ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEEP LEARNING ARCHITECTURES FOR EXPLORING SPATIO-TEMPORAL PATTERNS FROM EEG DATA FOR EMOTION DETECTION

ANDHAVARAPU BHANUSRI[1], PROF. MOGALLA SHASHI[2]

[1]Research Scholar, Department of Computer Science and Systems Engineering, Andhra University College of Engineering, Andhra University, India

[2]Professor, Department of Computer Science and Systems Engineering, Andhra University College of Engineering, Andhra University, India

E-mail: [1]bhanusriandhavarapu@gmail.com, [2]smogalla2000@yahoo.com

ABSTRACT

Human emotions are crucial across various domains by closing the divide between human and current technologies, thus fostering better understanding on critical mental health conditions. Traditional approaches rely mainly on facial expressions along with body gestures and are met with limited success as they miss underlying or repressed emotions. Electro Encephalo Gram (EEG) signals offer a direct glimpse into brain activity, making them a promising avenue for fool-proof emotion recognition. However, the complex temporal dynamics of EEG data pose challenges for classical machine learning algorithms, which often fail to capture spatial and temporal patterns effectively. To overcome these issues, we introduced a novel model called Spatio-Temporal Difference Identification - Convolution Neural Network (STDI-CNN). Our model efficiently captures the complex temporal dynamics of EEG data, indicating brain activity across time in various lobes, by utilizing deep learning approaches, especially a combination of CNN and sequential neural network architectures. Extensive experiments on the SEED EEG dataset demonstrate the efficacy of the proposed STDI-CNN model, achieving an impressive accuracy of 98.52%. Additional tests using CNN-LSTM and CNN-BiLSTM models also yielded strong performance, with accuracy rates of 97.04%. This surpasses current SOA models and highlights the potential of STDI-CNN in extracting meaningful patterns from EEG signals for emotion recognition. Our work reduces the gap by featuring a significant step forward in harnessing EEG signals to build well informed emotionally intelligent system that fosters prior detection and improved diagnosis for neurological disorders.

Keywords: *EEG Emotion Recognition, CNN For Emotion Recognition, Fusion of CNN and LSTM, Spatial and Temporal Patterns, SEED Dataset.*

1. INTRODUCTION

Recognizing human emotion is critical in various domains and has impact and relevance in determining human behavior, cognition and interaction with the environment. External indicators of emotion, such as facial expressions and vocal intonations, are non-physiological signals that help convey emotions. Distortions in these non-physiological signals, whether deliberately or accidentally, may be obstructive to the correct identification of an individual's mood [1], [2]. It is therefore crucial to appreciate and guarantee the purity of such signs in recognizing emotions and subsequent applications.

The brain is the primary integrative organ and the source of initiative in generating and mobilizing affective responses at the physiological level in the body. Electro Cardio Gram (ECG), Electro Muscular Gram (EMG), EOG, and EEG are the main physiological signals that contain information concerning the body. In contrast to the analysis of outer emotional experience, the EEG allows the study of neurological processes in the brain and is a better indicator of a person's emotional state. Unlike other physiological measurements, EEG has a higher spatial and temporal density which may be more informative for understanding the time course of affective processes.

However, in the medical field, EEG signals are used in diagnosing neurological disorders including epilepsy, Parkinson's, seizures [3], and Alzheimer's [4] and in modern usage for personalized teacher-student interaction [5], to understand customer's unhindered preferences on a broad range of product images [6]. This could make a substantial contribution to the fields of Assistive technology and Affective Computing [7], [8].

The EMG and ECG are physiological signals that show emotions along with the EEG, which may have noise from brain cells. Hence, EEG signals are extensively used for emotion detection, and many researchers have conducted experiments to analyze EEG signals by applying various algorithms.

The general procedure of classifying human emotions from the EEG signals, as illustrated in figure 1, begins with the recording of raw EEG data from human brain lobes at a high frequency. The EEG headset records brain signals, and an accompanying waveform represents the raw data. Next, we pre-process the signal to reduce high-frequency noise; we decimate the raw EEG signal and then send it through a band-pass filter. And then eliminate undesired artifacts from the EEG data, like muscular contractions or eye blinks, along with it, and make sure that the signal is noise-free. Later, the pre-processed signal has to be decomposed into frequency bands and then features are extracted. Thereafter apply any classifier which accurately classifies the human emotion.

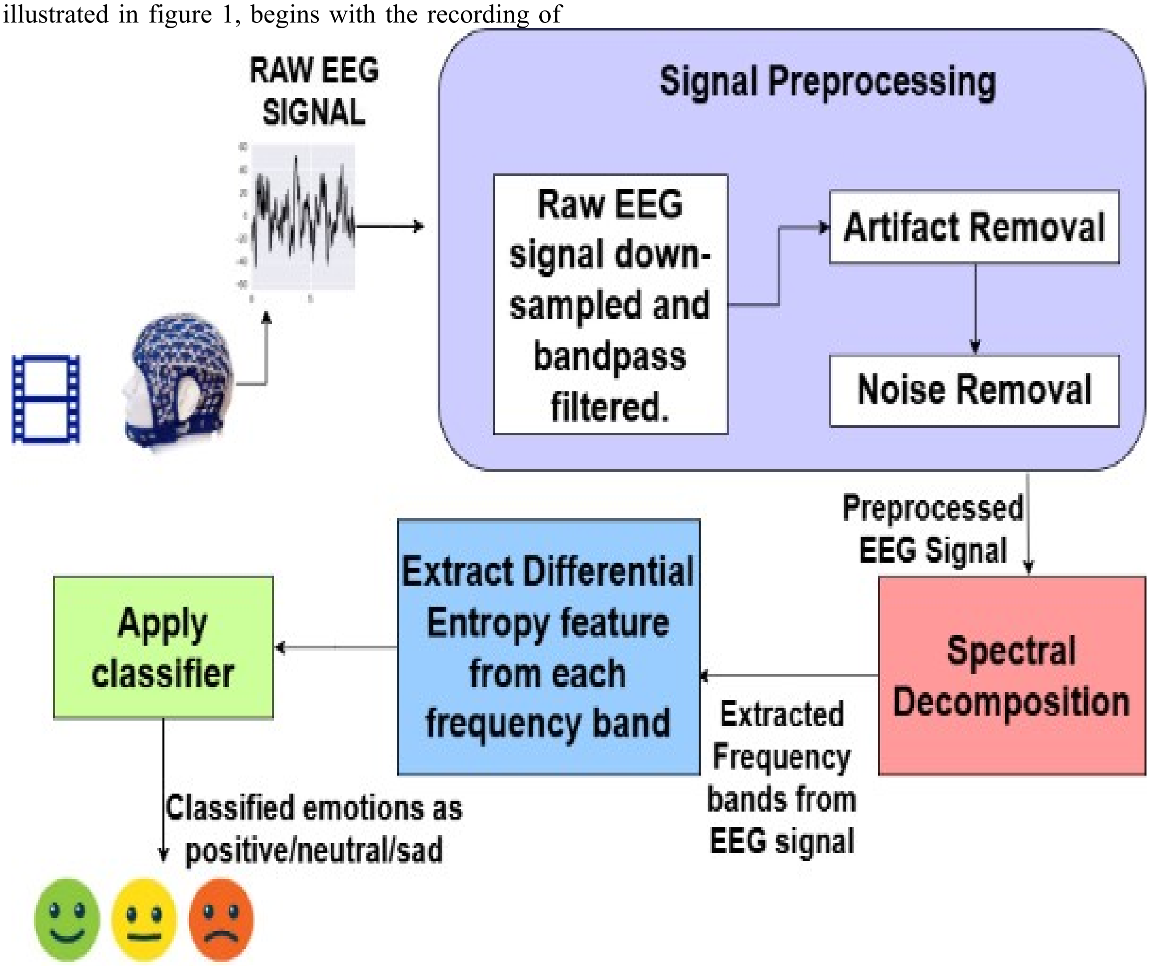

Figure 1: Procedure of emotion recognition from EEG recordings

Machine learning (ML) algorithms are hindered by the fact that feature selection has to be done explicitly or manually, where important features are identified for inclusion based on domain knowledge and statistical analysis, which may overlook spatial latent features in the dataset. The conventional machine learning algorithms have been challenged to extract prominent features and identify patterns of emotions because of the transient nature of EEG signals. The temporal behaviour and low Signal to Noise Ratio (SNR) feature of EEG signals adds to the complexity of feature extraction with traditional ML approaches [9].

Deep Learning (DL) architectures have proven capacity to process spatial and temporal

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

information from raw data; Convolution Neural Networks (CNNs) are expressly crafted to process spatial data, and the Long Short-Term Memory (LSTM) can effectively process temporal data. The spatiotemporal feature extraction process is crucial in identifying emotions from EEG signals. However, with deep learning (DL) algorithms, all the features hidden in the dataset can be extracted, and the electrode selection is automated based on the best input selection procedure of the specified neural network architecture. Despite this automation, the design and training of deep learning models still involve manual decisions such as selecting architecture and topology, tuning hyper parameters, and preprocessing data; sometimes manual feature engineering remains valuable, particularly with structured data or when domain expertise provides crucial insights beyond what the model can capture autonomously.

1.1 Related Work

Nouman et al., [11] works with non-physiological signals for emotion recognition. In their work, they considered facial edges but also traced eyeballs for detecting emotions through Hough circle transform.

Mu Li et al., [12] stated that the gamma frequency band is most relevant to emotional states. Techniques like LDS smoothing and minimal-redundancy maximal-relevance (MRMR) algorithms enhance classifier accuracy and efficiency. Chen Wei (2020) et al., [13] also stated that the higher frequency bands such as beta and gamma are most prominent while extracting emotions from the EEG signals. Ruo-Nan Duan et al., [14] proposed a DE feature as the best feature among the existing features of SEED for emotion recognition. G. Li et al., [15] proposed an experiment-level Batch-Normalization (BN) to capture individual differences for recognizing emotions on the SEED dataset. The author intends to work with fewer electrode channels instead of using all. There are cases, where working with unwanted data, the model quality will be significantly affected. The author tested PSD and DE features of SEED to get improved results out of it. Chunawale. A et al., [16] stated that feature selection and extraction of EEG signals plays a crucial role while applying machine learning techniques. Here, they preferred PSD features for processing EEG data and achieved 96.42% accuracy.

On DE features of SEED dataset, MLP (Multi-Layer Perceptron) and CNN has been implemented by Mohith Kumar et al., [17], CNN's positioning in this model is to learn from the spatial structure of placing different EEG channels on the scalp and, therefore, capture subtle patterns in brain activity. This feature enables CNN to outcompete MLP models in EGG data analysis procedures, which do not automatically include spatial information. Another research done by Mitul Kumar Ahirwal et al. [18] was for recognizing emotions using EEG signals, the work mainly covered pattern recognition and classification. Time domain, frequency domain, and statistical entropy are analysed with great detail, coupled with a methodical comparison of various classifiers including ANN, demonstrated as the best classifier with a mean accuracy of 93.75 % when using entropy-based features.

Li et al., [19] addressed the over fitting challenge in traditional machine learning algorithms while recognizing emotions through EEG signal analysis. Hierarchical Convolutional Neural Network (HCNN) outperforms especially elucidates nuances in classifying emotions within high-frequency Beta and Gamma wave bands. Comparative evaluations confirm HCNN's superiority with accuracy of 86.2% and 88.2% on the former, marking a significant advancement in affective computing.

R. K. Jeevanet et al., [20] look into EEG Based Emotion Recognition through LSTM-RNN and LSTM-CNN models. EEG signals feel discrete wavelet transforms for spectral filtrate and wavelet-based filtering for the division of the spectra band is used for classifying. Subjects watch objects for 1200ms which are images and music associated with babies or scenes. EEG signals are collected using a 32-channel configuration recorded at 350 Hz sampling rate. The LSTM + CNN models were better than conventional methods with 64% accuracy while the LSTM or RNN took 4 to 5 hours for training the model however it had an accuracy of 63.61%.

To enhance the performance of EEG-based emotion recognition, S Hwang et al., [21] used topology preserved DE images for classifying emotions on the SEED dataset by applying CNN along with LSTM. Shen et al. [22] put forward a new approach called the 4D convolutional RNN (4D-CRNN). Finally, a CRNN model, LSTM, is used to learn from these features and demonstrates

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the feasibility of adopting various EEG features for recognising emotions. T. Song et al. [23] suggested a new model, namely DGCNN for the specific task of emotion recognition from EEG signals. It supplies a dynamic adjacency matrix which makes the model have the ability to learn the inherent relativity among EEG channels dynamically.

Kulkarni, D. et al., [24] applied Bi-LSTM and Improved-RNN on the DEAP dataset and also, they contributed in creating a new DOSE dataset similar to DEAP. Yuling Luo et al., [25] spatiotemporal correlations have been identified in the encoded data and applied to SNN (Spiking Neural Network) to classify emotion on both SEED and DEAP datasets. Xiangyu Juet et. al., [26] another model called Temporal-Difference Minimising Neural Network (TDMNN) is designed based on the premise that emotional activity exhibits temporal stability, considering the fact that the temporal variation of emotion is relatively slower than changes in physiological signs. This model used the Maximum Mean Discrepancy (MMD) statistic to assess and minimize differences in EEG features over time intervals which led to the enhancement of the model performance in recognizing emotions.

Jianhua Wang et al. [27] reported that CNNs have outperformed LSTMs for classification of EEG signals for BCI. This is attributed to the CNNs ability of capturing spatial features from the input data by striving to overcome the problems encountered when training large signals for classification. Despite their success in working with sequential data, it was found that using LSTMs in cases where location matters in classification may not yield a great accuracy. Specifically, this study shows that CNNs are more appropriate for this particular use than LSTMs are.

Due to the rapid fluctuations in the EEG signals, traditional machine learning algorithms have limited capability to capture minute changes of nonstationary nature in EEG signals. Hence, identifying spatiotemporal patterns in the EEG signals is challenging. Designing a novel deep-learning architecture optimized for identifying spatiotemporal features essential for emotion recognition relies on appropriate hyper parameter settings to capture important information from EEG signals.

The main contributions of our research as follows:

1) Proposed a novel framework STDI-CNN to overcome subject variability.
2) The automated feature selection mechanism for identification of optimal frequency bands employed in STDI-CNN model has been impactful on improving the predictive accuracy.
3) Framework helps in identifying complex spatio-temporal correlations which tends to highlight the performance of the proposed model.

In our proposed workflow, firstly we will be discussing the dataset and its preprocessing methodology. Later we experimented with proposed models CNN-LSTM, CNN-BiLSTM, STDI-CNN. STDI-CNN outperformed the other proposed models as well as SOA models, which will be analysed in the results and discussion section. Lastly, we have summarised the findings of this paper and outlined potential areas for improvement in the existing research.

2. PROPOSED FRAMEWORK

Deep learning models were used not just to overcome subject variability but also to get beyond the drawbacks of conventional machine learning methods, which can pick up on the minute variations in nonstationary EEG data and capture complex spatiotemporal patterns that are necessary for emotion recognition. The deep learning model has the potential to identify the important temporal and spatial correlations in the data, however, is largely dependent on the proper hyper-parameter choices, including learning rate, number of layers, filter sizes, and batch size. By carefully adjusting these settings, the deep learning architecture can extract the important information from EEG signals more effectively, improving performance in tasks linked to emotion perception.

2.1. Dataset: SEED (Shanghai Jiao Tong University Emotion EEG Dataset) [28]:

The SEED dataset was generated from EEG signals. EEG is a nonintrusive neuroimaging technique that records electrical activity in the brain and captures physiological activity. Table 1 describes the components of the experimental data collection for the SEED dataset.

Table 1. Parameters of the Stimuli and Description

Parameters of the stimuli	Values of the parameters	Description
Video clips	15 Chinese (4-minute video clips)	The film clips for the experiment were chosen based on: (i) The total duration of the experiment is kept moderate to prevent participant fatigue. (ii) Selected clips should be easily understood without requiring additional context. (iii) The chosen clips are curated to evoke a single, target emotion, ensuring consistency in emotional response.
Subjects	15	Seven males and eight females participated in the experiment. The mean and standard deviation of their age are 23.27 and 2.37 respectively.
Number of experiments on each subject	3	Each experiment consists of 15 trials. Each trail is associated with only one emotion and constitutes three experiments conducted on each subject, with a maximum gap of one week. Thus, the total number of experiments done is 675, which is equivalent to 15*3*15 (subjects * experiments * trials)
Signal Frequency	1000 Hz.	The signal was decimated to 200 Hz and used a 0 to 75 Hz bandpass frequency filter. The filtered frequency was separated into five frequency bands. δ [1-3Hz], θ [4-7Hz], α [8-13Hz], β [14-30Hz], and γ [31-50Hz].
Number of electrode channels	62	62 electrode channels were used at prescribed locations on the scalp to extract EEG signals [28].
Number of features	6	Asymmetry (ASM), Differential Asymmetry (DASM), Rational Asymmetry (RASM), Power Spectral Density (PSD), Differential Entropy (DE), and Differential CAUdality (DCAU). Not all electrodes are utilized by all the features.
Feature smoothing techniques	2	Linear Dynamic System (LDS), Moving average window(movingAve). Each feature was smoothed using both of the techniques. Now, the features extracted from the EEG signals turned 12.
No. of Emotions	3	Negative as -1, Neutral as 0, Positive as 1. These labels are encoded to 0, 1, and 2, respectively.

The SEED dataset adheres to a standardized method for electrode placement in EEG recordings widely used in neuroscientific research and clinical practice, which is the International 10-20 electrode system using the ESI (EEG Source Imaging) NeuroScan system at a sampling rate of 1000 Hz. SEED considers the video stimuli to record EEG signals.

The SEED dataset provided 45.mat files, each .mat file contains data related to one experiment. Data was extracted from the recorded EEG signals with one second of Hanning Window (HW) and 0% overlapping between two consecutive HWs [10]. As shown in Figure 1, the Differential Entropy (DE) feature was extracted from the pre-processed EEG signal because it is most prominent in the said features [13, 14] shown in Table 1. The proposed method uses the DE features where each trial comprises of various sequence lengths 235, 233, 206, 238, 185, 195, 237, 216, 265, 237, 235, 233, 235, 238, 206, associated with the encoded labels 2, 1, 0, 0, 1, 2, 0, 1, 2, 2, 1, 0, 1, 2, 0 respectively. DE feature signal has been decomposed into 5 frequency bands such as δ, θ, α, β, γ. Experiment-wise data has been given to the data-pre-processing of the proposed framework.

2.2. Data Pre-processing

Significant variations were observed in the recorded EEG signals from session to session across experiments. The main cause of these differences is the transient nature of EEG signals, which are susceptible to a number of variables like cognitive moods, and minor physiological changes that occur between sessions. In order to neutralize these variations, the signal distributions within each experiment should be standardized to minimize the effect of experiment-to-experiment fluctuations. The pre-processing of the data is shown in Figure 2.

Journal of Theoretical and Applied Information Technology
31st August 2025. Vol.103. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

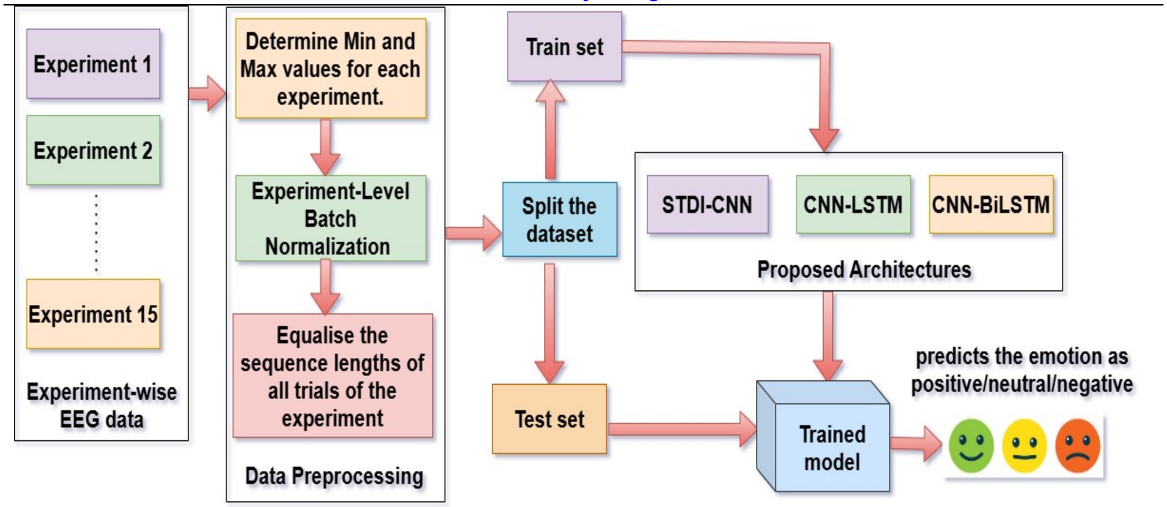

Figure 2: Proposed Framework

Experiment-wise data is pre-processed in 3 steps as depicted in Figure 2. After determining the minimum and maximum values of features, batch normalisation is applied for each experiment, and then the lengths of the sequences are equalised based on the maximum duration of the experiment's trial. Later in the proposed framework, the dataset has been split into train and test sets to generate a model with the proposed STDI-CNN architecture to predict emotions.

Let the dataset X[D] ∈ RE*T*C*F be the EEG data generated as sequences in T trails of E experiments described by F frequency bands and C channels where |E| = 3, |T| = 15, |F| = 5, and |C| = 62. X[D] is 50Hz down-sampled with a nonoverlapping sliding window of 1-second. The sampling points in each experiment is denoted by X[i][j] where j = 1,2,3…….,3394. All the sequences contained in X[D] are experiment-level batch-normalized. Later, data would be reshaped trial-wise and fed to the proposed STDI-CNN architecture. A trial T contains X[T]∈ RS*C*F where each trail X[T] = X[i] (i = 1,2,3…675), S data points S[ts]= (1,2,...,265) indicating the time-stamps in a sequence possibly extended by zero-padding. F denotes the frequency bands F[i] (i=1,2..,5), and C denotes channels C[i] (i=1,2,..,62) in each trial X[T]. As a result, the dataset is reshaped to Z with dimension (675,265,310) where F*C=310 to be fed as input for model building using different deep learning architectures.

We normalized the data Z[i] (i = 1,..,E*T) using experiment-level batch normalization to enable more trustworthy comparison of EEG signals across multiple trails. Batch normalization reduces the discrepancies owing to nonstationary nature while highlighting consistent subject-specific trends. As a result, this method enhanced the model's generalization and robustness when processing EEG signals from several sessions for the same participant. Experiment-level batch normalization is applied to account for variations in individuals' human behaviour and physiological responses.

The sequence lengths of trials of the experiment are different from one another. To address this issue, as in Figure 2, the maximum sequence length is identified and each sequence is extended by padding zeros up to the maximum sequence length to ensure consistency. The dataset is rearranged trial-wise; each trial out of the total 675 trials consists of data points belonging to five frequency bands. The pre-processed data is employed to train models on proposed architectures.

2.3. Proposed STDI-CNN (Spatio-Temporal Convolution Neural Network) model

The authors proposed a novel STDI-CNN model as illustrated in Figure 3 provides an automated feature selection mechanism for identification of optimal frequency bands and subsequent channels, which contributes to classify emotions efficiently. This capability leads to improvement in the model performance for emotion recognition. Input feature (IF) map is Z[i] with

Journal of Theoretical and Applied Information Technology
31st August 2025. Vol.103. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dimensions (trials, time-steps, features) = (675, 265, 310).

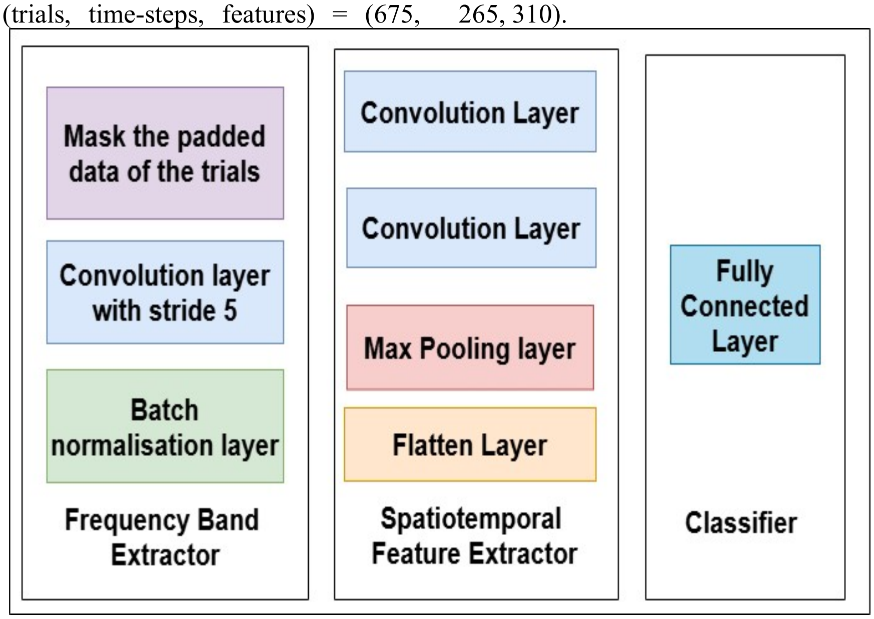

Figure 3: Proposed STDI-CNN model

The STDI-CNN model comprises three components such as a frequency band extractor, a spatiotemporal feature extractor, and a classifier is capable of recognizing spatial differences along with temporal.

2.3.1 Frequency band extractor

As illustrated in Figure 3, the Band Extractor is responsible for extracting efficient frequency bands from each electrode channel, which contain latent features of human emotions. It comprises 3 layers such as masking, convolution, and a batch-normalization layer respectively.

A masking layer was primarily incorporated to ignore the extended portion of sequences padded with zeros at the preprocessing phase. This layer effectively distinguishes between real and padded samples to allow downstream processes to disregard the padded values during convolution.

Let S[i] denote the i^{th} sample in the sequence, where i=1,2,...,265.

The masking function M(Si) can be defined as:

M(Si)= 1 if and only if Si is a real sample
0 if and only if Si is a padded sample

The purpose of M(Si) function is to yield 1 for the real samples and 0 for the remaining padded samples.

The mathematical notation for the masking layer is represented as:

Masking layer = [M(S1), M(S2),..., M(S255), M(S256),..., M(S265)] =[1,1,...,1,0,0,...,0]

The architecture of the frequency band extractor was illustrated in Figure 4 as 2D convolution layer processes the masked feature map M(S) with a kernel matrix of (5*5) which processes five frequency bands and selects one for each channel. The stride is 5 so that the kernel moves to the frequency bands of the next channel till it completes all the channels, resulting in a total of 62 features out of 310 raw features.

The batch normalization layer is applied to adjust and scale the activations of a neural network layer by normalizing them to center around a mean of 0 and standardize to a deviation of 1, to make it feasible for the network to learn and adapt during training.

2.3.2 Spatiotemporal feature extractor

As shown in Figure 3 Spatiotemporal Feature Extractor in succession to the frequency band extractor module has two consecutive 2D-convolution layers with a unit stride and 5*5 kernel to extract spatiotemporal features. A 2D-max-pooling layer with a kernel of 5 is used to regularize the model parameters to avoid overfitting the model. Followed by a flatten layer employed to change the 3D feature vector into a 1D feature vector to be fed to the classifier module. The architecture of the spatiotemporal feature extractor was illustrated in Figure 4.

2.3.3 Classifier

As shown in Figure 3, the classifier is the final module of the proposed STDI-CNN model, designed to prognosticate based on the processed data. The classifier module is made up of a dense layer with three output neurons, each representing an emotion class that is processed using a 1D feature vector obtained from the spatiotemporal feature extractor module. The soft-max activation function is employed to transform raw outputs of the dense layer into probabilities, making it ideal for multiclass classification. Categorical cross-entropy loss function computes the divergence among predicted probabilities and true labels, penalizing incorrect predictions. Soft-max and categorical cross-entropy collectively ensure effective learning and reliable multiclass classification performance. The architecture of the classifier was illustrated in Figure 4.

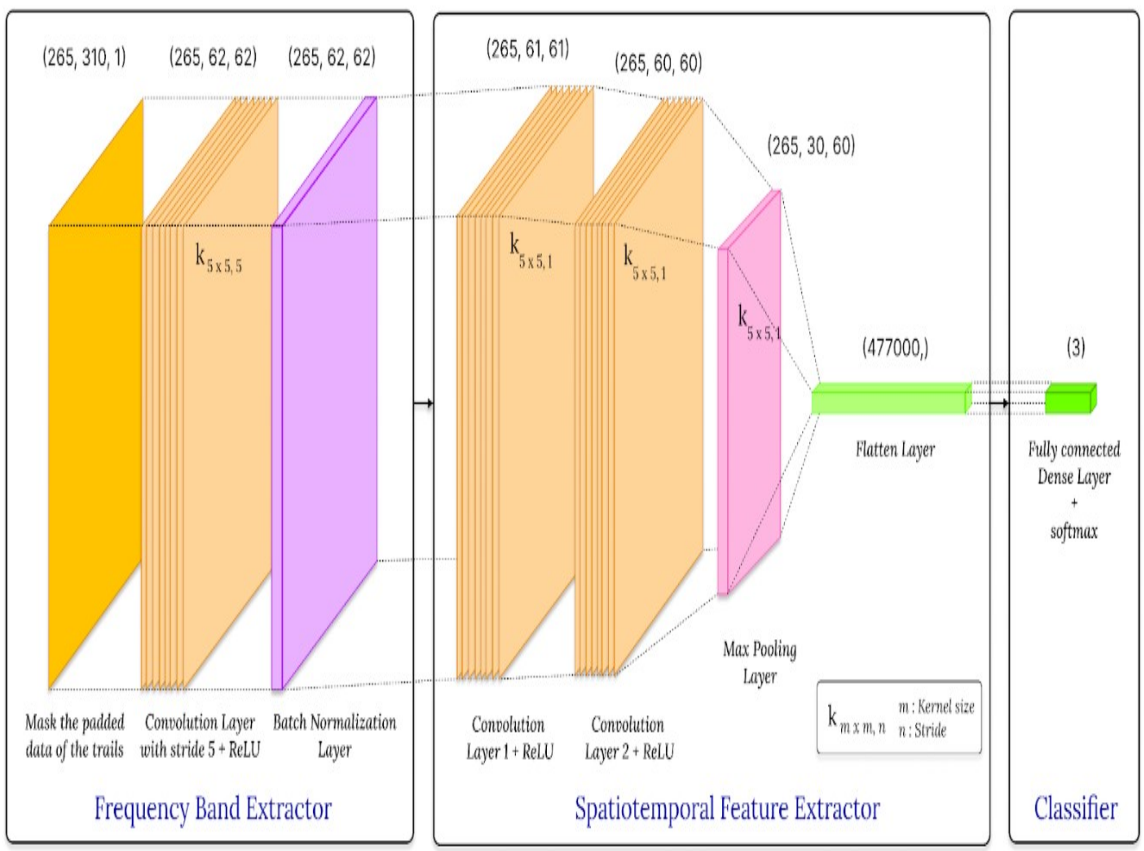

Figure 4: Proposed STDI-CNN Architecture

The detailed layered process of the proposed STDI-CNN model was depicted in the architecture. Relu activation function has been applied on all the convolution layers and soft-max function applied on the dense layer of the classifier module for multi-class classification. Categorical cross-entropy loss function assesses in reducing the distance between observed and predicted labels. To further enhance training efficiency and stability, the Adam (Adaptive Moment Estimation) optimizer is

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

employed to dynamically adjust learning rate using 1st and 2nd moment estimates of the gradient for model stability and convergence. The proposed model has been trained for a maximum of 20 epochs with initial learning rate as 0.01.

2.4. Proposed Fusion models

The fusion of CNN with LSTM and CNN with Bi-LSTM is also explored. The fusion of CNN with LSTM network and CNN with Bi-directional LSTM (Bi-LSTM) has shown superior performance compared to individual recurrent neural networks such as RNN, GRU, and LSTM alone. CNNs are best in identifying spatial patterns, making them well-suited for processing EEG signals. While LSTMs are effective at capturing temporal patterns and the bidirectional nature of LSTM enables it to analyze dependencies in both ways, leveraging past and future context for a better understanding of the data. The input feature map was processed through CNN architecture, extracted node embedding from the CNN architecture, and those embeddings were adjusted to map the input feature map of 4-layered LSTM and Bi-LSTM architecture to categorize the emotions.

3. RESULTS AND DISCUSSION

Experimentation done on the SEED to assess the performance of proposed architectures using a suite of standard metrics including precision, recall, F1-score, prediction accuracy, and loss. These metrics offer valuable insights into various facets of a model's performance. While our problem is a multi-class classification issue, precision, recall, and the F1-score are calculated for each class in the model independently. These metrics are then aggregated to derive overall precision, recall, and F1-score.

The authors compared the performance of the newly proposed STDI-CNN model with SOA models involving CNN, LSTM, and other deep learning architectures well established in the literature on emotion recognition from EEG signals. This comparative analysis also involves the simpler fusion architectures proposed in this paper named CNN-LSTM and CNN-BiLSTM.

Recurrent Neural Networks (RNNs) are identified as the potential solution for working with EEG data in the context of emotion recognition because of their capability of processing temporal dynamics in EEG signals as evidenced by the performance. The LSTM and Bi-LSTM networks in RNN family provide methods to tackle the vanishing gradient problem to capture long term dependencies which make these networks even more suitable for EEG based emotion recognition. Use of CNN's architecture has proven results for real-time EEG applications, particularly when the relevant information is localized within specific temporal or spatial regions of the brain signals.

Hence the fusion of CNN-LSTM and CNN-BiLSTM is explored in this research work. The fusion architectures developed in this paper shared a common CNN architecture and they yielded approximately equal accuracy which is clearly higher than most of the SOA architectures for emotion recognition.

The TDMNN presented by Xiangyu Ju et al., is the only one existing architecture found to be a close competitor for the fusion models on prediction accuracy. However, the decrease in loss from 10.68% with LSTM to 8.91% with Bi-LSTM signifies a significant enhancement in the model's predictive capability, instilling greater confidence in its outcomes. From this loss reduction, it could be realized that the Bi-LSTM architecture is well suited to capture temporal features of the EEG signals thereby advancing predictions.

It can be observed from the results provided in table 2 that the proposed STDI-CNN model outperformed among all SOA architectures and fusion models, CNN-LSTM and CNN-Bi LSTM proposed in this paper. Specifically, the STDI-CNN model yielded 98.52% accuracy and 5% loss as mentioned in table 3 which reflects a better performance compared to the fusion models which are its closest competitors.

The automated feature selection mechanism for identification of optimal frequency bands employed in STDI-CNN model has been impactful on improving the predictive accuracy by approximately 1.5% over that could be achieved by the best model among the SOA architectures. At the same time, the model loss is also kept at 5% which is comparatively low. Similarly, the detailed performance comparison in terms of precision, recall and F1-score in addition to accuracy and loss for the proposed models is presented in table 3.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From the results tabulated in table 2 and table 3 is evident that the STDI-CNN model has been more effective in capturing the intricate temporal dynamics of the EEG signal as it could extract meaningful features for accurate emotion recognition.

Table 2. Comparison of proposed models with State-of-the-art models

Existing/Proposed models	Reference paper	Accuracy
DBN - 2015	[10]	86.08%
CNN - 2022	[15]	62.04%
LSTM-2022	[15]	64.84%
MLP-2022	[15]	78.16%
HCNN-2017	[19]	88.2%
CNN+LSTM -2019	[21]	89.88%
SRU-2020	[13]	90%
Spiking Neural Network-2020	[25]	96.67%
CNN-2022	[17]	93.81%
TDMNN-2023	[26]	97.2%
CNN-LSTM	proposed	97.04%
CNN-BiLSTM	proposed	97.04%
STDI-CNN	proposed	98.52%

Table 3: Performance metrics of Proposed models.

Model No.	Proposed models	Accuracy	Loss	Precision	Recall	F1-score
1	CNN-LSTM	97.04%	10.68%	97%	97%	97%
2	CNN-BiLSTM	97.04%	8.91%	97%	97%	97%
3	STDI-CNN	98.52%	5.43%	98.4%	98.6%	97.54%

4. CONCLUSION AND FUTURE SCOPE

The proposed STDI-CNN model is able to predict emotions precisely and has proven the best accuracy of 98.52%, increased approximately by 1.5% above the SOA models. The model handles the nonstationary nature of EEG data efficiently while considering subject variability and also minimizes extraneous spatial features while preserving the inherent spatial characteristics to capture differences in temporal dynamics. The proposed model has the potential to automatically extract spatiotemporal features from the complex EEG signals without human intervention, achieved through appropriate hyper-parameter tuning at each layer in the seven-layered architecture. The Experimental findings demonstrate that the model outperforms SOA methods in recognizing emotions on the benchmark dataset SEED. The proposed model can acknowledge inexpressible emotions of patients suffering from neurological disorders which facilitates better decision making for better diagnosis. The emotion can be better expressed through our model for enhanced BCI.

Despite being trained on less data of 15 subjects, the proposed model has proven better results. Although larger data corresponds to better generalizability. But acquiring larger EEG dataset is challenging; to handle this challenge we can explore transfer learning in the future research as it supports model-building by refining pre-trained models as per the characteristics of target domain with limited data. Additionally, federated learning can increase the scope for decentralized model building in a constrained environment to foster deeper understanding of human emotions. Furthermore, future work can extend towards attention-based transformer models which better captures the temporal dynamics of EEG data. Altogether, future avenues can resolve these limitations to foster a better understanding of human emotions.

REFERENCES:

[1] Paweł Tarnowski, Marcin Kołodziej (2017) et al., "Emotion recognition using facial expressions", in proc. ICCS. https://www.sciencedirect.com/science/article/pii/S1877050917305264

[2] Zhijuan Shen, Jun Cheng, Xiping Hu, Qian Dong (2019) "EMOTION RECOGNITION BASED ON MULTI-VIEW BODY GESTURES" IEEE International Conference on Image Processing, DOI: 10.1109/ICIP.2019.8803460

[3] Md. Nurul Ahad Tawhid(2023) et al., "Automatic and Efficient Framework for Identifying Multiple Neurological Disorders From EEG Signals" IEEE Transactions on Technology and Society. DOI:10.1109/TTS.2023.3239526

[4] Smith K. Kare (2023) et al., "Adazd-Net: Automated adaptive and explainable Alzheimer's disease detection system using EEG signals", Vol-278. https://doi.org/10.1016/j.knosys.2023.110858

[5] Tianhao Gu (2023) et al., "Frame-Level Teacher–Student Learning With Data Privacy for EEG Emotion Recognition" IEEE Transactions on Neural Networks and Learning Systems Volume-34, Issue: 12. DOI:10.1109/TNNLS.2022.3168935

[6] Mahendra Yadava (2017) et al., "Analysis of EEG signals and its application to neuromarketing" Multimed Tools Appl 76:19087–19111. https://doi.org/10.1007/s11042-017-4580-6

[7] Xin Hu (2019) et al., "Ten Challenges for EEG based Affective Computing", Brain Science Advances, volume-5, issue-1, pages:1-20. https://doi.org/10.1177/2096595819896200.

[8] Ngoc-Dau Mai (2021) et al., "Affective Computing on Machine Learning-Based Emotion Recognition Using a Self-Made EEG Device", Sensors, 21, 5135. https://doi.org/10.3390/s21155135

[9] Xiang Li (2022), "EEG Based Emotion Recognition: A Tutorial and Review", ACM Computing Surveys, Volume 55, Issue 4, Article No.: 79, Pages 1 - 57, https://doi.org/10.1145/3524499

[10] Wei-Long Zheng (2015) et al., "Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. "IEEE Trans. on Autonomous Mental Development, Volume: 7, Issue: 3. 10.1109/TAMD.2015.2431497

[11] Nouman, A., Iqbal, H., & Saleem, M. K. (2023). An emotion recognition system: bridging the gap between humans-machines interaction. IAES International Journal of Robotics and Automation (IJRA), 12(4), 315. https://doi.org/10.11591/ijra.v12i4.pp315-324

[12] Mu Li (2009) et al., "Emotion classification based on gamma-band EEG" Annual International Conference of the IEEE Engineering in Medicine and Biology Society. DOI: 10.1109/IEMBS.2009.5334139

[13] Chen Wei (2020) et al., "EEG-based emotion recognition using Simple Recurrent Units network and ensemble learning", Biomedical Signal Processing and Control, ScienceDirect, Elsevier. https://doi.org/10.1016/j.bspc.2019.101756

[14] Ruo-Nan Duan (2013) et al., "Differential Entropy Feature for EEG-based Emotion Classification", 6th Annual Int. IEEE EMBS Conf. on Neural Engineering, San Diego, California. DOI: 10.1109/NER.2013.6695876

[15] G. Li (2022) et al., "An EEG Data Processing Approach for Emotion Recognition", IEEE Sensors Journal, Vol no. 22, No. 11. doi: https://doi.org/10.1109/JSEN.2022.3168572.

[16] Chunawale. A (2024) et al., "Electroencephalogram based human emotion classification for valence and arousal using machine learning approach", Indonesian Journal of Electrical Engineering and Computer Science, 33(2), 920. https://doi.org/10.11591/ijeecs.v33.i2.pp920-931

[17] Mohith Kumar (2022) et al., "Human emotion recognition from EEG signals: model evaluation in DEAP and SEED datasets", CEUR-WS proc., vol-3368.

[18] Mitul Kumar Ahirwal (2018) et al., "Emotion Recognition System based on EEG Signal: A Comparative Study of Different Features and Classifiers", Proc. of IEEE 2nd Int. Conf. on Computing Methodologies and Communication (ICCMC). DOI: 10.1109/ICCMC.2018.8488044

[19] Li (2018) et al., "Hierarchical Convolutional Neural Networks for EEG-Based Emotion Recognition.", Springer Journal of Cognitive Computation, Vol no. 10, pages: 368-380. https://doi.org/10.1007/s12559-017-9533-x

[20] R. K. Jeevan (2019) et al., "EEG-based emotion recognition using LSTM-RNN machine learning algorithm", IEEE 1st Int. Conf. on Innovations in Information and Communication Technology (ICIICT). doi: 10.1109/ICIICT1.2019.8741506

[21] S Hwang (2019) et al., "Learning CNN features from DE features for EEG-based emotion recognition", Pattern Analysis and Applications, 23:1323–1335. https://doi.org/10.1007/s10044-019-00860-w

[22] Shen (2020) et al., "EEG-based emotion recognition using 4D convolutional recurrent neural network", Cognitive Neurodynamics, vol. no. 14, pages:815-828. https://doi.org/10.1007/s11571-020-09634-1

[23] T. song (2018) et al., "EEG Emotion Recognition Using Dynamical Graph Convolutional Neural Networks", IEEE Transactions on Affective Computing.vol. 11, no. 3, pp. 532-541, doi: 10.1109/TAFFC.2018.2817622

[24] Kulkarni, D. D., Dixit, V. V., & Deshmukh, S. S. (2024). Emotion detection using EEG:

hybrid classification approach. Indonesian Journal of Electrical Engineering and Computer Science, 35(1), 459. https://doi.org/10.11591/ijeecs.v35.i1.pp459-466

[25] Yuling Luo (2020) et al., "EEG-Based Emotion Classification Using Spiking Neural Networks", IEEE Access, 2169-3536. DOI:10.1109/ACCESS.2020.2978163

[26] Xiangyu Ju (2023) et al., "EEG-based emotion recognition using a temporal-difference minimizing neural network", Cognitive Neurodynamics. https://doi.org/10.1007/s11571-023-10004-w

[27] Jianhua Wang (2019) et al., "Classification of EEG signal using convolutional neural networks", 14th IEEE Conference on Industrial Electronics and Applications (ICIEA). DOI: 10.1109/ICIEA.2019.8834381

[28] SEED (The SJTU Emotion EEG Dataset): A dataset collection for various purposes using EEG signals. https://bcmi.sjtu.edu.cn/home/seed/seed.html