants that is different from the training group. According to [35] there exists a physiological linkage with emotion recognition, which makes the recognition depend on age, culture, and gender. Both the structure of the training and testing data and the physiological linkage make the performance accuracy in a subject-independent task lower than a subject-dependent task. A summary of the comparison is provided in Section V.

A stratified 8-fold cross-validation is applied to the DEAP dataset. Again, 7 folds are used for training and validation, and the remaining fold is used for testing. The average accuracy is found after the classifier is trained 8 times. Leave-one-out-cross-validation (LOOCV) or Leave-one-subject-out-cross-validation (LOSOCV) were also considered for this dataset and are discussed in Section V.

The window size taken to create a sample is $N = 4$ seconds. For every 60 seconds of video, the samples are gathered from all 32 channels by sliding a window of size N with $N/2$ overlap.

The choice of the shift size for each dataset was made experimentally after examining different shift values. The window shifts that were tested are $N/4$, $N/2$, N, and $3N/2$. Early stopping was applied by monitoring the validation loss to avoid overfitting.

The number of folds for cross-validation was chosen to be $k = 8$ for both datasets after exploring various folds of sizes $k = \{5, 8, 10\}$. A fold of size $k = 8$ was found to yield the best results in terms of accuracy, computational cost, and being a divisor of the number of subjects.

B. DATA AUGMENTATION

One of the key challenges in machine learning algorithms in general and deep neural networks specifically is not having sufficient training data to properly perform a classification task [36]. Training with small datasets might cause the model to be highly biased to the data in the training set, making the model perform poorly on the validation or testing set, as it cannot generalize what it learned to unseen samples. These models suffer from overfitting. Regularization, dropout, batch normalization, and data augmentations are some of the methods used to tackle the problem of overfitting [37]. Image augmentation technique is introduced to help improve the classification performance by creating more robust models with the ability to generalize. Data augmentation refers to suitably generating data by creating new samples to expand the training dataset. It is done by performing transformations on the original images while preserving the label which is invariant to certain perturbations. In this work, horizontal and vertical shift augmentation is used to expand the training set by 4 times. This translation is done by moving the image along the X or Y directions, specifically, a shift of 1 or both 1 and 2 pixels is done on the image while preserving the image size. For 1D-CDR model, images are shifted by a specified number of pixels to the right, left, up, and down. The deleted rows or columns are simply replaced by the previous row or column respectively. For 2D-IDR model, the shift is done over the first 2 dimensions only. The third dimension which corresponds to the energy or entropy of the frequency bands is filled according to the feature value it has at that specific location.

C. CNN NETWORK STRUCTURE

In recent years, the use of deep learning solutions has become very popular in many applications. Deep learning-based methods have repeatedly shown improved performance compared with other classical machine learning algorithms on a wide variety of problems [38]. Specifically, CNNs have rapidly become a methodology of choice for image-related tasks, including, medical images processing [39]. CNNs are a specific kind of feedforward deep neural networks. Their architecture is characterized by arranging convolutional layers, pooling layers, and fully-connected layers. In our study, the input to the network is arrays of data containing energy and/or entropy values of subband signals which can be viewed as images.

A sequential model is built for SAD classification task as shown in Fig. 5. The first layer in this model is a 2D convolution layer with kernel size 3×3, 64 output filters, and ReLu activation over the outputs. It is then followed by a similar 2D convolution layer and a max-pooling layer with a pool size of 2. A dropout with rate=0.25 is performed. The input is then flattened and fed to a fully connected (FC) layer with 128 output dimension and ReLu activation. Another dropout is performed with rate=0.2 followed by a final fully connected layer with Softmax activation and output dimension that equals 2 corresponding to the two labels in the recognition task. Dropout is a regularization method that is used to reduce over-fitting. A similar configuration is used for the emotion recognition task, where the first layer is a 2D convolution layer with kernel size 3×3, 32 output filters, followed by a max-pooling layer of size 2×2. Another convolution layer follows with 64 output filters and a max-pooling layer with a pool size of 2. Dropout with rate=0.45 is performed. The output is then flattened and fed to a fully connected layer with Relu activation and 64 output dimension. Another dropout is performed with rate=0.25 followed by a fully connected layer with Softmax activation and output dimension that equals 2. It should be noted that batch normalization was applied to the input from both datasets and the convolutional layers to

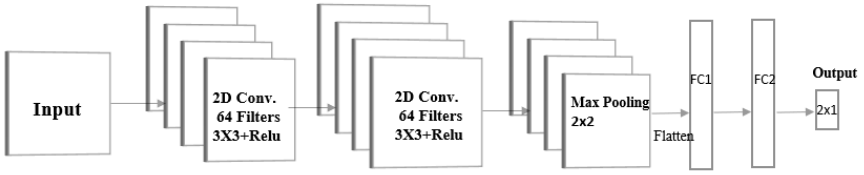


FIGURE 5. Convolutional neural network structure.

reduce the internal covariate shift [40]. We started from a filter size of 3×3 for the convolutional layers as the input has a relatively small-sized dimension. The filter size, the number of layers, and feature maps are further increased, but the structures specified above achieved better results.

V. RESULTS

In the study, inputs are constructed based on two models. These models are investigated for their classification performance using different classifiers, different types of feature values, and two different datasets. 1D-CDR model is constructed with no consideration of the spatial locations of EEG electrodes by simply concatenating the channels of the frequency bands. 2D-IDR model, on the other hand, factors in the electrode spatial configuration over the scalp.

The inputs to the classifiers of SAD and DEAP datasets are built using two groups of features: energy of the frequency bands, or a combination of energy and entropy. In each sliding window, the energy and entropy contents of the corresponding frequency band of each channel can be extracted according to Eqs. (1) to (4). For instance, if only the energy features are considered in the analysis, then for each channel we will have B features, where B is the number of frequency bands used in the analysis. The features can then be used to construct the samples following the configuration of the desired model.

A. SAD DATASET RESULTS

The main results for SAD dataset classification are presented using confusion matrices. In this case, accuracy is defined as the ability to correctly classify a subject. The analysis is subject-independent, and the confusion matrices for 1D-CDR and 2D-IDR models can be seen in Table 3. For each model, the top part represents inputs built using the energy of the frequency bands as features, and the bottom part represents inputs built using both energy and entropy. In the fourth column, the first entry includes the number of actual SAD patients predicted as positive (patients), the number in the second entry is actual patients predicted as negative (healthy subjects), and so on.

We observe that regardless of the number of features used, the accuracy and F1-score for the proposed approach based on 2D-IDR model are higher. We conclude that 2D-IDR model that factors in the location of the electrodes provides superior performance, achieving a maximum classification accuracy of 92.19% compared with 84.38% for 1D-CDR model. Another important observation is that the number of features influences the accuracy, and it is higher when using both entropy and energy features.

TABLE 3. Confusion matrices for SAD.

			Actual Positive	Actual Negative	Accuracy	F1 Score
1D-CDR	Energy	Predicted Positive	26	6	81.25%	81.25%
		Predicted Negative	6	26		
	Energy & Entropy	Predicted Positive	28	6	84.38%	84.85%
		Predicted Negative	4	26		
2D-IDR	Energy	Predicted Positive	29	4	89.06%	89.23%
		Predicted Negative	3	28		
	Energy & Entropy	Predicted Positive	30	3	92.19%	92.31%
		Predicted Negative	2	29		

TABLE 4. Average performance on SAD dataset(%).

	Energy		Energy & Entropy	
Classifier	1D-CDR	2D-IDR	1D-CDR	2D-IDR
kNN (k=5)	70.31	73.43	70.31	75
kNN (k=3)	70.31	76.56	71.87	78.13
SVM	73.43	80.5	76.56	81.25
CNN	81.25	89.06	84.38	92.19

In this task, classification using SVM with radial basis function (RBF) and kernel parameter $\sigma = 0.4$ was also investigated. The optimum values of the hyper-parameters were selected with a grid search method. In addition, k-NN classifiers with $k = 3$ and $k = 5$ were investigated but were clearly outperformed by the proposed CNN network in terms of overall classification accuracy, as seen in Table 4 for data using energy and entropy features.

It should be noted that 2D-IDR model gave significantly higher accuracy than 1D-CDR model, regardless of the classifier type or feature values, thereby establishing the importance of 2D representation of the spatial configuration of EEG sensors.

B. DEAP DATASET RESULTS

In this subsection, the binary valence and arousal states, low/high valence (LVHV) and low/high arousal (LAHA), are estimated for the case of the DEAP dataset. The classification is evaluated under two study cases, subject-dependent and subject-independent. The classifiers used for the emotion recognition task are the CNN network described in Section IV-C, SVM, kNN classifier with $k = 3$ neighbors, and kNN with $k = 5$ using both 8-fold cross-validation and Leave-one-out-cross-validation (LOOCV). The average performance of 1D-CDR and 2D-IDR models is evaluated using the energy of decomposed frequency bands as features, or both energy and entropy.

TABLE 5. Subject-independent average performance - DEAP(%).

		Energy		Energy &Entropy	
Emotion	Classifier	1D-CDR	2D-IDR	1D-CDR	2D-IDR
Valence	kNN (k=3)	73.87	78.19	75.55	81.65
	kNN (k=5)	71.92	76.4	73.83	80.73
	SVM	75.36	81.52	78.13	83.38
	CNN	82.36	87.64	84.77	**91.85**
Arousal	kNN (k=3)	71.93	76.06	73.8	80.64
	kNN (k=5)	69.87	75.41	71.33	77.58
	SVM	74.16	77.06	75.82	81.43
	CNN	82.61	85.94	83.94	**91.06**

TABLE 6. Subject-dependent average performance - DEAP(%).

		Energy		Energy &Entropy	
Emotion	Classifier	1D-CDR	2D-IDR	1D-CDR	2D-IDR
Valence	kNN (k=3)	79.03	82.85	81.56	85.93
	kNN (k=5)	76.45	81.03	77.24	81.92
	SVM	80.27	84.96	81.68	85.31
	CNN	84.63	88.97	88.02	**94.48**
Arousal	kNN (k=3)	77.86	82.55	79.91	84.76
	kNN (k=5)	77.12	80.05	77.94	82.05
	SVM	76.34	80.93	79.75	83.04
	CNN	83.72	90.26	86.07	**93.66**

1) The subject-independent case: First, a stratified 8-fold cross-validation is applied with 7 folds correspond to 28 participants and the last fold corresponds to the 4 remaining participants. Table 5 represents the performance of subject-independent task for valence and arousal recognition using different classifiers. Then, for the CNN classifier only, LOSOCV was implemented where the test set included samples from videos belonging to only one subject. This resulted in valence accuracies of 83.18% and 90.23% for 1D-CDR model and 2D-IDR model, respectively. For arousal recognition, accuracies of 82.95% and 89.35% were achieved.

2) The subject-dependent case: As in the case of the subject-independent case, 7 folds are used for training and validation, and the remaining fold is used for testing. Table 6 summarizes the accuracies for the subject-dependent task on valence and arousal. Furthermore, LOOCV cross-validation was used for subject-dependent valence and arousal classification tasks using the CNN classifier. An average valence recognition accuracies of 86.46% and 92.25% were achieved for 1D-CDR and 2D-IDR, respectively. As for arousal recognition, accuracies of 84.55% and 90.88% were achieved.

Overall, it is observed from the mentioned tables that 2D-IDR model outperforms 1D-CDR model in all cases, regardless of feature selection, classifier, or nature of the study case (subject dependent/independent). For instance, in subject-independent valence recognition, the highest average accuracy is 91.85% for 2D-IDR model and only 84.77% for 1D-CDR model. Furthermore, by comparing the average accuracies presented in the tables, it is concluded that including the entropy in the feature selection scheme improves the performance of the classifiers.

Another important observation is the noticeable differences in performance between the subject-dependent and subject-independent cases. The manner of construction of the training and testing data in these two cases makes the subject-dependent classifier achieve higher accuracies. However, the lack of generalization of such classifiers is the price one has to pay. Moreover, using LOOCV or LOSOCV for cross-validation resulted in a higher computation load compared to 8-fold cross validation.

TABLE 7. Statistical analysis of classification accuracy differences between the compared models for different classifiers.

	Classifier			
	kNN(k=3)	kNN(k=5)	SVM	CNN
Arousal	p=0.002	p=0.0078	p=0.000	p=0.012
Valence	p=0.063	p=0.026	p=0.000	p=0.001

Due to the non-normal distribution of the classification accuracies, Wilcoxon Signed Rank test was conducted to investigate the statistical significance. We define the null hypothesis to be that the 2D-IDR model does not provide performance improvement over the 1D-CDR model and the alternative hypothesis to be that 2D-IDR model does provide performance improvement. A significance level of 0.05 was considered. The upper-tailed paired Wilcoxon Signed Rank test was performed to compute an exact p-value using a set of 8 measurements, discarding all pairwise zero differences and without continuity correction. Table 7 summarizes the statistical significance of classification accuracy differences between 1D-CDR and 2D-IDR models for subject-dependent approach. The analysis is performed for valence and arousal classification using energy and entropy features for all four classifiers used to evaluate the performance. The results indicate that 2D-IDR model yielded significantly higher accuracy than that of 1D-CDR model in all cases investigated. While 1D-CDR is less complex than 2D-IDR, the latter aimed to boost the classification accuracies while maintaining the computational time within a reasonable range. 2D-IDR model is more efficient in terms of classification performance and it resulted in a noteworthy increase of 5–8% average accuracy within a reasonable time. Specifically, for 2D-IDR model, the CNN classifier had higher computational costs compared with the other classifiers, but it achieved better classification performance. This model can be integrated in real-time applications to improve the analysis of different EEG-based recognition and classification tasks. To further improve the computing efficiency, the analysis can focus on specified frequency bands with less features or smaller image-like grid representation.

The EEG channels' locations are represented as an image of size $K \times K$ in 2D-IDR model. There are 34 channels in SAD dataset, and 32 in DEAP. To fill out the empty pixels, a few interpolation schemes are tested, as mentioned in Section III-E. For both models, the Inverse Average Weighted interpolation technique is used as it yielded the best results in the classifier performance. The average accuracies of different interpolation methods tested using CNN classifier are presented in Table 8 for SAD dataset with energy and entropy features.

TABLE 8. Average performance of different interpolation methods using CNN.

Interpolation method	Average recognition perf.(%)
Zero Padding	70.31
Nearest Neighbor	75
Bilinear	76.56
IDW with 0 BP	84.37
Cubic B-spline	89.06
IDW with NN BP	92.19

TABLE 9. Average performance using different image sizes K×K using CNN.

Image size	SAD dataset	DEAP (Valence)	DEAP (Arousal)
K=10	90.65%	88.28%	86.79%
K=15	92.19%	91.85%	91.06%
K=20	89.06%	92.17%	89.75%
K=25	85.93%	80.93%	82.16%

Regarding the choice of the image size, different values of K pixels are considered in the analysis and the average accuracies of 2D-IDR model classification performance are represented in Table 9 for the CNN classifier for both datasets in subject-independent cases.

It should be noted that for both datasets, a smaller number of m out of M electrodes was also used in the analysis. Features from $m = 10$ and $m = 20$ electrodes were selected to assess the performance of both models using a reduced number of electrodes. In all cases, regardless of the classification task, 2D-IDR model still outperformed 1D-CDR model. However, using all 34 channels for SAD and 32 for DEAP gave superior results in terms of classification accuracies.

C. EEG HEMISPHERIC ASYMMETRY

It is well known that the anatomy, behavior, and function of the two cerebral hemispheres of the brain are not identical [41]. In recent years, many studies have been conducted to learn the asymmetric differences between the two hemispheres in EEG data. For instance, Li *et al.* [42] proposed a novel bi-hemispheric discrepancy model in emotion recognition and obtained deep representations of all the EEG channels' signals. This inspired us to investigate the efficiency of our method on the left and right hemispheres separately. Specifically, electrodes on either the right or left regions of the brain were employed in the analysis process of the right or left hemispheres, respectively. Electrodes in the cross-region or in the middle were eliminated when assessing the recognition performance using data from either hemisphere. The overall experimental results showed the following: 1) The performance was inferior to the previous models which utilized data collected from channels on both sides of the scalp. 2) In both datasets, whilst utilizing data associated with either left or right hemisphere only, 2D-IDR model still outperformed 1D-CDR model with an increase of 5% in average classification accuracy. 3) For SAD, models based on data collected from the right hemisphere exhibited an improved performance compared to the left hemisphere, which may indicate increased brain activity in the right hemisphere for people suffering from social anxiety. 4) For emotion recognition models, the left hemisphere performed better than the right hemisphere.

D. COMPARISON WITH EEGNET

To further emphasize the significance of the results achieved by our model, the classification performances of both datasets were investigated using EEGNet. EEGNet is a compact convolutional neural network with successful generalized architecture for EEG-based brain-computer interfaces [43]. The first block in the network is a temporal convolution, followed by a depthwise convolution to learn frequency-specific spatial filters. The second block is a separable convolution block, which is a combination of a depthwise convolution and a pointwise convolution. The last block consists of dense layers and softmax classification at the end. For performance analysis using EEGNet, only cross-subject classification approaches are considered. There are F1 2D convolutional filters of size (1,64), with filter length 64 chosen to be half the sampling rate of the data, which is 128Hz in all cases. For the depthwise convolution, there are D filters of size (C, 1) to learn the spatial filters, where C is the number of channels. In the separable convolution block, F2 pointwise filters are used. Finally, dropout rate is chosen to be 0.25, as the training set sizes are larger in cross-subject tasks compared to within-subject tasks. Different combinations of $F1 = \{8, 4, 16\}$, $D = \{1, 2, 4\}$, and $F2 = \{8, 16, 32, 64, 128\}$ are considered. The best average classification performances obtained for SAD, Valence and arousal in cross-subject analysis using EEGNet are 75.5%, 72.35%, and 74.76% respectively, substantially lower than what was achieved with 2D-IDR model. Since EEGNet is considered a successful generic EEG-based architecture, it was important to provide a comparison between the methods.

E. EFFECT OF DATA AUGMENTATION

The performance results listed above belong to classifiers with larger training sets, which were expanded using data augmentation. The significance of creating a more robust model by increasing the training dataset is reflected in the higher classification accuracies of training with augmented samples achieved in both models. Wilcoxon Signed Rank test was conducted using augmentation as a factor, and $p<0.05$ was achieved indicating the significance of data augmentation. The results are summarized in Table 10 for 1D-CDR and 2D-IDR models, where the third column represents the performance of CNN using the original training set, and the fourth column using the augmented training set.

F. COMPARISON WITH STATE-OF-THE-ART MODELS

The overall results of the experiments above establish the importance of 2D spatial configuration of EEG data in learning robust representations from the samples created using 2D-IDR model. This model gave significantly higher accuracy than the 1D-CDR model, regardless of feature values,

TABLE 10. Data augmentation effect.

	Classification Accuracy(%)		
	recognition task	no augmentation	with augmentation
1D-CDR	SAD	79.68	84.38
	Arousal	79.53	83.94
	Valence	82.74	84.77
2D-IDR	SAD	89.06	92.19
	Arousal	87.27	91.06
	Valence	87.15	91.85

classifier type or recognition task. This is in line with our predictions that integrating the spatial configuration of the sensors in the 2D-IDR model as prior information to the classifier enhances its convergence and recognition performances. Furthermore, the superiority of 2D-IDR model was anticipated as shown in [44]. This work proposed a global feature learning method that encapsulates the multi-channel EEG signals into gray-level images. Their results demonstrated that spatial characteristics are beneficial for recognizing emotions. Consequently, the model can be applied to either supervised or unsupervised EEG-based machine learning frameworks. Samples created using 2D-IDR model can be used in unsupervised learning methods where the prior spatial-information of the electrodes can be utilized in a variety of detection or clustering analyses of EEG data.

Our proposed approach improved the emotion recognition performance when compared with other spatial or state-of-the-art methods using the same DEAP dataset, as can be seen in Table 1 in Section II. Specifically, for 3D-based frameworks, a recent study of emotion recognition using DEAP dataset was conducted by Salama *et al.* [45]. The authors proposed a 3D-CNN framework with ensemble learning and achieved recognition accuracies 96.13% and 96.79% for valence and arousal classes respectively in the subject-dependent case. This approach combined two 3D-CNN based classifiers for the face data and the input video data. Then a third model is created based on the fusion chunks from the EEG, and face modalities. Although the recognition accuracy of this work was slightly higher than that of our work, it should be noted that this method uses not only the EEG data but also associated video data as well. Furthermore, it is very specific for the emotion recognition task and is clearly more computationally expensive. In a different study [46], the authors created a 3D feature and convolutional network from EEG signals by extracting time-domain features for each channel and reorganizing them as a 3D feature matrix according to positions of electrode sensors with no interpolation of the non-electrode locations and without accounting for the frequency bands. Here, the authors reported accuracies of 85.53% and 85.88% for valence and arousal in DEAP dataset, respectively.

In SAD, to the best of our knowledge, no EEG studies to date utilized deep learning methods for social anxiety diagnosis and classification. However, the results demonstrate the effectiveness of 2D-IDR model and the importance of the sensors' spatial configuration.

Finally, as mentioned in Section I, a preliminary version of this work was presented at IEEE EMBS Conference on Neural Engineering [1]. Here we investigated whether our proposed method was applicable in analyzing tasks other than SAD diagnosis. We used the publicly available DEAP dataset. The use of a different dataset required additional investigation of the previous study [1] to find the best configuration of the models' frameworks. This included an adjustment of frequency bands selection based on the dataset and the task being performed. It also required analyzing DEAP dataset in subject-dependent and subject-independent cases and comparing them with state-of-the-art techniques. Also, we investigated the choice of the image size k used in the 2D-IDR model for both datasets, the number of channels used in both models for optimal results, and the performance of both models using fewer sensors. Furthermore, we evaluated the efficacy of the proposed method under hemispheric asymmetry assumptions, we introduced new classifiers, and we compared the performance with compact and generalized EEG classifiers like EEGNet. In this work, we also incorporated data augmentation techniques and considered additional information content used in feature selection which included entropy of the frequency bands. All these changes resulted in a more robust performance and better accuracy than the previous model in [1], as can be seen in the 92.19% SAD accuracy reported in this work which exceeded the 87% accuracy reported in [1].

VI. CONCLUSION

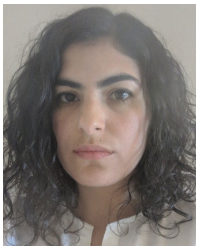
In this paper, a new EEG-based classification approach was proposed. Unlike many other studies, this approach factored in the spatial configuration of the EEG sensors in a novel image-like representation in the signal analysis. A model that takes advantage of the knowledge of the locations of the electrodes was created to construct the EEG dataset. In order to assess the effectiveness of the model, two EEG datasets were used for analysis: SAD for patient/control classification and DEAP for emotion recognition. Overall results showed that the performance of various classifiers based on this model was 5%-8% higher in accuracy, compared with the same classifiers which ignored the configuration. In addition, the use of the entropy along with the energy as relevant features in the EEG based classification task reduced the error rate for the different classifiers (CNN, SVM, kNN) in both datasets. It is also found that data augmentation for the training set is very important to further enhance the performance. This improvement is especially noticeable for 2D-IDR model. The effectiveness of the proposed novel image-like interpolated data representation capturing the spatial configuration is reflected in the results which prove its superiority over other approaches, including those which also considered the spatial topology of the sensors in their analysis. The superior performance of 2D-IDR model was consistently achieved, regardless of the classifier type, features, and most importantly,

the nature of the EEG dataset. For future work, the model can be used in EEG-based unsupervised learning frameworks. In addition, the spatial mapping technique can be enhanced and applied to other EEG-related tasks.

REFERENCES

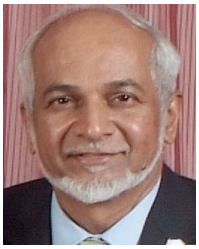
[1] L. S. Mokatren, R. Ansari, A. E. Cetin, A. D. Leow, O. Ajilore, H. Klumpp, and F. T. Y. Vural, "EEG classification based on image configuration in social anxiety disorder," in *Proc. 9th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Mar. 2019, pp. 577–580.
[2] S. Sanei and J. A. Chambers, *EEG Signal Processing*. Hoboken, NJ, USA: Wiley, 2013.
[3] M. Abo-Zahhad, D. of Electrical, E. Electronic EngineeringAssiut UniversityAssiut, S. M. Ahmed, and S. N. Abbas, "A new EEG acquisition protocol for biometric identification using eye blinking signals," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 6, pp. 48–54, May 2015.
[4] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, Mar. 2019.
[5] S. Siuly, Y. Li, and Y. Zhang, "Significance of eeg signals in medical and health research," in *EEG Signal Analysis and Classification: Techniques and Applications*. Cham, Switzerland: Springer, 2016, pp. 23–41.
[6] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *J. neural Eng.*, vol. 4, no. 2, p. R1, 2007.
[7] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, Jun. 2018, Art. no. 031005.
[8] T. A. Richards. (2017). *The Social Anxiety Association*.[Online]. Available: http://socialphobia.org/
[9] F. Leichsenring and F. Leweke, "Social anxiety disorder," *New England J. Med.*, vol. 376, no. 23, pp. 2255–2264, 2017.
[10] S. G. Hofmann and M. W. Otto, *Cognitive Behavioral Therapy for Social Anxiety Disorder: Evidence-Based and Disorder Specific Treatment Techniques*. Evanston, IL, USA: Routledge, 2017.
[11] H. C. Kraemer, D. J. Kupfer, D. E. Clarke, W. E. Narrow, and D. A. Regier, "DSM-5: How reliable is reliable enough?" *Amer. J. Psychiatry*, vol. 169, no. 1, pp. 13–15, 2012.
[12] D. A. Moscovitch, D. L. Santesso, V. Miskovic, R. E. McCabe, M. M. Antony, and L. A. Schmidt, "Frontal EEG asymmetry and symptom response to cognitive behavioral therapy in patients with social anxiety disorder," *Biol. Psychol.*, vol. 87, no. 3, pp. 379–385, Jul. 2011.
[13] 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.
[14] Y. Liu and O. Sourina, "Real-time subject-dependent EEG-based emotion recognition algorithm," in *Transactions on Computational Science XXIII: Special Issue on Cyberworlds*. Berlin, Germany: Springer, 2014, pp. 199–223.
[15] A. Harrewijn, M. J. W. Van der Molen, and P. M. Westenberg, "Putative EEG measures of social anxiety: Comparing frontal alpha asymmetry and delta–beta cross-frequency correlation," *Cognit., Affect., Behav. Neurosci.*, vol. 16, no. 6, pp. 1086–1098, Dec. 2016.
[16] V. Miskovic, D. A. Moscovitch, D. L. Santesso, R. E. McCabe, M. M. Antony, and L. A. Schmidt, "Changes in EEG cross-frequency coupling during cognitive behavioral therapy for social anxiety disorder," *Psychol. Sci.*, vol. 22, no. 4, pp. 507–516, Apr. 2011.
[17] A. Al-Ezzi, N. Kamel, I. Faye, and E. Gunaseli, "Review of EEG, ERP, and brain connectivity estimators as predictive biomarkers of social anxiety disorder," *Frontiers Psychol.*, vol. 11, May 2020.
[18] L. Piho and T. Tjahjadi, "A mutual information based adaptive windowing of informative EEG for emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 722–735, Oct. 2020.
[19] J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang, and Y. N. Zhang, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," *IEEE Access*, vol. 7, pp. 44317–44328, 2019.
[20] H. Chao, H. Zhi, L. Dong, and Y. Liu, "Recognition of emotions using multichannel EEG data and DBN-GC-Based ensemble deep learning framework," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–11, Dec. 2018.
[21] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh, "Internal emotion classification using EEG signal with sparse discriminative ensemble," *IEEE Access*, vol. 7, pp. 40144–40153, 2019.
[22] Y. Li, J. Huang, H. Zhou, and N. Zhong, "Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks," *Appl. Sci.*, vol. 7, no. 10, p. 1060, Oct. 2017.
[23] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion recognition from multiband EEG signals using CapsNet," *Sensors*, vol. 19, no. 9, p. 2212, May 2019.
[24] Y. Wang, Z. Huang, B. McCane, and P. Neo, "EmotioNet: A 3-D convolutional neural network for EEG-based emotion recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7.
[25] Y. Li, W. Zheng, L. Wang, Y. Zong, and Z. Cui, "From regional to global brain: A novel hierarchical spatial-temporal neural network model for EEG emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Jun. 14, 2019, doi: 10.1109/TAFFC.2019.2922912 .
[26] J. Jin, R. Xiao, I. Daly, Y. Miao, X. Wang, and A. Cichocki, "Internal feature selection method of CSP based on L1-norm and Dempster-Shafer theory," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 21, 2020, doi: 10.1109/TNNLS.2020.3015505.
[27] G. Gratton, M. G. H. Coles, and E. Donchin, "A new method for off-line removal of ocular artifact," *Electroencephalogr. Clin. Neurophysiol.*, vol. 55, no. 4, pp. 468–484, Apr. 1983.
[28] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Technol.*, vol. 2, no. 2, pp. 1–11, 2002.
[29] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, pp. 386–394, Sep. 1997.
[30] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wavelet transform," *J. Neurosci. Methods*, vol. 123, no. 1, pp. 69–87, Feb. 2003.
[31] J. Kevric and A. Subasi, "Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system," *Biomed. Signal Process. Control*, vol. 31, pp. 398–406, Jan. 2017.
[32] P. Jahankhani, V. Kodogiannis, and K. Revett, "EEG signal classification using wavelet feature extraction and neural networks," in *Proc. IEEE John Vincent Atanasoff Int. Symp. Modern Comput. (JVA)*, Oct. 2006, pp. 120–124.
[33] B. A. Eckstein, "Evaluation of spline and weighted average interpolation algorithms," *Comput. Geosci.*, vol. 15, no. 1, pp. 79–94, Jan. 1989.
[34] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. Med. Imag.*, vol. 18, no. 11, pp. 1049–1075, 1999.
[35] J. A. Soto and R. W. Levenson, "Emotion recognition across cultures: The influence of ethnicity on empathic accuracy and physiological linkage," *Emotion*, vol. 9, no. 6, p. 874, 2009.
[36] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart augmentation learning an optimal data augmentation strategy," *IEEE Access*, vol. 5, pp. 5858–5869, 2017.
[37] J. Wang and L. Perez, "The effectiveness of data augmentation in image classification using deep learning," in *Proc. Conf. Neural Netw. Vis. Recognit*, 2017.
[38] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, Jun. 2014.
[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167
[41] A. W. Toga and P. M. Thompson, "Mapping brain asymmetry," *Nature Rev. Neurosci.*, vol. 4, no. 1, pp. 37–48, Jan. 2003.
[42] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song, "A novel bi-hemispheric discrepancy model for EEG emotion recognition," *IEEE Trans. Cognit. Develop. Syst.*, early access, Jun. 1, 2020, doi: 10.1109/TCDS.2020.2999337.
[43] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013.
[44] H. Chao, L. Dong, Y. Liu, and B. Lu, "Improved deep feature learning by synchronization measurements for multi-channel EEG emotion recognition," *Complexity*, vol. 2020, Mar. 2020, Art. no. 6816502.

[45] E. S. Salama, R. A. El-Khoribi, M. E. Shoman, and M. A. Wahby Shalaby, "A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition," *Egyptian Informat. J.*, Aug. 2020.
[46] H. Chao and L. Dong, "Emotion recognition using three-dimensional feature and convolutional neural network from multichannel EEG signals," *IEEE Sensors J.*, vol. 21, no. 2, pp. 2024–2034, Jan. 2021.



LUBNA SHIBLY MOKATREN (Student Member, IEEE) received the B.S. degree in electrical engineering from the Technion, Israeli Institute of Technology, Haifa, Israel, in 2013. She is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Illinois at Chicago (UIC), Chicago, IL, USA.

From 2013 to 2015, she worked as a Software Engineer with Amdocs, Israel. Since 2015, she has been a Research Assistant and a Teaching Assistant with the Electrical and Computer Engineering Department, UIC. Her research interests include nonlinear analysis and modeling of multichannel signals using machine learning techniques.



RASHID ANSARI (Life Fellow, IEEE) received the B.Tech. and M.Tech. degrees in electrical engineering from the Indian Institute of Technology, Kanpur, India, and the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton, NJ, USA, in 1981.

He is currently a Professor with the Department of Electrical and Computer Engineering, University of Illinois at Chicago. His research interests include in the general areas of signal processing and communications, with recent focus on image and video analysis, multimedia communication, and machine learning for healthcare and biomedical applications.

Dr. Ansari was a member of the Digital Signal Processing Technical Committee of the IEEE Circuits and Systems Society and the Image, Video, and Multidimensional Signal Processing Technical Committee. He was a member of the organizing and program committees of several past IEEE conferences. He has served in editorial roles for professional society journals. He served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. He was on the Editorial Board of the *Journal of Visual Communication and Image Representation*. He served on the organizing and executive committees of the Visual Communication and Image Processing (VCIP) Conferences and served as the General Chair for a VCIP Conference.

AHMET ENIS CETIN (Fellow, IEEE) received the Ph.D. degree in systems engineering from the University of Pennsylvania, Philadelphia, PA, USA, in 1987.

From 1987 to 1989, he was an Assistant Professor with the Electrical Engineering Department, University of Toronto, Toronto, ON, Canada. From 1989 to 2017, he was on the faculty of Bilkent University, Ankara, Turkey. He is currently a Research Professor with the Department of Electrical and Computer Engineering, University of Illinois at Chicago. He has five U.S. patents and has written more than 100 scholarly journal articles.

Dr. Cetin is the Editor-in-Chief of *Signal, Image, and Video Processing* (Springer-Nature). He is one of the founders of Oncam-Grandeye, which designs and produces panoramic wide-angle security cameras. He has consulted Visioprime, Honeywell Video Systems, Seagate Technologies, Volant Aerial, and Bell Communications Research (Bellcore).

ALEX D. LEOW (Member, IEEE) received the Ph.D. degree in biomathematics from the University of California at Los Angeles (UCLA), Los Angeles, CA, USA, in 2003.

She received a Residency training in psychiatry from the Neuropsychiatric Institute, UCLA, and a Master's Research Fellowship training from the Laboratory of Neuro Imaging (LONI), UCLA. She has been a Medical Doctor and is a board-certified in adult psychiatry (the American Board of Psychiatry and Neurology), since 2012. In addition to managing a successful practice in neuropsychopharmacology, she has extensively published in the field of computational neuroimaging and its application in neuropsychiatry. She is currently on the faculty of the Departments of Psychiatry and Bioengineering, University of Illinois at Chicago, Chicago, IL, USA.

Dr. Leow was a recipient of the 2009 NARSAD Young Investigator Award. (Based on document published on 19 June 2013).

OLUSOLA A. AJILORE received the degree in biology *(magna cum laude)* from Harvard University, Cambridge, MA, USA.

He went on to do his M.D./Ph.D. degree at Stanford University, Stanford, CA, USA, where he studied the deleterious effects of stress hormones on the brain. He joined the research track residency at UCLA, where he transitioned into neuroimaging in major depression. He is currently using novel neuroimaging techniques to better understand the pathophysiology of mood disorders. He is also a Psychiatrist in Chicago, Illinois, and an Associate Professor with the Department of Psychiatry, University of Illinois–Chicago.

HEIDE KLUMPP (Member, IEEE) received the Ph.D. degree from the University of Georgia, Athens, GA, USA, in 2006.

She is currently an Assistant Professor and a Clinical Psychologist with the Department of Psychiatry, University of Illinois at Chicago. Her broad interests include to understand brain pathophysiology in anxiety and depression in the context of emotion, attention, and cognitive processes for clinical translation.

Dr. Klumpp is a member of the Society of Biological Psychiatry (SOBP), American Psychological Association (APA), and Anxiety and Depression Association of America (ADAA).

FATOS T. YARMAN VURAL (Senior Member, IEEE) received the B.Sc. degree from the Technical University of Istanbul, Istanbul, Turkey, in 1973, the M.Sc. degree from Bogazici University, Istanbul, in 1975, and the Ph.D. degree from Princeton University, Princeton, NJ, USA in 1981.

She returned to Turkey in 1982 and worked with Turkish Research Council, as a Researcher, from 1982 to 1985. She worked as an Assistant Professor with Drexel University and was a Research Fellow with the Massachusetts Institute of Technology, Cambridge, MA, USA, from 1985 to 1987. She was the General Manager of YapiTel Inc., from 1987 to 1992. During her courier, she developed couple of dozens of projects, including, transcription of the Ottoman Archives and Hittite plates, smart buildings, fault detection and crop classification from satellite image, and noise reduction from old recordings. Since 1992, she has been a Full Professor with the Department of Computer Engineering, Middle East Technical University, where she was the Chairperson of the Computer Engineering Department from 1996 to 2000 and the Wise President from 2000 to 2008. She chaired more than 100 National conferences and workshops. She also chaired International Symposium on Computer and Information Systems (ISCIS). She has published more than 100 articles in Machine Learning, Pattern Recognition, Image Processing, and Computer Vision in English and Turkish.