A Fuzzy Ensemble-Based Deep learning Model for EEG-Based Emotion Recognition

Trishita Dhara[1] · Pawan Kumar Singh[1,2] · Mufti Mahmud[3,4,5]

Abstract

Emotion recognition from EEG signals is a major field of research in cognitive computing. The major challenges involved in the task are extracting meaningful features from the signals and building an accurate model. This paper proposes a fuzzy ensemble-based deep learning approach to classify emotions from EEG-based models. Three individual deep learning models have been trained and combined using a fuzzy rank-based approach implemented using the Gompertz function. The model has been tested on two benchmark datasets: DEAP and AMIGOS. Our model has achieved 90.84% and 91.65% accuracies on the valence and arousal dimensions, respectively, for the DEAP dataset. The model also achieved accuracy above 95% on the DEAP dataset for the subject-dependent approach. On the AMIGOS dataset, our model has achieved state-of-the-art accuracies of 98.73% and 98.39% on the valence and arousal dimensions, respectively. The model achieved accuracies of 99.38% and 98.66% for the subject-independent and subject-dependent cases, respectively. The proposed model has provided satisfactory results on both DEAP and AMIGOS datasets and in both subject-dependent and subject-independent setups. Hence, we can conclude that this is a robust model for emotion recognition from EEG signals.

Keywords Fuzzy ensemble · Electroencephalogram · Emotion recognition · Gompertz function · DEAP · AMIGOS

Introduction

Over the last two centuries, there has been a lot of biological sciences research driven by the demand for new healthcare treatments and ongoing efforts to understand the biological underpinnings of illnesses [1, 2]. Recent developments in the life sciences have made it possible to investigate biological systems holistically and to gain access to the molecular minutiae of living things like never before. Nevertheless, it is extremely difficult to draw meaningful conclusions from such data due to the presence of inherent complexity of biological systems along with larger dimension, variety, and noise present in them [3]. As a result, new equipment that are accurate, dependable, durable, and capable of processing large amounts of biological data are needed. This has inspired many researchers in the life and computer sciences to adopt an interdisciplinary strategy to clarify the workings and dynamics of living things, with notable advancements in biological and biomedical research [4]. As a result, numerous artificial intelligence approaches, particularly machine learning, have been put forth over time to make it easier to identify, categorise, and forecast patterns in biological data [5].

Deep learning, a subset of machine learning, extracts more meaningful and complementary features from a large training dataset primarily without human intervention. Learning data representations by introducing increasingly complex degrees of abstraction is the fundamental idea behind deep learning. Nearly all levels work on the principle that more abstract representations at a high level are

✉ Mufti Mahmud
muftimahmud@gmail.com

Trishita Dhara
trishitadhara123@gmail.com

Pawan Kumar Singh
pawansingh.ju@gmail.com

1 Department of Information Technology, Jadavpur University, Jadavpur University Second Campus, Plot No. 8, Salt Lake Bypass, LB Block, Sector III, Salt Lake City, Kolkata-700106, West Bengal, India

2 School of Science and Technology, Nottingham Trent University, Clifton, Nottingham NG11 8NS, UK

3 Department of Computer Science, Nottingham Trent University, Clifton, Nottingham NG11 8NS, UK

4 Medical Technologies Innovation Facility, Nottingham Trent University, Nottingham NG11 8 NS, UK

5 Computing and Informatics Research Centre, Nottingham Trent University, Nottingham NG11 8 NS, UK
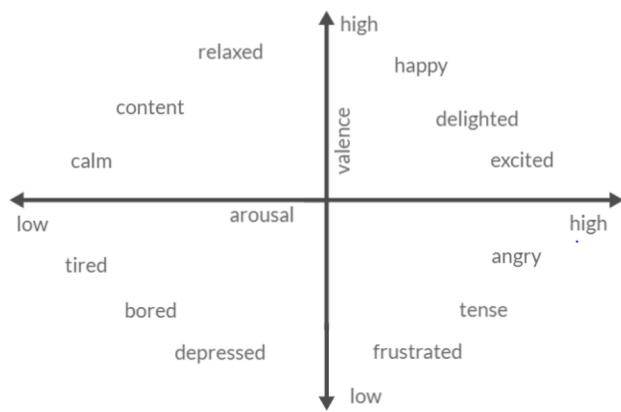
Fig. 1 Two-dimensional valence-arousal space

defined with respect to lesser abstract representations seen at low levels [6]. Because it enables a system to understand and learn complicated representations straight from the raw data, this kind of hierarchical learning process is particularly potent and useful in a wide range of disciplines [7, 8].

Emotions play a vital role in our daily lives. Hence, emotion recognition is an important part of human–computer interactions. Emotion can be recognised from speech, facial expressions, and physiological signals. However, emotion recognition from physiological signals is most reliable as humans can deliberately conceal or fake their emotional expression through speech or gesture [9, 10]. Hence, emotion recognition from EEG signals has drawn the attention of many researchers over the past decade.

Our study has classified two dimensions of emotion: valence and arousal. These are the two most important parameters for describing human emotions. Valence indicates whether an emotion is positive or negative. Arousal indicates the level of arousal developed in our bodies due to emotion. Figure 1 categorises some common emotions according to their valence and arousal values.

Electroencephalogram (EEG) is a medical procedure during which electrodes with thin wires are pasted on our scalp and detect voltage fluctuations in brain neurons. These EEG signals capture the electrical activities of our brain and can be used to detect human emotions [10]. As mentioned earlier, emotion recognition from physiological signals is more reliable and accurate. Moreover, EEG can continuously detect changes in human emotions. Hence, it can be used for patient monitoring [11, 12]. In our study, we have analysed EEG signals to predict

human emotions accurately. The steps involved in the process of EEG-based emotion recognition are represented in Fig. 2.

There are two ways to perform EEG-based emotion recognition: subject-dependent and subject-independent approaches [13]. The subject-dependent approach generally gives higher accuracy than the independent approach, but the former requires the model to be trained for each subject [14]. In our study, we have implemented both subject-dependent and independent approaches.

The main contributions of the work can thus be summarised as follows:

- This study proposes a fuzzy ensemble approach for emotion recognition from EEG signals. This fuzzy-based approach has been applied for the first time in this domain, achieving remarkable accuracy.
- The proposed model has been tested on two standard benchmark datasets: DEAP [15] and AMIGOS [16]. Both are benchmark datasets in the realm of EEG-based emotion recognition. It is to be noted that AMIGOS is the largest existing dataset in this domain.
- The model also gives a satisfactory performance in both subject-dependent and subject-independent approaches. Hence, the model is flexible and can be applied to either approach as the situation demands.
- The proposed model also shows impressive classification accuracy when tested for both valence and arousal dimensions.

The rest of the paper has been organised as follows: the literature analysis related to EEG-based emotion recognition has been discussed in the "Related Works" section, whereas the benchmark datasets used for our experiment are described in the "Dataset" section. The proposed fuzzy ensemble-based deep learning methodology is presented in the "Methodology" section, whereas the overall results obtained by the proposed model are explained in the "Results" section. Lastly, the conclusion followed by some future scope is mentioned in the "Conclusion" section.

Related Works

Many research works have been conducted on EEG-based emotion recognition over the past years. Initially, supervised machine learning algorithms were implemented. But later,

Fig. 2 Pipeline for emotion recognition from EEG signals

the focus shifted to mostly deep learning-based approaches to obtain state-of-the-art accuracies.

Yoon and Chung [17] used fast Fourier transform features and implemented a Bayesian function and perception convergence algorithm. They achieved an accuracy of 70.9% for the valence dimension on the DEAP dataset [15]. Dabas et al. [18] implemented machine learning models like support vector machine (SVM) and Naïve Bayes and obtained an accuracy of 58.90% and 78.06% on the DEAP dataset [15].

Liu et al. [19] performed emotion recognition on the DEAP dataset. They used different types of features like time-domain features such as mean and standard deviation, frequency-domain features like power spectral density (PSD), and time–frequency domain features like discrete wavelet transform (DWT). They used the random forest and K-nearest neighbour (KNN) model and achieved an accuracy of 66.17% for arousal. You and Liu [20] extracted time-domain features from a 5-s slice of EEG signals and implemented an autoencoder neural network. They achieved an accuracy greater than 80% on the DEAP dataset. Salama et al. [21] implemented a 3D-convolutional neural network (CNN) model to recognise emotion in the DEAP dataset. They achieved an accuracy of 87.44% and 88.49% for valence and arousal dimensions, respectively.

By implementing shallow depth-wise parallel CNN, Zhan et al. [22] achieved an accuracy of 84.07% and 82.95% on arousal and valence, respectively, on the DEAP dataset. Allghary et al. [23] achieved an accuracy of 85.65%, 85.45%, and 87.99% on arousal, valence, and liking, respectively. They proposed an LSTM model for emotion classification on the DEAP dataset. Wichakam et al. [24] used the band power feature and SVM model to classify emotion on the DEAP dataset. They reached an accuracy of 64.9% for valence and 66.8% for liking. They selected 10 channels for the recognition task and demonstrated that increasing the number of channels to 32 does not improve the performance. Parui et al. [25] proposed the XGBoost classifier model. They extracted several features from the EEG signals from the DEAP dataset and optimised them. The model achieved accuracies of 75.97%, 74.206%, 75.234%, and 76.424% for the four dimensions, respectively.

Aggarwal et al. [26] combined XGBoost and LightGBM models for emotion recognition on the DEAP dataset. They achieved an accuracy of 77.1% for the valence dimension. Bagzir et al. [27] decomposed EEG signals into gamma, beta, alpha, and theta bands by applying DWT to extract the frequency spectrum characteristics of each frequency band. Then, a KNN, an SVM, and an artificial neural network were used for classification. The model achieved accuracies of 91.1% and 91.3% on valence and arousal, respectively.

Few recent works have developed robust models and tested them on multiple datasets. Siddharth et al. [28] developed a multi-modal emotion recognition model by fusing the features from different modalities and then classifying them. They have evaluated their model on multiple datasets like DEAP [15], AMIGOS [16], MAHNOB-HCI [29], and DREAMER [30] datasets. Ante Topic et al. [31] used holographic feature maps; they also selected optimal channels by implementing ReliefF and neighbourhood component analysis (NCA). The holographic feature maps were fed as input to CNN, and finally, the output of CNN was passed to the SVM classifier. They have evaluated their model on DEAP [15], AMIGOS [16], SEED [32], and DREAMER [30] datasets. They have got highest accuracies on the DREAMER dataset, which is 90.76%, 92.92%, and 92.97%, respectively, on valence, arousal, and dominance dimensions. Singh et al. [33] extracted spectrogram features from 14 EEG channels and used a CNN model to classify emotions on the AMIGOS dataset. The model achieved 87.5% and 75% accuracy on valence and arousal dimensions. Garg et al. [34] used FFT and wavelet transform to extract features from EEG signals and implemented a deep neural network (DNN) model for emotion recognition. On the AMIGOS dataset, the method achieved 85.47%, 81.87%, 84.04%, and 86.63% for valence, arousal, dominance, and liking. Zhao et al. [35] used a 3D CNN model to recognise emotions from EEG signals on DEAP and AMIGOS datasets. They evaluated two-class (low/high arousal, low/high valence) and four-class (HAHV, HALV, LAHV, LALV) classifications. The model achieved 96.61%, 96.43% for the two-class classification task; 93.53% for the four-class classification on the DEAP dataset; 97.52%, 96.96% for the two-class classification task; and 95.86% for the four-class classification task on the AMIGOS dataset.

Motivation and Research Gap

Over the past years, the research works have notably improved performance/accuracy for the task of EEG-based emotion recognition [36, 37]. However, most of the existing works have a few common drawbacks. Firstly, most models have been tested on a particular dataset only. Such models can be data-dependent and may not be robust. Secondly, almost all existing works have been performed in either subject-independent or subject-dependent conditions. There is a lack of research work that focuses on both approaches simultaneously. Thirdly, most of the work done for EEG-based emotion recognition focuses on machine learning-based models. Some of the researchers have developed customized deep learning models for solving this problem. However, these models are simplistic such that they are unable to deal with the complexity of the problem resulting in low classification accuracy. Keeping in mind the above mentioned gap, this work motivates to propose a fuzzy ensemble-based deep learning model (by ensembling

three different complementary deep learning models such as a hybrid of CNN and LSTM models, a hybrid of CNN and GRU models, and ID-CNN model) for solving EEG-based emotion recognition problem. It is to be noted that this fuzzy-based approach has been applied for the first time in this domain. The work also attains satisfactory classification results for AMIGOS dataset, a benchmark largest dataset in this domain. The most highlighting aspect of this work is that the proposed work achieves impressive results for both subject-independent as well as subject-dependent cases. Additionally, the work has also been tested on both valence and arousal dimensions.

Datasets

The proposed model has been implemented on two datasets — DEAP [15] and AMIGOS [16], both developed by the Queen Mary University of London, UK.

The DEAP dataset consists of EEG signals of 32 subjects. Each subject watched 40 music videos of 1-min duration containing different emotions, while their EEG was recorded. Each subject also rated the level of valence, arousal, liking, and dominance for each video. The data is stored in 32 files, one for each participant. Each file consists of a total of 40 channels, out of which 32 channels contain EEG data. The data is pre-processed and available in both Python (.dat) and MATLAB (.mat) formats. Our experiment is conducted using the .dat files. The dataset also contains frontal face video recordings for 22 subjects which can be used for the multi-modal emotion recognition task.

The AMIGOS dataset can be used for a multi-modal study of mood and affective responses of individuals to different videos. The dataset consists of EEG, electrocardiogram (ECG), and galvanic skin response (GSR) recordings of 40 participants while they watched 16 videos. The experiment was performed both individually and in groups. Each subject assessed different parameters like valence, arousal, and familiarity. Video recordings of frontal entire body and depth are also available. The EEG signals are pre-processed and available in both Python and MATLAB format. For our experiment, we have used the python files.

Methodology

Datasets

This work has utilised two standard benchmark datasets to validate the proposed method. These datasets are described in the following subsections.

DEAP Dataset

The dataset contains EEG recordings of 32 channels, of which 14 channels were selected for our experiment. The selected channels are Fp1, AF3, F3, F7, T7, P7, Pz, O2, P4, P8, CP6, FC6, AF4, and Fz [38]. Valence and arousal dimensions have been considered for our study. The labels of both the dimensions had continuous values between '1' and '9'. The labels for each dimension were categorised into two classes — high and low. We used to label '1' for high and label '0' for low. So, any value below '5' has been labelled '0', and any value above '5' has been assigned label '1'.

Pre-processing Fast Fourier transform (FFT) has been implemented to extract features from the EEG signals. Several studies have shown that FFT gives better performance than traditional feature extraction methods [39, 40]. FFT transforms a signal from the time domain to the frequency domain. Since there is a probability of quick detection of emotions, the raw signals were segmented into 2-s temporal windows with 1-s overlap. The FFT is computed for each such segment of raw data. The frequency bands considered in the present work are 4–8 kHz, 8–12 kHz, 12–16 kHz, 16–25 kHz, and 25–45 kHz.

AMIGOS Dataset

The dataset contains EEG recordings of 40 subjects, out of which some subjects did not participate in both short and long video experiments. To maintain consistency in the data, we have chosen only those subjects who have watched both long and short videos. Out of 17 channels, we have selected 14 channels that contained EEG data– AF3, F3, F7, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4.

Pre-processing The pre-processing is done similarly as the DEAP dataset. Raw data is broken into segments, and the FFT features are extracted with the same frequency bands used for the DEAP dataset. The valence and arousal labels are classified into two classes. If the value is greater than '5', the assigned label is '1'; else '0'.

Candidate Models

Hybrid of CNN and LSTM Models

Our first candidate model is a hybrid of CNN and LSTM models. The CNN layers extract spatial features from the signals, and the LSTM part extracts temporal features.

Fully connected layers follow the LSTM layers, and then the final prediction is obtained. All the model hyperparameters have been summarised in Table 1.

Hybrid of CNN and GRU Models

The second model implemented, in this work, is similar to the first one; however, GRU layers have been used in place of LSTM layers. The GRU has a simplified structure than the LSTM model, and it usually takes lesser time to train. All the model parameters have been summarised in Table 2.

1D-CNN Model

The third model consists of 1D-CNN layers followed by fully connected layers. Numerous researches have proven that CNN is very effective in feature extraction tasks from images and other data. Hence, we have chosen CNN as our feature extraction, followed by dense layers to get the

predictions. All the model parameters have been summarised in Table 3.

Proposed Model

This paper proposes an ensemble learning approach for emotion detection from EEG signals. We have trained three individual models and combined them using the fuzzy ensemble technique and max voting ensemble. Each model has been trained individually for 50 epochs for both datasets before combining them. For all the models, adam optimiser has been used with a learning rate of 0.001. For the subject-independent approach, the entire data is split into train, test, and validation set in the ratio of 60:20:20. We have also tried with other ratios, but this was the optimal split; hence, the 60:20:20 ratio has been finalised. For the subject-dependent approach, the model is trained and tested separately for each subject, and then we take the average of the results. For each subject, the data was split into train, test, and validation sets in the ratio 60:20:20. The proposed model is illustrated in Fig. 3.

Table 1 Hyperparameters of the first candidate model

Layer 1: 1D convolutional layer	
Number of units	64
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Dropout	0.2
Layer 2: 1D convolutional layer	
Number of units	128
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Dropout	0.2
Layer 3: 1D convolutional layer	
Number of units	256
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Layer 4: LSTM layer	
Number of units	256
Layer 5: LSTM layer	
Number of units	128
Layer 6: LSTM layer	
Number of units	64
Layer 7: dense layer	
Number of units	32
Activation	ReLU
Layer 8: dense layer	
Number of units	1
Activation	Sigmoid

Table 2 Hyperparameters of the second candidate model

Layer 1: 1D convolutional layer	
Number of units	64
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Dropout	0.2
Layer 2: 1D convolutional layer	
Number of units	128
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Dropout	0.2
Layer 3: 1D convolutional layer	
Number of units	256
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Layer 4: GRU layer	
Number of units	256
Layer 5: GRU layer	
Number of units	128
Layer 6: GRU layer	
Number of units	64
Layer 7: dense layer	
Number of units	32
Activation	ReLU
Layer 8: dense layer	
Number of units	1
Activation	Sigmoid

Table 3 Hyperparameters of the third candidate model

Layer 1: 1D convolutional layer	
Number of units	64
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Dropout	0.2
Layer 2: 1D convolutional layer	
Number of units	128
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Dropout	0.2
Layer 3: 1D convolutional layer	
Number of units	256
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Layer 4: 1D convolutional layer	
Number of units	512
Kernel size	3
Activation	ReLU
Pooling: max pooling	Size = 2
Flatten	
Layer 7: dense layer	
Number of units	512
Activation	ReLU
Layer 8: dense layer	
Number of units	256
Activation	ReLU
Layer 7: dense layer	
Number of units	128
Activation	ReLU
Layer 8: dense layer	
Number of units	1
Activation	Sigmoid

Fuzzy Ensemble Using Gompertz Function

The proposed ensemble method generates fuzzy ranks of the different models using the Gompertz function. It fuses the decision scores adaptively from those models to make the combined prediction on the test set. In the hard voting ensemble, all the models are given the same priority, which can be a disadvantage if there is a weak classifier. However, this disadvantage is overcome to some extent in this fuzzy approach as weightage is assigned dynamically based on the confidence measure. This approach has been applied to other problems in previous studies [41–44]. The Gompertz function has been applied extensively for studying COVID-19 in recent years [41, 45–47]. In our study, we have implemented the re-parameterised Gompertz function for modelling our fuzzy ensemble.

Algorithm: For each model, we get individual confidence (c), which is normalised to get a normalised confidence value.

Let there be x candidate models and n number of classes. In our case, x is 3, and n is 2.

$$\sum_{i=1}^{n} c_i^j = 1.0 \quad \forall j \in 1, 2, ..x \tag{1}$$

These confidence scores are used for calculating the fuzzy rank using the Gompertz function.

$$r_i^j = 1 - \exp\{-\exp \exp\left(-2.0 * c_i^j\right)$$
$$\forall i \in 1, 2.., n, \forall j \in 1, 2, .., x \tag{2}$$

Lower the value of rank indicates better confidence scores. Let M^j denote the top M ranks for a particular class c. If the rank does not belong to the top M ranks, then two penalty values are calculated, $P1_i^j$ and P2. P1 is calculated by putting $c_i^j = 0$ in Eq. (2), and P2 is 0. Next, we calculate two more factors, the rank-sum (S^j) and the complement of confidence score factor (F^j) in the following way:
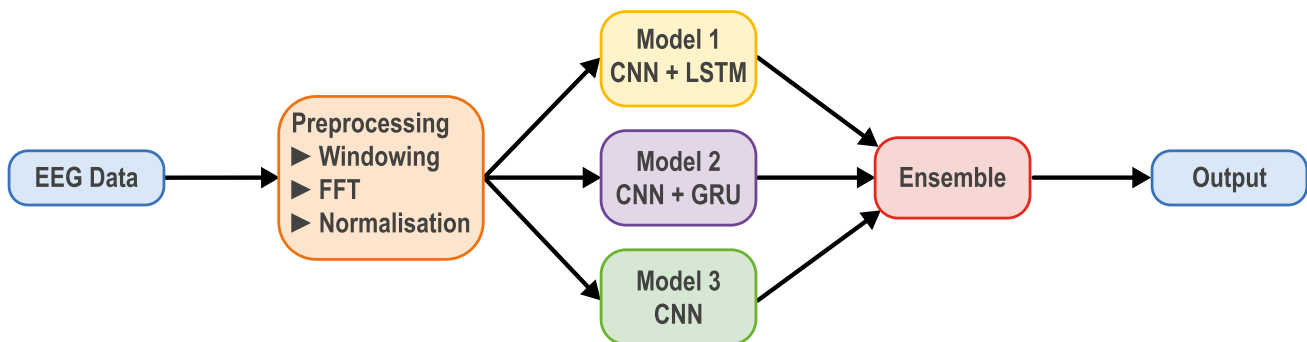
Fig. 3 Schematic diagram of our proposed ensemble learning approach for emotion detection from EEG signals

Table 4 Comparison of performance of different models for the subject-independent approach on the DEAP dataset

Models	Valence		Arousal	
	Accuracy score	**F1 score**	**Accuracy score**	**F1 score**
1D-CNN + LSTM	90.09	90.98	91.39	92.61
1D-CNN + GRU	89.18	90.11	90.21	91.57
1D-CNN only	87.9	88.98	88.56	90.22
Gompertz Fuzzy Ensemble	**90.84**	**91.65**	**91.72**	**92.9**

The bold values signify the highest accuracies achieved by the proposed Gompertz fuzzy ensemble model

$$S^j = \sum_{j=1}^{x} r_i^j \quad \text{if} \quad r_i^j \in M^j \qquad\qquad F^j = {}^1/x \sum_{j=1}^{x} c_i^j \quad \text{if} \quad r_i^j \in M^j$$

$$S^j = \sum_{j=1}^{x} P1_i^r \quad \text{otherwise} \qquad (3) \qquad F^j = {}^1/x \sum_{j=1}^{x} P2_i^c \quad \text{otherwise} \qquad (4)$$

(a)
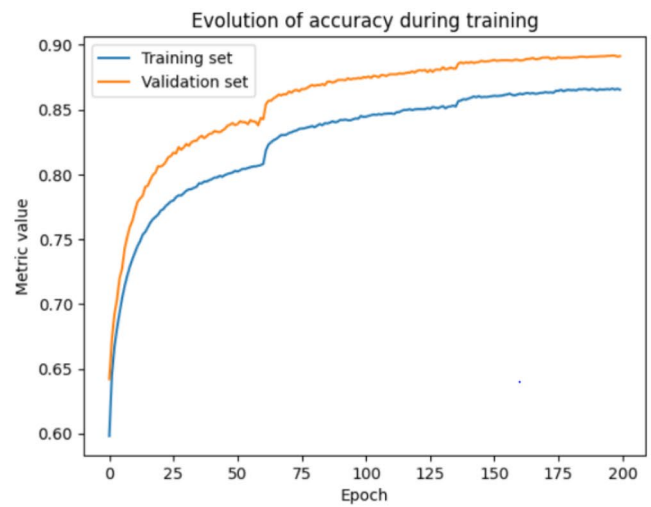
(b)
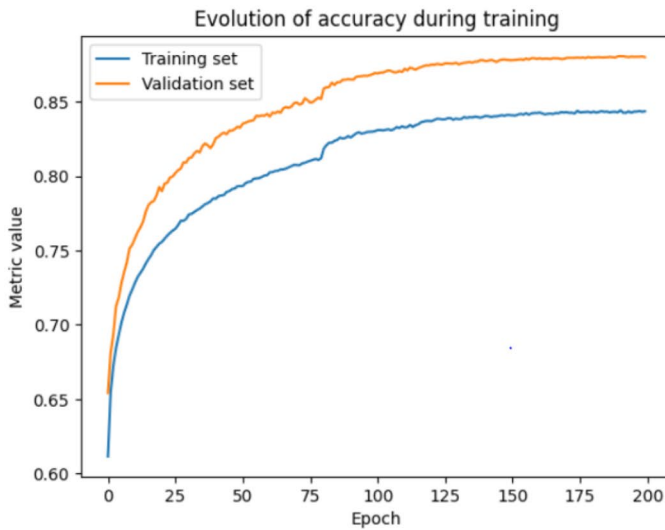
(c)

Fig. 4 Graphs showing the variation of training as well as validation accuracies with different epoch sizes for valence dimension on DEAP dataset for **a** model 1 (1D CNN + LSTM model), **b** model 2 (1D CNN + GRU model), and **c** model 3 (1D CNN only)
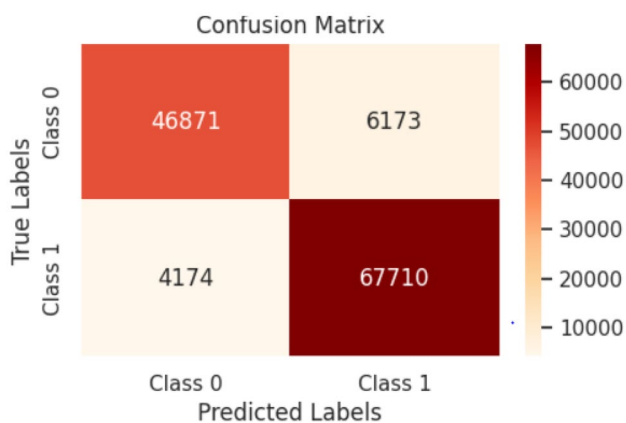
Here, $P1_i^r$ and $P2_i^c$ are defined as the penalty terms imposed on pattern class i, if it does not belong to the top M class ranks. The final score (SC^j) is the product of S^j and F^j

$$SC^j = S^j \times F^j \tag{5}$$

Finally, the resultant class (c) is the class with a minimum SC^j value which gives the final decision score of the proposed ensemble model.

Results

All programs have been run on the Google Colab platform. The GPU utilised for running the programs is Tesla T4 provided by the platform. The performance metrics used for
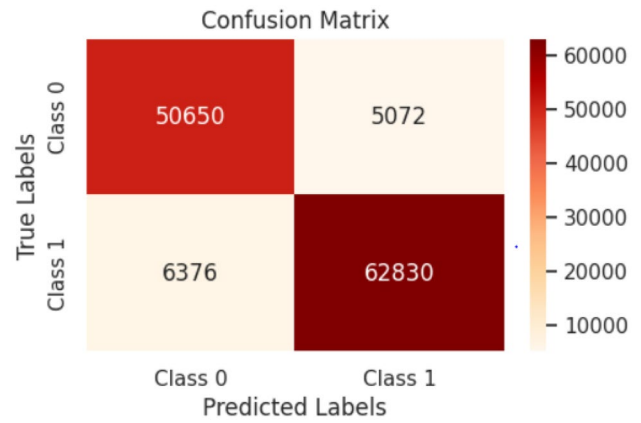
Fig. 5 Confusion matrix obtained from the proposed fuzzy ensemble model for valence dimension on DEAP dataset
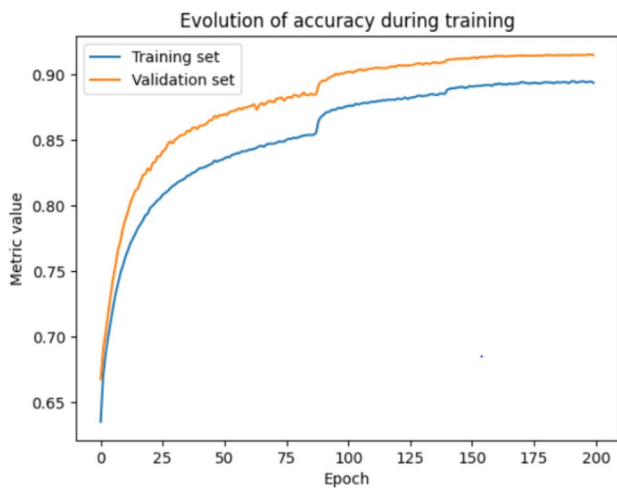
(a)
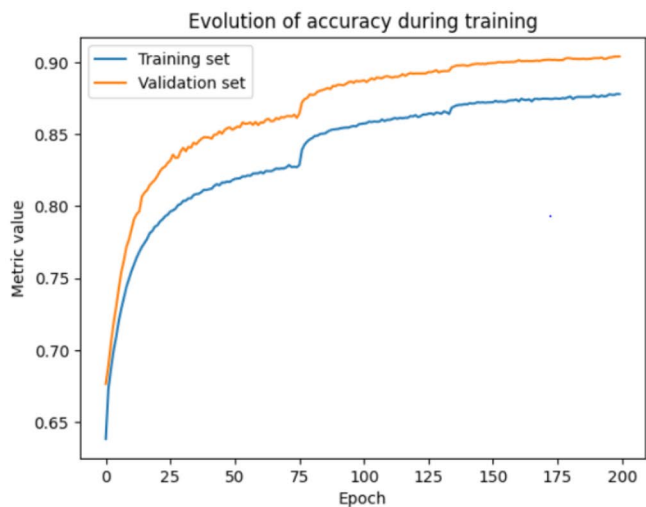
(b)
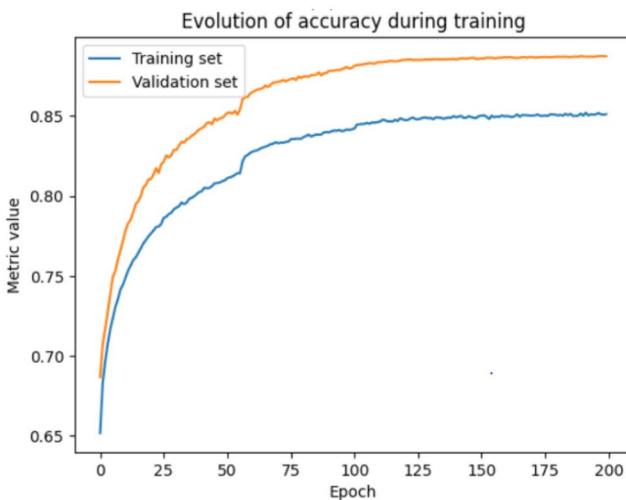
(c)

Fig. 6 Graphs showing the variation of training as well as validation accuracies with different epoch sizes for arousal dimension on DEAP dataset for **a** model 1 (1D CNN + LSTM model), **b** model 2 (1D CNN + GRU model), and **c** model 3 (1D CNN only)

Confusion Matrix

Fig. 7 Confusion matrix produced by the proposed fuzzy ensemble model for arousal dimension on DEAP dataset

The confusion matrix obtained by the proposed fuzzy ensemble model is represented in Fig. 5. From Fig. 5, it can be seen that the ratio of wrong predictions to correct predictions is approximately 10% for both high and low valence.

Arousal Dimension The variation of training and validation accuracies with epochs for the three candidate models have been shown in Fig. 6. It can be observed from Fig. 6 that the accuracy has become more or less stable near 200 epochs which indicates the training has been completed.

The confusion matrix obtained by the proposed fuzzy ensemble model is represented in Fig. 7. From Fig. 7, it can be seen that the misclassification rate is slightly higher for low arousal as compared to high arousal.

Subject-Dependent Approach to DEAP Dataset

For the subject-dependent approach, we have taken the average of our test results for each of 32 subjects, and those

evaluating our model are accuracy and f1-score; they have been defined as follows:

$$\text{Accuracy} = \frac{\text{True Postive} + \text{True Negative}}{\text{True Postive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (6)$$

$$\text{F1} - \text{score} = \frac{\text{Recall}^{-1} + \text{Precision}^{-1}}{2} \quad (7)$$

average test results for the different models have been presented in Table 5. In this approach, the ensemble model outperforms the individual models for both the arousal and valence dimensions.

DEAP Dataset

Subject-Independent Approach to DEAP Dataset

The evaluation metrics for different models on the DEAP dataset for the subject-independent approach have been summarised in Table 4. It can be seen that the ensemble model outperforms the individual models.

Valence Dimension The variation of training and validation accuracies with epochs for the different models has been shown in Fig. 4. It can be observed from the following curves that both training and validation accuracies increase with epoch rapidly in the beginning, then the rate of increase decreases, and it almost flattens near 200 epochs, which indicates that the model has been trained.

AMIGOS

Subject-Independent Approach

The evaluation metrics for different models on the DEAP dataset for the subject-independent approach have been summarised in Table 6. The ensemble model surpasses the accuracies achieved by the individual models, and we have obtained state-of-the-art accuracies for both valence and arousal dimensions.

Valence Dimension The variation of training and validation accuracies with epochs for the three different candidate models has been shown in Fig. 8. From the graphs shown in Fig. 8, it can be observed that the accuracy increases sharply

Table 5 Comparison of performance of different models for the subject-dependent approach on the DEAP dataset

Models	Valence		Arousal	
	Accuracy score	F1 score	Accuracy score	F1 score
1D-CNN+LSTM	94.11	94.41	94.39	94.64
1D-CNN+GRU	94.53	94.78	94.84	95.06
1D-CNN only	94.45	94.7	94.68	94.88
Gompertz Fuzzy Ensemble	**95.78**	**95.99**	**95.97**	**96.14**

The bold values signify the highest accuracies achieved by the proposed Gompertz fuzzy ensemble model

Table 6 Comparison of performance of different models for the subject-independent approach on the AMIGOS dataset

Models	Valence		Arousal	
	Accuracy score	F1 score	Accuracy score	F1 score
1D-CNN + LSTM	96.05	96.92	93.21	93.74
1D-CNN + GRU	96.59	96.90	97.03	96.19
1D-CNN only	97.00	97.44	97.34	96.60
Gompertz Fuzzy Ensemble	**98.73**	**99.00**	**98.39**	**98.53**

The bold values signify the highest accuracies achieved by the proposed Gompertz fuzzy ensemble model

initially, and then the curve becomes flat near 50 epochs. Hence, the models are not trained for further epochs.

The confusion matrix outputted by the proposed fuzzy ensemble model for valence dimension is represented in Fig. 9. It can be observed from Fig. 9 that the ratio of incorrect predictions to correct predictions for low valence is around 0.024 and that of high arousal is about 0.006.

Arousal Dimension The variation of training and validation accuracies with epochs for the different candidate models has been shown in Fig. 10. The training and validation

(a)

(b)

(c)
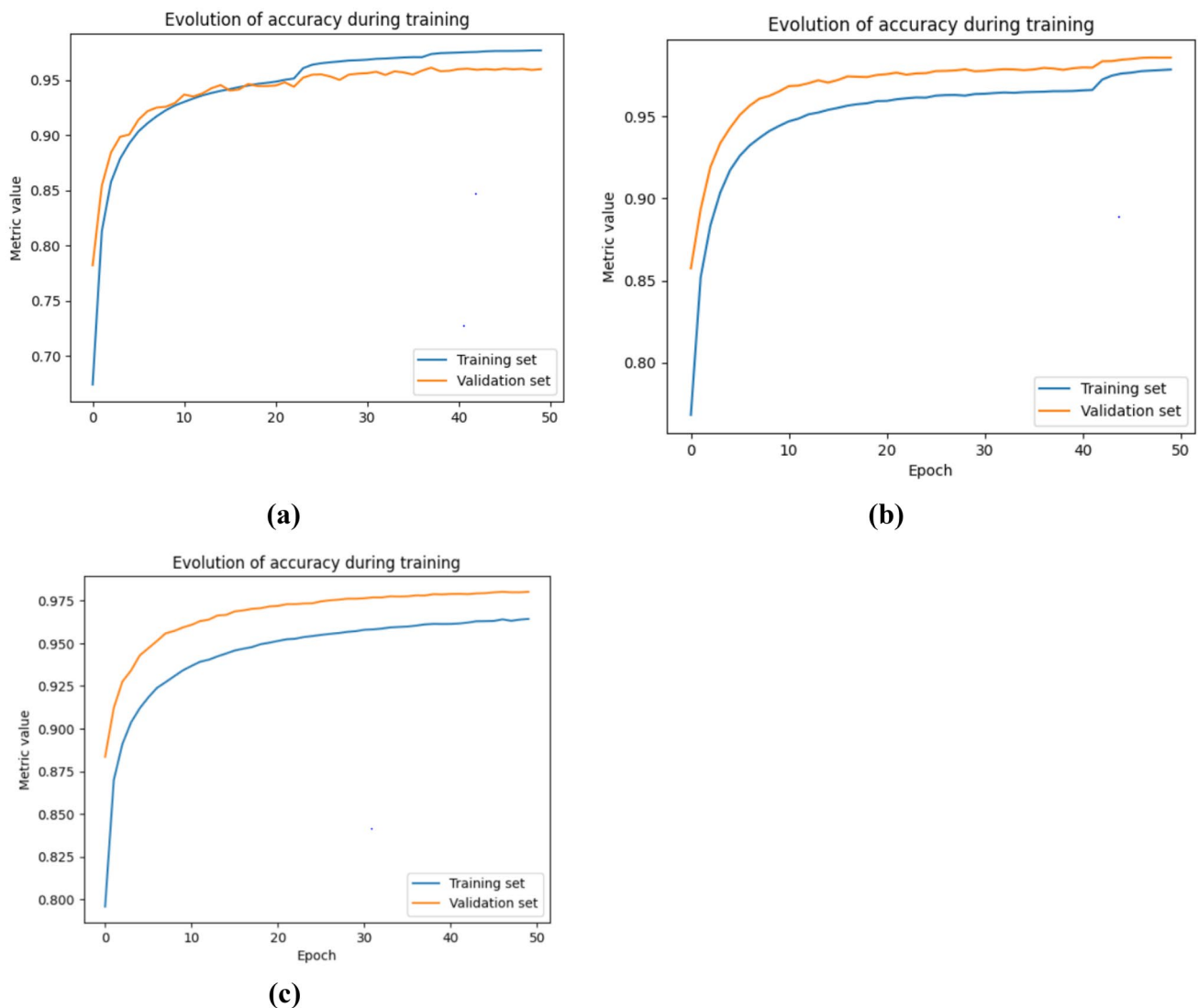
Fig. 8 Graphs showing the variation of training as well as validation accuracies with different epoch sizes for valence dimension on AMIGOS dataset for **a** model 1 (1D CNN + LSTM model), **b** model 2 (1D CNN + GRU model), and **c** model 3 (1D CNN only)
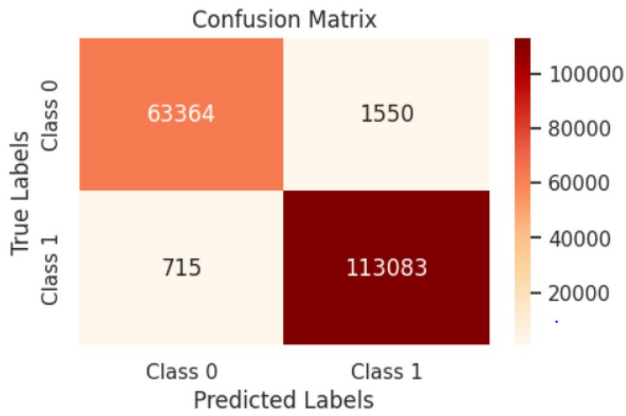
Fig. 9 Confusion matrix obtained by the proposed fuzzy ensemble model for valence dimension on AMIGOS dataset

accuracies reached a more or less constant value at 50 epochs for each model; hence the models were trained for 50 epochs only.
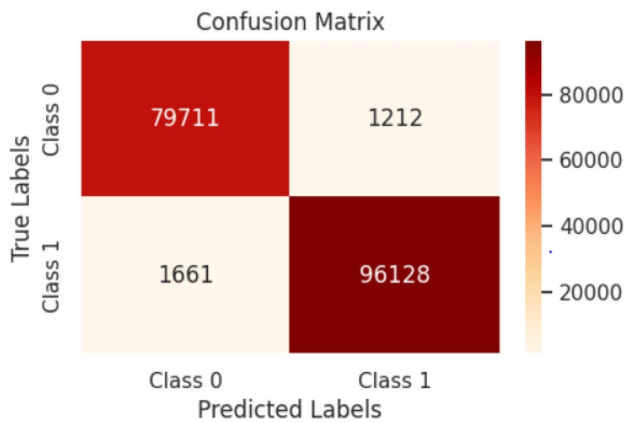
The confusion matrix produced by the proposed fuzzy ensemble model for arousal dimension is represented in Fig. 11. The misclassification percentage for low arousal is approximately 1.5% and that of high arousal is approximately 1.7%.

Subject-Dependent Approach

For the subject-dependent approach, we have taken the average of our test results for each of 32 subjects, and those average test results for the different models have
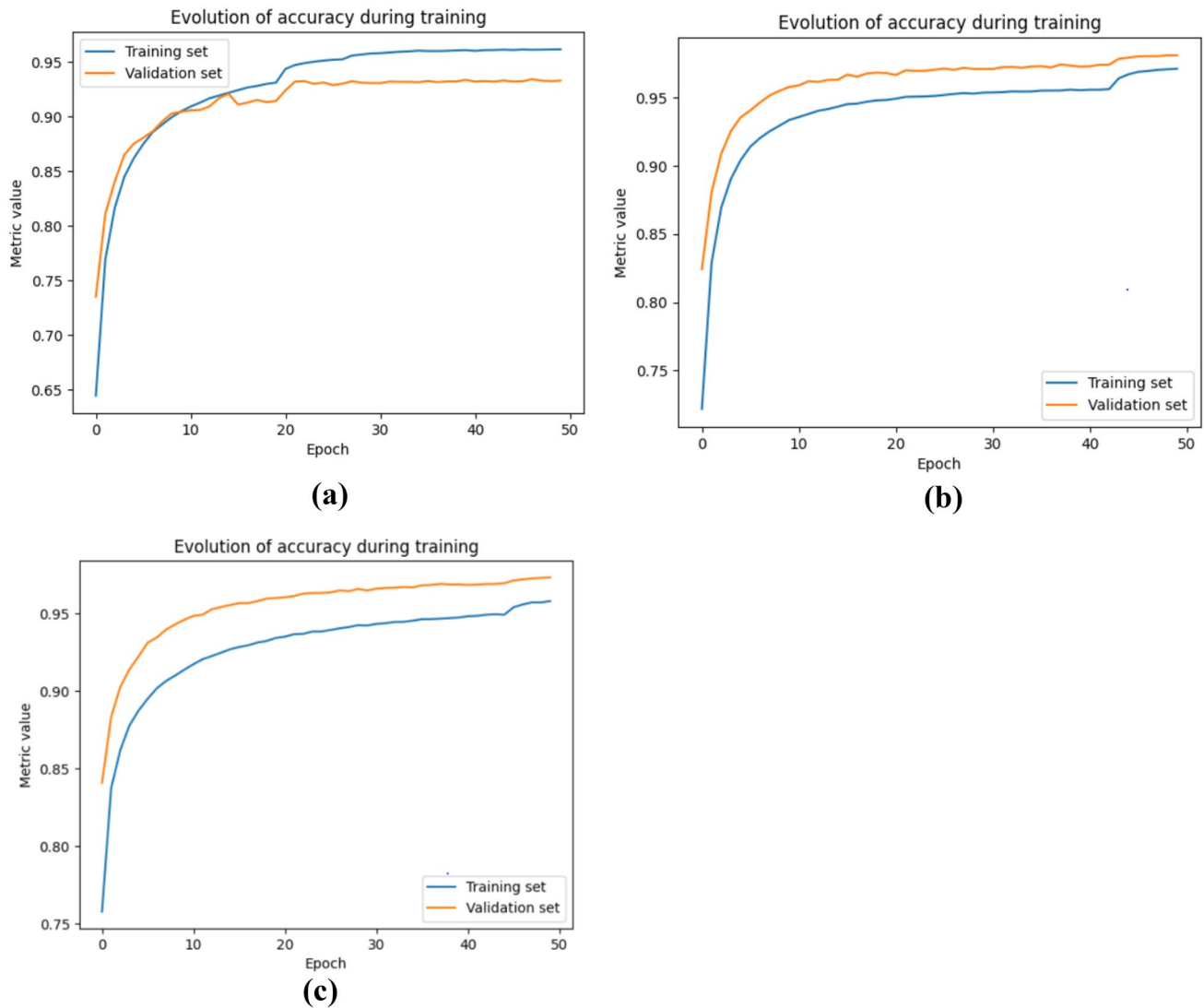
Fig. 10 Graphs showing the variation of training as well as validation accuracies with different epoch sizes for arousal dimension on AMIGOS dataset for **a** model 1 (1D CNN + LSTM model), **b** model 2 (1D CNN + GRU model), and **c** model 3 (1D CNN only)

Confusion Matrix

Fig. 11 Confusion matrix obtained by the proposed fuzzy ensemble model for arousal dimension on AMIGOS dataset

been presented in Table 7. It is evident from Table 7 that the proposed fuzzy ensemble-based deep learning model has given better results in case of both valence as well as arousal dimensions.

Comparison to Existing Works

The performance of our model has been compared to previous models in Table 8. It can be observed from Table 8 that our proposed fuzzy ensemble-based deep learning model has outperformed almost all the existing models for both DEAP and AMIGOS datasets.

Table 7 Comparison of performance of different models for the subject-dependent approach on the AMIGOS dataset

Models	Valence		Arousal	
	Accuracy score	F1 score	Accuracy score	F1 score
1D-CNN+LSTM	97.12	91.10	95.22	94.53
1D-CNN+GRU	98.21	91.76	97.05	97.03
1D-CNN only	97.37	92.86	96.56	96.44
Gompertz Fuzzy Ensemble	**99.38**	**93.87**	**98.66**	**98.54**

The bold values signify the highest accuracies achieved by the proposed Gompertz fuzzy ensemble model

Table 8 Performance comparison of our fuzzy ensemble-based deep learning model with some state-of-the-art deep learning models proposed by previous researchers

Author	Dataset	Model	Valence accuracy	Arousal accuracy
Koelstra et al. [15]	DEAP	PSD+Naïve Bayes	0.6200	0.5760
Dabas et al. [18]		SVM, Naïve Bayes	0.5890	0.7806
Salama et al. [21]		CNN	0.8744	0.8849
Zhan et al. [22]		CNN	0.8295	0.8407
Allghary et al. [23]		LSTM	0.8545	0.8565
Parui et al. [25]		XGBoost	0.7597	0.74206
Bagzir et al. [27]		DWT+KNN, SVM, ANN	0.9110	0.9130
Siddharth et al. [28]		CNN+Extreme Learning Machine	0.7109	0.7258
Ante Topic [31]		R-HOLO-FM+CNN+SVM	0.8326	0.8385
Zhao et al. [35]		3D CNN	0.9643	0.9661
Proposed (Subject Independent)		**Fuzzy Ensemble**	**0.9084**	**0.9172**
Proposed (Subject Dependent)		**Fuzzy Ensemble**	**0.9578**	**0.9597**
Miranda et al. [29]	AMIGOS	PSD,SPA+SVM	0.5760	0.5920
Siddharth et al. [28]		CNN+Extreme Learning Machine	0.8302	0.7913
Ante Topic [31]		R-HOLO-FM+CNN+SVM	0.8854	0.9151
Singh et al. [33]		Spectrogram+CNN	0.8750	0.7500
Garg et al. [34]		Wavelet energy+DNN	0.8547	0.8187
Zhao et al. [35]		3D CNN	0.9696	0.9752
Proposed (Subject Independent)		**Fuzzy Ensemble**	**0.9873**	**0.9839**
Proposed (Subject Dependent)		**Fuzzy Ensemble**	**0.9938**	**0.9866**

The bold values signify the highest accuracies achieved by the proposed Gompertz fuzzy ensemble model

Conclusion

This paper proposes a fuzzy ensemble-based deep learning approach to classify emotion from EEG signals. Emotion recognition from EEG signals is a very challenging task [48], and getting accurate predictions is crucial as it has applications in the medical domain [49, 50]. The proposed model has achieved state-of-the-art results on the benchmark AMIGOS dataset, which is the largest dataset in this domain. The proposed model achieved accuracies of 98.73% and 98.39% on the valence and arousal dimensions, respectively, for the subject-independent setup, while for the subject-dependent setup, the accuracies attained are 99.38% and 98.66%, respectively, on the valence and arousal dimensions. Our model has also achieved satisfactory results for both subject-dependent and subject-independent approaches on the standard DEAP dataset. For the subject-independent approach, we have obtained accuracies of 90.84% and 91.72%, respectively, on the valence and arousal dimensions. For the subject-dependent approach, the accuracies obtained are 95.78% and 95.97% on the valence and arousal dimensions, respectively. It is to be noted that the running time of the proposed fuzzy ensemble-based model are found to be approximately 446 s for the DEAP dataset and 759 s for the AMIGOS dataset. This proves that our proposed model produces results while utilising significantly lesser time. It is true that the Gompertz function is difficult and mathematically expensive as we need to calculate the fuzzy measure for each individual candidate model as well as groups of models. Sometimes, ensembling is also found to be expensive in terms of both time and space. Even ensemble methods reduce the model interpretability due to increased complexity. However, the proposed ensemble model produces impressive accuracies for both subject-dependent and subject-independent cases. This is the most highlighting part of the present approach. One important limitation of the present work is that our proposed model has been designed for inputs taken from EEG signals only and so the model may not perform well for multi-modal inputs.

In the future, we would like to develop a multi-modal model for emotion recognition by combining EEG signals and video recording as inputs. Due to resource constraints, we could not develop the multi-modal model in this paper. Also, future work could focus on converting EEG signals to the image domain and then training classifiers for emotion recognition. This will reduce the input size, and hence model efficiency will increase.

Declarations

References

1. Farah L, Hussain A, Kerrouche A, Ieracitano C, Ahmad J, Mahmud M. A highly-efficient fuzzy-based controller with high reduction inputs and membership functions for a grid-connected photovoltaic system. IEEE Access. 2020;8:163225–37. https://doi.org/10.1109/ACCESS.2020.3016981.

2. Mahmud M, Kaiser MS, Hussain A, Vassanelli S. Applications of deep learning and reinforcement learning to biological data. IEEE Trans Neural Netw Learn Syst. 2018;29(6):2063–79. https://doi.org/10.1109/TNNLS.2018.2790388. (PMID: 29771663).

3. Sumi AI, Zohora MF, Mahjabeen M, Faria TJ, Mahmud M, Kaiser MS. fASSERT: a fuzzy assistive system for children with autism using Internet of things. In: , et al. Brain Informatics. BI 2018. Lect Notes Comput Sci. 2018;11309. Springer, Cham. https://doi.org/10.1007/978-3-030-05587-5_38

4. Chen T, Su P, Shen Y, Chen L, Mahmud M, Zhao Y, Antoniou G. A dominant set-informed interpretable fuzzy system for automated diagnosis of dementia. Front Neurosci. 2022;16:867664. https://doi.org/10.3389/fnins.2022.867664. PMID: 35979331; PMCID: PMC9376621.

5. Kaiser MS, Chowdhury ZI, Mamun SA, et al. A neuro-fuzzy control system based on feature extraction of surface electromyogram signal for solar-powered wheelchair. Cogn Comput. 2016;8:946–54. https://doi.org/10.1007/s12559-016-9398-4.

6. Mahmud M, Kaiser MS, Rahman MM, et al. A brain-inspired trust management model to assure security in a cloud based IoT framework for neuroscience applications. Cogn Comput. 2018;10:864–73. https://doi.org/10.1007/s12559-018-9543-3.

7. Mammone N, Ieracitano C, Adeli H, Morabito FC. AutoEncoder filter bank common spatial patterns to decode motor imagery from EEG. IEEE J Biomed Health Inform. 2023;27(5):2365–76. https://doi.org/10.1109/JBHI.2023.3243698. (Epub 2023 May 4 PMID: 37022818).

8. Morabito FC, Ieracitano C, Mammone N. An explainable artificial intelligence approach to study MCI to AD conversion via

HD-EEG processing. Clin EEG Neurosci. 2023;54(1):51–60. https://doi.org/10.1177/15500594211063662.

9. Mahmud M, Kaiser MS, McGinnity TM, et al. Deep learning in mining biological data. Cogn Comput. 2021;13:1–33. https://doi.org/10.1007/s12559-020-09773-x.

10. "Electroencephalogram (EEG)", https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/electroencephalogram-eeg. Accessed on 2022–05–12.

11. Zheng W. Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis. IEEE Transactions on Cognitive and Developmental Systems. 2017;9(3):281–90. https://doi.org/10.1109/TCDS.2016.2587290.

12. Jatupaiboon N, Pan-ngum S, Israsena P. Real-time EEG-based happiness detection system. Sci World J. 2013;2013:618649. https://doi.org/10.1155/2013/618649.

13. Dhara T, Singh PK. "*Emotion recognition from EEG data using hybrid deep learning approach*", In: Proc. of 7th Int Conf Emerging Appl Inf Technol. (EAIT-2022), 2022.

14. Ghosh S, Kim S, Ijaz MF, Singh PK, Mahmud M. "*Classification of mental stress from wearable physiological sensors using image-encoding-based deep neural network*", In: Biosensors, MDPI Publishers, Vol. 12, No. 1153, pp. 1–15, 2022.

15. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nij-holt A, Patras I. DEAP: a database for emotion analysis using physiological signals. IEEE Trans Affect Comput. 2012;3:18–31.

16. "AMIGOS: a dataset for affect, personality and mood research on individuals and groups (PDF)", J.A. Miranda-Correa, M.K. Abadi, N. Sebe, and I. Patras, IEEE Transactions on Affective Computing, 2018.

17. Yoon HJ, Chung SY. EEG-based emotion estimation using Bayesian weighted log-posterior function and perceptron convergence algorithm. Comput Biol Med. 2013;43(12):2230–7.

18. Dabas H, Sethi C, Dua C, Dalawat M, Sethia D. "Emotion classification using EEG signals," in Proc. ACM Int. Conf. Comput Sci Artif Intell. ACM, 2018;380–384

19. Liu J, Meng H, Nandi A, Li M. "Emotion detection from EEG recordings," in Proc. IEEE Int. Conf. Nat Comput Fuzzy Syst Knowl Discovery. IEEE, 2016;1722–1727.

20. You SD, Liu C. "Classification of user preference for music videos based on EEG recordings," in Proceedings of the IEEE 2nd Global Conference on Life Sciences and Technologies (LifeTech), Kyoto, Japan, March 2020.

21. Salama S, El-Khoribi RA, Shoman ME, Shalaby MA. "Eeg based emotion recognition using 3D convolutional neural networks," Int. J Adv Comput Sci Appl. 2018;vol. 9, no. 8.

22. Zhan Y, Vai MI, Barma S, Pun SH, Li JW, Mak PU. "A computation resource friendly convolutional neural network engine for EEG-based emotion recognition." In 2019 IEEE Int Conf Comput Intell Virtual Environ Meas Syst Appl (CIVEMSA) 2019;1–6. IEEE.

23. Alhagry S, Fahmy AA, El-Khoribi RA. Emotion Recognition based on EEG using LSTM recurrent neural network. Emotion. 2017;8(10):355–8.

24. Wichakam I, Vateekul P. "An evaluation of feature extraction in EEG-based emotion prediction with support vector machines," in Proc. IEEE Int. Conf. Joint Conf Comput Sci Software Eng. IEEE, 2014;106–110.

25. Parui S, Bajiya AKR, Samanta D, Chakravorty N. Emotion recognition from EEG signal using XGBoost algorithm. In 2019 IEEE 16th India Council Inter Conf (INDICON). 2019;1–4. IEEE.

26. Aggarwal S, Aggarwal L, Rihal MS, Aggarwal S. "EEG based participant independent emotion classification using gradient boosting machines," in Proceedings of the IEEE 8th Int Adv Comput Conf (IACC). 2018;266–271. Greater Noida, India.

27. Bazgir O, Mohammadi Z, Habibi SAH. "Emotion recognition with machine learning using EEG signals," in Proceedings of the 25th National and 3rd Int Iranian Conf Biomed Eng (ICBME), 2018;5. Qom, Iran.

28. Siddharth, Jung T-P, Sejnowski TJ. "Utilising deep learning towards multi-modal bio-sensing and vision-based affective computing," in IEEE Trans Affective Comput. 2022;13(1):96–107. https://doi.org/10.1109/TAFFC.2019.2916015.

29. MAHNOB-HCI dataset. https://mahnob-db.eu/hci-tagging/.

30. Katsigiannis S, Ramzan N. "DREAMER: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices". Ieee J. Biomed Health Inf Jan. 2018;22(1):98–107.

31. Topic A, Russo M, Stella M, Saric M. Emotion recognition using a reduced set of EEG channels based on holographic feature maps. Sensors. 2022;22:3248. https://doi.org/10.3390/s22093248.

32. Zheng WL, Lu BL. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks, accepted by IEEE Trans Auton Ment Dev (IEEE TAMD). 2015;7(3):162–175.

33. Singh G, Verma K, Sharma N, Kumar A, Mantri A. "Emotion recognition using deep convolutional neural network on temporal representations of physiological signals." 2020 IEEE Int Conf Mach Learn Appl Network Technol (ICMLANT). 2020;1–6.

34. Garg S, Behera S, Patro KR, Garg A. "Deep neural network for electroencephalogram based emotion recognition." IOP Conf Series: Mater Sci Eng 1187. 2021.

35. Zhao Y, Yang J, Lin J, Yu D, Cao X. "A 3D convolutional neural network for emotion recognition based on EEG signals." 2020 Int Joint Conf Neural Networks (IJCNN). 2020;1–6.

36. Torres EP, Torres EA, Hernández-Álvarez M, Yoo SG. EEG-based BCI emotion recognition: a survey. Sensors. 2020;20:5083. https://doi.org/10.3390/s20185083.

37. Alarcão SM, Fonseca MJ. "Emotions recognition using EEG signals: a survey," in IEEE Transact Affective Comput. 2019;10(3):374–393. https://doi.org/10.1109/TAFFC.2017.2714671.

38. Zheng W, Lu B. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. IEEE Trans Auton Ment Dev. 2015;7(3):162–75. https://doi.org/10.1109/TAMD.2015.2431497.

39. Acharya, D. et al. Multi-class emotion classification using EEG signals. In: Garg D, Wong K, Sarangapani J, Gupta SK. (eds) Advanced Computing. IACC 2020. Commun Comput Inf Sci. 2021;1367. Springer, Singapore. https://doi.org/10.1007/978-981-16-0401-0_38

40. Murugappan M, Murugappan S. "Human emotion recognition through short time electroencephalogram (EEG) signals using fast Fourier transform (FFT)," 2013 IEEE 9th Int Colloq Signal Process Appl. 2013;289–294. https://doi.org/10.1109/CSPA.2013.6530058.

41. Kundu R, Basak H, Singh PK, et al. Fuzzy rank-based fusion of CNN models using Gompertz function for screening COVID-19 CT-scans. Sci Rep. 2021;11:14133. https://doi.org/10.1038/s41598-021-93658-y.

42. Basheer S, Nagwanshi KK, Bhatia S, Dubey S, Sinha GR. FESD: an approach for biometric human footprint matching using fuzzy ensemble learning. IEEE Access. 2021;9:26641–63. https://doi.org/10.1109/ACCESS.2021.3057931.

43. Amir Ziafati, Ali Maleki, Fuzzy ensemble system for SSVEP stimulation frequency detection using the MLR and MsetCCA. J Neurosci Methods. 2020;338(108686):0165–0270. https://doi.org/10.1016/j.jneumeth.2020.108686.

44. Ghosh M, Guha R, Singh PK, et al. A histogram based fuzzy ensemble technique for feature selection. Evol Intel. 2019;12:713–24. https://doi.org/10.1007/s12065-019-00279-6.

45. Spanakis, Marios, Zoumpoulakis M, Athina E, Patelarou, Evridiki Patelarou, Nikolaos Tzanakis. "COVID-19 epidemic: comparison of three European countries with different outcome using gompertz function method." Pneumon 33. 2020;1–6.

46. Akira Ohnishi, Yusuke Namekawa, Tokuro Fukui, Universality in COVID-19 spread given the Gompertz function, Progress of Theoretical and Experimental Physics, 2020;2020(12):123J01. https://doi.org/10.1093/ptep/ptaa148

47. Tjørve KM, Tjørve E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: an addition to the unified-Richards family. PLoS ONE. 2017;12: e0178691.

48. Wirawan IMA, Wardoyo R, Lelono D. "The challenges of emotion recognition methods based on EEG signals: a literature review." Int J Electr Comput Eng. 2022;12.

49. Dai Y, Wang X, Zhang P, Zhang W, Chen J. Sparsity constrained differential evolution enabled feature-channel-sample hybrid selection for daily-life EEG emotion recognition. Multimedia Tools and Applications. 2018;77:21967–94.

50. Su Y, Hu B, Xu L, Cai H,Moore P, Zhang X, Chen J. "EmotionO+: physiological signals knowledge representation and emotion reasoning model for mental health monitoring." 2014 IEEE Int Conf Bioinf Biomed (BIBM). 2014;529–535.