Journal of Information Systems Engineering and Management
2025, 10(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Improvised SimCLR Algorithm for Emotion Recognition using EEG Signal

Preethi V[1], Elizabeth Jesi V[2]

[1,2]Department of Networking & Communications, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai-603203, India.
[1]pv9600@srmist.edu.in, [2]jesiv@srmist.edu.in

ARTICLE INFO

Received: 25 Sept 2024

Revised: 30 Nov 2024

Accepted: 09 Dec 2024

ABSTRACT

The purpose of this endeavor is to create a system that is capable of accurately identifying each emotion that youngsters experience. In order to make children's life better, there is a need for study in the field of interpreting the mental condition of children, which is still developing. On the other hand, voice or image-based emotion recognition algorithms provide false findings due to the fact that children and adolescents will not easily expose their hidden emotions to anyone. The ability to articulate what is going on in their lives can be difficult for some children. People who are affected by this experience depression, which hinders them from living their life as they normally would. EEG waves are utilized to identify children's emotions more precisely than previous SOTA approaches. By doing so, it is possible to determine precisely the state of mind, which is something that cannot be ignored. The findings of this study include convolutional neural network (CNN) models that are merged with Bi-LSTM-SimCLR contrastive learning techniques for the purpose of achieving efficient emotion recognition. The DEAP, which is widely considered to be the most valuable EEG benchmark dataset, is utilized for classification purposes in this study. Additionally, its labels, valence and arousal, dominance and liking, are utilized. In order to represent the dataset, we make use of the Fast Fourier Transformation for the frequency domain features. SimCLR models achieve a flawless accuracy of 97.02% for EEG based image and 95.51% for EEG signal correspondingly, surpassing the performance of every other model that was considered to be state-of-the-art in the past.

Keywords: DEAP dataset, EEG, Contrastive learning, Valence, Dominance, Liking, Arousal, CNN, Bi-LSTM, SimCLR.

INTRODUCTION

Electroencephalogram, sometimes known as EEG, is a type of electrical signal that is produced by the human brain's outermost layer. The signal in question demonstrates characteristics that are nonlinear and non-smooth. Significant difficulties have been encountered by researchers in the process of effectively identifying EEG signals and extracting useful information from those signals. A number of researchers have proposed algorithms for the extraction of features that make use of conventional approaches [9] for the purpose of emotion recognition. A number of deep learning techniques have also been developed for the purpose of emotion recognition. These techniques include convolution neural networks [10], deep learning networks [9], long and short - term memory neural networks (LSTM) [7], graph convolutional networks [11], and attention mechanisms [12]. In recent years, graph convolution neural networks (GCNN), and attention mechanisms have been increasingly prominent because to the increasing number of applications they have in the area. Nevertheless, we have to deal with the problem of improving the fusion of these unique networks and applying them in EEG-related sentiment identification. This is a task that we are currently facing. The implementation of an approach for emotion recognition that is founded on a deep learning model is something that we recommend doing in order to address the problems that have been stated above [13]. In order to improve the spatial related knowledge included in the EEG data, the first step of the procedure entails making use of the similar spatial design. After that, the EEG source that has been filtered is put through a continuous wavelet transform, which ultimately results in a time-frequency map [13]. Specifically, this map is used as the starting point for the deep model, and this combines convolutional features. In addition to incorporating a multi-headed automatic-attentive mechanism on top of convolutional neural networks and bi-directional long-term and short-term memory networks. The hybrid model makes use of several convolutional neural networks (CNNs) to derive feature information, acquires knowledge about upcoming and past time series, and extracts temporal characteristics by utilizing the Bi-LSTM algorithm. Furthermore, it enhances the feature data by allocating value to emotion characteristic using a multi-headed automatic-attention process. This results in an improvement of the feature information. In the end, we carried out exhaustive tests on the DEAP the database, and the outcomes of those experiments indicated that the method demonstrates greater classification performance in comparison to the classification methods that are currently available [13].

Several methodology have been suggested in the fascinating quest to comprehend and categorize human emotions. An exemplary model is the Circumplex 2D model [1]. This model employs a novel methodology that utilizes a two-dimensional framework to map emotions, taking into account Valence and Arousal. Valence quantifies the inherent attractiveness of an emotion, predicting it as +ve or -ve. Conversely, Arousal measures the degree of excitement or intensity linked to the emotion. A more complex 3D model called Pleasure, Arousal, and Dominance (PAD), or Valence, Arousal and Dominance (VAD), gained prominence but only valence and arousal are considered in the study conducted [2]. Entering the domain of control, adding dominance evaluates how much a person feels in control of, or dominated by, a specific emotion [2].

Identifying the key features that fluctuate with changes in emotional state is the primary and difficult task in recognizing human emotion. For emotion recognition techniques based on shallow and deep learning, the following extracted EEG features are used: Time-domain features encompass statistical measures. The frequency domain features of EEG typically encompass a greater amount of pertinent information. The primary techniques utilized are Power Spectral Density (PSD), Fast Fourier Transform (FFT), and the Short Time Fourier Transform (STFT) [3] [4]. The Wavelet transform is an effective method for analyzing data in both the time and frequency domains [5]. It can be categorized into two types: Continuous Wavelet Transform [6] and Discrete Wavelet Transform [2]. This work employed the frequency domain approach, with a specific focus on power spectral density (PSD) analysis. Spectral density analysis (PSD) is a commonly employed method for investigating the spread power in various frequency components in an EEG signal. Figure 2 shows the PSD of subject 1 in various frequency components using welch's periodogram. It provides valuable data about the frequency of the data. The characteristic of the EEG signal is suggested by the detection of frequency pattern. Furthermore, analysis of PSD enables the measurement of the proportional contributions of power for various frequency components, which offers important insights for the subsequent examination and understanding of the emotional and mind pressure of the participants.

The detection of emotions through the use of EEG data has been the subject of study carried out by a number of researchers, who have utilized a variety of techniques for modeling, extraction of features, and preliminary processing [7]. On the other hand, the efficiency of an approach is not judged exclusively based on how accurate it is. Performance of an ideal system can take into account critical metrics such as resilience, execution speed, comprehension with little data, and training with restricted parameterized models [8]. In addition to recognition accuracy, efficiency of the best system can also take into consideration these metrics.

This research focussed on which deep learning model is efficient in recognizing emotion from EEG signal and image? Does contrastive learning improve the recognition of accuracy of deep learning technique? Does Augmentation increased dataset and improved generalization of the model?

The main contributions of this paper are:

1. Our proposal offers a method for utilizing EEG signals and image as data sources for modeling the strategy that we propose. The CNN-BiLSTM-SimCLR model is utilized for this work. By taking this technique, the model's capacity to derive informative features is improved, which in turn leads to an increase in recognition accuracy.

2. Extensive tests were conducted that our deep learning model achieved a higher level of accuracy than other models while taking into account the EEG image rather than the EEG signal as input and evaluates and compares their overall performance.

LITERATURE SURVEY

The initial electroencephalogram (EEG) is utilized as source signal, a convolutional neural network and a support vector machine, Bi-LSTM model are utilized for the purpose of acquiring features and fusion, and in the end, the electrode channel strength are distributed through the utilization of the system's attention. The procedure consists of each and every one of these steps. When everything is said and done, the signals that are produced by the electroencephalogram (EEG) are categorized according to the many types of feelings in this literature [24]. Within the scope of this investigation, we provide a novel method for the identification of emotions that is founded on deep learning models. After the electroencephalogram (EEG) data has been filtered with the common spatial pattern (CSP), and is then transformed into a time-frequency map using the continuous wavelet transform (CWT).

This process is repeated until the desired result is achieved. Specifically, the network makes use of this information as input, and the deep learning model is the one responsible for extracting and classifying features. A model known as CNN-BiLSTM-MHSA is a model that combines CNN, Bi-LSTM, and multi-head automatic-attention into a single entity. It is possible for the network to gain the continuous and spatial data of EEG, extract deep features using CNN, interpret upcoming and previous time series using BiLSTM, and increase identification accuracy using MHSA by rearranging values to emotion characteristics is adapted in this work [13].

SimCLR is a framework for visualized contrastive learning that is simple to use, and this study serves as an introduction to the framework. The contrastive self-supervised machine learning approaches that have been developed in recent times have been simplified by deleting specific structures and memory banks. This has occurred concurrently. A thorough investigation into the fundamental components of the system is carried out in order to get a grasp of the process by which activities involving contrastive forecasting acquire representations that can be utilized. (3) Contrastive learning is superior to supervised learning in terms of producing better results with larger batches and a greater number of training steps to complete [25]. Our findings demonstrate that the composition of data augmentation is necessary for predictive work, the use of nonlinear transformations enhances the quality of representation, and contrastive learning outperforms supervised learning in terms of the outcomes it produces.

METHODS AND MATERIALS

3.1 Preprocessing

EEG signals are inherently noisy due to the addition of numerous artifacts during the recording process, including head, muscle, and eye movements. Thus, eliminating this noise turns into an essential phase in the analysis. Although the analysis may be hampered by the large amount of artifacts and unwanted noise in raw data, the DEAP dataset team has already supplied a pre-processed copy of their original EEG data. Figure 1 shows the EEG data. This removes all notable noise from it and facilitates analysis. The initial 512 Hz of this EEG data has been reduced to 128 Hz. Eye artifacts are removed using the blind source division technique. A band pass spectrum filter with a bandwidth of 4.0 Hz to 45.0 Hz is in place. The data is averaged through adherence to widely used references. The Geneva order is adhered for the EEG channels since the EEG data has been captured twice. Every single trial's 60s of data and the 3s pre-trial starting point comprise a total of 63s of trial data. The 3 seconds of pre-trial baseline are later eliminated. Lastly, the order of each video display is changed to match its experimental 2D. To get the expected emotion detection model, an initially processed group of data from the previously identified DEAP Dataset was used [16].

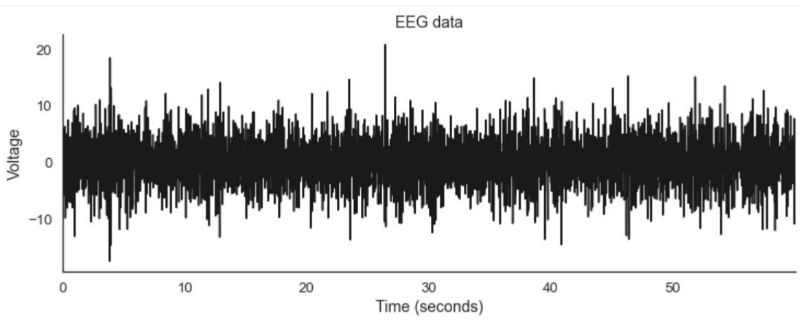

Figure 1: EEG data

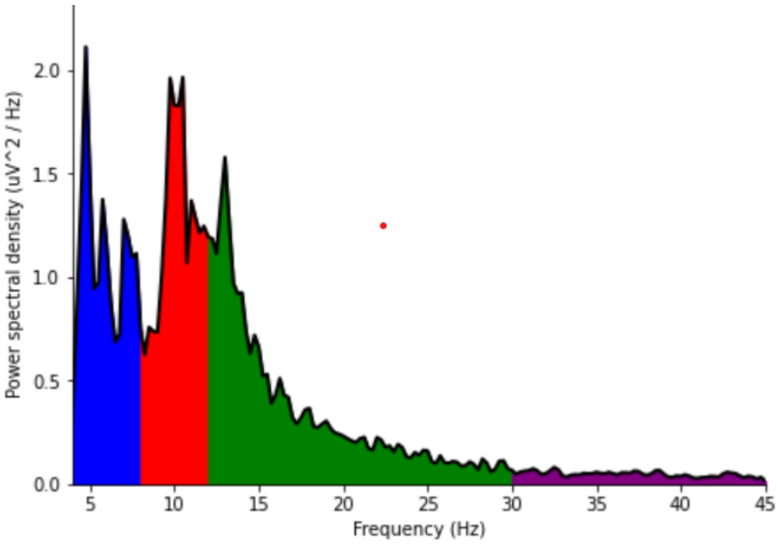

Figure 2: Power Spectral Density of a subject

3.2 Conversion of Signal to Image

Images utilizing the power spectrum of EEG data, which includes information about time, frequency, and electrodes, can be obtained through the utilization of CWT. In comparison to the Fourier transform, it is equipped with a greater capacity to identify certain MI patterns in EEG signals [26].

A signal can be analyzed in terms of its time and frequency simultaneously by employing a technique known as continuous wavelet transform. Coefficients, which are functions of scale or frequency and time, are the components that are produced by the CWT output. When compared to the initial signal, each scaled wavelet undergoes a shift in time throughout the whole length of the signal. This process is repeated for each of the scales, which ultimately results in coefficients that are functions of the scale parameter and the shift parameter of the wavelet itself. A signal that has 1500 samples and is processed using 20 scales will result in 30,000 coefficients. This is to put things into perspective.

We want to be able to extract features from EEG signals that would be helpful in classifying events that correspond to waves that are collected from the brain cells [27].

3.3 Augmentation

3.3.1 Image augmentation

Two approaches are presented in the study. The first is called "VH-BC+," which blends BC+ with the outcomes of "Vertical Concat" and "Horizontal Concat." The second approach, called "VH-Mix," combines linearity with the advantages of these approaches. For each of the two results, the method generates two intermediate images, which are then fed into the mix function. The image's lower right corner is derived from x_2, while the top left and fewer right areas are derived from x_1. Different mixing coefficients are used to blend the upper and lower left corners [17]. When λ1, λ2, and λ3 ~ Beta (α, α) are combined, we obtain: x (r, c) =

$$
\begin{array}{ll}
x_1(r, c), & \text{if } r \le \lambda_1 H \wedge c \le \lambda_2 W \\
\lambda_3 x_1(r, c) + (1 - \lambda_3)\, x_2(r, c), & \text{if } r \le \lambda_1 H \wedge c > \lambda_2 W \\
(1 - \lambda_3)\, x_1(r, c) + \lambda_3 x_2(r, c), & \text{if } r > \lambda_1 H \wedge c \le \lambda_2 W \\
x_2(r, c), & \text{if } r > \lambda_1 H \wedge c > \lambda_2 W
\end{array} \tag{1}
$$

Simply the probable percentage of the value that a random pixel takes from x_1 vs x_2.

$$\widetilde{x} = p(x_1-\mu_1) + (1-p)(x_2-\mu_2) / \text{SRT}(p_2+(1-p_2)) \tag{2}$$

Where $p = 1/((1+(\sigma_1/\sigma_2).(1-\lambda/\lambda))$

3.3.2 Signal augmentation

Generating artificial signals from brain waves is a popular use of GAN in EEG. There are many potential uses for this EEG generation technology, including improving EEG following brain traumas. In this instance, we further examine the learning signals derived from max activation technology using GAN. We conjecture that additional evidence of the comparison among the training signal as well as the electroencephalogram signal will be provided, if the discriminator is unable to identify the artificial EEG. One way to conceptualize the adversarial process is as a min-max game with a loss function:

$$\min_{G}\max_{D} L\,(D, G) = E_{i\sim qdata(i)}\,[\log D(i)] + E_{i\sim v(i)}\,[\log\,(1 - D(G(j)))] \quad (3)$$

First of all the formula above indicates that the parameter that discriminates is $D(i;\theta_d)$ and the training data distribution is qdata. It is able to assess the likelihood from qdata. The noise distribution is represented by v_j, and the learning parameters are θ_d. In this case, the generator is indicated by $G(j;\ \theta_g)$, and the Gauss noise distribution is used. With deep neural networks, it can produce synthetic data from brain waves from gaussian noise by utilizing the opposite transformation parameter θ_g. In the event that the data's label is taken into account, the generator and discriminator receive the auxiliary label information [18], which causes the loss function to alter as shown by equation (3).

3.4 Contrastive learning through CNN-BiLSTM proposed framework

The Input given for the proposed work is EEG signal. Here we consider signal as input in first experiment and converting signal to image in second experiment. Figure 3 shows the flow of CNN-Bi-LSTM-SimCLR.

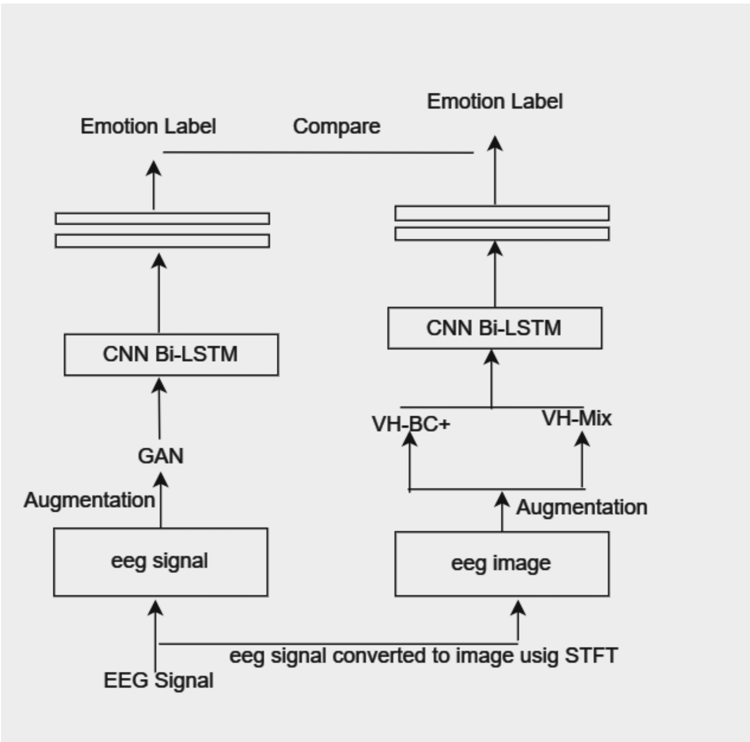

Figure 3: Flow of proposed work

3.4.1 First experiment

A GAN with an EEG signal as its input is utilized in order to acquire two perspectives that are capable of blindly modifying any specific data instance. For the purpose of achieving a high level of efficiency, the two related representations of an analogous instance are referred to as x_i and x_j, respectively. These views are considered to be a positive pair as well as a negative pair. The raw data is processed by a network of convolutional neural networks that is two-dimensional and has a convolutional kernel that has a memory capacity of 128. After that, the information that was collected from the sequence is then put into the Bi-LSTM deep network, which contains layer of LSTM network with 256 hidden units. This is done after batch normalization (BN), the ReLu, and pooling using a kernel size 3 [13]. A backward layer is incorporated into the Bi-LSTM model in order to gain information about future emotional states. By integrating the gating system and bidirectional features in a smooth manner, the Bi-LSTM makes it possible for the two LSTM units to remember and process additional data [19]. The framework receives the time series as input, the forward layer correlates image data from earlier batches to the information that is currently available, the backward layer establishes a connection to the future, and the predicted outcome is considered the output.

F (•) is an encoder that is based on neural networks and is capable of extracting representation vectors from augmented data samples. Without any restrictions, our system is capable of supporting a wide range of network structure configurations. For the purpose of achieving $z_i = f(y_i) = \text{ResNet}(y_i)$, where $z_i \in R_d$ is the outcome after the average pooling, we opt for ease and utilize the frequently utilized ResNet [20]. By mapping representations to the dimension where contrastive loss is utilized, a tiny neural network projection head denoted by the letter g (•) is utilized. To obtain $z_i = g(z_i) = W_{(2)}\, \sigma\, (W_{(1)}\, z_i)$, where σ is a ReLU nonlinearity, we employ a multilayer perceptron (MLP) that consists of a single hidden layer.

3.4.2 Experiment 2

The aforementioned system is utilized, besides the fact we use an EEG image through CWT transformation as data and apply VH-Mix and VH-BC+ augmentation.

EXPERIMENT ANALYSIS

4.1 Dataset

In this work DEAP dataset is used which is downloaded from DEAP dataset website that provides data in three format such as in original. bdf format that includes EMG, EOG and other interference, or preprocessed data in .dat format or as csv file [21], [22]. Here we used preprocessed data which is available in s32 file and downsized data to 128 Hz, eliminating EMG artifacts using 4.0 to 4.5 Hz band pass frequency filter as well as non-emotional facts. Using a threshold of 5, high and low valence, high and low arousal, were used to study different emotional states [14].

4.2 CNN-BiLSTM-SimCLR Model

In this, Adam is used to build a loss function for back propagation, cross-entropy loss is used, a popular loss function for multi-class applications. With a greater depth of the two layer bidirectional LSTM, fewer iterations (epoch = 50 - 200) are needed to find the best converge, and dropout = 0.5 is used in the training network to reduce the impact of the over-fitting issue. A CNN with a 128-kernel single-layer architecture powers the convolutional layer. The EEG input sequence's frequency domain features are extracted using a convolutional network, and the time domain features are analyzed by an LSTM unit. Lastly, the projection head is employed to raise the accuracy of classification. This approach uses a bidirectional LSTM layer with 256 units. To choose the 256 units that produce the best results, the hidden layer, which has 128 vs 512 units, is also examined. Out of 1, 2, or 3, the projection head mechanism chooses the heads that have the most impact. This section examines how the amount of projection heads in the basic encoder, or parameter Head, affects performance. The execution of the suggested CNN-BiLSTM-SimCLR on the DEAP file for various epoch is shown in Table 1. This shows that the CNN-BiLSTM-SimCLR (image) effectiveness is greatly seen suggesting that the method is reasonably resilient. Furthermore, the model produces more competitive identification results with Head = 2 than at any other parameter. Thus, we set the CNN-BiLSTM-SimCLR system to projection Head = 2 in order to achieve higher performance. The influence of various parameter configurations is why multiple tests are conducted to determine the most appropriate values [13]. Figure 4 shows the comparison of accuracies under different epoch.

Table 1: Accuracy of different models under each epoch

Epoch	CNN-LSTM-SimCLR(signal)	CNN-LSTM-SimCLR(image)	CNN-Bi-LSTM-SimCLR(signal)	CNN-Bi-LSTM-SimCLR(image)
10	35.58	43.77	45.54	47.33
50	61.64	63.88	75.56	77.12
100	77.28	82.09	85.79	89.41
200	87.14	89.81	95.51	97.02

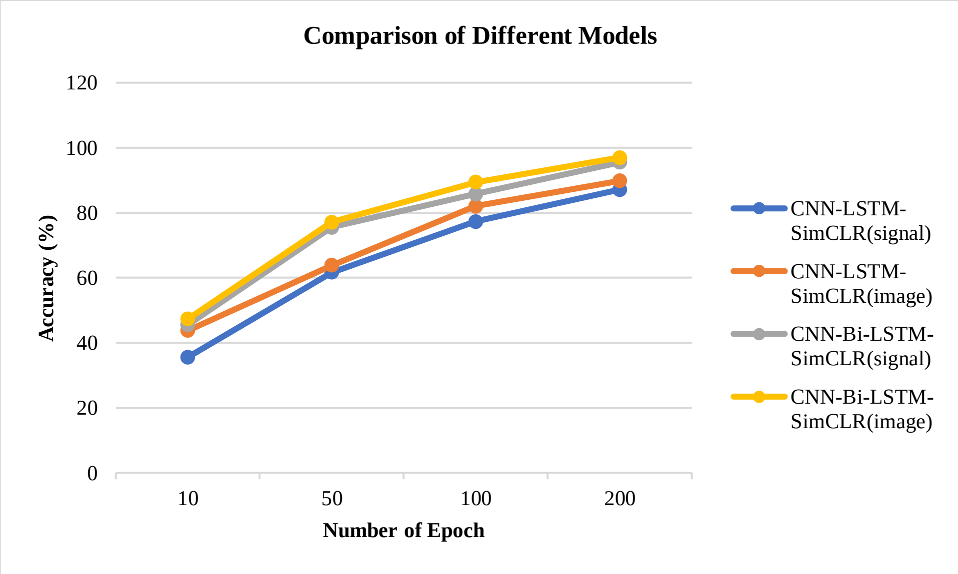

Figure 4: Comparison of accuracies under different epoch

4.3 Evaluation

The effectiveness of the model is done by calculating the accuracy, precision, recall rate, F1-score, and Matthews's correlation coefficient (MCC). Accuracy shows both positive and negative scenarios, estimating the right amount as a percentage of totals. Precision tells the forecast's outcome determines the percentage of the data in which the prediction is accurate. Recall rate calculates on real samples, this is the ratio of expected correct samples to the overall number of actual samples. Neutralized accuracy and recall measures make up the F1-score. An ideal subject prediction is 1else 0 and complete difference then its -1. In essence, MCC is a coefficient that explains the relation among the exact value and the predicted value [13]. Table 2 shows the Accuracy, Precision, Recall, F1 Score and MCC values for EEG signal and EEG image input compared with other models. Figure 5 shows the validity of the Model.

$$\text{Accuracy} = TP + TN/\ (TP + FP + TN + FN) \quad (4)$$

$$\text{Precision} = TP/\ (TP + FP) \quad (5)$$

$$\text{Recall} = TP/\ (TP + FN) \quad (6)$$

$$\text{F1 score} = 2 * (\text{precision}*\text{recall})/\ (\text{precision} + \text{recall}) \quad (7)$$

$$MCC = TP * TN - FP * FN/\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)} \quad (8)$$

Table 2: Validity of the model

Model	Accuracy	Precision	Recall	F1 score	MCC
CNN-Bi-LSTM	93.1	93.16	93.1	93.09	89.69
CNN-LSTM	94.69	94.7	94.69	94.69	92.04
DSCNN-Bi-LSTM	91.1	91.3	91.1	91.08	86.76
DSCNN-LSTM	94.03	94.04	94.03	94.03	91.05
CNN-Bi-LSTM-SimCLR(signal)	97.02	97.03	97.02	97.02	94.04
CNN-Bi-LSTM-SimCLR(image)	95.51	95.55	95.51	95.51	91.86

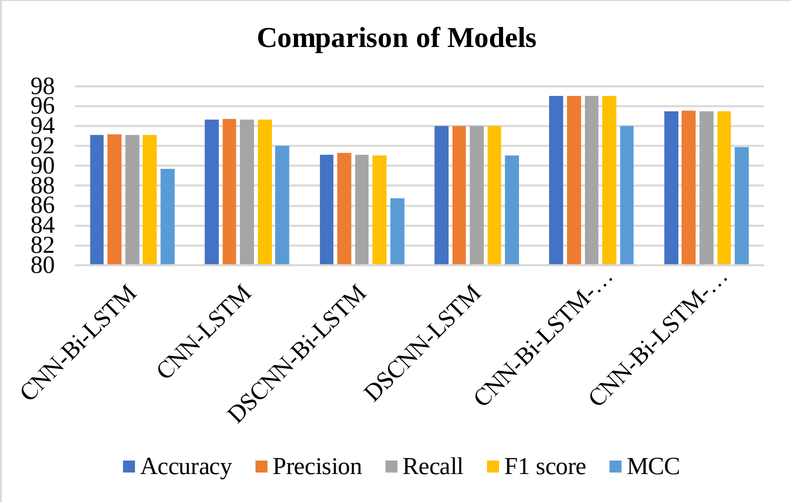

Figure 5: Comparisons of Metrics in different Model

4.4 EEG Emotion Detection First Experiment Evaluation

We made an effort to improve the input long-short time memory network before the Bi-LSTM by introducing CNN beforehand. This was done because CNN has the ability to assist in signal smoothing and minimize the total length of the EEG signal, which ultimately led to a considerable improvement in their performance. CNN modified the electrical brain signal after it was gathered and then put the analyzed data into BiLSTM in order to get a higher degree of accuracy in sentiment categorization. This was done rather than directly utilizing the raw information. In addition, we incorporate a down sampling layer, also known as a pooling layer, into the CNN network in order to reduce the total number of parameters and speed up the computing process. There will be no change to the length of the data. It is then possible for us to make use of BiLSTM in order to acquire a more profound comprehension of the temporal features that CNN goes through. BiLSTM is able to learn the future as well as the past information of the time series, detect the influence of the past and the future on the present, and merge the sentiment features obtained prior to and following as the last sentiment features. This is in contrast to LSTM, which can only take into account sentiment information over a shorter time span. Whether or not BiLSTM makes advantage of the relationship between labels, such as whether or not high valence influences low valence, or the association among other criteria that are presented in more categories, is something that is possible. The accuracy of emotion recognition is improved by the use of generalization of contrastive learning, which achieves this by giving priority to the essential information concerning EEG emotion traits [23].

The pre-processed Python version of the raw data was selected for the scope of this investigation since it was more suitable. Following the extraction of the first 32 EEG channels, statistics were down sampled to a frequency of 128 Hz. Therefore, there are approximately 7680 instances in a minute, which is equal to sixty seconds, while there are 128 samples in a second. Nevertheless, it is important to note that the data should contain 8064 samples that are 128 × 63 samples, and the duration of each test should be a standard of three seconds. Data augmentation is one method that can be utilized to generate additional data. In this article, a video that is one minute long is divided into twelve portions that are each five seconds long. This considerably increases the number of samples and, as a result, the probability of making a forecast or prediction. The data was initially presented in a format that was 1280 × 32 × 8064 which was divided 12 pieces. The information is obtained by a CNN-BiLSTM-SimCLR network that has a batch size of 128 for the data. Following this, a threshold value of 5 is utilized in order to differentiate between the high and low states. The high label is assigned a value of 1 for values that are greater than 5, while the low label is assigned a value of 0 for values that are lower than 5. Within the framework of the validating split technique, this study determines that the training and validation ratio is 7:2, with the remaining data being utilized as a test model.

4.5 Experiment 2 Evaluation

EEG images are used as input data for the CNN-BiLSTM-SimCLR technique, which is put into practice. The resulting images are used as the data input for CNN-BiLSTM-SimCLR, which then generates a time-frequency map that is 20 x 65 x 3. For the purpose of improving the quality of image feature extraction, the CNN presented in this study is built as a two-dimensional convolution of two layers. The image is then placed through the pooling layer after being sent as an input to CNN in the form of padding. Optimization is carried out by the optimizer by the utilization of back-propagation of stochastic gradient descent and cross-entropy. With a learning rate of 0.0001, Adam is still utilized with a range of fifty to two hundred epochs. When the training accuracy is compared to the tests that were conducted utilizing the time series of the source EEG waveform and image, there is an improvement of around 1.51%. The percentage of recognition for an EEG image is 97.02%, whereas the rate of recognition for an EEG signal is 95.51%.

Table 3: Values of Valance and Arousal of CNN-Bi-LSTM-SimCLR (EEG image)

	High Arousal High Valence	High Arousal Low Valence	Low Arousal Low Valence	Low Arousal High Valence
High Arousal High Valence	97.02%	1.42%	1.63%	1.43%
High Arousal Low Valence	3.40%	91.77%	3.51%	1.32%
Low Arousal Low Valence	4.44%	1.43%	92.53%	1.60%
Low Arousal High Valence	3.71%	1.49%	3.10%	91.70%

4.7 Experimental Comparison

This research compares the suggested CNN-BiLSTM-SimCLR integration model, using DEAP dataset. Two CNN-BiLSTM-SimCLR models—one with an EEG signal and the other with an EEG image—are compared. The DEAP dataset is utilized to test the models' efficacy and temporal complexity during training and testing. Every experiment's dataset was split into 7:2:1 patterns before being put to the test on the server. Table 3 and 4 shows the confusion matrix of EEG signal and EEG image.

Table 4: Values of Valance and Arousal of CNN-Bi-LSTM-SimCLR (EEG signal)

	High Arousal High Valence	High Arousal Low Valence	Low Arousal Low Valence	Low Arousal High Valence
High Arousal High Valence	95.51%	1.43%	1.63%	1.43%
High Arousal Low Valence	3.40%	91.77%	3.51%	1.32%
Low Arousal Low Valence	4.44%	1.43%	92.53%	1.60%
Low Arousal High Valence	3.71%	1.49%	3.10%	91.70%

CONCLUSION

It was suggested that the CNN-BiLSTM-SimCLR model be used for this work in order to identify the emotions that are hidden in youngsters. This was done for the goal of classifying emotions. For the purpose of this work, the DEAP dataset was utilized, which results in improved performance. Due to the fact that EEG data for children is extremely difficult to get, this method has the potential to be utilized in the future for the purpose of achieving efficient emotion detection in children. The evaluation was carried out using only valence and arousal data, which can be enhanced by including dominance into the equation in order to obtain dominant emotions that are maintained for a longer period of time. A CWT converted image was used for the experiment, and the results showed that the CWT converted image had a higher recognition rate than the EEG signal. The experiment was carried out in both using deep model.

REFERENCES

[1] Russell, J. A. (1980). A circumplex model of affect. Journal of personality and social psychology, 39(6), 1161.

[2] Mahmud, R., Rahman, M. M., Rassel, M. A., Monayem, F. B., Sayeed, S. J. B., Islam, M. S., & Islam, M. M. (2021). Post-COVID-19 syndrome among symptomatic COVID-19 patients: A prospective cohort study in a tertiary care center of Bangladesh. PloS one, 16(4), e0249644.

[3] Chaudhary, A., Kolhe, S., & Kamal, R. (2016). An improved random forest classifier for multi-class classification. Information Processing in Agriculture, 3(4), 215-222.

[4] Lee, G., Kwon, M., Sri, S. K., & Lee, M. (2014). Emotion recognition based on 3D fuzzy visual and EEG features in movie clips. Neurocomputing, 144, 560-568.

[5] Mohammadi, Z., Frounchi, J., & Amiri, M. (2017). Wavelet-based emotion recognition system using EEG signal. Neural Computing and Applications, 28, 1985-1990.

[6] Bostanov, V. (2004). BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE Transactions on Biomedical engineering, 51(6), 1057-1061.

[7] 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.

[8] Fares, A., Zhong, S., & Jiang, J. (2018, December). Region level bi-directional deep learning framework for eeg-based image classification. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 368-373). IEEE.

[9] Robinson, N., Vinod, A. P., Ang, K. K., Tee, K. P., & Guan, C. T. (2013). EEG-based classification of fast and slow hand movements using wavelet-CSP algorithm. IEEE Transactions on Biomedical Engineering, 60(8), 2123-2132.

[10] Tabar, Y. R., & Halici, U. (2016). A novel deep learning approach for classification of EEG motor imagery signals. Journal of neural engineering, 14(1), 016003.

[11] Cui, F., Wang, R., Ding, W., Chen, Y., & Huang, L. (2022). A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition. Mathematics, 10(4), 582.

[12] Kim, Y., & Choi, A. (2020). EEG-based emotion classification using long short-term memory network with attention mechanism. Sensors, 20(23), 6727.

[13] 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.

[14] Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing, 3(1), 18-31.

[15] Mandhouj, B., Cherni, M. A., & Sayadi, M. (2021). An automated classification of EEG signals based on spectrogram and CNN for epilepsy diagnosis. Analog integrated circuits and signal processing, 108(1), 101-110.

[16] Akter, S., Prodhan, R. A., Pias, T. S., Eisenberg, D., & Fresneda Fernandez, J. (2022). M1M2: deep-learning-based real-time emotion recognition from neural activity. Sensors, 22(21), 8467.

[17] Summers, C., & Dinneen, M. J. (2021, July). Nondeterminism and instability in neural network optimization. In International Conference on Machine Learning (pp. 9913-9922). PMLR.

[18] He, C., Liu, J., Zhu, Y., & Du, W. (2021). Data augmentation for deep neural networks model in EEG classification task: a review. Frontiers in Human Neuroscience, 15, 765525.

[19] Zheng, X., & Chen, W. (2021). An attention-based bi-LSTM method for visual object classification via EEG. Biomedical Signal Processing and Control, 63, 102174.

[20] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Identity mappings in deep residual networks. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14 (pp. 630-645). Springer International Publishing.

[21] Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing, 3(1), 18-31.

[22] Gaertner, M., Sauter, D., Baumgartl, H., Rieg, T., & Buettner, R. (2021, August). Multi-Class Emotion Recognition within the Valence-Arousal-Dominance Space Using EEG. In AMCIS.

[23] Kwon, Y. H., Shin, S. B., & Kim, S. D. (2018). Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system. Sensors, 18(5), 1383.
[24] Huang, Z., Ma, Y., Wang, R., Li, W., & Dai, Y. (2023). A model for EEG-based emotion recognition: CNN-BI-LSTM with attention mechanism. Electronics, 12(14), 3188.
[25] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.
[26] Chakraborty, P., Akter, S., Hasin, B., & Ahmed, S. F. (2021). Electroencephalography based brain controlled grasp and lift (Doctoral dissertation, Brac University).
[27] https://medium.com/@Adewale_Adekunle/grasp-and-lift-eeg-classification-using-cwt-of-electroencephalography-eeg-signals-in-python-a-4269acfeea71
[28] https://pywavelets.readthedocs.io/en/latest/ref/cwt.html