IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Towards Innovative Neural Network Paradigms: Enhanced EEG Emotion Recognition Through Hybrid STANN-3DCANN Deep Architectures

Geethanjali P [1] , Metun [2] , Debstuti Biswas [3] , Midhilesh Momidi [4] , Deepak Naidu Sarika [5]

[1] Undergraduate Student Researcher, Bangalore Institute of Technology, Bangalore, India
[2] Graduate Student, New York University, New York, USA
[3] Undergraduate Student (ECE), University of Calcutta, Kolkata, India
[4] Senior Machine Learning Engineer, Dell Technologies, Bangalore, India
[5] Undergraduate Student (CSE), Parul University, Gujarat, India

Abstract—Emotion recognition from electroencephalography (EEG) signals has become a pivotal aspect of affective computing. This research proposes the concatenation of two novel deep neural network architectures to advance the state-of-the-art in EEG-based emotion classification. The first model termed Hybrid STANN with Graph-Smooth Signals, employs a unique combination of spatiotemporal encoding and recurrent attention network blocks. Graph signal processing tools are applied as a preprocessing step for spatial graph smoothing, enhancing the interpretability of physiological representations. The model outperforms existing methods on the DEAP dataset for emotion classification. Additionally, its robustness is demonstrated through successful transfer learning from DEAP to DREAMER and the Emotional English Word (EEWD) datasets, showcasing its effectiveness across diverse EEG-based emotion classification tasks. The second model, named 3DCANN: Spatio-Temporal Convolution Attention Neural Network, addresses the dynamic nature of EEG signals in emotional states. The 3DCANN model features a spatiotemporal feature extraction module and an EEG channel attention weight learning module. By effectively capturing the dynamic relationships and internal spatial relations among multi-channel EEG signals, the model surpasses state-of-the-art performance on the (SEED) Dataset. The integration of dual attention learning and SoftMax classification enhances the model's ability to discern intricate patterns in EEG signals, resulting in superior emotion recognition accuracy. Both proposed models contribute to EEG-based emotion recognition by introducing innovative architectural elements and demonstrating their efficacy through comprehensive evaluations of diverse datasets. This research opens avenues for further exploration in physiological data-driven affective computing applications.

Index Terms—Emotion Recognition, Graph Filtering, Spatio-Temporal Encoding, 3D Convolution Attention Neural Network, Dual Attention Learning, Transfer Learning.

I. INTRODUCTION

Deep learning (DL) models typically demand a substantial number of parameters for training, presenting a departure from the parameter requirements of traditional machine learning methods [1]. Consequently, a substantial volume of labeled data becomes imperative for the effective training of DL models. Within the domain of electroencephalography (EEG) tasks, the integration of DL poses a specific challenge owing to the scarcity of labeled data about analogous tasks [2]. Mitigating this challenge, transfer-based methodologies have been proposed, utilizing a pre-existing dataset of considerable magnitude, denoted as the source dataset. However, the incongruities between target and source domains necessitate a fine-tuning process to align the network with the nuances of the target data. This research endeavors to elucidate the transferability of knowledge acquired through our proposed hybrid architecture in the context of akin emotion classification tasks. Leveraging transfer learning and subsequent fine-tuning, we explore the generalizability of our model by evaluating its performance across various tuning layouts using constrained EEG data. As the target data, we scrutinize EEG signals sourced from the publicly available DREAMER dataset [3] and an Emotional English Word dataset (EEWD) recorded at Sabanci University. These datasets encompass diverse stimuli types, providing insights into the adaptability of our proposed model in the realm of cross-modal emotion learning [4]. Through this investigation, we aim to contribute nuanced perspectives on the efficacy and adaptability of our proposed hybrid architecture in scenarios where labeled data for target tasks is limited.

Multichannel electroencephalography (EEG) serves as a valuable tool for recording neural activity in cortical regions, capturing spectral and rhythmic characteristics of brain signals [5]. In comparison to alternative non-invasive recording methods, EEG stands out for its superior temporal resolution, allowing the acquisition of brain signals on a millisecond timescale. This temporal precision, coupled with its user-friendly nature, positions EEG as a practical choice for tasks associated with cognitive and affective reactions. Nonetheless, the efficacy of EEG is tempered by challenges such as a low signal-to-noise ratio (SNR) and suboptimal spatial resolution, distinguishing it less favorably when compared to magnetic resonance imaging (MRI) and functional near-infrared spectroscopy (fNIRS) [6], [7]. Consequently, leveraging EEG signals for downstream tasks presents inherent challenges due to these limitations, prompting a careful consideration of the trade-offs involved in selecting the appropriate neuroimaging modality for specific applications.

The analysis of electroencephalogram (EEG) signals often involves the examination of specific frequency bands, including theta (θ: 4-8 Hz), alpha (α: 8-12 Hz), beta (β: 12-29 Hz), and gamma (γ: > 30 Hz). Early endeavors in EEG-based emotion classification typically encompass two primary stages: feature extraction and the implementation of a supervised machine learning framework. In a study by Wang et al. [8], the performance of three distinct features—power spectral density (PSD), wavelet entropy, and nonlinear dynamical features—was evaluated using a kernel support vector machine (SVM). Zheng et al. [9] delved into the identification of critical frequency bands and channels, employing differential entropy (DE), DE asymmetry, and PSD features. Their investigation encompassed the assessment of different features utilizing K-nearest neighbors (K-NN), SVM, and deep belief networks (DBN). In a separate work [10], the authors introduced a method to compute spectral and temporal entropies by decomposing EEG data through Fourier-Bessel series expansion-based empirical wavelet transform. This was followed by the computation of K-NN and Shannon entropies post multi-scaling operations in the spectral and temporal domains. A novel rhythmic sequencing approach was proposed in [11] to discern optimal rhythmic features from the sequence of multi-channel EEG data. Finally, the work by [12] addressed the challenge of the distribution shift between training and test data through the application of transfer learning, providing a strategy to mitigate the impact of this shift on the performance of EEG-based emotion classification models. These diverse approaches underscore the evolution and multifaceted nature of methodologies applied to extract informative features and enhance the effectiveness of supervised machine learning models in the domain of EEG-based emotion classification.

The primary objective of this study is to devise an innovative deep architecture aimed at augmenting the efficacy of current algorithms employed in EEG-based emotion recognition. Recognizing the established effectiveness of Convolutional Neural Networks (CNNs) in capturing spatial representations and Recurrent Neural Networks (RNNs) in capturing temporal dependencies across diverse domains [13], we leverage the temporal and spatial structures inherent in EEG data acquired from multiple electrodes over the scalp. Successfully analyzing these structured time series necessitates the assimilation of information extracted from

both spatial structures and temporal dynamics [14]. In response to this imperative, we propose the integration of a Hybrid End-to-End Spatio-Temporal Attention Neural Network (STANN) with graph-smooth signals, offering a comprehensive treatment of both aspects within a unified architecture with 3DCANN. The pivotal contribution of this work lies in the design of STANN, comprising two parallel blocks: the spatio-temporal encoding block and the recurrent attention network block. Acknowledging the intricate nature of brain signals and their time-varying characteristics, we introduce a novel pre-processing step involving the application of the graph Fourier transform (GFT) and low-pass graph filtering [15]. This innovative approach aims to enhance the representation and analysis of EEG data by effectively incorporating spatial and temporal structures, thereby advancing the state-of-the-art in EEG-based emotion recognition.

The significant contributions of our research are succinctly outlined as follows:

1. Innovative Deep Architecture: Introducing a novel deep architecture designed to effectively capture both spatial and temporal information inherent in time-series data, this research is specifically tailored for EEG emotion classification. The proposed hybrid network seamlessly integrates the Spatio-Temporal Attention Neural Network (STANN) and attentive temporal information within the 3D Convolution Attention Neural Network (3DCANN).

2. Spatial-Spectral Consideration: This paper addresses the interrelation among neighboring EEG electrodes, harnessing their spatial-spectral characteristics to enhance the representation of EEG data. A crucial aspect of our approach involves the application of low-pass graph filtering, a mechanism employed to enforce graph smoothness in the spatial domain.

3. Integration of Transfer Learning: The research incorporates the Transfer Learning paradigm within the 3DCANN-STANN approach. This strategic integration extends the adaptability and generalizability of the proposed architecture, allowing knowledge transfer from a source dataset, such as DEAP, to improve performance on target datasets like the DREAMER, (SEED) Dataset and the Emotional English Word dataset (EEWD). This not only demonstrates the versatility of our model but also underscores its potential applicability in scenarios where labeled data for target tasks is limited.

II. LITERATURE REVIEW

The conventional model for EEG emotion recognition typically involves two distinct stages: feature extraction and classification. Obtaining the most discriminative features is crucial in the context of EEG signal characteristics. Common feature extraction approaches analyze EEG features individually from the time domain, frequency domain, and time-frequency domain. For instance, a study [31] evaluates ten time-domain EEG signal features for negative emotion recognition, highlighting the significance of the first-order difference feature in emotion classification. However, reliance on a single feature often results in suboptimal classification outcomes for emotional states. To enhance classification performance, researchers [32] introduce a comprehensive approach where nine time-frequency EEG features are collectively input into a random forest classifier for emotion recognition. This approach significantly improves upon the limitations of single-feature methodologies. In instances involving single-channel EEG signals, an approach [33] integrates features from both the time domain and frequency domain into a neural network, thereby leveraging the strengths of both traditional features and neural networks. Another avenue explored by researchers involves the calculation of statistical features of EEG signals to enhance the discriminability of emotional features for efficient emotion recognition. In a specific instance [34], an emotion classification algorithm is proposed, incorporating the fusion of late positive potential (LPP) and support vector machines (SVM). This amalgamation achieves the integration of statistical features and frequency-domain features, contributing to the effective processing of EEG data for emotion recognition.

Electroencephalography (EEG) represents a complex, nonlinear, and high time-resolution signal, making it challenging to fully encapsulate all details solely through traditional time-domain feature extraction techniques. Signal decomposition, involving the division of data into different frequencies using distinct transformations, proves to be a valuable strategy. Applying different processing methods to various frequencies effectively enhances EEG signal classification [35]. Widely employed transforms such as wavelet transform, empirical mode decomposition (EMD), and variational mode decomposition (VMD) find

application in EEG data processing. For instance, Zhang et al. [36] propose an EEG signal feature extraction method based on EMD and autoregressive model (AR), validating its reliability and stability in emotion recognition. In another study [37], discrete wavelet transform (DWT) is utilized to extract information from a small number of EEG channels and frequency bands. Subsequently, support vector machine (SVM) and k-nearest neighbor classifiers (KNN) are employed to recognize EEG features, demonstrating commendable classification performance. Signal decomposition technologies contribute to optimizing the performance indices of EEG signal classification algorithms within the machine learning paradigm. However, manual feature extraction falls short in capturing high-level semantic information and multi-level details embedded in EEG signals. Deep learning emerges as a promising solution to this limitation, offering a pathway to extracting intricate information from EEG signals that may not be attainable through traditional methods.

Undoubtedly, the integration of spatio-temporal joint features in EEG yields superior signal representation performance, making spatio-temporal network models [38] a focal point in emotion recognition research. The effectiveness of Convolutional Neural Networks (CNNs) is intricately tied to the structure and parameter configuration of the network. Consequently, researchers often leverage adjustments to the network structure to enhance classification performance. The Two-Dimension Convolution Neural Network (2DCNN) applies the convolution layer to a two-dimensional feature map, capturing only the spatial information of the EEG signal. In contrast, the three-dimensional convolution Neural Network (3DCNN) [39] introduces a time dimension through an additional convolution layer on the three-dimensional feature map. This augmentation allows the extraction of more comprehensive spatio-temporal information from EEG data. Unlike 2DCNN, 3DCNN not only ensures synchronization of the time dimension but also extracts detailed information from both time and space domains, facilitating effective fusion of temporal and spatial features. In EEG signal processing, traditional 2D convolution may overlook the amplitude fluctuation within each period, capturing only the connection between the EEG channel and the amplitude. Conversely, utilizing 3D convolution for continuous-time EEG signals enables the capture of time-varying fluctuation information and the exploration of associated inter-channel dynamics. This distinction underscores the enhanced capability of 3DCNN in mining intricate spatio-temporal patterns in EEG data.

Moreover, scientific investigations have elucidated that distinct functions are attributed to each part of the cerebral cortex, and cooperative functional distribution is observed across various regions of the brain. To emulate this intricate brain activity, the attention mechanism (AM) has been proposed. AM can learn and establish the mapping relationship between different brain regions and emotional tasks. This allows it to emphasize disparities in spatial and temporal features associated with different emotions. For instance, a study introduced channel attention-based emotion recognition networks (CAERN) [40], demonstrating the efficacy of the feature extraction method based on AM in acquiring emotion recognition features. In a similar vein, Jia et al. [41] presented a spatial-spectral-temporal attention 3D dense network (SST-EmotionNet), exploring the correlation between emotional EEG signals and temporal changes. The application of AM in EEG emotion recognition is deemed feasible. In this paper, we formulate a 3D Convolution Attention Neural Network (3DCANN) to classify EEG emotions. Within 3DCANN, 2D EEG data undergoes conversion into a 3D representation, enabling the extraction of deep spatio-temporal information through the 3DCNN network. Subsequently, the spatio-temporal features are amalgamated with the weights obtained from dual attention learning. Finally, the fused features are input into the softmax classifier for emotion classification. This architecture leverages the strengths of the attention mechanism to enhance the discernment of emotional patterns within EEG signals.

Furthermore, Petrantonakis et al. [42] have substantiated that distinct brain areas within the cerebral cortex execute specific functions. Remarkably, changes in emotional states exhibit discernible variations in particular brain regions, notably the prefrontal and temporal lobes. This underscores the varying degrees of correlation between different brain regions and emotion recognition tasks. Opting for brain regions exhibiting high correlation holds practical significance. Consequently, our research explores the selection of channels in emotional state investigations, aiming to alleviate the computational load of the emotion recognition model by utilizing a limited number of EEG channels. The judicious selection of EEG channels not only streamlines computational requirements but also provides valuable insights for the development of innovative portable EEG equipment, offering scientific contributions to advancements in the field.

III. RESEARCH METHODOLOGY

With the widespread adoption of EEG processing and advancements in computer science, the intersection of emotion research involving Convolutional Neural Networks (CNNs) and EEG has seen gradual development. In this context, we introduce a novel CNN architecture termed 3DCANN. Comprising both a 3D Convolutional Neural Network (3DCNN) and an Attention Mechanism (AM), 3DCANN facilitates both temporal and spatial analyses. The 3DCNN component is employed for efficient spatiotemporal analysis, while the AM delves into studying the mapping relationship between EEG emotion features and specific brain regions. Through this innovative fusion of 3DCNN and AM, 3DCANN emerges as a robust framework capable of realizing efficient classification across various emotions. The temporal and spatial insights provided by 3DCNN, coupled with the mapping relationship elucidated by the AM, collectively contribute to the enhanced capabilities of 3DCANN in deciphering and classifying intricate emotional patterns from EEG data. With the widespread adoption of EEG processing and advancements in computer science, the intersection of emotion research involving Convolutional Neural Networks (CNNs) and EEG has seen gradual development. In this context, we introduce a novel CNN architecture termed 3DCANN. Comprising both a 3D Convolutional Neural Network (3DCNN) and an Attention Mechanism (AM), 3DCANN facilitates both temporal and spatial analyses. The 3DCNN component is employed for efficient spatiotemporal analysis, while the AM delves into studying the mapping relationship between EEG emotion features and specific brain regions. The proposed architecture of the STANN-3DCANN attention network is illustrated in Figure 1.

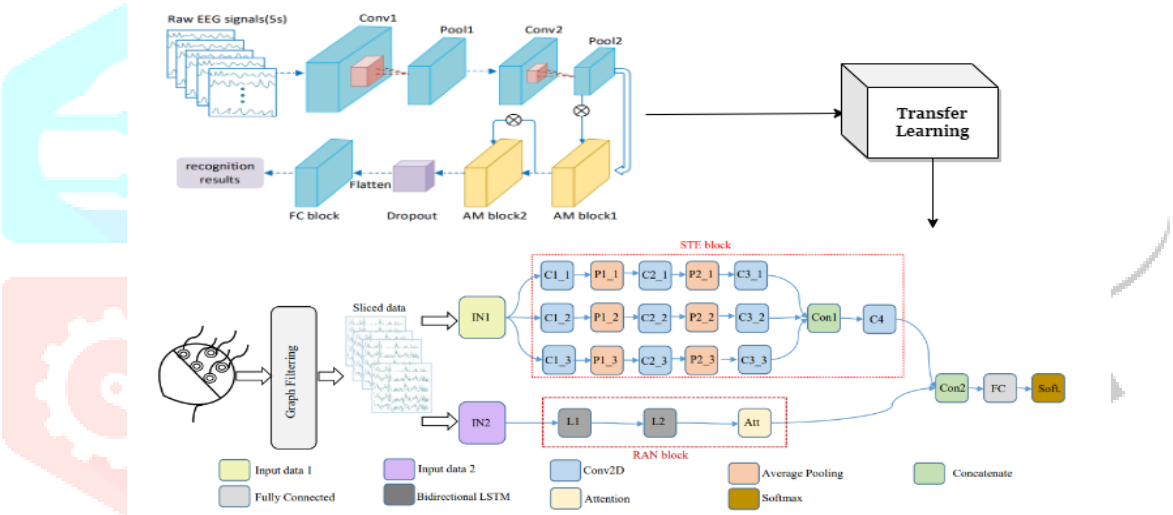

Fig. 1. Overview of the proposed 3DCANN-STANN architecture.

In the 3DCANN framework, spatial features of EEG signals are extracted through the 3D Convolutional Neural Network (3DCANN). The attention mechanism (AM) is then leveraged to enhance emotion-related electrodes while suppressing emotion-independent electrodes in the first AM block module, achieved through the learning of the weight matrix. Subsequently, the second AM block module strengthens or weakens the influence of different electrodes in emotion recognition, contributing to the attainment of highly discriminative high-dimensional EEG features. Given the high complexity and limited sample size of EEG data, a dropout layer with a ratio of 0.2 is incorporated in 3DCANN to mitigate the risk of overfitting within the neural network. The output of the dropout layer is then fed into the Flatten layer, enabling the realization of probability estimation for various emotional states through three fully connected layers. The Rectified Linear Unit (ReLU) activation function is employed in the first two layers of the fully connected network, while the SoftMax activation function is applied in the final layer to estimate the probabilities associated with different emotional states. The spatial-temporal feature extraction and attention mechanism of the network are expounded upon in the subsequent sections.

The linear transformation $W \in R^{n*m}$ is used to transform the m - dimensional vector into the n - dimensional vector, while bias is denoted as $b \in R^{n}$. In 3DCANN, the last fully connected layer determines the emotional state corresponding to EEG signals through SoftMax activation function, which can be expressed as

$$p_k = \frac{\exp(w_k^T x)}{\sum_j \exp(w_k^T x)}$$

where x represents a sample, k represents an emotional state. $\sum_j(w_k^T x)$ is the normalization term and the sum of the guaranteed probabilities is 1. The SoftMax classifier has unique advantages when dealing with natural language, multi-dimension signals, and color image data, and is popularized in deep learning models.

Exploring structural and functional connectivity of the brain [16] and tracking the relative spatial positions of EEG nodes could be in use of decoding responses elicited from sensory stimuli [17]. To exploit GSP tools, we need to define an underlying graph. Thus, we calculate the pairwise Euclidean distances of EEG electrodes and build the graph accordingly. In this way, we solely require the Cartesian coordinates of electrodes while the classical common spatial pattern (CSP) filtering depends on individual subjects or tasks. Let us consider an undirected, weighted graph $G(V, E, A)$, where $V = \{1, 2, \ldots, n\}$ is the set of nodes or EEG channels, $E \subseteq V \times V$ is the set of edges or spatial connections, and $A \in R\, n \times n$ is the adjacency matrix. The edge weights Aij are inversely proportional to the pairwise Euclidean distances between node i and j.

For each electrode, K-NN is considered to construct the adjacency matrix while keeping it symmetric to represent the brain topology [18]. To avoid a densely-connected graph, we set K to 2 and 4. In the literature, a 2-NN topology was motivated by separating the graph into frontotemporal and parieto-occipital networks [19]. The 4-NN graph brings engagement with central areas as well. Figure 2 shows the final adjacency matrices for these K values and their corresponding scalp topology

Since the electrodes installed in adjacent locations detect electrical activities of common sources [20], we apply smoothing via lowpass graph filtering concerning the defined graph. This will ensure the similarity of the behavior among neighbor electrodes. Low frequencies in the graph correspond to small eigenvalues. Considering $\tilde{h} = [\tilde{h}1, \tilde{h}2, \ldots, \tilde{h}n]$ the ideal lowpass filter with bandwidth $w \in \{1, 2, \ldots, n\}$ where $\tilde{h}i = 0$ if $i > w$, the GFT coefficients corresponding to the low frequencies with respect to G are given by:

$$\tilde{X}_{low} = \tilde{h}\tilde{X}$$

where $\tilde{h}$ is equal to one for $w = [1, n/2]$ and zero otherwise. While we choose the filter bandwidth as $n/2$, users can choose a different number depending on the level of smoothing they want to apply. A smaller bandwidth would lead to more smoothing.

The Proposed Hybrid STANN-3DCANN Network:

In several sequence-based applications such as semantics analysis, natural language processing, and medical imaging, certain time steps of the input data might contain the most discriminative information and attention mechanism addresses this issue by focusing on specific time steps [21]. In this mechanism, the most discriminative task-related features are calculated by multiplication of outputs of hidden states by trainable weights. The output of the attention layer, v, is computed as below:

$$v = \sum_i \alpha_i h_i$$

$$\alpha_i = \frac{\exp(Wh_i + b)}{\sum_j \exp(Wh_i + b)}$$

where h_i denotes the output of the i_th LSTM layer, and W and b are trainable parameters.

Transfer Learning:

To examine the performance of the proposed network, the features of STE and RAN block are fused and fed to the dense layer. Next, the encoded representation is fed to the final SoftMax classifier. The cross-entropy loss, L, is calculated as follows:

$$L = \sum_i Y_i \log \tilde{\tilde{Y}}_i$$

where Yi is the ground truth emotion label for each data sample and $\widetilde{Y}_i$ is the predicted label. Finally, the weights and the biases are trained with batch gradient descent.

To investigate the possibility of using this trained network in similar EEG-based emotion recognition tasks, we propose and implement a transfer learning (TL) approach. The goal of TL is to test our model ability in real-life conditions where the available amount of labeled data is not sufficient. TL helps to improve the learning capability of the target data by leveraging the knowledge of the source domain. In this study, we investigate transferring the learned model parameters assuming that individual models across different datasets with similar tasks should share some parameters. Firstly, the model is fully trained using sufficient labeled data (source dataset). Second, we peruse different schemes to tune the pre-trained network via the target dataset. The source and target datasets involve EEG-based emotion recognition experiments with different stimuli.

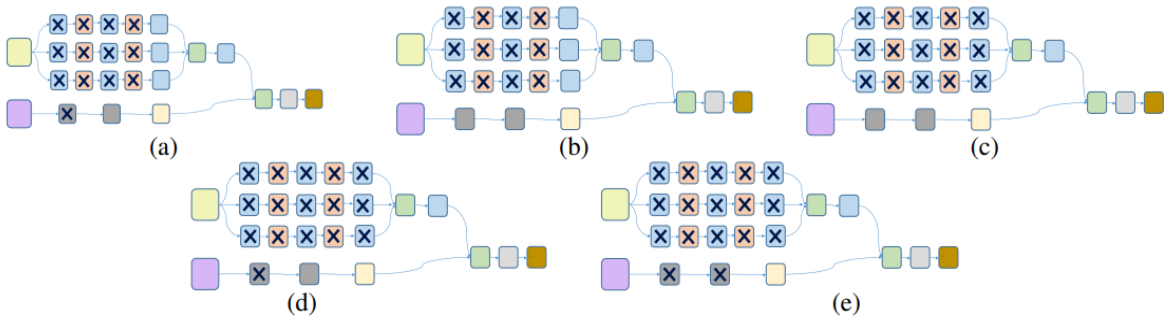

Fig. 2. The proposed transfer learning scheme

Figure 2 demonstrates five different schemes we consider in the calibration (fine-tuning) session. The TL schemes in our STE blocks are inspired by observations in CNN-based TL frameworks in computer vision where usually the later network layers are retrained, as the earlier layers are responsible for generic features [22]. In our work, we consider different retrainable cells in both STE and RAN blocks. Going from scheme (a) to scheme (e), we change the status of exactly one layer either to retrainable or non-retrainable (frozen) at each step. In each scheme, blocks marked with a cross are left unchanged during fine-tuning of the network. The number of retrainable parameters in scheme (a) to scheme (e) is equal to 239100, 311420, 280295, 207975, and 53735, respectively.

Due to the variations in inter-dataset samples, we use a small part of the new data (target data), N, to calibrate and fine-tune our pre-trained model. Since the amount of calibration data is limited, we scale down the initial learning rate to avoid clobbering in initialization [23]. Scaling down the initial learning rate η with the scale α can be defined as:

$$\Phi^{i+1} = (1-\alpha)\phi^{i} + \alpha(\phi^{i} - \eta\frac{\partial\mathcal{L}}{\partial\Phi})$$

where Φ i is the trainable parameters at the i_th iteration and L is the cross-entropy loss function. Here, α is set to 0.1.

IV. RESULTS & DISCUSSION

The 3D convolution attention neural network includes two parts: 3DCNN feature extraction and EEG channel attention weight learning. The spatial and temporal characteristics of different emotional states were obtained by 3DCNN and the channel attention is used to learn the weight matrix of different electrodes affecting the emotional state.

TABLE 1 PARAMETERS SETTING OF 3DCANN

Layer	Type	Output/Shape	Kernel Size
1	Conv3D	5*62*200*32	5*5*5
2	Permute	1*64*50*62	1*1*2
3	Dense	1*64*50*62	2*1*2
4	MaxPooling3D	2*62*100*32	2*1*2
5	Multiply	1*62*50*64	1*2*2
6	Dropout	1*62*50*64	2*1*2

It is noteworthy that the transformation of 2D EEG into 3D EEG is accomplished using a sliding window, a technique that preserves the temporal information of EEG signals to the maximum extent. By amalgamating the spatio-temporal feature extraction capabilities of 3D Convolutional Neural Networks (3DCNN) with the attention mechanism, the 3DCANN method adeptly enhances the comprehensiveness of EEG features. Furthermore, it effectively captures and reflects the intricate relationships between different channels and emotional states. This approach leverages the synergistic benefits of both 3DCNN and the attention mechanism, contributing to an optimized representation of EEG data that encapsulates both spatial and temporal dimensions for improved emotion recognition. In the comparison with state-of-the-art solutions, the performance of the proposed 3DCANN-SSTANN method for valence and arousal classification on wide-band EEG data from the DEAP dataset is evaluated against several contemporary approaches, including DCCA [24], ECLGCNN [25], DGCNN [26], CVCNN [27], CRAM [28], ACRNN [29], and S-EEGNet [30]. DCCA computes Differential Entropy (DE) features across four frequency bands, followed by the extraction of representations for two modalities through multiple stacked layers of nonlinear transformations. ECLGCNN integrates graph convolutional neural networks with Long Short-Term Memory (LSTM) units, employing DE of windowed EEG data as input. DGCNN1 computes DE features and employs a 2400-feature vector as the input for graph CNN processing. CVCNN combines raw EEG and normalized EEG data with Power Spectral Density (PSD) features. CRAM extracts spatio-temporal information and attentive temporal dynamics in a cascaded format, utilizing a CNN layer with a fixed kernel and filter size. ACRNN utilizes an attention technique to assign different weights to each EEG channel, followed by CNN processing for spatial feature extraction. SEEGNet applies Hilbert–Huang transform preprocessing before feeding the data to a separable CNN. Casc-CNN-LSTM employs hybrid convolutional recurrent neural networks by transforming 1D EEG vector sequences into 2D mesh-like matrices. Table 2 provides a comprehensive comparison of the proposed SS4-STANN method on wide-band data against the state-of-the-art methods from recent literature. The results, reported with subject-dependent 10-fold cross-validation, underscore the superior performance of the proposed approach. In instances where a 4-fold cross-validation was performed by other authors, the proposed method continues to exhibit superiority.

TABLE 2 COMPARISON OF THE PROPOSED 3DCANN-STANN METHOD WITH RECENT STATE-OF-THE-ART SOLUTION FOR VALENCE AND AROUSAL CLASSIFICATION

Method	Valence (%)	Arousal (%)
S-EEGNet [30]	89.9 %	88.3 %
ACRNN [29]	89.9 %	88.3 %
CVCNN [27]	88.8 %	86.9 %
DGCNN [26]	92.5 %	93.5 %
ECLGCNN [23]	90.5 %	90.6 %
DCCA [24]	84.3 %	85.6 %
Proposed method [Hybrid STANN-3DCANN]	96.4 %	97.5 %

Evaluating Transfer Learning Performance To show that the learned representations based on 3DCANN-STANN have the generality to be applied in similar tasks that have limited labeled data, we perform TL in cross-dataset TL. To make a fair comparison among the different amounts of calibration data, we set 10% of the target data samples as a test set, and the calibration data is selected among the rest 90% of the data.

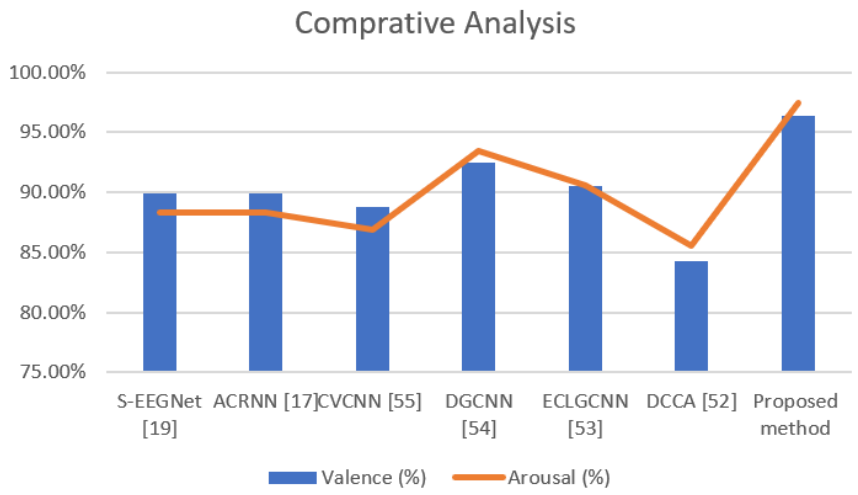

Fig 3: Custom Combination Chart showing comparative analysis

For the cross-dataset Transfer Learning (TL), our proposed model is initially trained on the entire DEAP dataset, and subsequently fine-tuned on new EEG emotion recognition datasets that were collected with different stimuli. Specifically, we choose to apply TL on the publicly available DREAMER dataset and the Emotional English Word dataset (EEWD). To ensure consistency in dataset characteristics, including EEG electrodes, frequency bands of interest, and the sliding window for segmentation, we adopt the following settings. We select fifteen common electrodes that are present across all datasets, namely AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4. The same data-slicing process is applied to the DREAMER dataset. Given that the trial length in DREAMER varies from trial to trial, we opt to select the last 60 seconds of each trial and segment it into 1-second data samples. This ensures uniformity in the data representation across datasets, facilitating the effective transfer of learned features from the source (DEAP) to the target datasets (DREAMER and EEWD).

To assess the transfer learning capability of models initially trained with EEG responses to video clips when applied to EEG signals elicited by written words, we focus on the Emotional English Word Dataset (EEWD). In this evaluation, we specifically address a binary valence classification scenario, as all the words in the EEWD experiment are selected from the high arousal group. For this binary valence classification, we designate trials with rating values lower or equal to 3 as indicative of negative valence, while trials corresponding to rating values higher or equal to 7 are considered as representing positive valence. This classification scheme allows us to effectively distinguish between negative and positive valence categories in the context of the EEWD dataset.

TABLE 3 TL RESULTS FOR DIFFERENT SCHEMES ON VALENCE AND AROUSAL CLASSIFICATION OF DREAMER USING THE PROPOSED 3DCANN-STANN MODEL (DEAP → DREAMER→SEED)

Amount of calibration data (N)	Valence	Arousal
1 trial per class	64.0 %	64.9 %
10% of data samples	75.0 %	81.0 %
20% of data samples	82.8 %	83.4 %
90% of data samples	83.5 %	87.3 %

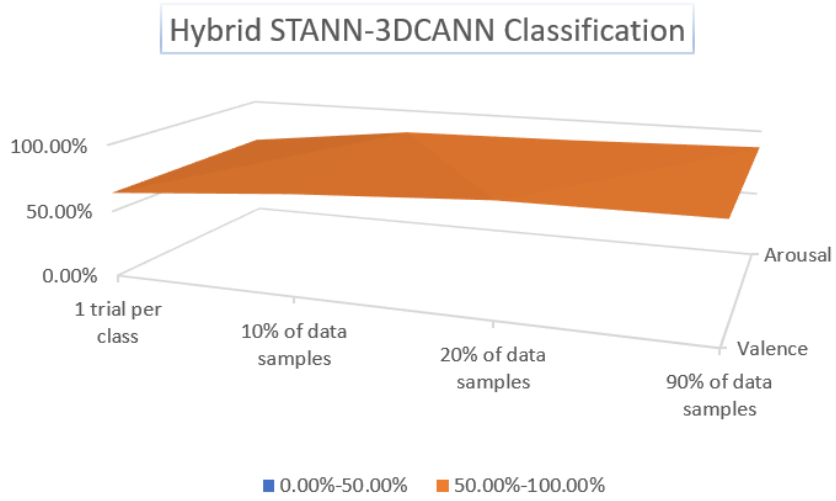

Fig. 4. Surface map of representative samples in the last dense layer ahead of the classification SoftMax layer, visualized using t-SNE.

To validate the efficacy of the proposed model and demonstrate the impact of Transfer Learning (TL) in segregating components related to each class, we randomly select EEG samples from one subject in the DREAMER dataset and visualize them using t-distributed Stochastic Neighbor Embedding (t-SNE) (Figure 4). The scatter plot in Figure 4 depicts the samples in the last dense layer before the classification SoftMax layer without TL and fine-tuning. Subsequently, the figure illustrates the outcomes after applying TL and fine-tuning. The initial representations before fine-tuning reveal a more mixed configuration of samples across different classes. However, with the integration of TL and fine-tuning, the representations become more distinct and separable, emphasizing the effectiveness of TL in enhancing the model's ability to differentiate components associated with each class. This visual analysis provides insight into the improved feature separation achieved through the proposed TL approach.

V. CONCLUSION

In this study, a novel deep-learning architecture for subject-dependent EEG-based emotion classification tasks was proposed. The 3DCANN-STANN involves a hybrid structure with parallel Spatio-Temporal Encoding (STE) and Recurrent Attention Network (RAN) blocks, complemented by an attention mechanism. In this study, an EEG research framework based on the three-dimensional convolution attention neural network (3DCANN) is presented, demonstrating its effective optimization for emotion recognition performance. This architecture adeptly captures both spatial and temporal information inherent in multi-channel EEG data, concurrently ensuring graph smoothness in the spatial domain through graph signal processing. The proposed approach demonstrated superior performance compared to state-of-the-art solutions, achieving classification accuracies exceeding 97.0% for valence, arousal, and dominance on the DEAP dataset. The exploration of critical frequency bands and regions further validated the efficacy of the proposed method. The 3DCANN method comprises a feature extraction component using 3D convolutional neural networks (3DCNN) and a channel attention weight learning module. Firstly, the 3D convolution in the 3DCNN module not only calculates features in the spatial dimension but also extracts dynamic information in the time domain, effectively capturing complex features in EEG data and demonstrating strong performance in biosignal recognition tasks. The incorporation of 3D max-pooling significantly reduces the dimension of the feature map, preventing overfitting. Secondly, to reduce feature redundancy and investigate the relationship between EEG channels and emotional states, the attention mechanism (AM) block is employed to learn the channel attention weight matrix. The attention weight distribution of 65 electrodes is obtained through the SoftMax activation function, providing insights into the connection between different brain regions and emotions. Additionally, the applicability of the learned representations across different EEG emotion recognition tasks was showcased, demonstrating cross-modal transferability potential. The model exhibited robust performance with both similar and different stimulus modalities, emphasizing the generalization capabilities of the trained architecture. Finally, based on the algorithm's evaluation results, 3DCANN optimizes the characteristic information of EEG signals and leverages hidden spatio-temporal characteristics to achieve optimal performance in emotion recognition tasks. For future research, investigations into the impact of integrating

different modalities alongside EEG in similar problem domains are planned. Furthermore, the aim is to extend the spatio-temporal feature learning concepts presented here to address challenges involving other modalities of physiological data, such as functional MRI (fMRI) or magnetoencephalography (MEG). This expansion could contribute to a more comprehensive understanding of emotional states across various neurophysiological domains.

REFERENCES

[1] K. Zhang, N. Robinson, S.-W. Lee, and C. Guan, "Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network," Neural Networks, vol. 136, pp. 1–10, 2021.

[2] M. Wronkiewicz, E. Larson, and A. K. Lee, "Leveraging anatomical information to improve transfer learning in brain–computer interfaces," J. Neural. Eng., vol. 12, no. 4, pp. 046027, 2015.

[3] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through EEG and ecg signals from wireless low-cost offthe-shelf devices," IEEE J. Biomed. Health. Inform., vol. 22, no. 1, pp. 98–107, 2017.

[4] S. Zhang, M. Chen, J. Chen, Y.-F. Li, Y. Wu, M. Li, and C. Zhu, "Combining cross-modal knowledge transfer and semi-supervised learning for speech emotion recognition," Knowl. Based Syst., vol. 229, pp. 107340, 2021.

[5] S. M. Alarcao and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," IEEE Trans. Affect. Comput., vol. 10, no. 3, pp. 374–393, 2017.

[6] T. Nguyen, S. Ahn, H. Jang, S. C. Jun, and J. G. Kim, "Utilization of a combined EEG/NIRS system to predict driver drowsiness," Sci. Rep., vol. 7, no. 1, pp. 1–10, 2017.

[7] M. S. Al-Quraishi, I. Elamvazuthi, T. B. Tang, M. Al-Qurishi, S. H. Adil, and M. Ebrahim, "Bimodal data fusion of simultaneous measurements of EEG and fNIRS during lower limb movements," Brain Sciences, vol. 11, no. 6, pp. 713, 2021.

[8] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," Neurocomputing, vol. 129, pp. 94–106, 2014.

[9] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," IEEE Transactions on Autonomous Mental Development, vol. 7, no. 3, pp. 162–175, 2015.

[10] A. Bhattacharyya, R. K. Tripathy, L. Garg, and R. B. Pachori, "A novel multivariate-multiscale approach for computing EEG spectral and temporal complexity for human emotion recognition," IEEE Sensors Journal, vol. 21, no. 3, pp. 3579–3591, 2020.

[11] J. W. Li, S. Barma, S. H. Pun, M. I. Vai, and P. U. Mak, "Emotion recognition based on EEG brain rhythm sequencing technique," IEEE Transactions on Cognitive and Developmental Systems, 2022.

[12] C. Yang, Z. Deng, K.-S. Choi, and S. Wang, "Takagi–sugeno–kang transfer learning fuzzy logic system for the adaptive recognition of epileptic electroencephalogram signals," IEEE Transactions on Fuzzy Systems, vol. 24, no. 5, pp. 1079–1094, 2015

[13] J. Chen, D. Jiang, Y. Zhang, and P. Zhang, "Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset," Computer Communications, vol. 154, pp. 58–65, 2020.

[14] S. Sartipi, M. Torkamani-Azar, and M. Cetin, "EEG emotion recognition via graph-based spatio-temporal attention neural networks," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2021, pp. 571–574.

[15] S. S. Saboksayr, G. Mateos, and M. Cetin, "Online discriminative graph learning from multi-class smooth signals," Signal Processing, vol. 186, pp. 108101, 2021.

[16] H.-J. Park and K. Friston, "Structural and functional brain networks: from connections to cognition," Science, vol. 342, no. 6158, 2013.

[17] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang, et al., "EEG based emotion recognition by combining functional connectivity network and local activations," IEEE Trans. Biomed. Eng., vol. 66, no. 10, pp. 2869–2881, 2019.

[18] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," Nature reviews neuroscience, vol. 10, no. 3, pp. 186–198, 2009.

[19] S. Itani and D. Thanou, "A graph signal processing framework for the classification of temporal brain data," in 2020 28th European Signal Processing Conference (EUSIPCO). IEEE, 2021, pp. 1180–1184.

[20] H. Higashi, T. Tanaka, and Y. Tanaka, "Smoothing of spatial filter by graph fourier transform for EEG signals," in Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific. IEEE, 2014, pp. 1–8.

[21] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[22] J. Teuwen and N. Moriakov, "Convolutional neural networks," in Hand-book of medical image computing and computer assisted intervention, pp. 481–501. Elsevier, 2020.

[23] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 1, pp. 142–158, 2015.

[24] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, "Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition," IEEE Transactions on Cognitive and Developmental Systems, 2021.

[25] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and lstm," Applied Soft Computing, vol. 100, pp. 106954, 2021.

[26] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," IEEE Trans. Affect. Comput., vol. 11, no. 3, pp. 532–541, 2018.

[27] J. Chen, P. Zhang, Z. Mao, Y. Huang, D. Jiang, and Y. Zhang, "Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks," IEEE Access, vol. 7, pp. 44317– 44328, 2019.

[28] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent EEG signal analysis," IEEE Signal Processing Letters, vol. 26, no. 5, pp. 715–719, 2019.

[29] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "EEG-based emotion recognition via channel-wise attention and self-attention," IEEE Trans. Affect. Comput., 2020.

[30] W. Huang, Y. Xue, L. Hu, and H. Liuli, "S-EEGNet: Electroencephalogram signal classification based on a separable convolution neural network with bilinear interpolation," IEEE Access, vol. 8, pp. 131636–131646, 2020.

[31] F. Feradov and T. Ganchev, "Ranking of EEG time-domain features on the negative emotions recognition task," Annual Journal of Electronics, vol. 9, pp. 26-29, 2015.

[32] T. D. Kusumaningrum, A. Faqih, and B. Kusumoputro, "Emotion Recognition Based on DEAP Database using EEG Time-Frequency Features and Machine Learning Methods," J. Phys.: Conf. Ser., vol. 1501, pp. 012020, Nov. 2019.

[33] A. Gómez, L. Quintero, N. López and J. Castro, "An approach to emotion recognition in single-channel EEG signals: a mother child interaction," J. Phys.: Conf. Ser., vol. 705, no. 1, pp. 012051, Oct. 2015.

[34] R. M. Mehmood and H. J. Lee, "A novel feature extraction method based on late positive potential for emotion recognition in human brain signal patterns," Comput. Elect. Eng., vol. 53, pp. 444-457, Jul. 2016.

[35] V. Bajaj, R. B. Pachori, "Human emotion classification from EEG signals using multiwavelet transform," Proc. Int. Conf. Med. Biometrics, pp. 125-130, Jun. 2014.

[36] Y. Zhang, S. Zhang and X. Ji, "EEG-based classification of emotions using empirical mode decomposition and autoregressive model," Multimedia Tools Appl., vol. 77, no. 20, pp. 26697-26710, Mar. 2018.

[37] Z. Mohammadi, J. Frounchi and M. Amiri, "Wavelet-based emotion recognition system using EEG signal," Neural Comput. Appl., vol. 28, no. 8, pp. 1985-1990, Jan. 2016.

[38] J. Yang, X. Huang, H. Wu and X. Yang, "EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network," Procedia Computer Science, vol. 174, pp. 491-504, Jul. 2020.

[39] P. C. Petrantonakis, L. J. Hadjileontiadis, "A novel emotion elicitation index using frontal brain asymmetry for enhanced EEG-based emotion recognition," IEEE Trans. Inf. Technol. Biomed., vol. 15, no. 5, pp. 737-746, Sep. 2011.

[40] X. Zhang, T. Du and Z. Zhang, "EEG Emotion Recognition Based on Channel Attention for E-Healthcare Applications," in International Conference on Multimedia Modeling. Springer, vol. 12573, pp. 159-169, Jan. 2021.

[41] Z. Jia, Y. Lin, X. Cai, H. Chen, H. Gou and J. Wang, "SST-EmotionNet: Spatial-spectral-temporal based attention 3D dense network for EEG emotion recognition," in Proc. 28th ACM Int. Conf. Multimedia (MM), pp. 2909–2917, Oct. 2020.

[42] S. L. Oh, J. Vicnesh, E. J. Ciaccio, R. Yuvaraj and U. R. Acharya, "Deep convolutional neural network model for automated diagnosis of schizophrenia using EEG signals," Appl. Sci., vol. 9, no. 14, pp. 2870, Jul. 2019.