



## Enhancement of Hybrid Deep Neural Network Using Activation Function for EEG Based Emotion Recognition



Jehosheba Margaret Matthew<sup>\*</sup>, Masoodhu Banu Noordheen Mohammad Mustafa<sup></sup>

Department of Biomedical Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai 600062, India

Corresponding Author Email: [jehosh17@gmail.com](mailto:jehosh17@gmail.com)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410428>

### ABSTRACT

**Received:** 11 December 2023

**Revised:** 15 April 2024

**Accepted:** 15 June 2024

**Available online:** 31 August 2024

#### Keywords:

*activation function, ELU, BCI, emotion recognition, ReLU, Leaky ReLU, Deep Neural Network*

Deep Neural Network (DNN) is an advancing technology that improves our life by allowing machines to perform complex tasks. Hybrid Deep Neural Network (HDNN) is widely used for emotion recognition using EEG signals due to its increase in performance than DNN. Among several factors that improve the performance of the network, activation is an essential parameter that improves the model accuracy by introducing non-linearity into DNN. The activation function enables non-linear learning and solves the complexity between the input and output data. The selection of the activation function depends on the type of data that is used for computation. This paper investigates the model performance with respect to various activation functions like ReLU, ELU, and Leaky ReLU on a hybrid CNN with a Bi-LSTM and CNN model for emotion recognition. The model was tested on the DEAP dataset which is an emotion dataset that uses physiological and EEG signals. The experimental results have shown that the model has improved accuracy when the ELU function is used.

## 1. INTRODUCTION

Affective computing is a branch of Artificial Intelligence (AI) that creates systems and devices to respond to human emotions naturally. The communication between the machine and the human is more effective and seamless. It is a computational study of an emotion that communicates to healthcare people like doctors, healthcare educators, medical administrators, and patients [1]. It has many applications such as developing assisting technology for people with disability, enhancing mental diagnosis and treatment, and improving human computer interaction. The biggest challenge in affective computing is the recognition of emotion since it varies across subjects and cultures and is complex. Usually, emotions were expressed in different ways like facial expressions, speech signals, and physiological signals.

Emotions represent the true state of a person's thoughts and behavioural responses and they are closely associated with the nervous system [2]. Emotion recognition through physiological signals is more reliable since the interpretation of emotions starts from the central nervous system [3]. The brain computer interface is a technique that acquires the Electroencephalogram (EEG) signal from the brain and is used for various purposes like assisting the individual with disabilities, limited motor functions, neurological injuries, mental health applications [4], and in neuroscience for emotion recognition [3]. BCI enables the machine and the environment to exchange information with the human. These signals are obtained by EEG from the brain cortex through various electrodes placed on the head. It gives more precise

and consistent results [5]. It is more commonly preferred by researchers since it is non-invasive, portable, and can be used in different environments [3]. It also plays an important role in illustrating the activity of brain regions in different emotional states [6].

Emotion recognition using EEG signals is carried out either using the machine learning method like Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) or the deep learning method. In a machine learning based model, the features are extracted manually and given to the classifiers like SVM, LDA, Random Forest, Naive Bayes, and other classifiers to classify different types of emotions. Since EEG signal has non-linear and high dimensions problems, it is difficult to solve the problem using the linear algorithm. In order to overcome this issue, deep learning was introduced to solve non-linear issues [7]. Deep learning is preferred to develop an accurate and automated system that can learn large amounts of data and recognize different emotional cues. It learns complex features from the raw data automatically which is essential to quantify different emotions. The performance of the deep learning model can be improved through various techniques like changing the network architecture, tuning parameters, preventing excessive dropout of neurons, augmentation, and preprocessing of data. The network architecture changed in many ways by changing the number of layers, learning rate etc. including the activation function of the layers.

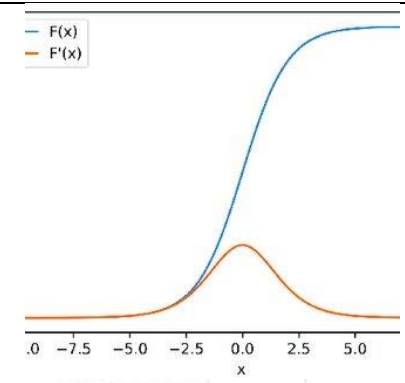
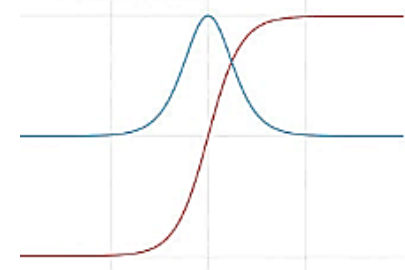
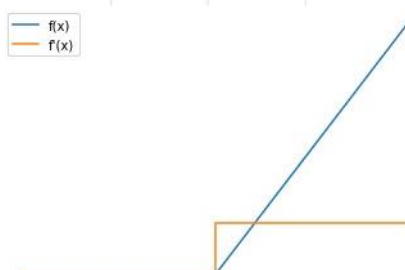
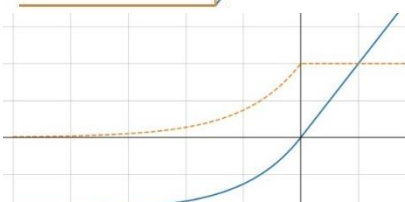
An activation function is a mathematical function that determines whether the neuron should be activated or not based on the weighted sum of its input. It introduces non-

linearity at the output of the neuron, which allows the model to solve complex and non-linear data. There are activation functions that operate on both positive and negative data like Tanh, sigmoid and ELU while some of them operate only on positive data like ReLU as shown in Table 1. Among them, the Tanh function is zero centred and therefore reduces the number of epochs to train the model and helps the back propagation process [8]. However, it cannot be used for multiple layers networks due to the vanishing gradient problem. As it leads to saturation with differentiation, the sigmoid function suffers from vanishing gradient problems and as it is not differentiable. ReLU and ELU do not suffer from saturation effect due to its linearity in the positive region. ReLU function also introduces sparsity in the model by removing the negative values which reduces the computational complexity of the model and improves the model's generalization performance. It performs well for multiple layers in the network but suffers from dead neurons due to its characteristics in the negative region. At such situations, ELU introduces exponential non linearity for smaller portion in the negative region. Due to this, it gives better performance than ReLU and Leaky ReLU [9]. It performs well for large datasets and gives fast convergence [10].

Therefore, this proposed work does the empirical study on different activation function applied to HDNN and CNN models and subsequently identifies the best fitting activation function.

Keelawat et al. [11] have found that the change in the activation function of the model has improved the accuracy of the model. Supporting this, Apicella et al. [12] found that, the choice of activation function has a great impact on the learning process and the shape of the error function. Therefore, in this proposed work different activation functions were experimented on the HDNN model and the best fitting activation function was determined. In this paper, ReLU, Leaky ReLU and ELU activation functions were used to evaluate the changes in model accuracy. The effectiveness and efficiency of selected activation functions are tested for two different architectures with DEAP EEG dataset for emotion classification in the proposed work. The effectiveness of an activation function here is defined to be the accuracy and value of the loss function computed on the test set. Efficiency on the other hand is defined as the prediction time, epochs required to converge and time required for a single epoch. The following research question and hypotheses have been devised for the proposed work:

**Table 1.** Activation functions and their corresponding equation and waveforms

Activation	Function and Gradient	Equation
Sigmoid		$f(x) = \frac{1}{1+e^{-x}}$
Tanh		$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
ReLU		$f(x) = \max(0, x)$
ELU		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$

Research Question: Does the use of activation functions results in noteworthy changes in efficacy of the model chosen with DEAP EEG dataset?

H1. Can the performance of the model attributed to mean gradients in each layer?

H2. Do weights play the major role in model performance?

H3. Can the performance of the model attributed to results from statistical analysis?

The paper is organized in the following way. Section 2 explains the survey reports and other research works that discusses the activation function. Section 3 elaborates on the model architecture and the process involved in the experiment. Section 4 describes the experimentation and the interpretation of results that were evaluated for various activation functions. Finally, Section 5 concludes of the research work.

## 2. LITERATURE REVIEW

Most of the papers that focused on emotion recognition using EEG signals have experimented with the DEAP dataset. The various research works that are available in the literature with DEAP dataset have been studied to understand the use of activation functions in the DNN model. The most prominent activation function that was used in the models was Tanh, sigmoid, and ReLU [13]. Tanh activation function is preferred for dataset like DEAP where there are negative values. It is good in introducing non-linearity to the model and helps to stabilize the gradients during training Very few articles have used Tanh, linear, and sigmoid functions in the middle layers.

In recent days, ReLU is the most commonly preferred activation function in the model layers for different applications [14], as it introduces non-linearity, prevents vanishing gradients, and computationally efficient to avoid overfitting. Generally, most of the CNN models independent of application, use ReLU and Leaky ReLU and the two works [15, 16] are among those. Mastromichalakis [15] have used Leaky ReLU function for COVID pneumonia dataset and concluded that it was the best alternative to ReLU for avoiding dying neuron problem. Jebadurai et al. [16] have used Leaky ReLU for generalization of their model.

Among the works considered specifically for EEG based application and most importantly emotion recognition application, Pan and Zheng [17] have used generative adversarial network for data augmentation and convolutional network as their model for EEG based emotion classification. Here in every convolutional layer, they have used ReLU after batch normalization. While they have reported increase in accuracy, it is not discussed that if it was due to ReLU and data augmentation method or only due to data augmentation. Similar to this there are multiple works in literature which have used ReLU or Leaky ReLU but have not attempted to analyse the impact of the activation function. The below Table 2 provides the details of such works. The work shows, though the actual reason for increase in accuracy was not discussed, there is a room for increase in accuracy due to change in activation functions like ReLU, Leaky ReLU and this is one of the reasons for the proposed work which analyses and identifies suitable activation function for the given application.

**Table 2.** Literature works available that used ReLU or Leaky ReLU as the activation function

Authors	Model/Application	Reported Accuracy in % and Inference	Activation Function Used
Chao et al. [18]	DNN model with capsules	Average accuracy of 67 for 2D-CNN and 65 for 1D-CNN	ReLU activation function in the CNN layer
Pandey et al. [19]	Variational mode decomposition feature extraction on a DNN to detect subject dependent emotions	61 for arousal and 62 for valence and the difference in accuracy is due to class imbalance	ReLU
Li et al. [20]	Hybrid DNN with the combination of CNN and LSTM networks	Average accuracy of 75.21	ReLU
Garg and Verma [21]	GoogleNet model for emotion recognition	83 for valence and 55 for arousal	RELU
Ashgar et al. [22]	Alexnet model with SVM	Average accuracy of 77 was obtained for DEAP dataset	ReLU
Xing et al. [23]	Stacked Auto-Encoder for feature extraction with LSTM model	81- valence and 74 - arousal	Linear and sigmoid activation
Ozdemir et al. [24]	CNN	Accuracy of 90 for 2class and 88 for 4 classes	ReLU
Acharya et al. [25]	LSTM and CNN models	87 for CNN and 88 for LSTM	ReLU
Yang et al. [26]	CNN model	90 for arousal and 89 for valence	ReLU

ELU is another activation function that is recently preferred by researchers for negative inputs. The use of ELU in various applications like emotion recognition through text [27], facial emotion recognition [28] and speech emotion recognition [29] has shown significant improvement in the model. Devi and Deepa [27] have compared the results of ELU and ReLU activation function for emotion recognition from twitter data. On comparing, they have found that ELU has positive merits over negative values of data in not allowing the mean activation to be nearer to zero. This has enabled to decrease

the gap between the gradients. It has also provided better accuracy than ReLU. Bejjagam and Chakradhara [28] have used ELU as the activation function for Facial Emotion Recognition since the inputs are negative and to overcome the dying problem of ReLU. Le et al. [29] have used the ELU activation function for speech based emotion recognition which has negative data as the inputs. The maximum accuracy of the model is estimated to be 90.54%. Abelwahab and Busso [30] have studied the activation function of the dense network. It was found that there was no significant change in the

performance on using ELU and ReLU for speech emotion recognition. Therefore, they have suggested increasing the number of layers and training data set which led to increase in performance by ELU. Few research works have used ELU as the activation function for EEG signals. Schirrmester et al. [31] have identified the degradation in the accuracy over frequency ranges on using ReLU over ELU for Motor Imagery based EEG signals and it happened due to the presence of negative values in the data. By experimenting with different batch size, epochs and learning rate, Nguyen et al. [32] concluded that ELU and its combinations like SELU and GELU performs better by avoiding vanishing gradient and dead state problems. Liu et al. [33] have proposed TSin as the activation and proved its efficiency based on training stability, convergence speed and precision.

León et al. [34] have observed that the model got stuck in learning due to dying neurons with ReLU and therefore they have used ELU as the activation function. Farahat et al. [35] have insisted on choosing Tanh or ELU as the activation over ReLU since its inability to compute negative inputs leading to poor results. Bai [36] have observed better performance with ELU since it pushes the average activation value closer to zero just like Batch Normalization and thus aiding in fast learning with reduced bias. Jang et al. [37] have used ELU in the graphical model for fast extraction of the features. Liang et al. [38] have used ELU as the activation function in the middle layers to improve the model fitting. With these observations, different activation functions were experimented in the proposed work in two different models i.e., HDNN and CNN.

### 3. METHODOLOGY

The role of the activation function is well known with respect to simple datasets like random dataset having normal distribution etc. However, as pointed out in literature review section, though the model uses different activation function and reports accuracy, the reason behind in that particular scenario is not well explained. Hence this paper considers the following performance criteria for deciding the effectiveness of the activation function, while considering the regular evaluation accuracy and loss metrics and computational efficiency for comparison.

1. The change of gradients with respect to each layer both in terms of mean and standard deviation;
2. The change of weights with respect to each layer both in terms of mean and standard deviation.

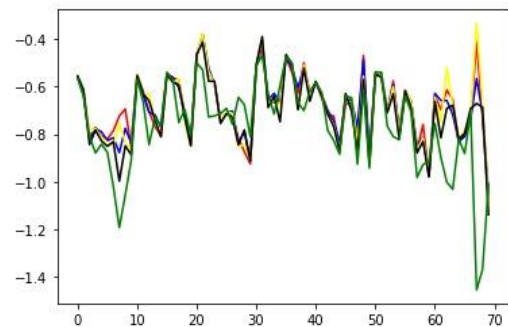
The gradients were obtained by running the model with user defined fit function instead of using the default fit function available with Keras. The weights were obtained by using

default fit function with callbacks for weight capture. Two different models have been tested with three activation functions i.e., ReLU, Leaky ReLU and ELU. The experimental setup used for running the experiment is given in Table 3. Also, in order to test the above hypothesis the work considers intra subject variations i.e., if subjects 1-5 are considered and augmented, it will be split into 70-20-10 ratio, 70% of the data will be used for training, 20 % will be used for validation and remaining 10% will be used for evaluation.

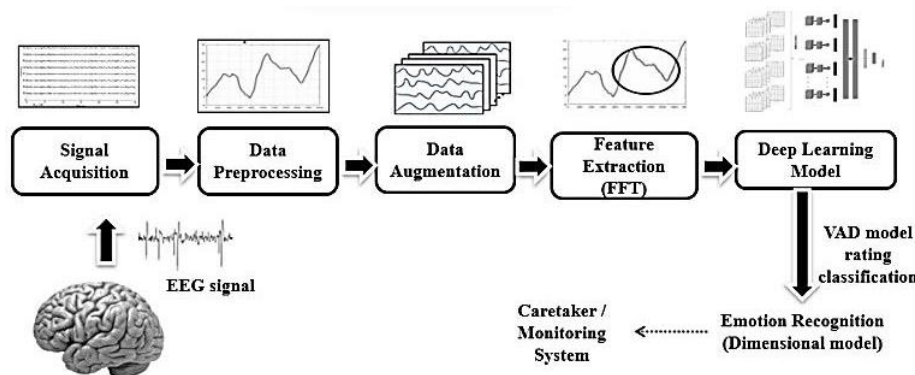
**Table 3.** Experimental setup

Name /Description	Version
CPU	Intel® Core™ i5
RAM	8 GB
OS	Windows 10
Python	Python 3.11.5
TensorFlow	TensorFlow 2.14.0
Scikit-learn	Scikit-learn 1.3.1
Anaconda	2021.05

The distribution of data within few features (from a total of 70 features) after normalization in the input layer shown in Figure 1 and it can be seen that, it has both positive and negative values. However, after the first layer of CNN, the distribution is unknown, i.e., each layer’s distribution changes during training and hence in order to make the distribution as normal, batch normalization can be used. Batch normalization forces the neurons to work in the linear region. Thus, Batch Normalization is very beneficial to neural networks but it comes with some additional Batch Normalization parameters to learn. Hybrid CNN with LSTM, as it is computationally expensive always; CNN model alone has been tested with and without Batch Normalization for identifying the significance of ELU. Batch Normalization weakens the dependency between the layers by recentering the inputs to each layer.



**Figure 1.** Distribution of data after normalization



**Figure 2.** Block diagram of EEG based emotion recognition

Each participant’s single emotion has been recorded for 63 seconds in DEAP dataset. However, considering the hardware implementation and augmentation, each 2 seconds data is taken as one sample and this is discussed in Augmentation section. As the model is tested only with one modality, intrasubject validation is considered in this work, i.e., out of n subjects’ data available, n-1 or n-2 subjects data were considered for training and evaluation. The left-out subject’s data will be considered for prediction.

The emotion recognition process undergoes a few steps and the architecture block diagram to predict such emotion is given in Figure 2. The input EEG signals are used to test the model. EEG signals can either be self-recorded signals or a dataset that consists of a large collection of EEG signals. Self-recorded signals refer to the collection of signals or data that have been recorded or captured by an individual user. Then the signals are pre-processed to remove the unwanted noise and make them ready for the next step. The most relevant information from the processed EEG signals is given to the DNN model for classification.

### 3.1 Dataset description

DEAP dataset is an emotion dataset that has the collection of EEG from different participants. In this dataset, the EEG signals were recorded from 32 participants at the sampling frequency of 512 Hz. The other peripheral signals like skin temperature, blood pressure, Electromyogram, and galvanic skin response were recorded from the channels separately. Each subject responded to 40 videos which were considered as the number of trials. After the video, each subject assessed the emotional value based on the emotion levels of valence, arousal, dominance, and liking on a scale of 1-9.

The original signal was downsampled to 128Hz to reduce the computational complexity and high frequency noise in the signal. The signal was also applied to a band pass filter of 4-45Hz to filter the signal and remove the artifacts from the signal. The Electrooculogram noise in the DEAP dataset was removed by Independent Component Analysis so that the data could represent the emotions of the subject clearly [23]. Electromyogram artifacts were removed by subtracting the temporal low frequency drift [39].

### 3.2 Augmentation and feature extraction of EEG

The pre-processed signal was augmented to increase the number of data samples. The data sequence was augmented with a window size of 2sec and an overlap of 0.125sec corresponding to 16 samples and the detailed description of this process is given in the previous study [40]. Therefore, the 63sec data sequence has generated 488 samples of data which is described in Figure 3. The augmented data was given as the input to the feature extraction module. The relevant five band features were extracted from the EEG dataset using Fast Fourier Transform (FFT) that convert the signal from the time domain to the frequency domain and they are theta, alpha, sigma, beta, and gamma [41] which are then split into training and testing and evaluation samples in the ratio of 70-20-10. After the split of the data samples, they were given to the deep learning models.

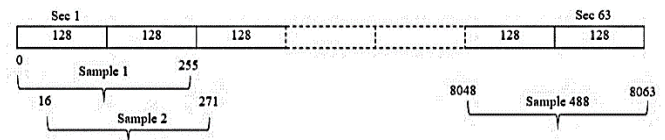


Figure 3. Augmentation of EEG data

The model parameters of hybrid Bi-LSTM and CNN are given in Table 4 comprises two neural networks namely CNN and Bi-LSTM. The extracted data is standardized and normalized to be more suitable for classification. In order to improve the input of the Bi-LSTM network, CNN is added in the first two layers of the model. CNN will process the signal by smoothing and reducing the sequence length [42] so that the processed data is fed into the Bi-LSTM network instead of the original data. Here the Bi-LSTM model sends both the positive and inverted sequence of EEG data to be trained [43]. The model is used to classify 4 classes based on valence, and arousal. In both the models, CNN is formed at the beginning of the model to extract spatial and temporal features of time series data. It can detect essential information from different positions with excellent accuracy [44]. It also converts the univariate data into a multi-dimensional dataset which refers to the extraction of multiple features from the EEG signal.

Table 4. Model Parameters of hybrid Bi-LSTM and CNN

Parameters of Hybrid Bi-LSTM Model						
Layer	Filter/No. of Neurons	Kernel	Activation	Max Pooling	Dropout	Batch Normalization
Conv1D	128	3	Yes	2	0.2	-
Conv1D	128	3	Yes	2	0.2	-
Bi-LSTM	256	-	-	-	0.2	-
Bi-LSTM	32	-	-	-	0.2	-
Flatten	-	-	-	-	-	-
Dense	128	-	Yes	-	0.2	-
Dense	4	-	Softmax	-	-	-
Parameters of CNN Model						
Layer	Filter/No. of Neurons	Kernel	Activation	Max Pooling	Dropout	Batch Normalization
Conv1D	128	3	Yes	2	-	Yes
Conv1D	128	3	Yes	2	-	Yes
Conv1D	64	3	Yes	2	-	No
Flatten	-	-	-	-	-	-
Dense	64	-	Yes	-	0.2	-
Dense	32	-	Yes	-	0.2	-
Dense	16	-	Yes	-	0.2	-
Dense	4	-	Softmax	-	-	-

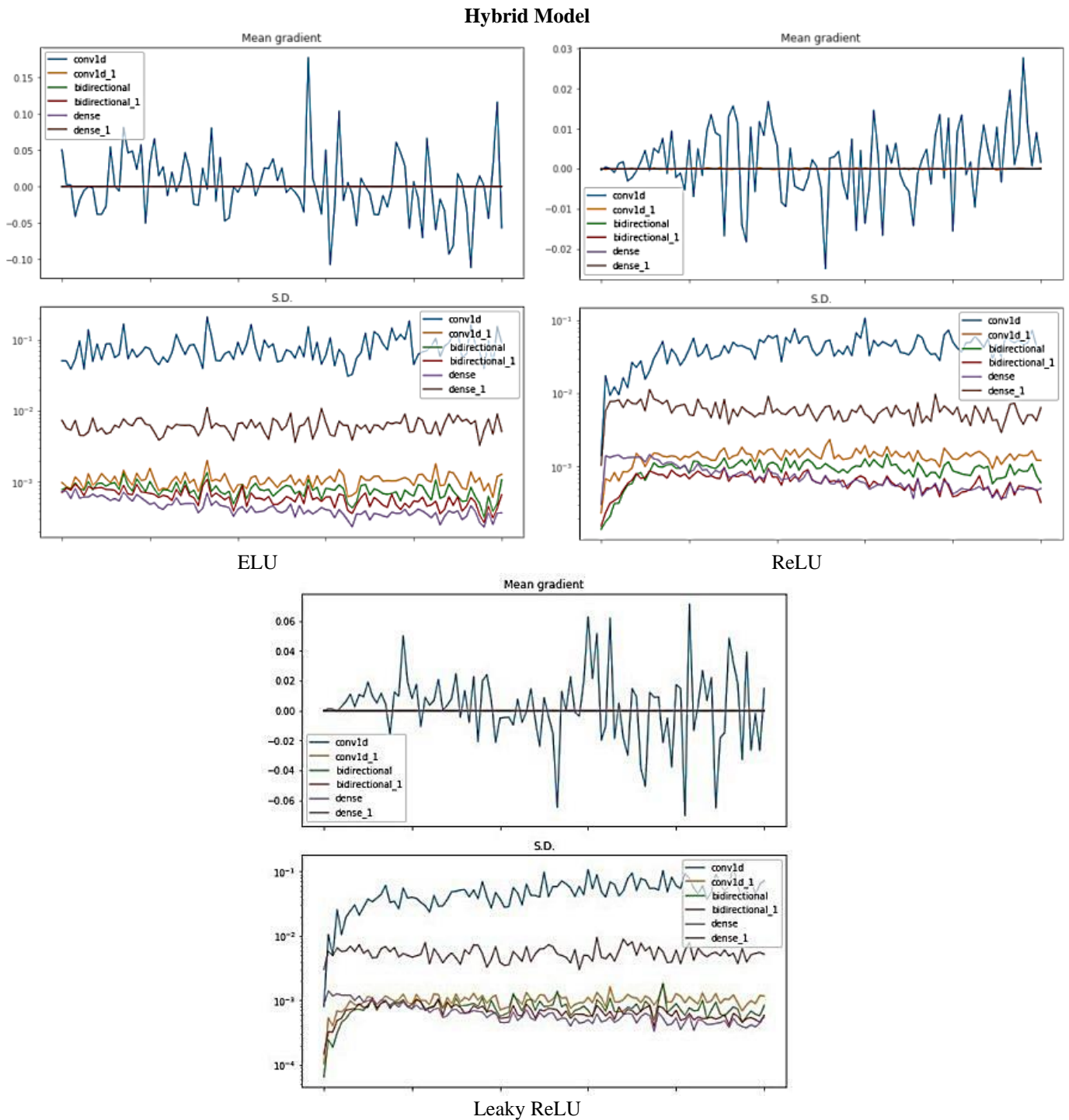
Pooling layer shrinks feature maps of large sizes into smaller feature sizes. In order to avoid the overfitting problem and also to decrease the complexity of the model, the dropout layer is added. The dropout technique is used for generalization where the neurons are dropped randomly while training at each epoch. Therefore, the feature selection is equal for all neurons and that makes the model learn different independent features [45]. At the output layer, softmax activation is used to generate the output based on the probability distribution [45]. The probability in the output is  $p \in \{0, 1\}$ . It is also called a log loss function. It is mainly used in multiclass classification problems.

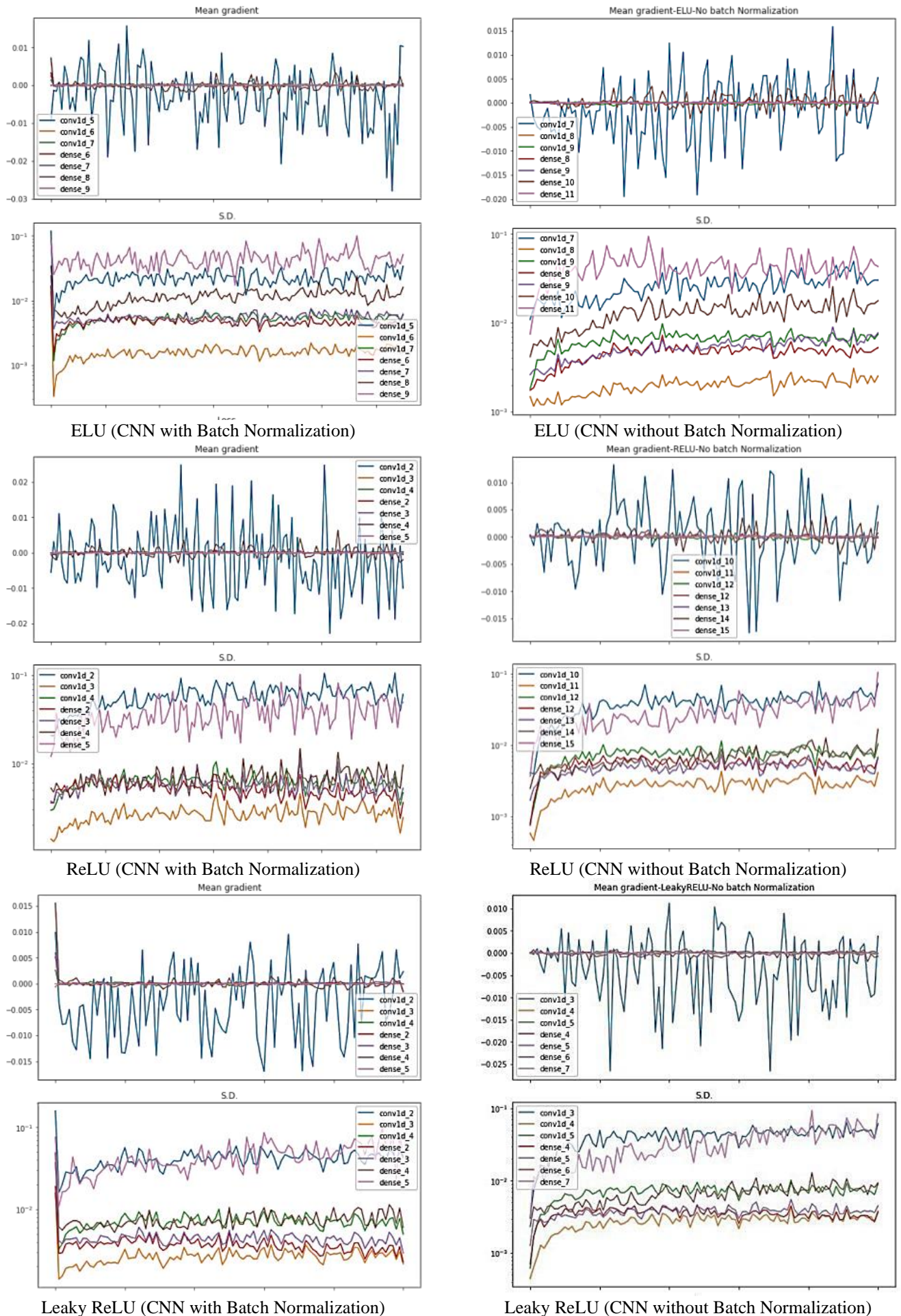
#### 4. RESULTS AND DISCUSSION

The results and discussions are discussed with respect to the hypothesis made in literature review section. According to this the experiment was conducted to tap the gradients for the following cases:

1. ELU, ReLU and Leaky ReLU are used with Hybrid Model
2. ELU, ReLU and Leaky ReLU are used with CNN Model and Batch Normalization after each CNN layer
3. ELU, ReLU and Leaky ReLU are used with CNN Model without Batch Normalization

The respective gradients are shown in Figure 4.





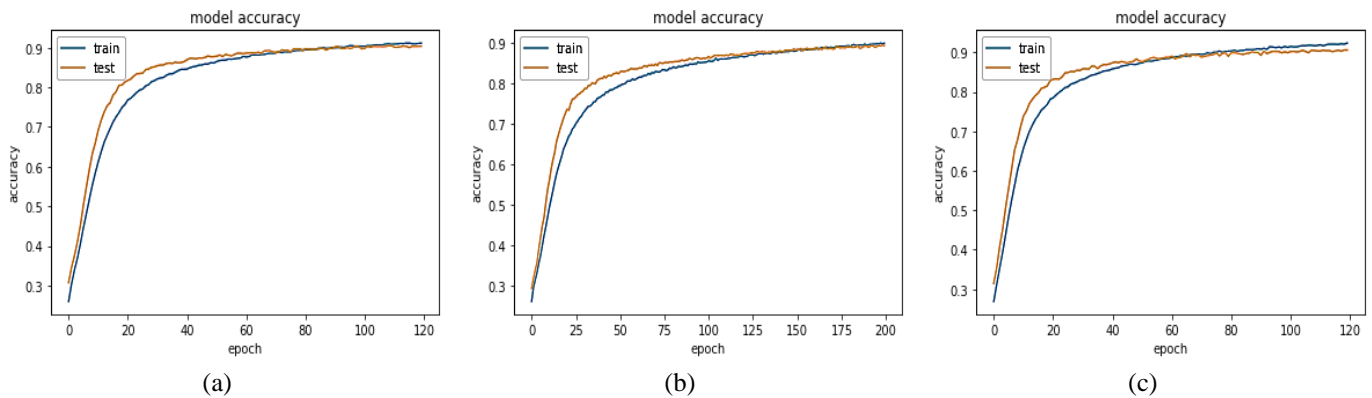
**Figure 4.** Gradient plot of different activation function

**Table 5.** Performance measure (effectiveness) of different models

Model	Activation Function	Evaluation Accuracy (%)	Evaluation Loss
Hybrid	Leaky ReLU	94.0	0.207
	ReLU	92.5	0.31
	ELU	94.3	0.209
CNN with Batch Normalization	Leaky ReLU	84.73	0.43017
	ReLU	83.22	0.4957
	ELU	86.15	0.3942
CNN without Batch Normalization	Leaky ReLU	85.66	0.4114
	ReLU	82.30	0.48747
	ELU	85.09	0.4189

**Table 6.** Performance measure (efficiency) of different models

Model	Activation Function	Training Time in Sec/Epoch	Prediction Time in Sec	Convergence Epoch Number
Hybrid CNN + LSTM	Leaky ReLU	247	1.82	120
	ReLU	233	1.45	150
	ELU	255	1.711	70
CNN model with Batch Normalization	Leaky ReLU	59	0.393	120
	ResLU	60	0.334	150
	ELU	62	0.471	70
CNN model with without Batch Normalization	Leaky ReLU	32	0.301	120
	ReLU	30	0.303	130
	ELU	33	0.310	70

**Figure 5.** Accuracy of Hybrid Bi-LSTM model for various activation functions: (a) Leaky ReLU (b) ReLU (c) ELU

#### 4.1 Hypothesis-1

It can be seen from the gradients Figure 4 that, in Hybrid model, ELU has its gradients more away from zero gradient (-0.15 to -0.2) which is the desirable factor compared to ReLU which are closer to zero (-0.03 to 0.03) meaning that it may lead to vanishing gradient problem. Leaky ReLU however has in between values. As vanishing gradient depends on the number of layers, the experiment has been conducted with CNN model of total 7 layers as given in the Table 4. This was preferred as LSTM is computationally more complex. Here it can be noticed that there is a difference between gradients flow between the model with Batch Normalization and model without. ReLU performs well with Batch Normalization i.e., the gradient distance from zero is doubled (0.015 to 0.03), whereas ELU almost similar between with Batch Normalization and without Batch Normalization. The difference between with and without Batch Normalization in case of ReLU was due to the normalization of the input to each layer done by Batch Normalization. Leaky ReLU which was designed for reducing the dying neuron problem did not perform better with Batch Normalization. Also, with respect to each activation function, compared to the hybrid model, the

gradients are little closer to zero and this is because of the increase in number of layers. In every case even if the changes are very less with respect to gradient flow and other parameters, ELU outperforms over all other activation function as shown in the below Table 5 and Table 6, i.e., ELU is the effective activation function.

The model computation included both training and testing of the samples. The model was computed for 120 epochs for all the activation function except ReLU which converges at 150 epochs. The Figure 5 depicts the accuracy of emotion classification with respect to various activation functions. It can be seen from the Figure 5(c) that fast convergence happens and it was attained in less than 50 epochs due to the improved gradient flow of the model facilitated by the ELU activation.

It is observed from Table 5 that ELU has 94% of accuracy i.e., ELU has performed well for both positive and negative data. The use of ELU as the activation function allows the negative values to be pushed closer to zero with lower computational complexity. Due to the reduced bias effect, the normal gradient and unit natural gradient gets closer resulting in a faster mean shift towards zero [9]. On the other hand, ReLU becomes inactive and stops learning for negative inputs. However, the distribution of negative data is less compared to

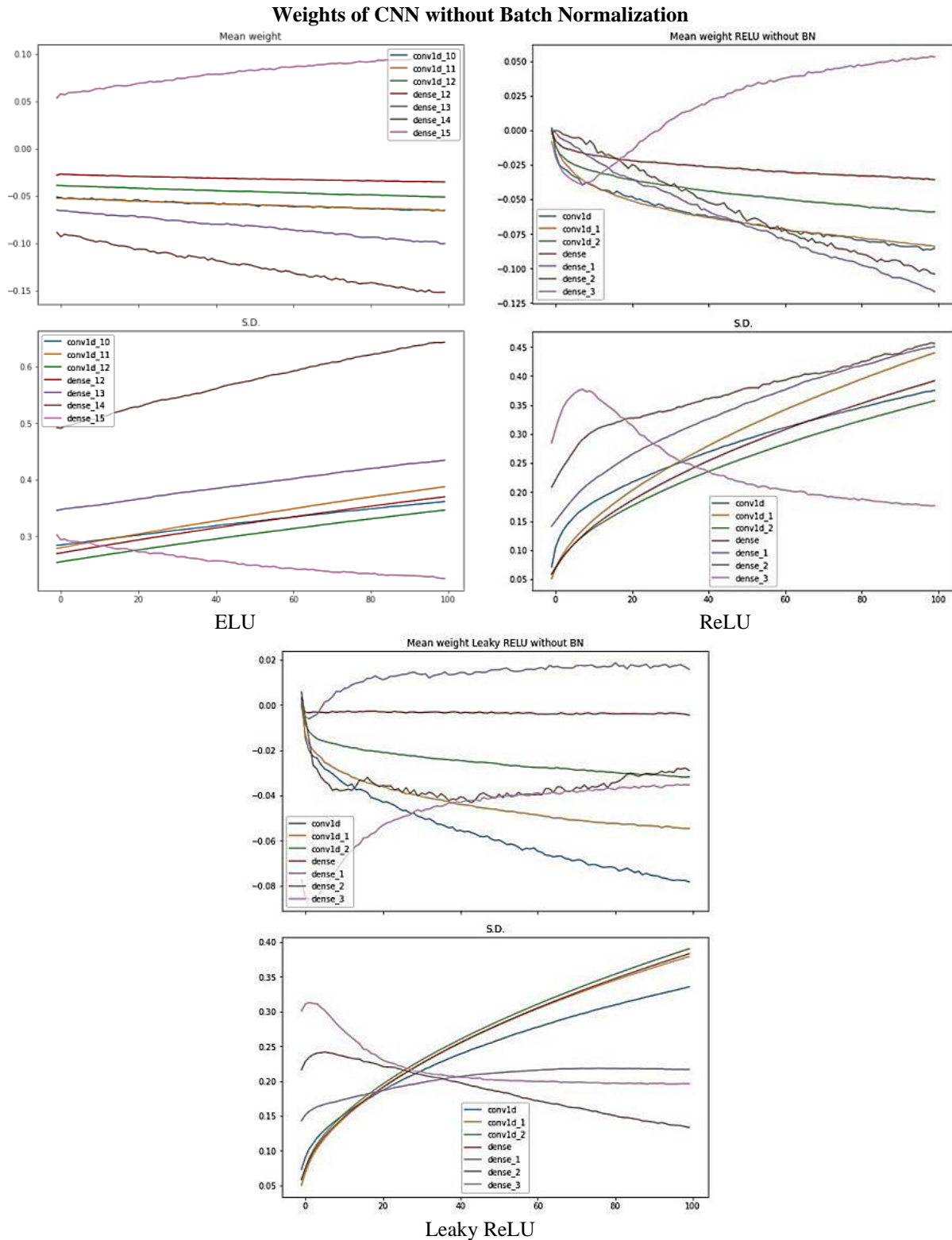


positive data and hence the efficiency is not affected more and it is 92.5%.

### 4.2 Hypothesis-2

During the model fitting, the weight grows by the learning algorithm. The weights will rise in size in order to handle the features of the samples given in the training data. The model with smaller weights is preferred than larger weights as larger weights tend to capture more specifics of the given data rather

than generalizing it [46]. Hence it is confirmed from the weight plots as shown in Figure 6, that ELU performs well than other two models where the weights are keep growing until the last epoch and their magnitudes also larger than ELU weights. However, the changes are very minimal i.e., ReLU weights changes from 0 to -0.075 and Leaky ReLU weights changes from 0 to -0.08 whereas ELU weights most constant around -0.05. Hence it is confirmed that ELU performs well compared to other activations functions considered in this work.



**Figure 6.** Mean Weight plots of different activation functions

### 4.3 Hypothesis-3

Among the two DNN models, hybrid Bi-LSTM showed better performance and thus statistical analysis has been done for the same and follows some additional combinational activation function also. This was done because not a major difference was observed with plain ELU, ReLU and Leaky ReLU as discussed earlier. For statistical analysis, Friedman chi-square test was used to determine the statistical difference between the models with different activation function. The results of the test are shown in Table 7 with the p-value of 0.001518 which is less than 0.05. Hence it is concluded that there are statistically significant differences between the models and the performance with null hypothesis can be rejected.

**Table 7.** Statistical test analysis results

<b>Friedman Test Statistic</b>	19.5488
<b>p-value</b>	0.001518

After rejecting the null hypothesis in the Friedman test, Dunn-Bonferroni test is carried out for pairwise comparisons in situations where multiple comparisons are made between the groups. For each pairwise comparison, the test results have provided a p-value. The p-values of Leaky ReLU with model 2 is 0.0106 and model 3 is 0.0247. Additionally, the p-values of ELU is significantly different from model 2 is 0.0205 and model 3 is 0.0445.

**Table 8.** Performance measure of the model for different activation functions

Model No.	Model	Activation Function	Accuracy	Loss	Precision	Recall	F1-Score
Model 1	Hybrid Bi-LSTM	Tanh+ ReLU	90.3%	0.353	0.89	0.89	0.89
Model 2	Hybrid Bi-LSTM	ReLU + ELU	89.3%	0.374	0.88	0.88	0.88
Model 3	Hybrid Bi-LSTM	Tanh + ReLU + ELU	89.6%	0.384	0.89	0.88	0.88
Model 4	Hybrid Bi-LSTM	ReLU	92.6%	0.312	0.91	0.92	0.91
Model 5	Hybrid Bi-LSTM	Leaky ReLU	94.0%	0.207	0.93	0.93	0.94
Model 6	Hybrid Bi-LSTM	ELU	<b>94.3%</b>	<b>0.209</b>	<b>0.94</b>	<b>0.93</b>	<b>0.93</b>

**Table 9.** Summary results of hypothesis tests

Hypothesis	Comparative Performance		
	ReLU	Leaky ReLU	ELU
<b>H1</b>	Fair	Fair	Good
<b>H2</b>	Fair	Fair	Good
<b>H3</b>	Fair	Fair	Good
<b>Effectiveness (Accuracy and Loss Function)</b>	Fair	Fair	Good
<b>Efficiency</b>	Good	Moderate	Moderate

The performance measures like accuracy, loss, precision, recall, and F1-score of activation function which was experimented on Hybrid Bi-LSTM is given in Table 8. From the Table 8, it is understood that the ELU activation function has better emotion recognition than other activation functions. Finally, it was decided that ELU is the best activation function for EEG based applications. The summarized results of each hypothesis are given in Table 9.

With each of the hypothesis results and effectiveness and efficiency terminology discussed, it is clear that ELU shows improved performance for any model combination. However, it is not clear that the performance improvement is exactly due to mean gradients, the mean weight parameters or any other parameters considered in this article. Hence, eventually the performance improvement due to the activation functions is not completed by the results discussed in this article. Further experiments need to be conducted with different value of learning rate and parameters of activation function such as  $\alpha$  in Leaky ReLU, ELU, for wide range of datasets.

### 5. CONCLUSIONS

A comparison of three different combinations of activation functions that were used for emotion recognition has been carried out. The comparison was done on a hybrid Bi-LSTM and CNN model. The potential use of ELU in DNN has given accuracy of 94% while the other models gave an accuracy of around 92%. The pragmatic choice of ELU as the activation

function is best suited for dataset like DEAP if the data has distributed values of both positive and negative values. ELU has better performance due to its ability to process negative inputs. It also gives fast convergence and avoids overfitting of data. However as seen there is no marked differences in the accuracies and hence much denser analysis i.e., activation function in combination with other tuning parameters has to be done in order to truly appreciate the effect of role of the activation functions in any DNN.

### REFERENCES

- [1] Yannakakis, G.N. (2018). Enhancing health care via affective computing. *Malta Journal of Health Sciences*, 5(1): 38-42. <https://doi.org/10.14614/HEALTHCOMP/9/18>
- [2] Margaret, M.J., Banu, N.M. (2022). A survey on brain computer interface using EEG signals for emotion recognition. In *AIP Conference Proceedings*, 2518(1): 040002. <https://doi.org/10.1063/5.0103476>
- [3] Rached, T.S., Perkusich, A. (2013). Emotion recognition based on brain-computer interface systems. *Brain-Computer Interface Systems-Recent Progress and Future Prospects*, 253-270. <https://doi.org/10.5772/56227>
- [4] Banu, N.M., Sujithra, T., Cherian, S.M. (2021). Performance comparison of BCI speller stimuli design. *Materials Today: Proceedings*, 45: 2821-2827. <https://doi.org/10.1016/j.matpr.2020.11.804>

- [5] Khanna, A., Gupta, D., Bhattacharyya, S., Hassanien, A. E., Anand, S., Jaiswal, A. (2021). International conference on innovative computing and communications. Proceedings of ICICC, 2. <http://doi.org/10.1007/978-981-16-3071-2>
- [6] Niemic, C.P. (2002). Studies of emotion: A theoretical and empirical review of psychophysiological studies of emotion. *Journal of Undergraduate Research*, 1(1): 15-18.
- [7] Liu, H., Zhang, Y., Li, Y., Kong, X. (2021). Review on emotion recognition based on electroencephalography. *Frontiers in Computational Neuroscience*, 15: 758212. <https://doi.org/10.3389/fncom.2021.758212>
- [8] Gustineli, M. (2022). A survey on recently proposed activation functions for Deep Learning. arXiv preprint arXiv:2204.02921. <https://doi.org/10.31224/2245>
- [9] Clevert, D.A., Unterthiner, T., Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289. <https://doi.org/10.48550/arXiv.1511.07289>
- [10] Dubey, S.R., Singh, S.K., Chaudhuri, B.B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 503: 92-108. <https://doi.org/10.1016/j.neucom.2022.06.111>
- [11] Keelawat, P., Thammasan, N., Numao, M., Kijirikul, B. (2021). A comparative study of window size and channel arrangement on EEG-emotion recognition using deep CNN. *Sensors*, 21(5): 1678. <https://doi.org/10.3390/s21051678>
- [12] Apicella, A., Isgro, F., Prevete, R. (2019). A simple and efficient architecture for trainable activation functions. *Neurocomputing*, 370: 1-15. <https://doi.org/10.1016/j.neucom.2019.08.065>
- [13] Liu, X.Y., Xing, Q., Zhang, H., Chen, F. (2023). A Novel Activation Function of Deep Neural Network. *Scientific Programming*, 2023(1): 3873561. <https://doi.org/10.1155/2023/3873561>
- [14] Madhu, G., Kautish, S., Alnowibet, K.A., Zawbaa, H. M., Mohamed, A.W. (2023). Nipuna: A novel optimizer activation function for deep neural networks. *Axioms*, 12(3), 246. <https://doi.org/10.3390/axioms12030246>
- [15] Mastromichalakis, S. (2020). ALReLU: A different approach on Leaky ReLU activation function to improve Neural Networks Performance. arXiv preprint arXiv:2012.07564. <http://arxiv.org/abs/2012.07564>.
- [16] Jebadurai, J., Jebadurai, I.J., Paulraj, G.J.L., Samuel, N.E. (2019). Super-resolution of digital images using CNN with leaky ReLU. *Int. J. Recent Technol. Eng*, 8(2): 210-212. <https://doi.org/10.35940/ijrte.B1034.0982S1119>
- [17] Pan, B., Zheng, W. (2021). Emotion recognition based on EEG using generative adversarial nets and convolutional neural network. *computational and Mathematical Methods in Medicine*, 2021(1): 2520394. <https://doi.org/10.1155/2021/2520394>
- [18] Chao, H., Dong, L., Liu, Y., Lu, B. (2019). Emotion recognition from multiband EEG signals using CapsNet. *Sensors*, 19(9): 2212. <https://doi.org/10.3390/s19092212>
- [19] Pandey, P., Seeja, K.R. (2022). Subject independent emotion recognition from EEG using VMD and deep learning. *Journal of King Saud University-Computer and Information Sciences*, 34(5): 1730-1738. <https://doi.org/10.1016/j.jksuci.2019.11.003>
- [20] Li, Y., Huang, J., Zhou, H., Zhong, N. (2017). Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. *Applied Sciences*, 7(10): 1060. <https://doi.org/10.3390/app7101060>
- [21] Garg, D., Verma, G.K. (2020). Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet based deep learning framework. *Procedia Computer Science*, 171: 857-867. <https://doi.org/10.1016/j.procs.2020.04.093>
- [22] Asghar, M.A., Khan, M.J., Fawad, X., Amin, Y., Rizwan, M., Rahman, M., Badnava, S., Mirjavadi, S.S. (2019). EEG-based multi-modal emotion recognition using bag of deep features: An optimal feature selection approach. *Sensors*, 19(23): 5218. <https://doi.org/10.3390/s19235218>
- [23] Xing, X., Li, Z., Xu, T., Shu, L., Hu, B., Xu, X. (2019). SAE+ LSTM: A new framework for emotion recognition from multi-channel EEG. *Frontiers in Neurorobotics*, 13: 37. <https://doi.org/10.3389/fnbot.2019.00037>
- [24] Ozdemir, M.A., Degirmenci, M., Izci, E., Akan, A. (2021). EEG-based emotion recognition with deep convolutional neural networks. *Biomedical Engineering/Biomedizinische Technik*, 66(1): 43-57. <https://doi.org/10.1515/bmt-2019-0306>
- [25] Acharya, D., Jain, R., Panigrahi, S.S., Sahni, R., Jain, S., Deshmukh, S.P., Bhardwaj, A. (2021). Multi-class emotion classification using EEG signals. In *Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India*, pp. 474-491. [https://doi.org/10.1007/978-981-16-0401-0\\_38](https://doi.org/10.1007/978-981-16-0401-0_38)
- [26] Yang, Y., Wu, Q., Fu, Y., Chen, X. (2018). Continuous convolutional neural network with 3D input for EEG-based emotion recognition. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia*, pp. 433-443. [https://doi.org/10.1007/978-3-030-04239-4\\_39](https://doi.org/10.1007/978-3-030-04239-4_39)
- [27] Devi, T., Deepa, N. (2021). A novel intervention method for aspect-based emotion using exponential linear unit (ELU) activation function in a deep neural network. In *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India*, pp. 1671-1675. <https://doi.org/10.1109/ICICCS51141.2021.9432223>
- [28] Bejjagam, L., Chakradhara, R. (2022). Facial emotion recognition using convolutional neural network with multiclass classification and Bayesian optimization for hyper parameter tuning.
- [29] Le, T.D.T., Van, L.T., Hong, Q.N. (2020). Deep convolutional neural networks for emotion recognition of Vietnamese. *International Journal of Machine Learning and Computing*, 10(5): 692-699. <https://doi.org/10.18178/ijmlc.2020.10.5.992>
- [30] Abdelwahab, M., Busso, C. (2018). Study of dense network approaches for speech emotion recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada*, pp. 5084-5088. <https://doi.org/10.1109/ICASSP.2018.8461866>
- [31] Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11): 5391-5420. <https://doi.org/10.1002/hbm.23730>
- [32] Nguyen, A., Pham, K., Ngo, D., Ngo, T., Pham, L. (2021). An analysis of state-of-the-art activation functions for

- supervised deep neural network. In 2021 International Conference on System Science and Engineering (ICSSE), Ho Chi Minh City, Vietnam, pp. 215-220. <https://doi.org/10.1109/ICSSE52999.2021.9538437>
- [33] Liu, X.Y., Xing, Q., Zhang, H., Chen, F. (2023). A Novel Activation Function of Deep Neural Network. *Scientific Programming*, 2023(1): 3873561. <https://doi.org/10.1155/2023/3873561>
- [34] León, J., Escobar, J.J., Ortiz, A., Ortega, J., González, J., Martín-Smith, P., Gan, J.Q., Damas, M. (2020). Deep learning for EEG-based Motor Imagery classification: Accuracy-cost trade-off. *Plos One*, 15(6): e0234178. <https://doi.org/10.1371/journal.pone.0234178>
- [35] Farahat, A., Reichert, C., Sweeney-Reed, C.M., Hinrichs, H. (2019). Convolutional neural networks for decoding of covert attention focus and saliency maps for EEG feature visualization. *Journal of Neural Engineering*, 16(6): 066010. <https://doi.org/10.1088/1741-2552/ab3bb4>
- [36] Bai, Y. (2022). RELU-function and derived function review. In *SHS Web of Conferences*, 144: 02006. <https://doi.org/10.1051/shsconf/202214402006>
- [37] Jang, S., Moon, S.E., Lee, J.S. (2021). EEG-based emotional video classification via learning connectivity structure. *IEEE Transactions on Affective Computing*, 14(2): 1586-1597. <https://doi.org/10.1109/TAFFC.2021.3126263>
- [38] Liang, Z., Zhou, R., Zhang, L., Li, L., Huang, G., Zhang, Z., Ishii, S. (2021). EEGFuseNet: Hybrid unsupervised deep feature characterization and fusion for high-dimensional EEG with an application to emotion recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29: 1913-1925. <https://doi.org/10.1109/TNSRE.2021.3111689>
- [39] Koelstra, S., Muhl, C., Soleymani, M., Lee, J.S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1): 18-31. <https://doi.org/10.1109/TAFFC.2011.15>
- [40] Jehosheba Margaret, M., Masoodhu Banu, N.M. (2023). Performance analysis of EEG based emotion recognition using deep learning models. *Brain-Computer Interfaces*, 10(2-4): 79-98. <https://doi.org/10.1080/2326263X.2023.2206292>
- [41] Al-Fahoum, A.S., Al-Fraihat, A.A. (2014). Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices*, 2014(1): 730218. <http://doi.org/10.1155/2014/730218>
- [42] Hu, Z., Chen, L., Luo, Y., Zhou, J. (2022). EEG-based emotion recognition using convolutional recurrent neural network with multi-head self-attention. *Applied Sciences*, 12(21): 11255. <https://doi.org/10.3390/app122111255>
- [43] Yang, J., Huang, X., Wu, H., Yang, X. (2020). EEG-based emotion classification based on bidirectional long short-term memory network. *Procedia Computer Science*, 174: 491-504. <https://doi.org/10.1016/j.procs.2020.06.117>
- [44] Rhanoui, M., Mikram, M., Yousfi, S., Barzali, S. (2019). A CNN-BiLSTM model for document-level sentiment analysis. *Machine Learning and Knowledge Extraction*, 1(3): 832-847. <https://doi.org/10.3390/make1030048>
- [45] Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8: 1-74. <https://doi.org/10.1186/s40537-021-00444-8>
- [46] Wang, H., Czerminski, R. and Jamieson, A.C. (2021). *Neural Networks and Deep Learning*. Einhorn, M., Löffler, M., de Bellis, E., Herrmann, A. and Burghartz, P. (Ed.) The Machine Age of Customer Insight, Emerald Publishing Limited, Leeds, pp. 91-101. <https://doi.org/10.1108/978-1-83909-694-520211010>