Ensemble Learning Model for EEG Based Emotion Classification

Sanjit Kumar Dash[1]([✉]), Sambit Subhasish Sahu[1], J. Chandrakant Badajena[1], Sweta Dash[1], and Chinmayee Rout[2]

[1] Odisha University of Technology and Research, Bhubaneswar, Odisha, India
`sanjitkumar303@gmail.com`
[2] Ajay Binay Institute of Technology, Cuttack, Odisha, India

Abstract. Emotion and feelings are recently becoming popular concepts in the everyday life. It not only affects human health but also plays an essential role in the decision-making processes. For this reason, emotion classification is one of the important aspects to deal with the problems like mental disorders, suicidal activities and judgmental process. Electroencephalogram (EEG) signal is one of the physiological signals which can be collected from the human brain activity while a person performing various mental and physical task. In this paper, the DEAP dataset has been implemented with the deep learning model for the classification process. In the process of developing models, for the extraction of the important features from the unprocessed EEG signals, Fast Fourier Transformation is used. Three ensemble deep learning models are tested and compared to get the best accuracy result for emotion classification. Furthermore, by the best model we can classify four emotional regions in the valance-arousal plane: HVHA, HVLA, LVHA and LVLA can be classified. The experimental results show that among all the three deep learning models, 1D-CNN-GRU achieved the highest training accuracy of 96.54% as compared to LSTM and 1D-CNN, which is the best model to classify emotions in this context.

Keywords: EEG · Emotion classification · Feature extraction · LSTM · GRU · Fast fourier transformation

1 Introduction

Now a days several industries have adopted the implementation of emotion recognition and classification on their employees. When the human resources fields start gain the emotional state of the employee, they could take the advantages from it to improve the quality of taking decision regarding their organization and it will also help to make judgment regarding employees' work. Emotion is physiological process triggered due to the environmental and personal conditions, which affects the mood, behavior, character and motivations. It also plays an important role in the daily conversation and behavior which affects the decision-making process, personal and professional workflow. Nowadays, the emotional state of a person takes a very crucial position in life, which is associated with their mental and physical health, daily lifestyle, decisions about the future and

many more. The Valence, Arousal, and Dominance (VAD) dimensions are another way to divide up emotion. A measure of pleasantness known as valence can originate from very good (pleasure) or very negative emotions (displeasure). Arousal, on the other hand, is the level of feeling that a situation elicits and can range from positive excitement to calmness (negative). Last but not least, dominance is the degree of control demonstrated in response to a stimulus [1]. The popular Circumplex Model of Affect is depicted in Fig. 1 and categories emotions into two groups: Low valence or negative valence emotions (such as fear, tension, rage, sadness, boredom, etc.) and High valence or pleasant valence emotions (happy, calm, delighted, excited, etc.).
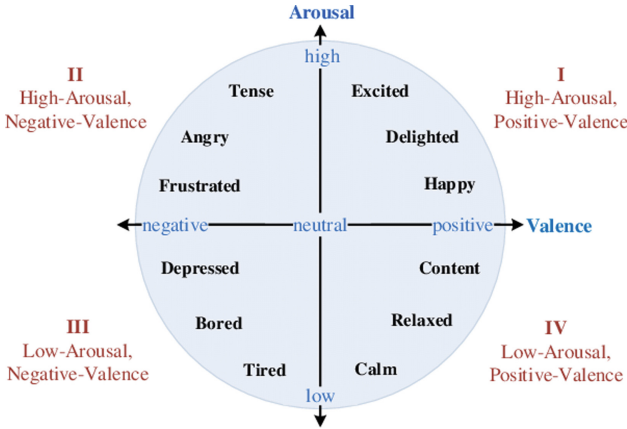

Fig. 1. Graphical representation of circumplex model

In the present scenario, it has become increasingly significant to analyze a person's emotional state which can be apprehended from social networking websites, blogging platforms like twitter, online review sites and so on. However, in order to facilitate emotion classification, the use of EEG data is an emerging field of study. The EEG is (Electroencephalogram) which is used on the scalp surface to record electric current using metallic electrodes and conductive media [2]. By using the electrodes, the different electric waves forming according to different mental states inside the brain could be collected as EEG signals. This can help to have a better understanding of a person's mental state.

Recently, various researchers are using different data sets like AMIGOS, DREAMER, Seed according to their field of experiment, but for emotion recognition DEAP dataset is highly recommended. The EEG dataset is not used for emotion classification but also different fields of BCI like wheel chair, Telemedicine, stock market prediction etc. In different studies different researchers are using different processing and feature extraction methods and different machine learning models. In our paper we are using Fast Fourier Transformation as the feature extraction model before feed into the Deep learning model. CNN and RNN are used for the classification purpose. We also propose 3 models, likely 1D-CNN, LSTM and a combination of two neural networks CNN and RNN as 1D-CNN-GRU.1D-CNN is chosen from CNN and LSTM and GRU

are chosen from RNN. All the models are tested in the training and testing process to check the accuracy and loss value and will be compared.

The remainder of the paper is structured as follows. Section 2 discusses the related works which helped us to get the fundamental knowledge about the research work. Section 3 describes System Model and Methodology. Section 4 describes the Dataset description used in the paper. In Sect. 5, the complete environmental setup and results were elaborated, including the model accuracy and data loss. Section 6 concludes the paper with detailed future work and limitations.

2 Related Works

The C-RNN deep learning hybrid model, which combines CNN & RNN, was introduced by Xiang Li et al. (2016) for the aim of recognizing emotions [3]. The data is pre-processed using continuous wavelet transformation & body production before being taught inside the model. For the arousal and valence dimensions, this test's overall performance is 74.12% and 72.06%, respectively. In their deep learning architecture, they used DEAP dataset [4] for emotion recognition, Alhagry et al. (2017) recommended LSTM. According to this method, the average accuracy for the arousal, valence, and liking classes is 85.65%, 85.45%, and 87.99%, respectively.

Lin et al. performed emotional state classification in CNN using end -to-end learning method and DEAP Dataset [5]. The datasets were turned into six grayscale images that included frequency and time information, and the previously identified characteristics were then taught using the AlexNet version. For arousal and valence, this study gives accuracy of 87.30% and 85.50%, respectively. Li et al. (2017) used the DEAP dataset to perform their emotion recognition task using a hybrid CNN and LSTM RNN (CLRNN) [6]. Already converted into a series of Multidimensional Feature Images is the dataset. With the hybrid neural networks suggested inside the study, every event may be accurately classified as having a common emotion 75.21% of the time. In order to accomplish emotion classification tasks, Acharya et al. (2020) examined the general performance of CNN and LSTM models and performed feature extraction using FFT [7]. The test outcome was excellent for LSTM and CNN model with an accuracy of 88.6% and 87.2% respectively for the liking emotion.

Zhang et al. (2020) discussed different deep learning models like CNN, DNN, LSTM, and combination of CNN and LSTM model along with their applications to the research field of EEG-based emotion classification [8]. This study made substantial use of the DEAP dataset, and many capabilities, including mean, maximum value, standard deviation, minimum value, skewness, and kurtosis, were retrieved from it. The CNN model with 90.12% accuracy and CNN-LSTM model with 94.17% accuracy shows excellent ability to complete this task. Anubhav et al. (2020) investigated the EEG signals with the goal of creating a headgear version for tracking real-time emotions [9]. From the DEAP dataset band energy and frequency domain were recovered, and the accuracies for valence and arousal dimensions were calculated with accuracy of 94.69% 93.13% respectively using LSTM.

A 2D-CNN structure was suggested by Dar et al. (2020) to systematic EEG indications for emotion recognition [10]. The DREAMER & AMIGOS datasets are utilized

for this test and before being input into CNN, each statistic is converted into 2D function matrix (PNG format). The multi-modal emotion reputation system also uses different peripheral physiological markers, such as ECG and GSR, in addition to EEG. Only 76.65% accuracy can be attained using the EEG modality, and multi-modal fusion is required to achieve the overall maximum accuracy of 90.8% and 99.0% for DREAMER and AMIGOS dataset respectively.

3   System Model and Methodology

Figure 2 shows the system model for emotion classification. There are 4 phases in the process, initially collected DEAP dataset with proper approval which is publicly available. Then it will be passed through the Feature Extraction process in the second phase, which extract the main features from the unprocessed EEG data. Thirdly the data will split into Train and Test data, training data will be passed through the three different models; 1D-CNN, LSTM,1D-CNN-GRU to check testing accuracy and all other parameters. In the end phase the best model will be implemented as the classification model for the classification of four emotional regions from valance arousal plane.
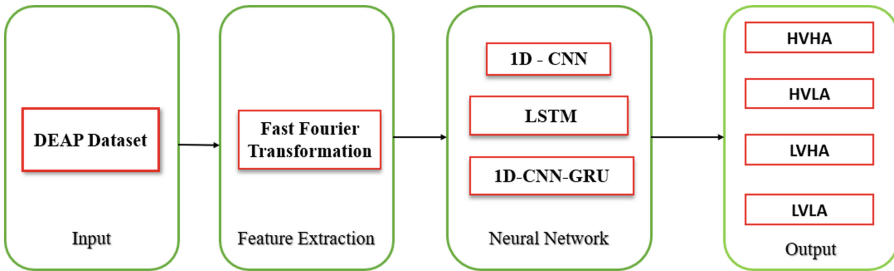

Fig. 2. System model

3.1   Feature Extraction

For feature extraction, Fast Fourier Transformation (FFT) was used which is best in all tradition feature extraction models to extract the important features for the classification purpose, which reduces the number of computation data needed for an experiment from actual data of size N and the final result was achieved (58560,70) dimension from (40,40,8064) which allows for faster training and higher accuracy.

The extracted features contain five frequency bands: Delta, Theta, Alpha, Beta and Gamma with frequency 1–4, 4–8, 8–14, 14–31, and 31–50 Hz respectively. The PyEEG python package was used to extract these five out of 70 characteristics in total. The signal domain is transformed from time to frequency on the x-axis using FFT. This is based on the idea of discrete fourier transform (DFT) on time series data. Calculating the DFT coefficients in an iterative approach will reduce both computational time and complexity. It also helps to reduce round-off errors in computations.

3.2 Deep Learning Model Implementation

In this section, three deep learning models have been implemented: 1D-CNN (One Dimensional Convolutional Neural Network), LSTM (Long Short-Term Memory), and a hybrid model of 1D-CNN and GRU (Gated Recurrent Unit). Preprocessed version of the DEAP dataset used in the models. The models were trained to categorise each emotion individually using different train-test splits, such as arousal, valence, dominance, and liking. The following is a description of model implementations using Keras (Chollet (2015) [11]):

A. 1D-CNN Model

1D-CNN is used to extract the significant features from the DEAP dataset. CNN works good on the time series data which are the 1D signals. In Conv1D the kernel slides along one dimension only. This is one of the key justifications for using 1D-CNN in our research. As shown in Fig. 3, out of the 10 classes, we employed three conv1D, three dense layers that were completely coupled, and one dense layer with SoftMax activation. The first convolution layer employs the Rectified linear unit (ReLU) as its activation function and has 164 filters and a kernel of size 3. After hyperparameter tuning, optimization with Grid Search and manual adjustments, the number and size of filters are determined. A shape of (70,1) with the same padding and stride of one is feed into Conv1D's first layer. To reduce network overfitting, dropout on the dense layer outputs is implemented with 0.2 dropout probability. Following this dense layer of 21 ReLU-activated neurons with a 0.2 dropout probability is a layer of 42 Tanh-activated neurons with a 0.2 dropout probability. At the end Dense layer of 10 neurons with a SoftMax activation function produces the network's final output.
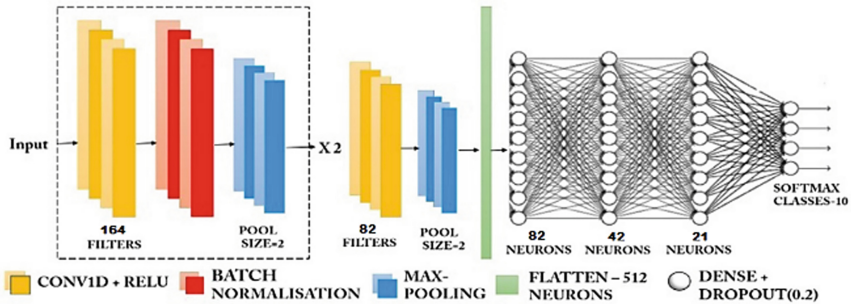

Fig. 3. 1D - CNN architecture

B. LSTM Model

Long Short-Term Memory Networks (LSTMs) is one type of recurrent neural network (RNN), first introduced in 1997 by Hochreiter and Schmidhuber [12]. It solves the problem of short-term memory. It has a built-in gate that can recognize which information in a sequence has to be kept and which information can be discarded.
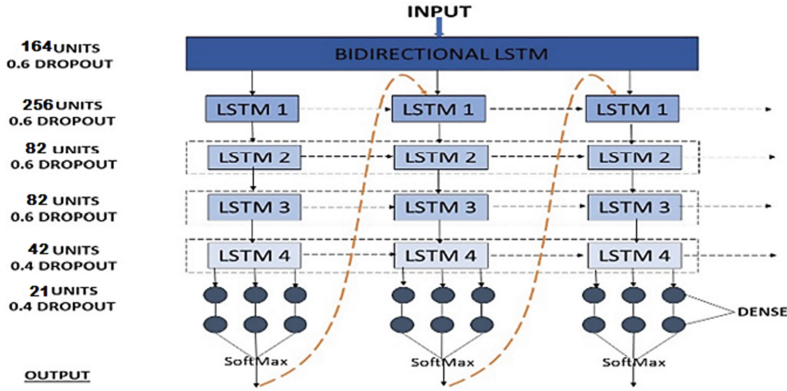
Fig. 4. LSTM architecture

Figure 4 shows the model architecture of LSTM, which consists of two dense layers, four LSTM layers, and one bi-directional layer. The initial bi-directional LSTM layer contains 164 units. It involves adding a second LSTM layer on top of the first one in the network. The first receives the input sequence, and the second receives a reverse copy to the next layer. The dropout layer, which has a probability of 0.6, comes next. Inputs are randomly set to 0 which helps to prevent overfitting. The following layer is a 256-neuron LSTM layer, followed by a 0.6 dropout layer. Two LSTM layers with 82 neurons each make up the next four layers, which are then followed by a dropout layer. There were 0.6 and 0.4 percent dropout rates, respectively. 42 neurons make up the final LSTM layer with a dropout layer of 0.4 following that 21 units of dense layers is applied. ReLU is the activation method employed in this case. The SoftMax activation function is then applied to dense layer of 10 classes, which gives a multiclass probability distribution. Using argmax find out the class output is done after knowing the probability of all the classes.

C. 1D-CNN-GRU Model

This variant is a hybrid model of the 1D-CNN and GRU of deep learning architectures. The network's first input size is 256 units, followed by two seconds of the time-stamped signal with 256 data points, 128 units of convolutional filters, and a kernel size of 3. Rectified linear unit (ReLU) activation function is used in the first convolution layer. The precise number and size of filters are identified after extensive hyperparameter optimization using Grid Search and manual adjustments. Conv1D's first layer receives as its input a shape of the form (70,1) with the same padding. The outputs of the first layer are normalized using a batch normalized layer, which has a mean value zero and a standard deviation value of one. The input is down sampled in the following layer, dropout of 0.2 after a Max pooling 1D layer with a pool size of 2 and the second convolutional layer, which is same as the first one. The implementation of GRU comes after the convolutional layers with an input length of 256 units and 32 units. At the end of every GRU layers a dropout layer of 0.2 has been set. A flattening operation is implemented to send the features into the

1D feature vector prior to dense layer. The dense layer is set to 32 units and ReLU is used as the activation function. To represent the 4 labels of classification 4 units of dense layer has been set. The activation function for the dense layer is SoftMax. In this model 379,594 units of trainable parameter used. The details model architecture of 1D-CNN-GRU is given in Fig. 5.
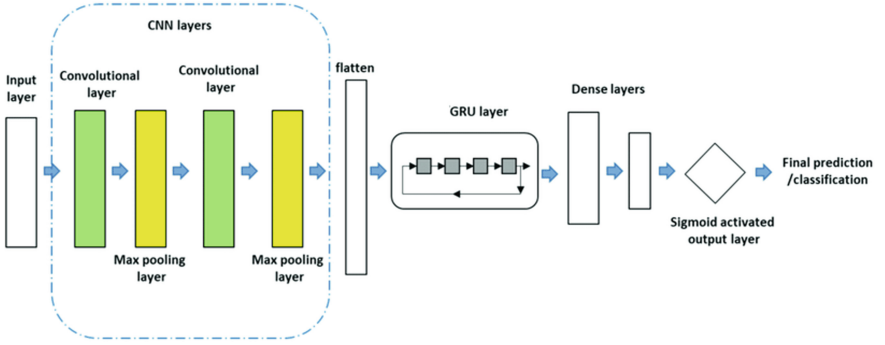

Fig. 5. 1D - CNN?+?GRU architecture

4 Dataset Description

The DEAP dataset [13] used in the experiment is publicly available for researchers for their experiments. This data set contains both EEG and EMG signals. To collect the data 32 participants as engaged. In the dataset the physiological recordings and participant evaluations of 32 individuals (s01-s32) are covered. These physiological clips are already in BioSemi.bdf format and have not been processed. For each of them, 40 films were presented which makes 40 channels in total. Depending on the rating, the emotion is either stronger or weaker; the stronger the emotion, the higher the rating. Table 1 contains each subject's information, which contains two arrays: data and label, which contains the array shape and the content of each file.

Table 1. Pre-processed dataset description

Array name	Array shape	Array contents
Data	40 ×40 ×8064	video/trial * Channel * Data
Labels	40 ×4	video/trial * label (valence, arousal, dominance, liking)

The data was filtered by a band-pass filter with a bandwidth of 4–45 Hz and down sampled by 128 Hz. In addition to these, the collection includes listings and links to YouTube music videos. Questions to ask before testing are contained in the participant questionnaire file, each trial or video was given a rating on a scale of 1 to 9. The dataset is

divided into four classes, which is subsequently labelled as: High-Valence Low-Arousal (HVLA), Low-Valence High-Arousal (LVHA), and Low-Valence Low-Arousal (LVLA) are four different types of arousal The Table 2 contains all the four classes and two labels, the threshold value for this classification is 5. If the value is greater than 5, then it will be classified as high and low if it is less than 5.

Table 2. Label classification

Label	HVHA	HVLA	LVHA	LVLA
Valence	>5	>5	<=5	<=5
Arousal	>5	?<=5	>5	<=5

In our paper we have selected 14 channels AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P7, P8, O1, O2 and 5 bands 4, 8, 12, 16, 25, 45 to reduce the computational cost as well as for the better result for emotion classification [14]. This entire channel selection process is based on the significance of the brain regions that make up emotional states. When we divided the label into the four categories of HVHA, HVLA, LVHA, and LVLA, the paper was considered to be complete. But for now, as a result, models will be compared on the basis of an increase in training accuracy while decreasing validation loss for the selection of a better classification model.

5   Experimental Setup and Results

Google Collaboratory was used to compute 1D – CNN, LSTM, GRU classifiers because, it uses Jupyter notebook service that needs no installation and gives unrestricted access to computing tools, such as GPUs. Python is the python version (3.7.13). The current version of TensorFlow is 2.8.0. Pandas: 1.3.5, numpy: 1.21.6, sklearn: 1.0.2, plotly: 5.5.0. A laptop is used to run the code. Nvidia K80s, T4s, P4s, and P100s are common GPUs seen in CoLab. To get an amazing result, hyperparameter should be tuned adequately. Conv1D layers are utilized in the CNN architecture because they are best suited for data in time series. Both maximum and average pooling were used, however maximum pooling produced better results, as predicted by the literature. For the CNN, LSTM, and 1D-CNN-GRU architectures, the finalized epoch size is 200 and batch size of 256. The models are trained on 80–20 train–test splits and 10-fold cross validation is also employed to determine the best metrics-accuracy. To update the weights during back-propagation, used Adam as the optimizer and category cross entropy as the loss function. In both cases, SoftMax performs the activation of the last layer. Separate decisions were made for each of the three models, including the number of layers, hidden layers, filter size, number of filters, pool size for the CNN model, and hidden neurons, dropout rates, and layers for the LSTM and GRU models separately. Both grid search and manual testing are used to finalize everything.

Overall performance of the three models can be elaborated and analyzed on the basic of accuracy and loss value. Table 3 contains test accuracy and test loss of each model. It can be inferred that, the 1D-CNN-GRU model architecture provide the best test accuracy of 96.54%, 41.3% of test loss with an 80–20 train test split as compared to the LSTM model architecture's accuracy of 89.6%, test loss of 39.6% with an 80–20 train test split and the 1D-CNN model architecture's accuracy of 90.65%, test loss of 42.2% with an 80–20 train test split. All these three experiments were designed to test the overall performance of each model for classification of emotion.

Table 3. Test accuracy and test loss comparison

Classifier	Test accuracy	Test loss
1D - CNN	90.65%	42.2%
LSTM	89.6%	39.6%
1D CNN?+?GRU	96.54%	11.4%

Above discussed models performed exceptionally well at generalizing findings since they classified each emotion with an accuracy rate of more than 80%. The loss function is categorical cross entropy. Overfitting on training data was avoided by the use of dropout layers, which also enhanced model results. Batch normalization layers also significantly affect model accuracy. By altering the number of epochs, data units and input layers the accuracy affects. When we implemented the GRU architecture we got a very less accuracy, but after implementing the hybrid model of 1D-CNN and GRU, it crossed all the levels.
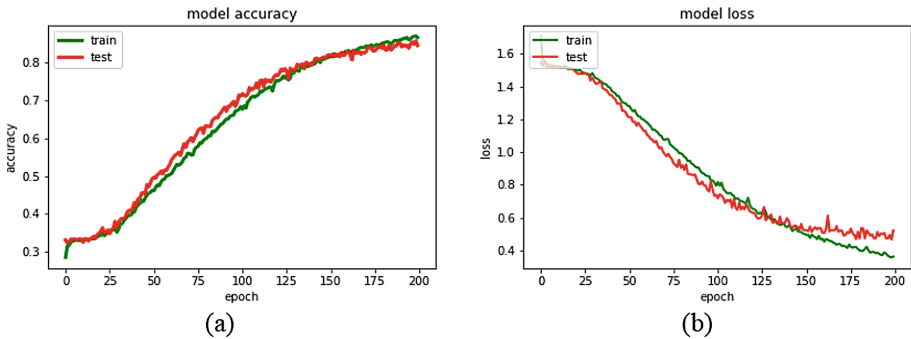
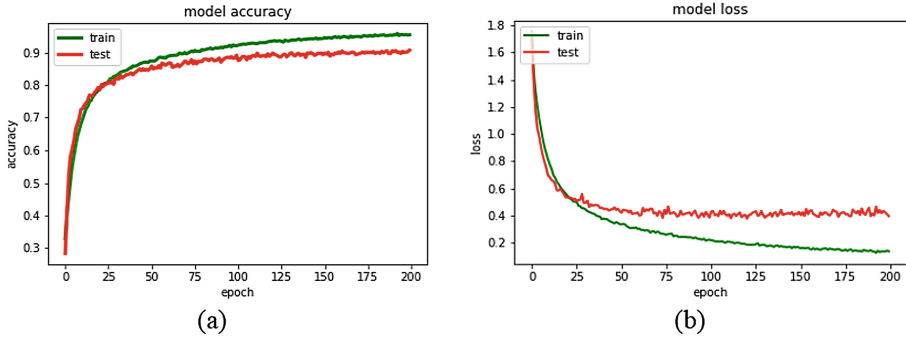**Fig. 6.** (a) Model accuracy and (b) Model loss of LSTM

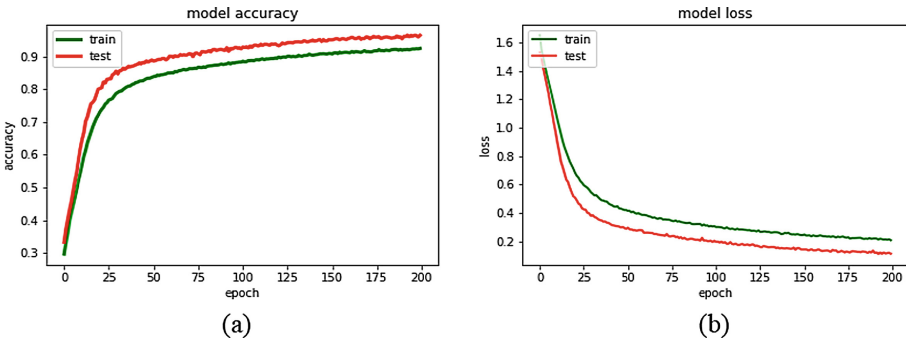**Fig. 7.** (a) Model accuracy and (b) Model loss of 1D-CNN

Fig. 8. (a) Model accuracy and (b) Model loss of 1D-CNN-GRU

Figure 6, Fig. 7, Fig. 8 shows model accuracy and model loss of LSTM, 1D-CNN, 1D-CNN-LSTM respectively. In model accuracy curves the train and test curve goes up award direction with increase of epoch, where as in model loss figure it goes down with increase of epoch. Curves were found to have a minor variation, The 1D-CNN model starts learning earlier than the LSTM model and where as it took the CNN model about 50 epochs to reach a stable point, the LSTM model took about 130 epochs. No overfitting took place, and the training process came to an end after 165 epochs. Each training period takes between 16 and 330 ms. Figure 9 shows test model of LSTM, 1D-CNN and 1D-CNN-GRU, which makes a curve between test loss and test accuracy, when the accuracy increases the value loss decreases by the increase of epoch.
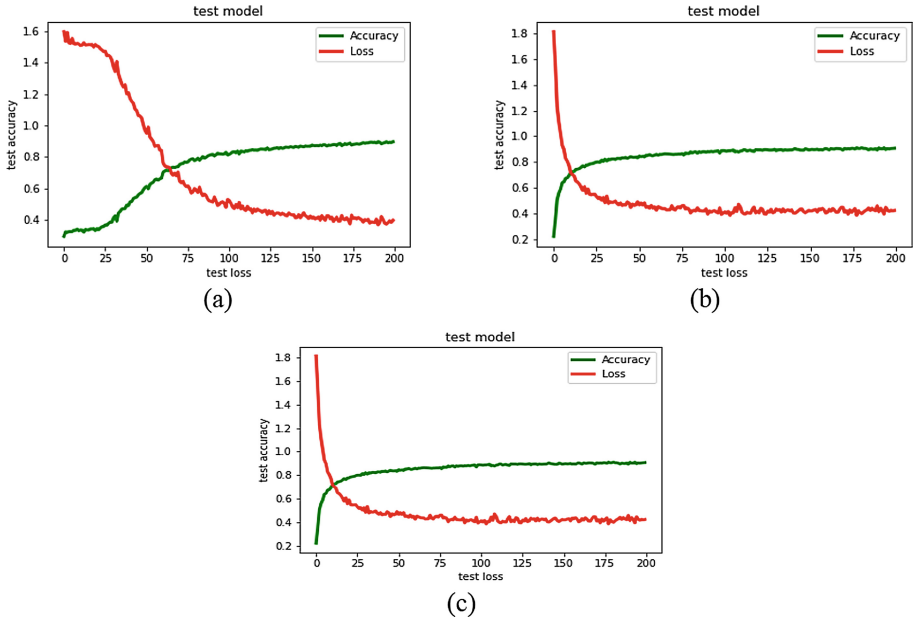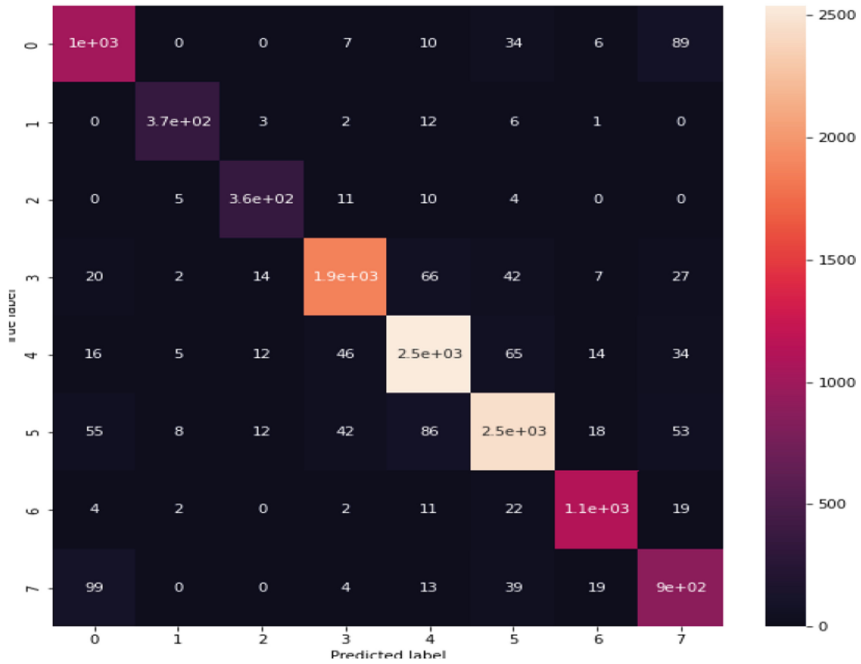
Fig. 9. Test model of (a) LSTM, (b) 1D-CNN, (c) 1D-CNN-GRU

Dropout and batch normalisation layers have a big impact on the model's accuracy. Additionally, we constructed a confusion matrix in Fig. 10 to explore the discrepancy between the predicted and actual value. Each model's F1-Score and recall value are close to one, which shows that this model has good quality.
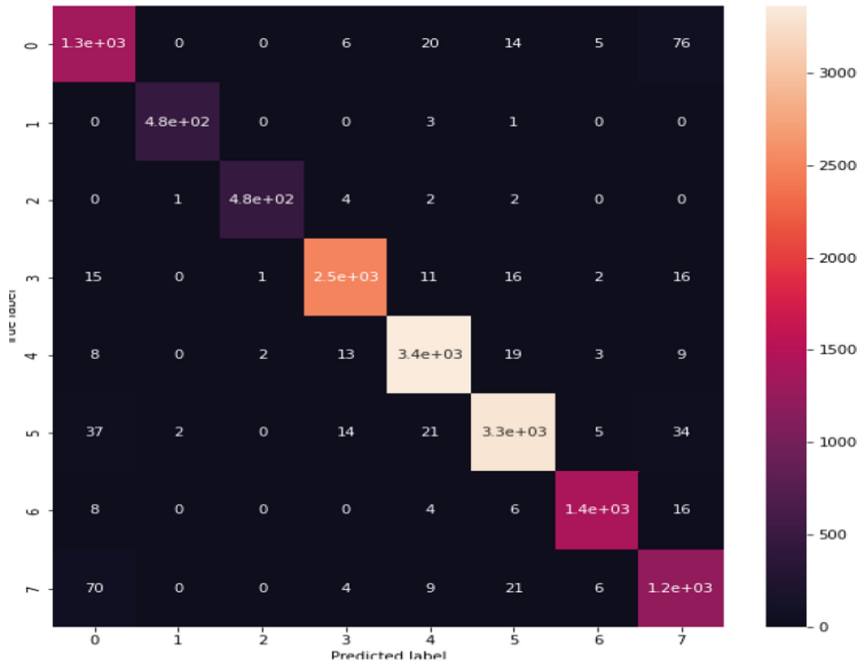
a.   **LSTM**

b. **1D-CNN**

Fig. 10. Confusion matrix of LSTM, 1D-CNN, 1D-CNN-GRU

c. 1D-CNN-GRU

Fig. 10. (*continued*)

6 Conclusion

In this paper we have discussed the simplest feature extraction method with the best classification model for emotion classification using DEAP dataset. In this paper, we have described three deep learning models: 1D-CNN, LSTM, and 1D-CNN-GRU. As compared to normal feature extraction strategies, FFT increased accuracy and extracted crucial characteristics. The 1D-CNN design has proven to be best model to extract EEG signal features. The 1D-CNN architecture has a classification accuracy of 90.8% which is somewhat better than the 89.2% classification accuracy of the LSTM model. From all the three models the 1D – CNN – GRU gives the best accuracy of 96.54%.

Bi-LSTMs were able to preserve data from the past as well as the future, which contributed to increase the accuracy in the LSTM model. Even though the emotion classification test showed that our model worked remarkably well, we still wish to assess it further in additional datasets like DREAMER and AMIGOS and enhance our model. More particular, while this version no longer undergoes testing on various people, it was best trained using the DEAP dataset only.

In future we would try to work on quick output of the EEG data processing in real-time online analysis systems, which limits calculation time. We will concentrate on a multi-task cascaded-hybrid LSTM and CNN model in the future, which will combine their features and improve the efficacy of the emotion classification model. The system to detection of emotion can improve human experiences by minimizing the gap between

computational technology and human emotions and allowing computers, BCI system and robots to receive emotional feedback in real-time. Even with the help of this technology, therapists can more thoroughly evaluate their patients and learn how to spot depression early and prevent it before any outward situation occurs.

References

1. Alarcao, S.M., Fonseca, M.J.: Emotions recognition using EEG signals: a survey. IEEE Trans. Affect. Comput. **10**(3), 374–393 (2017)
2. Teplan, M.: Fundamentals of EEG measurement. Measur. Sci. Rev. **2**(2), 1–11 (2002)
3. Li, X., Song, D., Zhang, P., Yu, G., Hou, Y., Hu, B.: Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In: 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 352–359. IEEE (2016)
4. Alhagry, S., Fahmy, A.A., El-Khoribi, R.A.: Emotion recognition based on EEG using LSTM recurrent neural network. Int. J. Adv. Comput. Sci. Appl. **8**(10), 355–358 (2017)
5. Lin, W., Li, C., Sun, S.: Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal. In: Zhao, Y., Kong, X., Taubman, D. (eds.) ICIG 2017. LNCS, vol. 10667, pp. 385–394. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-71589-6_33
6. Li, Y., Huang, J., Zhou, H., Zhong, N.: Human emotion recognition with electroencephalographic multidimensional features by hybrid deep neural networks. Appl. Sci. **7**(10), 1060 (2017)
7. Acharya, D., et al.: Multi-class emotion classification using EEG signals. In: Garg, D., Wong, K., Sarangapani, J., Gupta, S.K. (eds.) IACC 2020. CCIS, vol. 1367, pp. 474–491. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-0401-0_38
8. Zhang, Y., et al.: An investigation of deep learning models for EEG-based emotion recognition. Front. Neurosci. **14**, 622759 (2020)
9. Nath, D., Singh, M., Sethia, D., Kalra, D., Indu, S.: An efficient approach to EEG-based emotion recognition using lstm network. In: 2020 16th IEEE International Colloquium on Signal Processing and its Applications (CSPA), pp. 88–92. IEEE (2020)
10. Dar, M.N., Akram, M.U., Khawaja, S.G., Pujari, A.N.: CNN and LSTM-based emotion charting using physiological signals. Sensors **20**(16), 4551 (2020)
11. Chollet, F. (2017). Keras (2015)
12. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
13. https://www.eecs.qmul.ac.uk/mmv/datasets/deap/readme.html
14. Al-Qazzaz, N.K., Sabir, M.K., Ali, S., Ahmad, S.A., Grammer, K.: Effective EEG channels for emotion identification over the brain regions using differential evolution algorithm. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4703–4706. IEEE (2019)