Expert Systems With Applications 173 (2021) 114516

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A LSTM based deep learning network for recognizing emotions using wireless brainwave driven system

Aditi Sakalle [a], Pradeep Tomar [a], Harshit Bhardwaj [a], Divya Acharya [b,*], Arpit Bhardwaj [b]

[a] CSE, USICT Gautam Buddha University, Greater Noida, India

[b] Computer Science Engineering Department, Bennett University, Greater Noida, India

ARTICLE INFO

Keywords:
Emotion recognition
Deep learning
EEG
Empirical mode decomposition
LSTM

ABSTRACT

Positive and Negative emotions are experienced by the majority of individuals in their day-to-day life. It is important to control access of negative emotions because it may lead to several chronic health issues like depression and anxiety. The purpose of this research work is to develop a portable brainwave driven system for recognizing positive, negative, and neutral emotions. This research considers the classification of four negative class of emotions using genres sadness, disgust, angry, and surprise along with the classification of three basic class of emotions i.e., positive, negative, and neutral. This paper introduces a long short term memory deep learning (LSTM) network to recognize emotions using EEG signals. The primary goal of this approach is to assess the classification performance of the LSTM model. The secondary goal is to assess the human behavior of different age groups and gender. We have compared the performance of Multilayer Perceptron (MLP), K-nearest neighbors (KNN), Support Vector Machine (SVM), LIB-Support Vector Machine (LIB-SVM), and LSTM based deep learning model for classification. The analysis shows that, for four class of emotions LSTM based deep learning model provides classification accuracy as 83.12%, 86.94%, 91.67%, and 94.12% for 50–50, 60–40, 70–30, and 10-fold cross-validations. For three class of emotions LSTM based deep learning model provides classification accuracy as 81.33%, 85.41%, 89.44%, and 92.66% for 50–50, 60–40, 70–30, and 10-fold cross-validation. The generalizability and reliability of this approach are evaluated by applying our approach to publicly available EEG datasets DEAP and SEED. In compliance with the self-reported feelings, brain signals of 18–25 years of age group provided the highest emotional identification. The results show that among genders, females are more emotionally active as compared to males. These results affirmed the potential use of our method for recognizing positive, negative, and neutral emotions.

1. Introduction

Emotion recognition is growing as an evolving research area. Emotion recognition is applied in many application areas like online gaming (Kim et al., 2013), online shopping (Hossain et al., 2017), health care monitoring (San-Segundo et al., 2019; Bhardwaj et al., 2016). Emotions have a significant impact on human life as it impinges their psychological and physiological status (Acharya et al., 2020). Emotion is defined as the response to stimulation lasting for seconds or minutes as a result of that person experiences some feelings (Picard et al., 2001). Emotions are integral parts of human behavior, which are divided into two broad categories by Davidson (1994), i.e., positive emotions and negative emotions.

Positive emotions help in improving the quality and health of human life, while negative emotions directly impact the health and reasoning capability of humans. Negative emotions are also an influential factor in causing many mental health problems (Clifford et al., 1035; Acharya et al., 2020). Mental health issues like depression, stress, and anxiety are the result of the amassing of negative emotions for an extended period, which can even lead to self-destruction in many cases (Al-Shargie et al., 2016). Approximately 89% of the inhabitants in India report emotional instability as a contrast to the worldwide average of 86% (Deb et al., 2015). Therefore, there is a need to develop a new accurate emotion recognition approach that recognizes negative emotions, which helps in improving the quality of human life. In this research work, we are proposing an LSTM based deep learning network for the classification of

* Corresponding author.
E-mail addresses: aditi.sakalle@gmail.com (A. Sakalle), pradeep.tomar@gbu.ac.in (P. Tomar), hb151191@gmail.com (H. Bhardwaj), divyaacharya099@gmail.com (D. Acharya), arpit.bhardwaj@bennett.edu.in (A. Bhardwaj).

https://doi.org/10.1016/j.eswa.2020.114516
Received 26 February 2020; Received in revised form 20 November 2020; Accepted 14 December 2020
Available online 20 January 2021

positive, negative, and neutral emotion to recognize human emotions using EEG signals.

In general, emotion recognition methods are broadly classified into three categories. First, clinical method or subjective measures that contain self-assessment manikin (SAM) forms (Bradley et al., 1994). All conventional emotion recognition models rely on self-assessment forms and human inputs to analyze emotion, which is an explicit time taking process (Kartelj et al., 2012). Second, the method recognizes emotions based on physical measures. Physical measure includes images (Scott et al., 2015), speech (Albanie et al., 2018), audio (Kim and André, 2008), and facial expression (Zhang et al., 2016). To recognize emotions from images, ImageNet was used with the convolutional neural network (CNN) in which static images were taken as input, and the concept of transfer learning was applied (Ng et al., 2015). The third method is based on using physiological signals to recognize emotion.

Recently, recognizing emotions using periphery physiological signals has gained much attention in this field (Acharya et al., 2020). The physiological signal provides a better understanding of the participant's reactions generated at the time of experiments. In this category, brain signals are recorded using electroencephalograph (EEG) (Acharya et al., 2021; Lin et al., 2010), functional magnetic resonance imaging (fMRI) (Phan et al., 2002), and electrocorticography (ECoG) (Rao et al., 2018). Among these, EEG signals have gained more popularity in recent years.

The main challenge in the field of emotion recognition is controlling of physical response and tampering of subjective measures recorded to analyze emotions. The process used to analyze emotions in the literature suffers from low accuracy and data reliability issues. Also, the clinical method used to recognize emotion is a very time consuming explicit process, which leads to the need for automating the process of emotion recognition to improve the quality of human life. To overcome this issue, in this research work, EEG based analysis is proposed using the LSTM model to study the response of the brain towards emotional stimuli (Niemic, 2004). The proposed framework has a high potential to assist in the formulation of Brain-computer interactions based tools and systems for real-time applications. This EEG signals based framework can lead to an extremely effective and precise EEG system to monitor brain function tools. Which can allow disabled people, for example, to control their wheelchair or shift a cursor on the computer. Also, the clinical and psychiatric research community can take advantage of this as a framework to make an EEG based application which can evaluate the patient's cognitive states, determine lesion sites, and classify symptoms and analyze how the patient's brains improve over time.

In this research work, movie clips are used as an emotional stimulus. Movie clips targeting specific emotions have the ability to sway emotional and behavioral changes, which can help in the treatment of any mental health issues in human beings using emotions elicitation therapy (Liu et al., 2017). Many studies have shown that using movie therapy for recognizing positive, negative, and neutral emotions, the emotional health of a person can be improved as there is a significant drop in self-reported stress and anxiety (Lai and Li, 2011; Logemann et al., 2010).

To enhance the performance of EEG signals based classification the Long Short Term Memory (LSTM) model is proposed. In comparison with traditional RNN, the innovative part of LSTM networks is the implementation of "three gates" that address the problem of the vanishing gradient and enable the algorithm to more precisely control which information needs to be stored and what needs to be removed (Gers et al., 1999). The three gates are input gate, forget gate, and output gate. By regulating the LSTM network's learning rate, it is better suited for large sequences of data series. EEG signals are highly dynamic, and non-linear time series data and LSTM networks have the advantage of isolating temporal characteristics of brain activity of different states over CNN thus, making LSTM a preferred choice for positive, negative, neutral emotion classification using EEG signals. Our aim in this research work is to analyze EEG signals in response to emotional movie clips stimuli using a long short term memory (LSTM) based deep learning network. As per our knowledge, it is for the first time classification of negative emotions using EEG signals recorded by a wearable EEG device is done using LSTM. To show the supremacy of our LSTM based deep learning network, we also analyzed it over the two publicly available benchmark dataset DEAP (Koelstra et al., 2011) and SEED (Zheng and Lu, 2015) and compare our work with other state-of-the-art methods as shown in Table 13 (Bhatti et al., 2016; Chen et al., 2018; Mohammadi et al., 2017; Lalitha and Tripathi, 2016).

Empirical mode decomposition is used for feature extraction. We aim to enhance classification results on the state-of-the-art techniques by using LSTM based deep learning model for negative, positive and neutral emotion classification. To illustrate the dominance of our method, the partition scheme used is 50–50, 60–40, 70–30, and 10-fold cross-validation methods for the confusion matrix and the Mann Whitney test. Also, we evaluated the maximum, minimum, and average classification accuracy of our model. In terms of the social behaviour analysis age group based response to emotions, and gender-based responsiveness is analyzed. The contribution of this paper can be summarized as below:

- A novel LSTM based method is implemented for emotion recognition using EEG signals.
- Comparison of different LSTM architecture is performed.
- A new EEG signal dataset for emotional clips is created with a portable, computer-efficient four-channel EEG headset (MUSE 2).
- A human behaviour analysis is performed for age and gender-based responsiveness towards emotions.

The rest of the paper is organized as follows: Section 2 contains the related work in Section 3 the background approach is described. In Section 4, dataset description is provided. The proposed methodology is explained in Section 5. Performance measures are described in Section 6. Section 7 contains results which illustrate the experimental settings followed by Discussion in Section 8. The conclusion is provided in Section 9.

2. Related work

Researchers have developed many models in the past few years for recognizing negative, positive, and neutral emotions using EEG signals. In literature for analyzing emotions, two different emotion models are used generally. The first approach is based on recognizing emotions using discrete emotion models, where word descriptions are used to analyze emotion. Second is using a multi-dimensional model space for recognizing emotions. In this section work done in literature is discussed next.

Cheng and Liu (2008) stated a method to decompose EMG signals and remove maximum and minimum wavelet coefficients as the features using the discrete wavelet transformation method. The five emotions are then classified using an artificial neural network (ANN) classifier, and 82.29% is the correct classification rate.

Lee and Cho (2011) proposed a music mood classification system of the users which proves that a specific emotional state is evocated when participants are exposed to emotional stimuli like music and videos. A generic mood descriptor is defined in their methodology. Using SVM classifier accuracy achieved is 60.3%. The generalizability of the method is not proved as it is analyzed on only 10 participants.

Khosrowabadi et al. (2011) proposed a framework for classifying EEG to detect chronic mental stress. The feature extraction method used is the Higuchi's fractal dimension of EEG, Gaussian mixtures of EEG spectrogram to extract features from raw EEG signals. Classification is performed using the K-Nearest Neighbor (K-NN) classifier. Using the 8-channel EEG device they have created dataset and achieved classification accuracy 71.23%.

Wijeratne and Perera (2012) implemented a support vector machine (SVM) model to recognize three class of emotions in response to pictures. Results obtained with 60 training samples is 83% and for 12 testing samples is 17%, which shows the highest classification accuracy over other findings obtained from the evaluation of the face image classification in real-time.

Bastos-Filho et al. (2012) proposed a novel method for feature extraction using EEG signals. In their approach, they used features extraction methods like statistical characteristics, Power Spectral Density-based features (PSD), and High Order Crossings (HOC) based features for recognizing emotions. Two states of emotions were considered, i.e., calm and stress. DEAP dataset was used for emotion recognition. They achieved an accuracy of 70.1% and 69.59% using PSD and HOC method.

Wang et al. (2014) proposed a noise removing method from EEG signals so that feature extraction can be performed smoothly. Power spectral and wavelet features are extracted from the feature extraction process. Also, the non-linear dynamical feature is considered. To reduce the dimensionality of features, principal component analysis and linear discriminant analysis is used. They implemented SVM classifier for emotion recognition. For power spectrum based classification accuracy, an average accuracy of 82% is achieved.

Atkinson and Campos (2016) proposed an improved features and kernel selection-based approach to improving the accuracy of brain-computer-interface based emotion recognition approach. Their methodology considered a more comprehensive set of emotions. They used minimum-Redundancy-maximum-Relevance (mRMR) method for selecting a pertinent set of features. Genetic algorithm and support vector machine (GA-SVM) classifier was used to classify a multi-class of emotions, and they have considered two, three, and five emotions for experimental stimuli. Their classifier produces 73% accuracy for two class of emotions, and three class it is 62.33%, and for five class 46.69% accuracy was achieved.

Liu et al. (2016) proposed a multimodal deep learning model for recognizing emotions. Bimodal Deep AutoEncoder (BDAE) is used to classify emotions. The dataset used is DEAP and SEED datasets in their research work. Power spectral density is also used for the feature extraction process. Using the BDAE network, mean accuracies of 91.01% and 83.25% on SEED and DEAP datasets is achieved.

Bhatti et al. (2016) proposed a method to recognize four class of emotions in response to music which is divided into four categories, i.e., electronic rap, metal, rock, and hip-hop music. This paper used multi-layer perceptron classifier to recognize emotions where features like time domain, frequency domain, and wavelet domain are extracted. They achieved an accuracy of 78.11%.

Zhang et al. (2017) suggested using multi-task deep neural networks with shared hidden layers (MT-SHL-DNN) to recognize emotions. Airplane Behaviour Corpus (ABC) (Schuller et al., 2007) and The Berlin Emotional Speech Database (EMOD) is used to recognize emotions. In MT-SHL-DNN the cross task hidden layers are used to transform the features. The valence arousal based emotion model is classified using a multiple-output layer in this method. Classification accuracy achieved using multi-task deep neural network is 56.11% and 78.34%. Weights of cross-entropy loss and sum of squared errors need to be managed. Also, the generalizability of the method is not proved.

Zhuang et al. (2017) implemented emotion recognition based on empirical mode decomposition (EMD). The significance of each IMF is investigated that leads to a result that high-frequency component IMF1 has a significant effect on recognizing emotional states. SVM classifier is used in their work. DEAP dataset is used and the classification accuracy of 71.99% is achieved for two class of emotions.

Mert and Akan (2018) investigated the advance attributes of empirical mode decomposition (EMD) and its multi-variable extensions for recognition of emotion. Time-domain and frequency domain based EEG signals are considered as input for power ration, power spectral density, Hjorth parameters, and correlation. KNN classifier emotion model is used and achieved a classification accuracy of 72.87% and 75%.

Qing et al. (2019) also proposed emotional analysis method in their research work utilizing machine learning and EEG triggers with the activation mechanism. The dataset used for training and testing is DEAP and SEED, acquired from 32-channel and 62-channel EEG device. Using activation curve and classification accuracy the performance is measured. The soft voting strategy is selected to build the classifier from combining multiple classifiers(Decision Tree (DT), KNN and Random Forest as the base classifier). For DEAP classification accuracy achieved is 62.63% and for SEED it is 74.85% for four class of emotions.

Acharya et al. (2020) also proposed a genetic programming based emotion recognition network. They have created their dataset for 2 class of emotions positive and negative using a single-channel EEG device. To extract features Fast Fourier transformation technique is used and the classification accuracy of 85.97% is achieved using the 10-fold cross-validation partition method.

Cimtay and Ekmekcioglu (2020) proposed dense convolutional neural network to classify EEG signals and recognize emotions. Two class and three class of emotions are considered in their work named as positive–negative and positive–negative-neutral. DEAP and SEED dataset along with LUMED dataset is used to analyze the performance of the proposed CNN network. Using 10-fold cross-validation they have achieved 86.50% and 78.30% accuracy on SEED dataset for two and three class of emotions respectively. Also, when DEAP dataset is used for two class of emotion a classification accuracy of 72.80% is achieved.

Although various approaches have been proposed for feature extraction and emotion recognition using EEG signals, still many of the experimental results are not comparable directly for different steps of experiments.

In the literature for human emotions recognition using EEG signals, distinct methods have been suggested, but these methods are restricted and must be addressed. With expanding the number of classified emotions, the precision of the EEG-based emotion recognition strategy lessened. The other challenge of EEG-based emotional stability recognition is to acquire signals while the number of electrodes positioned on the human scalp. The other challenge of EEG-based emotional stability recognition is to acquire signals while the number of electrodes positioned on the human scalp. Those people who carry electrodes claim that it is necessary to create more comfortable models of EEG devices.

So, in this research, we are introducing a long short term memory based deep learning network for the classification of positive, negative, and neutral emotions in response to emotional clips using an easy to use, fast, and light weight porTable 4-channel EEG device. Also, by proposing an EEG based framework controlling of physical response and tampering of subjective measures recorded to analyze emotions can be prevented.

3. Background

This section describes the statutory background for this approach, which includes the feature extraction process Empirical Mode Decomposition (EMD) (Bajaj and Pachori, 2011). The brief description of EMD is described next.

3.1. Empirical mode decomposition

This section delineates the extraction process of features obtained from EEG signals to recognize three (Positive, Negative, and Neutral) and four class of emotions (sadness, disgust, anger, and surprise). Here, we use Empirical Mode decomposition (EMD) for feature extraction from an EEG signal. The EMD is an innovative technique for the analysis of nonlinear and non-stationary time series (Diez et al., 2009), such as EEG. The major advantage of EMD basis functions is that it is directly derived from the signal itself. IMF's can be both amplitude and frequency modulated. The EMD decomposes the original signal into a definable set of the adaptive basis of functions called the intrinsic mode functions. The EMD offer an advantage over other signal analysis methods, like spectral analysis or wavelet transform since EMD is adaptive to the signal, whereas, in Fourier and wavelet transforms the basis are fixed. Hence, EMD allows extracting better features from non-stationary signals, such as EEG.

Also, EMD has the advantage of utilizing more oscillation information. Compared to the time–frequency method like discrete wavelet transformation (DWT), EMD can decompose EEG signals automatically, getting rid of selecting the transform window first.

In general, the word decomposition implies the transformation into several essential parts of a complex method or a synthesized material (Bhardwaj et al., 2014; Bhardwaj et al., 2019; Bhardwaj et al., 2015). However, in a wider context in the field of signal analysis, this term is used not to describes breakdowns into primary components but to disrupt certain functions that did not exist when the initial data was also evolved, it covers the analysis of different sorts of sequences were covered. Huang et al. (1998) proposed Empirical Mode Decomposition (EMD) method to decompose functions into a superposition of natural modes, each of which could be easily analyzed for their immediate frequencies and bandwidths. Bajaj and Pachori (2011) proposed the use of Empirical mode decomposition for feature extraction from an EEG signal.

EMD is generally a way to break down a signal without dropping the time domain for nonlinear and non-stationary signals assessment. EMD decomposes any data into intrinsic mode functions (IMF) which are not analytically defined and determined by the analyzed sequence on its own, with lower-frequency oscillations in successive IMF than the previous one. At the end of this process, many bandwidth parameters would be generated, which are feature attributes to the hybrid genetic programming (HGP) classifier. The decomposition of an EEG signal by EMD includes the following steps:

- Calculation of IMF for each iteration using EMD on EEG signals.
- Estimation of two features, namely frequency parameter and Amplitude parameter using Hilbert transform applied on IMF's for each iteration.
- Generation of Bandwidth parameters viz. Frequency parameter and the Amplitude parameter.

3.2. Calculation of intrinsic mode functions (IMF)

The Intrinsic Mode Functions (IMF) are acquired using EMD, at each iteration. In this situation, the core features are explicitly obtained from the input data. EMD decomposes given data into intrinsic mode functions (IMF) and a residual function. The resulting IMF's from the EMD (Wu et al., 2007) shall satisfy the following requirements:

1. The number of IMF extrema (the sum of the maxima and minima) and the number of zero-crossings must either be equal or differ at most by one.
2. At any point of an IMF, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima shall be zero.

To extract IMF's, we use a process called sifting (Wu et al., 2007). This sifting process is repeated until all IMF's are extracted from the EEG signal. This process often stops when the residue, for instance, contains no more than two extrema. After decomposition, the original EEG Signal is depicted as the sum of IMF's and the final residue. Fig. 1 shows the IMF's generated by the EMD process on the 23.6 s EEG signal for a

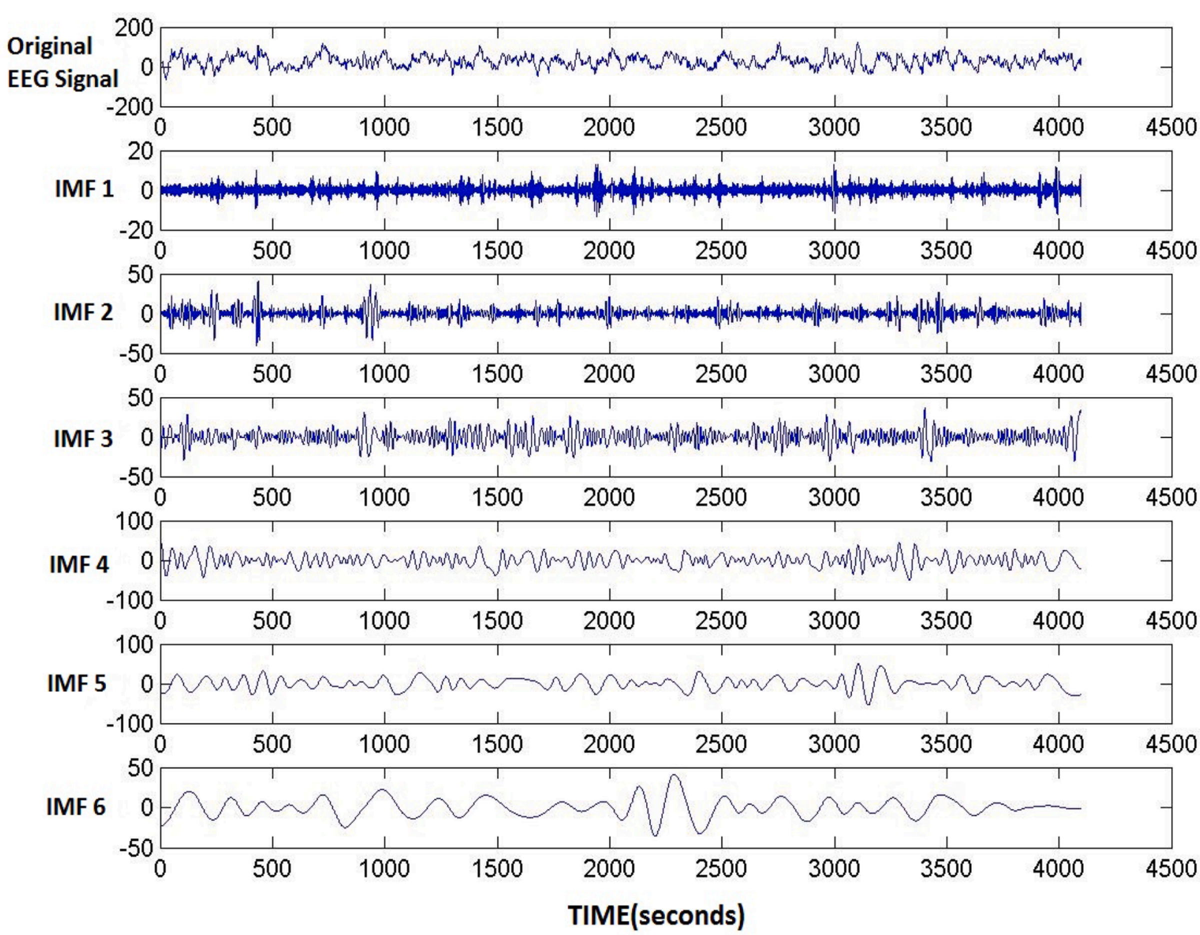


Fig. 1. Example of EEG signals from each of the five subsets (Z, O, N, F and S).

sample EEG signal. EMD can separate the inherent oscillatory forms of signals into a finite number of IMFs.

3.3. Analytical representation of EEG signal

Empirical mode decomposition successively separates the intrinsic oscillatory modes of signals into a finite number of IMFs. At the end of the decomposition, the original EEG Signal is analytically represented (Huang et al., 1998) as follows:

$$x(t) = \sum_{i=1}^{n} C_i(t) + R\left(t\right)$$

where n is the number of IMFs, C_i is the i^{th} IMF and $R(t)$ is the final residue. The analytic signal of any real IMF $A(t)$ is represented as:

$$A(t) = \sqrt{c^2(t) + c_H^2(t)}$$

Where $c(t)$ is the IMF and $c_H(t)$ refer to Hilbert transform of IMF. The instantaneous frequency $\omega(t)$ is defined as:

$$\omega(t) = \frac{d\varnothing(t)}{dt}$$

where $\varnothing(t)$ is instantaneous phase. Then we calculate the center frequency which can be defined as:

$$<\omega> = \frac{1}{E}\int \omega\left|Z(\omega)\right|^2 d\omega$$

where E is the energy of the analytic signal and $Z(\omega)$ is the Fourier transform of the analytic signal. The amplitude parameter and the frequency parameter are defined respectively.

$$B_{am}^2 = \frac{1}{E}\int\left(\frac{dA(t)}{dt}\right)^2 dt$$

$$B_{fm}^2 = \frac{1}{E}\int\left(\frac{d\varnothing(t)}{dt} - <\omega>\right)^2 A^2\left(t\right) dt$$

The total bandwidth of analytic IMF x(t) is defined as:

$$B = \sqrt{B_{am}^2 + B_{fm}^2}$$

These B_{am} and B_{fm} are taken as input features for the LSTM classifier. For feature extraction, we select ten IMF's from each EEG signal and remove the residue. Input to our LSTM classifier is a total of 20 extracted features from each EEG signal (two from each IMF's). This completes the feature extraction process from an EEG signal.

The calculated values are output by the onboard digital signal processing module, through the headset, to the application program. Data is collected every second and is analyzed in the temporal domain to detect and correct artifacts/ background noise as much as possible, without losing the accuracy of the original signal, using MUSE 2 proprietary algorithms. The headset gives us access to 20 features named as f1,f2,…f20. In this study, our dataset is extended to accommodate these features as well, as part of our feature set we extracted a total of 20 features from five frequency bands as Delta (1–4 Hz), Theta (4–7 Hz), Alpha (8–12 Hz), Beta (13–25 Hz), and Gamma (26–50 Hz)) and the 20 features are named as Delta_TP9, Delta_AF7, Delta_AF8, Delta_TP10, Theta_TP9, Theta_AF7, Theta_AF8, Theta_TP10, Alpha_TP9, Alpha_AF7, Alpha_AF8, Alpha_TP10, Beta_TP9, Beta_AF7, Beta_AF8, Beta_TP10, Gamma_TP9, Gamma_AF7, Gamma_AF8, Gamma_TP10.

4. Dataset

In this research work, the generated dataset is described along with 2 publicly available EEG dataset DEAP and SEED dataset. The description of each of the dataset is given below:

4.0.1. Generated dataset

The experiment was performed in a soundproof room, and the subject has to switch off all Bluetooth and wireless devices to prevent possible intrusion. The dataset was prepared by taking brainwave samples of participants described above by following the procedure as shown in Fig. 2.

Every subject wears the easy to use MUSE 2 headset. The 12-bit Raw-Brainwaves (3–100 Hz) are produced with a sampling rate of 512 Hz and the EEG output is produced in different frequency and morphological bands. A video is also recorded of participating subjects from the front side. Total of 50 sessions experimented in one set of the experiment for subjects. Fig. 3 shows the indenture process of EEG based experiment session. For one participant, this process is iterated four times for negative emotions and three times for positive, negative, and neutral class of emotion. To start a session, the hint of start is given 5s before each movie clip. After these subjects were then shown four clips from the list of stimuli dataset that contains sadness, disgust, anger, and surprise, positive, and neutral emotion targeting clips as shown in Table 1. The clips are shuffled randomly so that no thought is prepared. All film clips have been modified with the same resolution (720 × 576). On a 22-inch screen, clips are displayed randomly. The sound has been adjusted to the proper rate such that participants can easily use a speaker. They had their eyes about 0.6 meters from the middle of the frame. Table 1 contains a sample description of film clips shown.

After viewing every movie clip, we gathered the participant's emotions and reactions in the form of self-assessment form which was on a 10-point scale (1 = "not at all" 10 = "extremely") to assess the intensity of each self-reported emotions and 40 s is devoted to it as shown in the Fig. 3. A rest of 10s is given after every stimulus. Since our model used the standardized database of emotions induced by movies as the ground truth for labelling EEG signals we compared the self-assessments of each participant with the standardized database and discarded all cases in which the target emotion was not at least one point greater than other non-target emotions. The EEG based emotion recognition framework using LSTM based deep learning network model framework is shown in Fig. 2, which contains the explanation of collecting data process.

The proposed system is using the standardized database of movie-induced emotions as the ground truth to label the EEG signals. The participants were encouraged to answer all the questions based on their true feelings when watching each film excerpt, instead of their familiar feelings or general mood. Also, every participant's self-assessment is compared with the standardized database and rejected the cases where he or she did not rate the target emotion at least one point higher than the other non-target emotions. The data is collected in two different manner one where four different negative emotions are considered. In second case three basic emotion class are considered positive, negative (emotions considered are sadness, disgust, anger, and surprise), and neutral.

The test data was collected by showing a single clip targeting one emotion to the participants so that we can measure the accuracy of our model. Every clip shown generates raw EEG signal outputs. The raw signals collected are the 20 features that are subjected to EMD by collecting data from MUSE 2. In this study, five groups of features are extracted from preprocessed EEG signals from four channels (AF7, AF8, TP9, TP10) and five bands (Delta (1–4 Hz), Theta (4–7 Hz), Alpha (8–12 Hz), Beta (13–25 Hz), and Gamma (26–50 Hz)) so a total of 20 features (Delta_TP9, Delta_AF7, Delta_AF8, Delta_TP10, Theta_TP9, Theta_AF7, Theta_AF8, Theta_TP10, Alpha_TP9, Alpha_AF7, Alpha_AF8, Alpha_TP10, Beta_TP9, Beta_AF7, Beta_AF8, Beta_TP10, Gamma_TP9, Gamma_AF7, Gamma_AF8, Gamma_TP10) are extracted and referred as f1,f2,f3,…,f20 in this research work.

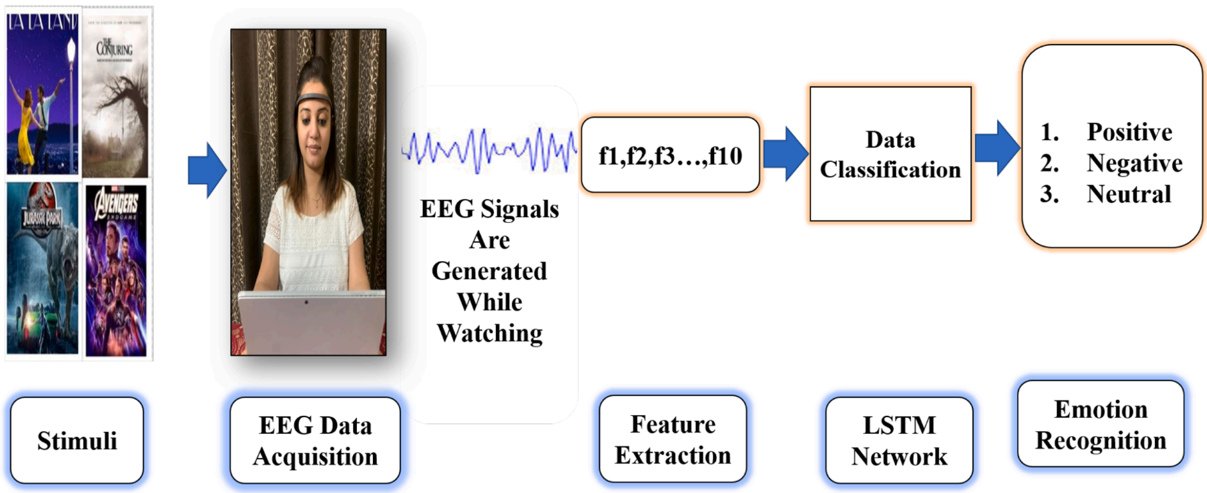


Fig. 2. EEG based emotion recognition framework using LSTM network.

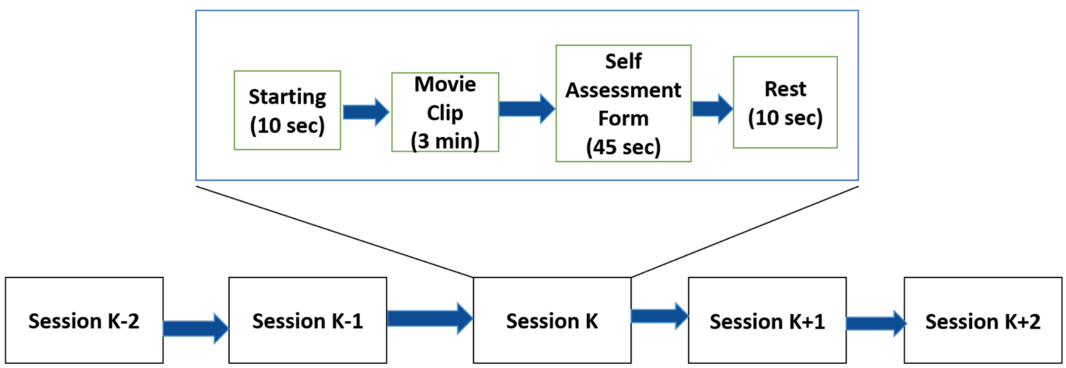


Fig. 3. EEG experiment indenture process.

Table 1
Sample description of genres of film clips shown to elicit target emotion

Genres	Film Name	Length (s)	Clip Content
Sadness	Irreplaceable you	72	A girl was diagnosed with cancer
	Adrift	123	A man lost consciousness when attacked by hurricane
Disgust	Saw	65	One person torture, kills the other one to live
	Raw	125	A vegetarian man having unquenchable hunger for human flesh
Anger	Joker	70	A man mistreated by the society takes revenge
	I Spit On Your Grave	114	A female writer being assaulted by four men
Surprise	Final Destination	70	Teenager having future vision of dying in flying plane
	Supernatural	126	Two brother's fighting evil supernatural creations
Positive	Harry Potter	70	Teenager flying on magical broom in a game
	Avengers	126	supernatural creations fighting and defeating demon
Neutral	Life of Pi	70	Teenager moving on boat in sea
	Interstellar	125	A team of explorers travel through a wormhole

4.0.2. DEAP dataset

DEAP stands for database for emotion analysis using physiological signals (Koelstra et al., 2011). It is an affective EEG benchmark dataset for analysis of emotions. The DEAP has 32 subjects data which is recorded using 32-channel EEG device under laboratory settings, and the placement of electrodes is done according to the international 10–20 system. For emotion elicitation one-minute long music videos are used a total of 40 music videos are used to elicitate emotions. After watching every single video participants filled a self-assessment form which is a two-dimensional emotional scale. Four-dimensions of the scale are High/Low arousal and High/Low valence.

Algorithm 1: Algorithm for Emotion Classification Using LSTM based Deep Learning Network
Input Raw EEG signals dataset and LSTM network parameter see Table 2.
Output Four and Three class of emotions recognized.
Begin
Initialization Inform consent by subjects
for all watch Stimuli **do**
  EEG device setup
  while subject watch clips **do**
    EEG data acquisition
  end while
end for
for all feature extraction **do**
  Load Raw EEG signals
  Apply EMD technique on raw EEG dataset
for all Calculate intrinsic mode functions **do**
  Apply Hilbert Transformation on IMF
  Generate frequency and amplitude parameter
  10 IMF calculated, Returns 20 features
end for
20 extracted features are f1,f2,f3,….f20
while For classifcation **do**
    Load and Pass Extracted Features to LSTM classifier
    Apply normalization using MinMaxScaler

(*continued on next page*)

(*continued*)

```
    Initialize LSTM parameters as Table 2
    Classify EEG signals
  end while
end fot
Return: Classified four class of negative emotion, Classified three class of
  emotions
End
```

4.0.3. SEED dataset

The SEED dataset is a benchmark EEG dataset collected from 62-channel EEG device under laboratory settings provided by Shanghai Jiao Tong University in 2015 (Zheng and Lu, 2015). It has 15 subjects EEG recordings as data. For every stimulus, participants watched 15 emotional clips, then a 45 s self-assessment form is filled by them. The elicitation materials contain clips to elicit positive, neutral, and negative emotions of subjects. The placement of electrodes is done according to the international 10–20 system.

5. Proposed methodology

This section describes the statutory background for positive, negative, and neutral emotion classification using LSTM based deep learning network. Starting with the description of experimental settings including stimuli, subjects description, dataset description, and description of the experimental procedure, followed by Signal pre-processing, feature extraction process empirical mode decomposition (EMD) (Zhao and He, 2013) and classification algorithm description. All the blocks for four and three class of emotions classification are described next.

5.1. Stimuli

In this research work, we have used 40 movie clips as elicitation material. In literature, it is already proven that videos are a reliable source, for emotion elicitation (Day et al., 2009). Movie clips contain audio and video content which helps subject in having a legitimate experience. In this research work, we have selected English language film excerpts that lasted 1 to 2 min and contained independent and integrated content to elicit a single target emotion. For elicitation of emotions, we have created a standardized movie clips database that is designed targeting four class of emotions, i.e., sadness, disgust, anger, and surprise, positive, and neutral emotions taken from IMDb (IMDb, 2018). Sample description of clips in each emotion category is shown in Table 1. All the movie clips chosen are producing positive, negative, and neutral emotions feelings. To ensure effectiveness, the order of clips was randomized. Also, the effectiveness of the film clips were examined by the participants in which they rated these movie clips on a Likert scale of 1–10 point (eg., 1 being “not at all sad” and 10 being “extremely sad”) against the intensity of emotions. The effectiveness of total 10 sets (each set is having four targeting clips) of emotional stimuli is also analyzed and average effectiveness achieved is 8.5 point.

5.2. Subjects

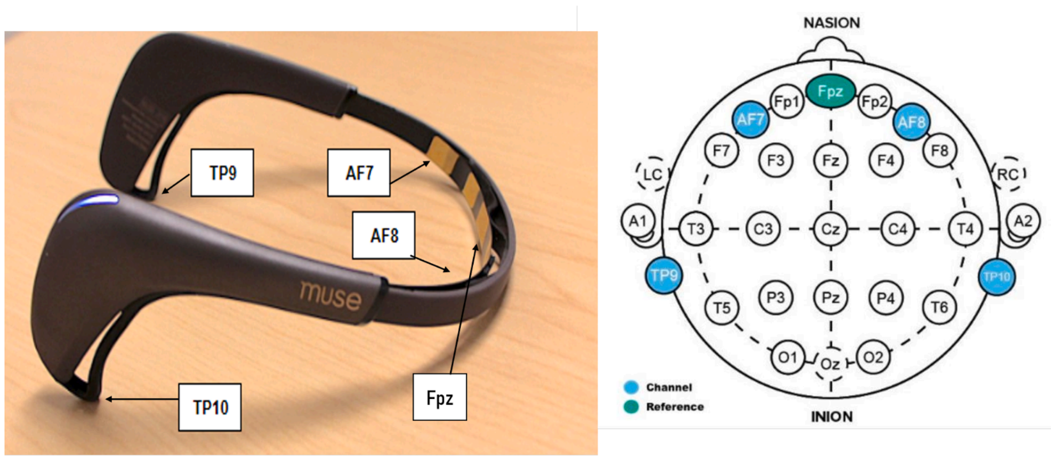
Fifty healthy subjects (25 males and 25 females) pool was considered for this research work. Participants have participated voluntarily in this research study and also signed an informed consent form. All participants are from a different culture and educational background divided into three age group 18–25 years, 26–35 years, 36–55 years. However, from the final analysis, 5 data samples were dropped due to equipment failure or excessive EEG signal artifacts. As a result, 45 samples (23 men and 22 women) were left as valid subjects. In this pool, 44 participants were right-handed, and 1 is left-handed having a self-reported normal or corrected-to-normal vision. The population considered for this research can be divided into two categories of students and academic professionals. The participants were mandated not to ingest tobacco or caffeine 24 h before the experiment. Initially, all the participants were informed about the scope in advanced and procedure of the study were followed according to the principle of Helsinki declaration. This research study is approved by the board of research and development at Bennett University, India. The experiment was performed two times with the same pool of participants in a gap of one week.

5.3. Device description

The device used to capture neuro-psychological signals is the MUSE 2 brain-sensing headset as shown in Fig. 4. It is easy to control, light in weight, commercially available, and affordable brainwave reading four-channel (two on the forehead and two behind the ears) EEG headset having sampling rate 256 Hz. It comprises of 7 EEG electrodes: 3 references and 4 input. The 3 reference electrodes are to be put on the forehead and the 4 input electrodes are to be placed two at the left and right side of the forehead and two behind the ears. The headband has a single accelerometer module capable of tracking 3-axis movement of the head. It is also, used to reject noisy data acquired due to head movements. It is an easy to wear portable EEG headband in which four dry electrodes are there positioned at AF7, AF8, TP9, and TP10 place according to the 10–20 international electrode placement system as illustrated in Fig. 4. Also, one reference electrode is placed on Fpz position. This device is made up of conductive silicon rubber for TP9 and TP10 and AF7 and AF8 material used is silver. The headband produces bipolar readings using Fpz as the reference for AF7, AF8, TP9, and TP10. For pairing purposes, it has a Static Headset ID. This is a wireless EEG headband having Bluetooth based pairing option to connect with any smart-phone using MUSE monitor application (APK) to record EEG data which can be transferred offline for processing further. It outputs five EEG bands (1–100 Hz) with a sampling rate of 256 Hz as delta (1–4 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (12–25 Hz), and gamma (25–55 Hz) which is taken as input for the LSTM based deep learning network for positive, negative, and neutral emotions classification using EEG signals.

5.4. Signal preprocessing

To remove the noise from recorded EEG signals before applying feature extraction technique and classification signals are pre-processed. Using MUSE 2 EEG device we have taken raw signals. An adjustable notch filter is applied to the raw EEG signals. The notch filter is used to reject a particular band or stop filtering a band having a narrow bandwidth. This filter is inverse of the bandpass filter. Recorded raw EEG signals are passed through a notch filter to remove frequencies between 45 Hz and 64 Hz. An over-sampling followed by down-sampling the signals to achieve a sampling frequency of 256 Hz, with 2uV (RMS) noise, is obtained as an output sampling rate in pre-processed EEG signals recorded from MUSE 2. Between the frontal electrodes and Fpz electrode on-board driven right leg (DRL) circuit is used to cancel the active noise. To remove noise from EEG signals DRL feedback circuit is used also it ensures proper skin contact between EEG electrodes. A decision tree is used to select the band in EEG signals the characteristics of EEG signals are considered as input for this tree i.e., variance, amplitude, and kurtosis. If any of this three input is having low values then they are considered as clean signal and passed on. Empirical mode decomposition as a feature extraction technique is applied to these clean raw EEG signals using a 256-sample window size with an overlap of 90% to give the required frequency bands. MUSE 2 gives five bands named as delta (1–4 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (12–25 Hz), and gamma (25–44 Hz) bands which are passed further to extract features.



(a) **(b)**

Fig. 4. (a) 4-channel MUSE 2 EEG headset (b) 10–20 International Electrode positioning system based electrode placement.

5.5. Classification

In literature different machine learning algorithms are used to recognize and classify emotions. In this research work, long short term memory (LSTM) model-based deep learning network is used to classify four class of negative emotions, and three class of positive, negative, and neutral emotions using EEG signals. LSTM model architecture is described next.

5.5.1. LSTM architecture

The recurrent neural network has an LSTM model as the eminent variation, which has shown its efficiency in the extraction of temporal information from long physiological signals (Bashivan et al., 2015; Zheng et al., 2018). In this section basic LSTM cell architecture and testing of three different LSTM architecture is explained as shown in Figs. 5 and 6.

Fig. 5 shows the basic LSTM architecture. LSTM network has three gates input gate, forget gate, and output gate. With the help of these gates, LSTM network decides which information is important to keep and which one to remove from the memory. The memory units present in the LSTM cell makes it able to remembers previous steps. LSTM cell takes in two inputs, output from the last hidden state and observation at time = t. Besides the hidden state, there is no information about the past to remember. The cell state (the horizontal line on the top of Fig. 5 (C_t)) is the key to this network. Forget gate is the first gate present in Fig. 5 that decide which information to keep and which one to drop from the cell state, this decision is made by the sigmoid layer.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

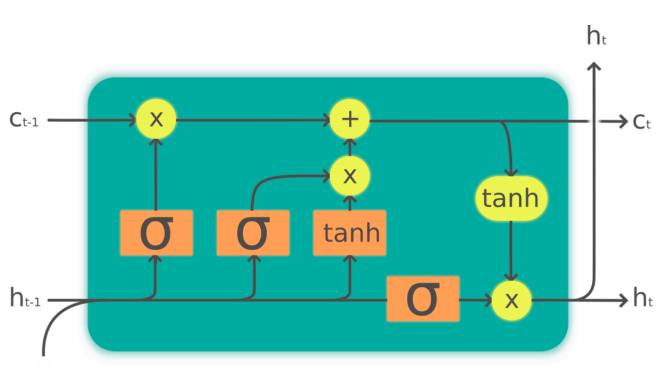


Fig. 5. Basic LSTM cells network architecture.

Input gate is the second gate that has a sigmoid layer to decide which values will be updated, and *tanh* layer which creates a vector of new updated values as shown in Eqs. (2) and (3).

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\overline{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{3}$$

Eq. (4) shows the updated cell state after calculating Eqs. (1), (2), and (3)

$$C_t = f_t * C_{t-1} + i_t * \overline{C}_t \tag{4}$$

At the end output of the current state is calculated for which inputs are updated cell state values and a sigmoid layer. This layer decides which parts of the cell state will be the final output as described in Eqs. (5) and (6).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

where σ is sigmoid activation function which squashes numbers into the range (0,1), hyperbolic tangent activation function tanh is also used which has output in the range of $(-1,1)$. W denotes the weight and W_f , W_i, W_c , W_o are the weight matrices, x_t is the input vector, previous hidden states are denoted by $h_t -1$. Bias vectors are represented by b_f,b_i, b_c, and b_o.

The EEG signals are long sequence which is difficult for a recurrent neural network to learn because of the factor that they are trained by back-propagation through time called as BPTT which leads to the problem of vanishing/exploding gradient. To overcome this issue, the RNN cell is replaced by a gate cell called an LSTM cell. LSTM has insensitivity towards gap length as it remembers values over arbitrary intervals which is a plus point in the classification of EEG signals. This advantage of LSTM makes it a preferred choice to adopt it for four and three class of positive, negative, and neutral class of emotion classification using EEG signals. The remember vector is usually called the forget gate. Also, LSTM has the advantage in learning long-term dependencies of time series and when it is used for EEG signals classification problem it explored the temporal correlations of EEG signals which lead to better results. Also, the reason behind the supremacy of the proposed model is that LSTM's are explicitly designed to avoid the long-term dependency problem.

Fig. 6 contains the three LSTM cell network architecture which is used in this research study for classifying four and three class of

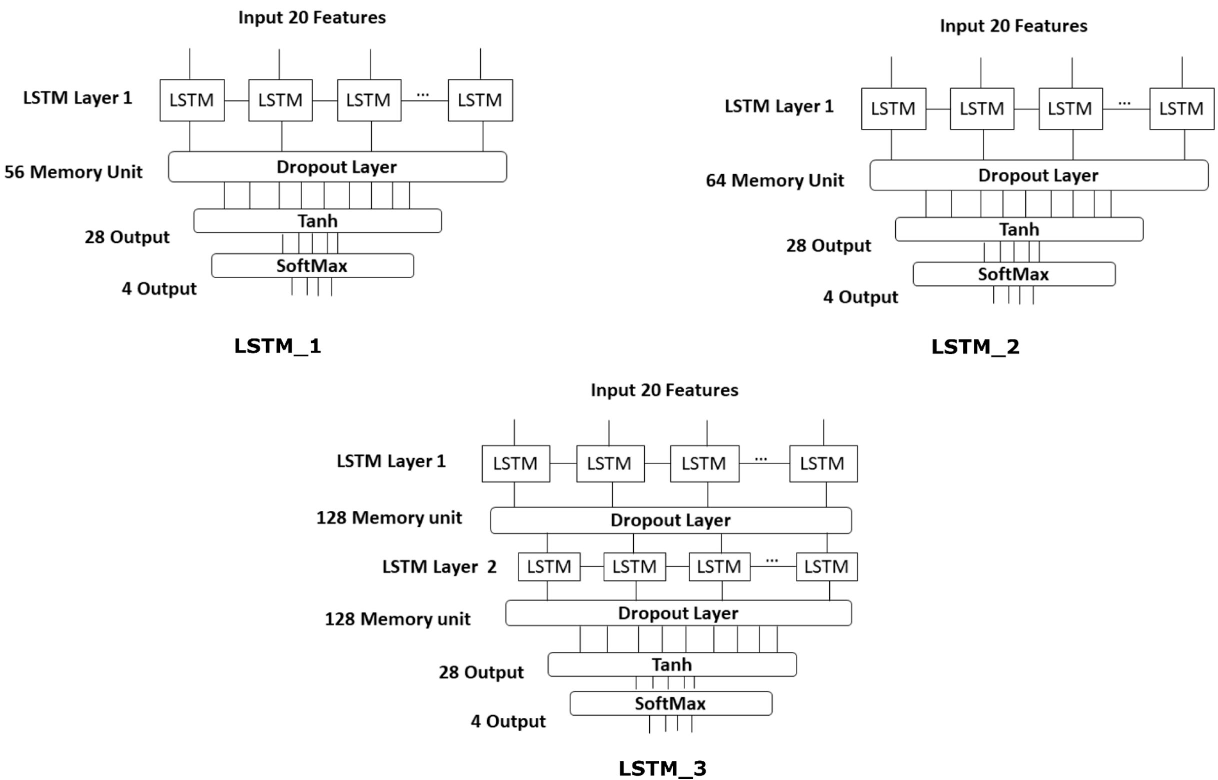


Fig. 6. Three LSTM cells network architecture.

emotions using EEG signals. These three LSTM architectures are containing simple to complex internal architecture named as LSTM_1, LSTM_2, LSTM_3 architectures respectively. These LSTM networks were built using Keras 2.0.9 upon Tensorflow 1.4.0 backend in Python 3.6. The simplest approach is implemented in LSTM_1 architecture having 56 memory unit in a single layer. The memory blocks are responsible for remembering things and manipulations to this memory is done through three major mechanisms, called gates. Three gates are used in the proposed LSTM architecture Forget gate, Input gate, and Output gate. The information removed in the proposed method's cell state is done by forget gate. If any information is not needed from the input EEG signal dataset, either it is less relevant information or not necessary to be grasped by the LSTM model it removes it by multiplying a filter. This step is necessary to optimize the LSTM network's performance.

In LSTM_2 architecture the LSTM layer used is single layer but the number of the memory unit is increased to 64. In LSTM_3 the number of LSTM layers are increased to 2 while each layer having 128 memory units. The dropout layer is also used within all proposed architecture with probability value taken as 0.2. The Dropout layer is added in the model between existing layers and applies to outputs of the prior layer that are fed to the subsequent layer as shown in Fig. 6. The LSTM model takes benefit from dropout regularization without sacrificing its valuable memorization ability (Zaremba et al., 2014). Also, by using the dropout of 0.2 it leads to train the model faster, reduce Overfitting, and make better predictions for the proposed LSTM deep learning model. LSTM_1, LSTM_2, LSTM_3 networks architecture is fully connected having an output of 28 unit using "*tanh*" as the activation function.

$$tanh\left(x\right) = \frac{2}{1+e^{-2x}} - 1 \tag{7}$$

Last and final dense layer is using "Softmax" as activation function and generating 4 output for multi-class classification i.e., for four class of emotions.

$$\text{Softmax}\left(x_i\right) = \frac{\exp(x_i)}{\sum_j \exp\left(x_j\right)} \tag{8}$$

To learn features from EEG signals LSTM cells and dropout layers are used. In our three LSTM architectures dropout layer was used to reduce the Overfitting by limiting units from co-adapting too much. For classification of positive, negative, and neutral emotions, the dense layer is used, loss function used in these network architectures is categorical cross-entropy and batch size is 10. Adaptive Moment Estimation (Adam) optimizer is used with a learning rate of 0.001. Initially, 20 features are extracted using EMD technique from raw EEG network named as f1,f2, …,f20. These 20 input features are given to the proposed LSTM network architecture. After loading of the dataset, normalization is applied to the input features of the dataset using MinMaxScaler function. This function normalizes each feature of the dataset so that every feature contribute in a maintained manner, as some feature has higher numerical value than others. It also reduces internal covariate shift which leads to the change in the distribution of network activations due to the change in network parameters during training. When normalization is applied to the proposed network it improves the training, as it reduces the internal covariate shift. Also, it helped the proposed network to fasten the optimization process as it doesn't allow weights to explode all over the place and restricts them to a certain range. An unintended benefit of normalization is that it helps the network in slightly implementing regularization also. The proposed LSTM network is initialized using the parameter defined in Table 2. The pre-processed extracted features are passed to the proposed LSTM network and for 100 epochs with a batch size of 10, the performance of our model is analyzed. As output value, the LSTM based deep learning model classifies the EEG signals into four and three class of emotions.

Each of the above mentioned LSTM architecture was evaluated on the generated dataset that we have created and the best architecture

Table 2
Parameter values used to create the LSTM model.

Parameter	Value
Optimizer	Adam
Learning Rate	0.001
Dropout Rate	0.2
Loss Function	Categorical cross entropy
Metrics	Accuracy
Batch size	10
Epochs	100

among them was applied on publicly available EEG dataset i.e., DEAP and SEED.

6. Performance measure

The efficiency, performance, and reliability of the LSTM based deep learning network classifier are deliberate by performing analytical analysis like confusion matrix, accuracy, F1-score, precision, recall sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Mann–Whitney test. Description of these performance measures are as follows:

6.1. Confusion matrix

The efficiency of a classifier on a group of samples for which the true values are recognized is called a confusion matrix. A multi-class confusion matrix for classification is described in the results for problem having four and three classes.

6.2. Accuracy

Accuracy is the measuring ability of the classifier to accurately categorize the inputs to the category they belong to.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + N} \times 100 \tag{9}$$

6.3. F1 score

F1 score is an important parameter for the analysis of unbalanced dataset based classifiers as some classes have less No. of samples. This method takes the weighted harmonic mean of precision and recall. It can be calculated using the following formula:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{10}$$

In unbalanced dataset minority classes have less number of samples. However, minority class contains important information like in case of brain signal dataset, which contains emotions. So, recall is significant. Hence, we have chosen the F1 score to give significant value to minority classes also. Recall and precision can be given as:

$$recall = \frac{TP}{TP + FN} \tag{11}$$

$$precision = \frac{TP}{TP + FP} \tag{12}$$

6.4. Specificity

Specificity is also called a true negative rate calculated as the ratio of the true negatives of a specific class to the sum of its True negatives and false positives. It can be calculated using the following formula:

$$Specificity = \frac{TN}{TN + FP} \tag{13}$$

6.5. Positive predictive value (PPV) and negative predictive value (NPV)

Positive Predictive Value is calculated as the ratio of the number of true positive to the number of positive calls. It is calculated using the following formula:

$$PPV = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})} \tag{14}$$

Prevalence is a statistical concept referring to the number of cases of class that are present in a particular population at a given time

Negative Predictive Value is calculated as the ratio of the number of true negative to the number of negative calls. It is calculated using the following formula:

$$NPV = \frac{\text{specificity} \times (1 - \text{prevalence})}{(1 - \text{sensitivity}) \times \text{prevalence} + \text{specificity} \times (1 - \text{prevalence})} \tag{15}$$

6.6. Mann–Whitney test

Mann and Whitney (1947) test, is used to present the analytical implications of results. In this two-tailed test, the median values are compared for the two samples. To analyze two different population, this test is used. If p is less than 0.001 ($p < 0.001$) value, the high significance of classifier is illustrated. Else, no significant difference is illustrated.

7. Results

This section presents the experimental results of the LSTM based deep learning model for positive, negative, and neutral emotion classification by the analysis of the EEG signal to validate the effectiveness of our method. Computing environment has Python (3.6) to implement three LSTM cell architecture i.e., LSTM_1, LSTM_2, LSTM_3, and other state of the art methods i.e., MLP, KNN, SVM, LIB-SVM, and LSTM_3 based deep learning model classifier framework as a classifier on a Windows-based operating system having core processor as Intel i7 7th gen of 3.4 GHz with 32 GB of RAM. The feature extraction technique used for all these classifiers is EMD. For the implementation of MLP, KNN, SVM, LIB-SVM, and LSTM_3 based deep learning model classifier parameters values are same as defined in Bhatti et al. (2016), Bastos-Filho et al. (2012), Koelstra et al. (2010), Lalitha and Tripathi (2016) respectively. All these algorithms are used to classify human emotions into four class of negative emotions, i.e., sadness, disgust, angry, and surprise. The standard sigmoid activation function is used for algorithms. We have implemented LSTM architectures using the parameter values as shown in Table 2. The parameters and configurations for these variables were introduced on the same device to ensure that the findings and comparisons presented were unequivocal and univocal.

The dataset is commonly divided into two distinct sets in the machine learning field, i.e., in the training set and testing set. In this study of positive, negative, and neutral emotion classification using EEG signals, generalized analysis of our approach is performed for which we have split the training and testing data into four distinct partitions to compare with current work in literature. The performance evaluation is done using 50–50, 60–40, 70–30, and 10-fold cross-validation technique. One of the methods is 50–50, where the classifier uses half samples for training and the remainder for testing. We also, split the data in training–testing proportions between 60–40 and 70–30 methods to illustrate how vital training information is in our strategy. A dataset is segregated into ten chunks of about the same size in the 10-fold cross-validation method. In this 10% or 1 chunk is used for testing, and the leftover is used for training. This method is reiterated for ten times so that a distinct input block will be used to test each session. Therefore, 100 runs are assessed. Also, for 100 epochs the results are assessed for

50–50, 60–40, 70–30, and 10-fold cross-validation analysis. Table 3 and Table 4 shows the information of dataset, the total number of samples in each class, and training–testing partition for four and three class of emotions. We analyze the accuracy and confusion matrix for four class of negative emotions and three class of positive, negative, and neutral emotions. Also, specificity, PPV, NPV, recall, precision, and F1-score value of all the models for four class of emotions using 50–50, 60–40, 70–30, and 10-fold cross-validation is calculated.

The EMD technique is applied to extract 20 features from EEG brain signal dataset to acknowledge humans positive, negative, and neutral emotions, and five distinct classifiers were used. Table 5 and Table 6 represents confusion matrix for four class of negative emotion and three class of positive, negative, and neutral class recognition of LSTM_3 based deep learning model classifier for 50–50, 60–40, 70–30, and 10-fold cross-validation. Each matrix feature shows four feasible findings, i.e. True positive (TP), True negative (TN), False positive (FP), False negative (FN). In the next subsection performance of three different LSTM based deep learning model is evaluated, comparison of LSTM model and traditional machine learning approaches are performed also, accuracy based comparison for four class of negative emotions and three class of positive, negative, and neutral class of emotions has shown the analysis of our best LSTM architecture i.e., LSTM_3 on the publicly available DEAP and SEED EEG datasets.

7.1. LSTM based deep learning model architecture evaluation

In this study, a deep learning algorithm is used to classify four and three class of positive, negative, and neutral emotions using EEG signals. The deep learning model is made up of two different units, called the computational unit and the classification unit. To progressively extract higher-level attributes from the input EEG signals having 20 features multiple layers are present in the computational unit. After analyzing data the redundancy in the input data is eliminated by selecting features which improve performance. The LSTM model has outperformed the traditional machine learning algorithms because it remembers the long-term dependencies between the time steps of sequential data that makes a higher probability of correctness in a short time (Bengio et al., 1994).

Fig. 7, shows the classification accuracy comparison of three different LSTM architectures for negative emotion classification using EEG signals. Fig. 7(a) part shows the classification accuracy for the proposed LSTM_1 model for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition minimum accuracy achieved is 32.22%, 38.78%, 40.68%, and 48.46%. The Maximum accuracy achieved for LSTM_1 is 40.45%, 45.88%, 50.46%, and 53.36% for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition. The average accuracy is 37.26%, 41.96%, 45.33%, and 49.83% for all the dataset partition method.

Fig. 7(b) shows the classification accuracy of the proposed LSTM_2 model for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition minimum accuracy achieved is 54.41%, 56.89%, 59.77%, and 61.25%.

The maximum accuracy achieved for LSTM_2 is 64.28%, 66.39%, 68.76%, and 71.46% for all the mentioned training–testing partition.

Table 3
Training and testing set partition for four class of emotions.

Training–testing partition (%)	No. of Records in the Dataset (18000)				
	Total Test Records	Sadness	Disgust	Anger	Surprise
50–50	9000	3000	3000	1500	1500
60–40	7200	3000	1800	1100	1300
70–30	5400	1900	1600	1000	800
10-fold cross validation	1800	700	500	400	200

Table 4
Training and testing set partition for three class (positive, negative, and neutral) of emotions.

Training–testing partition (%)	No. of Records in the dataset (30000)			
	Total test records	Positive	Negative	Neutral
50–50	15,000	7000	5000	3000
60–40	12,000	6000	4000	2000
70–30	9000	4000	3000	2000
10-fold cross validation	3000	1000	1100	900

Table 5
Confusion matrix for negative emotion classification by LSTM_3 model.

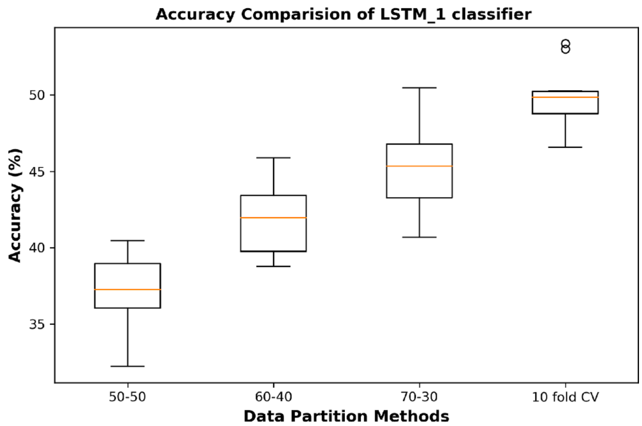
Validation technique		Classifier-LSTM_3			
		Sadness	Disgust	Anger	Surprise
50–50	Sadness	2300	200	250	250
	Disgust	200	2450	300	50
	Anger	50	300	1050	100
	Surprise	50	100	70	1280
		Sadness	Disgust	Anger	Surprise
60–40	Sadness	2310	200	240	250
	Disgust	70	1625	75	30
	Anger	50	60	950	40
	Surprise	20	80	120	1080
		Sadness	Disgust	Anger	Surprise
70–30	Sadness	1600	130	110	60
	Disgust	40	1430	50	80
	Anger	50	60	930	60
	Surprise	50	35	15	700
		Sadness	Disgust	Anger	Surprise
10-fold cross validation	Sadness	675	10	05	10
	Disgust	20	395	30	55
	Anger	10	05	378	07
	Surprise	11	09	10	170

Table 6
Confusion matrix for three class (positive, negative, and neutral) of emotion classification by LSTM_3 model.

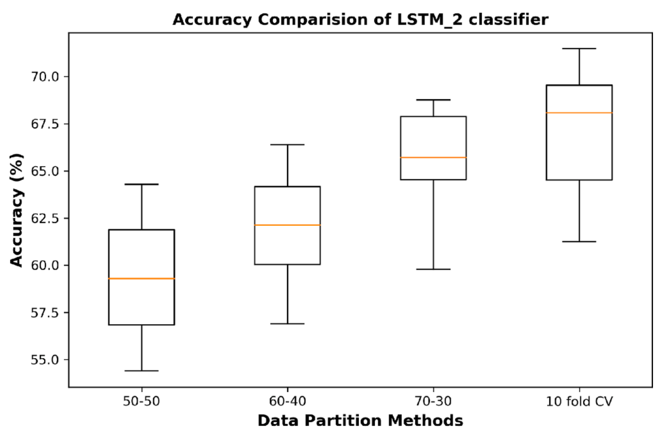
Validation technique		Classifier-LSTM_3		
		Positive	Negative	Neutral
50–50	Positive	5985	400	615
	Negative	340	3715	945
	Neutral	400	100	2500
		Positive	Negative	Neutral
60–40	Positive	4850	400	750
	Negative	200	3700	100
	Neutral	200	100	1700
		Positive	Negative	Neutral
70–30	Positive	3550	400	50
	Negative	100	2700	200
	Neutral	150	50	1800
		Positive	Negative	Neutral
10-fold cross validation	Positive	950	30	20
	Negative	20	980	100
	Neutral	30	20	850

The average accuracy is 59.29%, 62.12%, 65.70%, and 68.06% for all the dataset partition method.

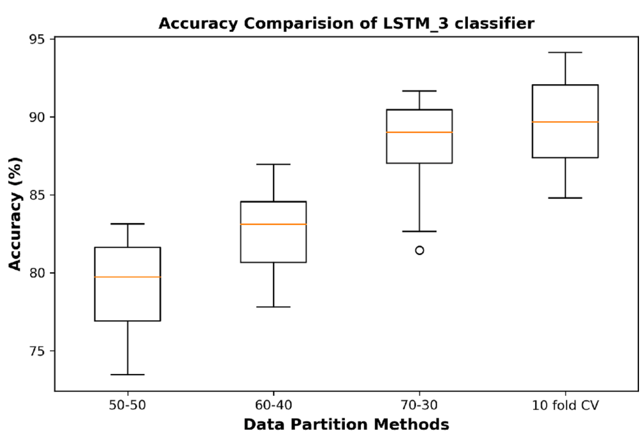
Fig. 7(c) shows the classification accuracy of the proposed LSTM_3 model in 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition minimum accuracy achieved is 73.45%, 77.78%, 81.45%, and 84.79%. The Maximum accuracy achieved for LSTM_3 is 83.12%, 86.94%, 91.67%, and 94.12% for all the mentioned training–testing partition. The average accuracy is 79.72%, 83.11%, 89.00%, and 89.67% for all the dataset partition method. Among these three



(a) LSTM_1 classification accuracy



(b) LSTM_2 classification accuracy



(c) LSTM_3 classification accuracy

Fig. 7. Three different LSTM model classification accuracy comparison.

architecture, LSTM_3 outperformed the other two models by achieving better accuracy and classification rate. the best LSTM architecture i.e., LSTM_3 is evaluated further on two on publicly available EEG dataset DEAP and SEED.

Among these three architecture LSTM_3 model outperformed the other two models by achieving average accuracy for four class of negative emotions as 78.67%, 82.85%, 86.30%, 89.83% for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition respectively. Also, this LSTM_3 model is implemented for the classification of positive, negative, and neutral class of emotions and achieved average class.

Table 5, shows the confusion matrix of our proposed LSTM_3 model for the classification of four class of negative emotions. Table 6, shows the confusion matrix of our proposed LSTM_3 model for the classification of three class of positive, negative, and neutral emotions. The TP and TN of our method are much higher than the other state-of-the-art methods, which show that our method is able to classify the correct samples properly.

Table 7 shows the recall, precision, F1-score, and class accuracy values of our proposed method. The recall is used to evaluate the performance of a classifier on minority class as it also contains essential information if a classifier has high recall value than it is able to classify minority class samples correctly. As, this dataset is also, unbalanced, which makes recall a critical parameter to evaluate the performance of our proposed LSTM based deep learning classifier. Table 7 shows that our method LSTM_3 based deep learning classifier has high recall values for 50–50, 60–40, 70–30, and 10-fold cross-validation which proves that our method is also classifying minority classes more accurately. As the number of records for the negative class anger and surprise is less, so it is important to give preference to the method having higher recall value for unbiased classification of emotions. Using the above results from Table 7, our method has produced outstanding results when compared with other methods which make it superior to other classifiers.

Table 7 also shows the precision value of our method. Precision evaluates how often the predictions made for the positive class are accurate. The findings demonstrate the evidence of comparability of our model with other techniques. As this dataset for negative emotion recognition is unbalanced so, accuracy is not a correct measure to evaluate the performance of unbalanced dataset classifier. Therefore, we also calculated the F1-score of our method. Table 7 represents the calculated F1-score using values from the confusion matrix of LSTM_3 method. The results mentioned above in the table shows that our methods F1-score is higher which proves that for unbalanced dataset classification, our method is more preferable than others.

Table 8 shows the comparison of specificity, PPV, and NPV values of our method for 10-fold cross-validation data partitioning scheme. Specificity evaluates the correctly recognized positive and negative emotions out of all the generated positive and negative emotions. It is noticeable that our method has a higher value of specificity value for each class classification. Which proves that it is able to classify negative emotions more accurately.

The evaluation results suggest that the proposed LSTM based deep learning model for classifying negative emotions is providing accurate

Table 7
Comparison of precision, recall, F1-score, and class accuracy of LSTM_3 model.

Validation technique	Classifier-LSTM_3				
	Class	Precision	Recall	F1-score	Class_Accuracy
50–50	Sadness	0.77	**0.88**	**0.82**	88.89
	Disgust	0.82	0.80	0.81	87.22
	Anger	0.70	0.63	0.66	88.11
	Surprise	**0.85**	0.76	0.81	**93.11**
60–40	Sadness	0.77	**0.94**	0.85	88.47
	Disgust	**0.90**	0.83	**0.86**	**92.85**
	Anger	0.86	0.68	0.76	91.74
	Surprise	0.83	0.77	0.80	92.50
70–30	Sadness	0.84	**0.92**	**0.88**	91.85
	Disgust	**0.89**	0.86	**0.88**	92.69
	Anger	0.85	0.84	0.84	93.61
	Surprise	0.88	0.78	0.82	**94.44**
10-fold cross validation	Sadness	**0.96**	**0.94**	**0.95**	**96.33**
	Disgust	0.79	**0.94**	0.86	92.83
	Anger	0.94	0.89	0.92	96.28
	Surprise	0.85	0.70	0.77	94.33

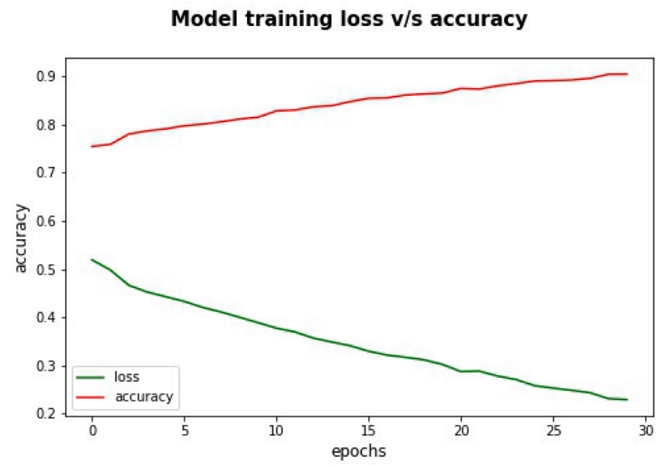
Table 8
Comparison of Specificity, PPV, and NPV of LSTM_3 model for 10-fold cross-validation partition scheme.

Validation technique	Classifier-LSTM_3			
	Class	Specificity (%)	PPV (%)	NPV (%)
		Mean ± Std	Mean ± Std	Mean ± Std
10-fold cross validation	Sadness	94.76 ± 2.96	94.27 ± 2.93	**96.74 ± 4.35**
	Disgust	95.62 ± 3.66	94.27 ± 3.29	83.31 ± 2.16
	Anger	**95.29 ± 4.86**	**94.50 ± 5.42**	92.47 ± 2.36
	Surprise	92.91 ± 2.36	89.47 ± 2.47	78.44 ± 3.22

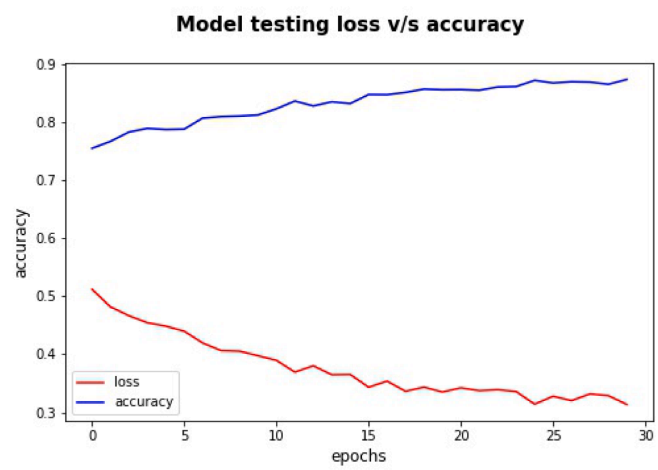
classification results. In Table 8 the positive prediction value and negative prediction values are also given which shows that the proposed method is able to recognize emotions accurately for four class of negative emotions. Along with this, the statistical analysis is also performed on the performance of LSTM based deep learning classifier, human behaviour is also analyzed of the population considered in this research. The population considered for this research work is non-clinical, means no subject is suffering from any mental-illness like stress, anxiety, depression, etc. The human behaviour is analyzed considering two parameters: age responsiveness and gender responsiveness. Next is the description of results obtained from statistical and human behaviour analysis.

The statistical result difference is illustrated in Table 9 with the two-tailed Mann Whitney test (Mann and Whitney, 1947). To compute the p-value comparison of classification accuracy, the Mann Whitney test is used. If p-values are higher than 0.05, the results do not differ significantly, and if the value of p is less than 0.001 than it is highly significant. From the interventions in Table 9, it is evident that for each training testing partition, the solution produced by our proposed approach is statistically distinct from MLP, KNN, SVM, and LIB-SVM. There is a significant difference in results when the p-values are compared for these classifiers with LSTM based deep learning classifier.

Fig. 8 presents the model training and testing loss versus accuracy plot for 10-fold cross-validation data partitioning schemes. Loss is defined as the difference between the predicted value by your model and the true value. The loss function used in our model is categorical cross-entropy. Accuracy is one of the metrics to measure the performance of our model. It is clear from Fig. 8 that the model has minimized the loss up-to 12.23% for 30 epochs in the testing phase of 10-fold cross-validation data partitioning scheme. Also, the results indicated that our model has achieved higher accuracy and the loss of our model is low in comparison with the accuracy which means that our model has achieved lower FP and FN. Fig. 8 results show that for the 10-fold cross-validation scheme the model has achieved average accuracy up-to 89.83% for 30 epochs in the testing phase of 10-fold cross-validation



(a) For 10 fold cross validation training



(b) For 10 fold cross validation testing

Fig. 8. LSTM_3 Model loss versus accuracy plot for 10-fold cross validation data partitioning.

scheme which indicates a higher rate of correct classification. Additionally, empirical testing found that 50 epochs for the training of units seemed best but further exploration is required to fine-tune this parameter.

In this study, we also perform experiments on three different age groups (18–25) years, (26–35) years, (36–55) years to analyze their response to negative emotions on targeted video clips. Each set of age participants are considered one by another to see which age group is responding more when their emotions are elicitate using stimuli. The results are presented in Fig. 9. It is clear from the results that age group

Table 9
p-value comparison for LSTM_3 based deep learning model using Mann-Whitney test

	Training-Testing Partition	MLP		KNN		SVM		LIB-SVM	
		p-value	Significance	p-value	Significance	p-value	Significance	p-value	Significance
	50-50	2.76×10^{-11}	Highly Significant	2.97×10^{-11}	Highly Significant	1.18×10^{-8}	Highly Significant	1.15×10^{-8}	Highly Significant
	60-40	3.652×10^{-11}	Highly Significant	4.623×10^{-11}	Highly Significant	1.27×10^{-8}	Highly Significant	1.15×10^{-8}	Highly Significant
LSTM_3	**70-30**	3.203×10^{-11}	Highly Significant	3.987×10^{-11}	Highly Significant	1.121×10^{-5}	Highly Significant	1.127×10^{-5}	Highly Significant
	10-fold cross validation	3.236×10^{-11}	Highly Significant	4.673×10^{-11}	Highly Significant	1.066×10^{-8}	Highly Significant	1.002×10^{-8}	Highly Significant

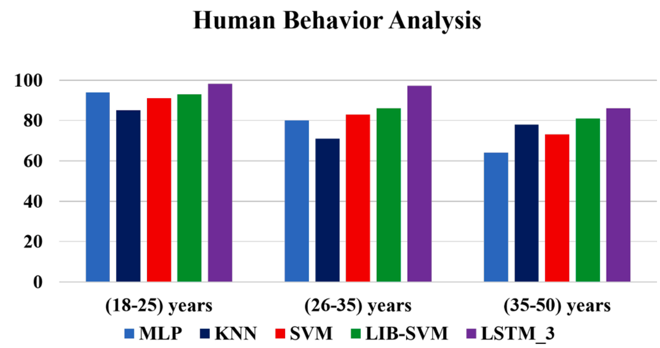


Fig. 9. Age groups based classification comparison of negative emotions.

from (18–25) years has the highest average emotion recognition accuracy, i.e., 93.12% for the five classifiers. Which leads to the conclusion that emotions are at a peak in the age group of (18–25) years. For the age group above (26–35) years, emotion recognition has an average accuracy of 81.41% and age group between (36–50) years has an accuracy of 76.45%, which is shown in Fig. 9. The emotions from clips sensed very well in people aged (18–25) years and (26–35) years. The identification level of emotions is not as high for the community of adults of the group 35–50 years age. The age group of 35–50 years is associated with selective decrements in the recognition of negative emotions. Another demonstrated reality in our studies that the intensity of emotions in the age bracket of 18–25 years is at its peak. A feature that our research has proven is that the strength of expressing emotions in age groups of 18–25 is at its pinnacle.

Fig. 10 shows a pie graph of the difference in gender-based responsiveness of subjects for targeted genres. From the results, we analyze that females are more responsive as compared to males and express their emotions. When Sad and disgust genres are shown to the participants, females responsiveness is more as compared to males. While, when the angry genre is shown to the participant's responsiveness of males is more as compared to females. Although all these genres belong to the negative class of emotion, there is still a difference in the responsiveness of participants. It is also verified from the results that sadness and angry emotions are easiest to classify by the classifiers used on the other side independently. However, we find out that among these targeting genres surprise genre's responsiveness for both the genders are approximately equal. In other words, we can say that females are more overt than males in responding to negative emotion targeting visual stimuli.

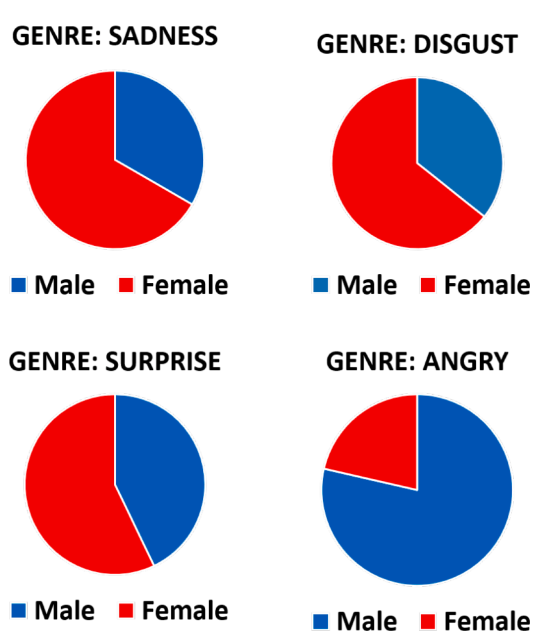


Fig. 10. Gender based responsiveness on targeted genres.

7.2. Comparison with other datasets

The approach suggested was also used to demonstrate the generality of our method for other datasets. The dataset used for comparison is DEAP (Koelstra et al., 2011) and SEED (Zheng and Lu, 2015) which are larger EEG dataset as compared to our dataset. Several trials on both DEAP and SEED datasets were intended and performed to verify the efficiency and generalization of our method. Table 10 shows the comparative statistical analysis of EEG datasets. Based on No. of subjects participated to give data, the gender of participation pool, EEG device having different no of electrodes, stimuli type and duration, how many class of emotions are used to create DEAP, SEED, and our dataset a statistical analysis is performed in the Table 10 for four and three class of emotions. Table 11 shows the comparison of classification accuracy of our LSTM based deep learning model for classification of four and three class of emotions on two publicly available datasets and two of our dataset. Datasets considered are DEAP dataset of 32 participants with four class of emotions, and SEED dataset of 15 participants with three class of emotions, and our dataset of 45 participants with four and three class of emotions. It is observed from the result that for four class of emotions classification for 50–50, 60–40, 70–30, and 10-fold cross-validation on our dataset LSTM based deep learning model has maximum classification accuracy as 83.12%, 86.94%, 91.67%, 94.12% and mean accuracy as 78.67%, 82.85%, 86.30%, and 89.89%. On DEAP dataset for four class of emotions classification 50–50, 60–40, 70–30, and 10-fold cross-validation LSTM based deep learning model have max. classification accuracy as 80.66%, 85.42%, 86.22%, and 91.38% and mean accuracy as 79.02%, 81.16%, 84.38%, and 88.42%. In case of SEED dataset for three class of emotions classification 50–50, 60–40, 70–30, and 10-fold cross-validation LSTM based deep learning model has maximum classification accuracy as 83.26%, 86.94%, 87.29%, and 89.34% and mean accuracy as 82.12%, 83.29%, 85.96%, and 87.22%. For three class of emotions classification for 50–50, 60–40, 70–30, and 10-fold cross-validation on our dataset LSTM based deep learning model has maximum classification accuracy as 81.33%, 85.41%, 89.44%, 92.66% and mean accuracy as 78.96%, 83.14%, 87.63%, and 88.49%. These results show that the LSTM based deep learning model is able to achieve better accuracy for other publicly available benchmark datasets too. The classification accuracy achieved by LSTM based deep learning model for classifying four and three class of emotions for 50–50, 60–40, 70–30, and 10-fold cross-validation on these four datasets in Table 11 shows the credibility and generalizability of our approach.

However, our dataset is collected from a 4-channel device while the DEAP dataset is collected from the 32-channel device, and SEED dataset is collected from the 62-channel device. As the number of channels increases the number of features extracted also increases. So, when a model classifies data collected from 4-channel, 32-channel, and 62-channel, its classification accuracy tends to be reduced gradually. The results show that our model is able to achieve very good accuracy even with four-channel EEG device, this proves the supremacy of our approach.

So we have reason to believe that our model's classification result is also reliable on other datasets, that also confirms the generalization of our model. The experimental results indicate that our method LSTM based deep learning model achieved better results than other available state-of-the-art methods and also on publicly available EEG datasets too. These findings demonstrate that our method is generalized and has the potential to handle complex real-life problems also.

8. Discussion

In this section, a detailed discussion is done on how the proposed LSTM based deep learning model performs in comparison with

Table 10
Statistical analysis of EEG datasets.

Dataset (Year)	No. of Subjects (M/F)	No. of electrodes	Stimuli Duration (s)	Stimuli	No. of emotions	Emotions
DEAP (2011)	32 (16/16)	32	60 s	Videos	4	High/Low Valence and High/Low arousal
Our Dataset (2020)	45 (23/22)	4	60–120 s	Videos	4	Sadness, Disgust, Anger, Surprise
SEED (2015)	15 (7/8)	62	240 s	Videos	3	Positive, Negative, Neutral
Our Dataset (2020)	45 (23/22)	4	60–120 s	Videos	3	Positive, Negative, Neutral

Table 11
Comparison of LSTM_3 model performance on other EEG datasets

	LSTM_3 classifier Accuracy on other EEG datasets							
	Our Dataset (Four Class)		DEAP Dataset (Four Class)		Our Dataset (Three Class)		SEED Dataset (Three Class)	
Partition	Max.	Mean	Max.	Mean	Max.	Mean	Max.	Mean
50–50	**83.12**	78.67	80.66	**79.02**	81.33	78.96	**83.26**	**82.12**
60–40	**86.94**	**82.85**	85.42	81.16	85.41	83.14	**86.94**	**83.29**
70–30	**91.67**	**86.30**	86.22	84.38	**89.44**	**87.63**	87.29	85.96
10-fold CV	**94.12**	**89.89**	91.38	88.42	**92.66**	**88.49**	89.34	87.22

traditional machine learning algorithms.

To demonstrate the advantages of introducing deep learning into four class of emotion classification a comparison with popular traditional classification algorithms is performed, using the same set of features as in the LSTM-based methodology. The other state-of-the-art approach considered for comparison are MLP, KNN, SVM, LIB-SVM. From the interventions in Table 12,13, it is evident that for each training testing partition, the solution produced by our proposed approach is statistically distinct from MLP, KNN, SVM, and LIB-SVM. Table 12 contains the classification accuracy comparison for four negative class of emotions. It contains the maximum, average, and minimum accuracy for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition respectively. For all classifiers, the feature extraction technique used is EMD and the same number of features are passed. The EMD offer an advantage over other signal analysis methods, like spectral analysis or wavelet transform since EMD is adaptive to the signal, whereas, in Fourier and wavelet transforms the basis are fixed. Hence, EMD allows extracting better features from non-stationary signals, such as EEG. This is also one of the causes that leads to a better performance of our model. The parameters and configurations for these variables were introduced on the same device to ensure that the findings and comparisons presented were unequivocal and univocal. It is clear from the Table 12 that among traditional machine learning classifier LSTM_3 model outperformed the classification accuracy by achieving maximum accuracy for four class of emotions as 83.12%, 86.94%, 91.67%, 94.12% for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition respectively. The second best classifier after LSTM based deep learning model is LIB-SVM it has achieved maximum accuracy for four class of emotions as 78.88%, 80.44%, 82.59%, 83.61% for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition.

Table 12
Classification accuracy comparison for four class of emotions.

Method	Validation Technique	Accuracy		
		Max	Avg	Min
MLP	50–50	70.21	69.03	68.56
KNN	50–50	64.10	63.22	61.73
SVM	50–50	71.44	70.11	69.78
LIB-SVM	50–50	78.88	76.51	75.26
LSTM_3 Classifier	50–50	**83.12**	**78.67**	**73.45**
MLP	60–40	72.48	71.22	70.12
KNN	60–40	68.62	65.61	64.24
SVM	60–40	74.86	72.54	71.89
LIB-SVM	60–40	80.44	78.92	76.73
LSTM_3 Classifier	60–40	**86.94**	**82.85**	**77.78**
MLP	70–30	75.94	72.43	71.84
KNN	70–30	70.09	69.91	67.26
SVM	70–30	77.48	74.28	73.33
LIB-SVM	70–30	82.59	80.44	78.43
LSTM_3 Classifier	70–30	**91.67**	**86.30**	**81.45**
MLP	10-Fold CV	77.27	71.21	70.59
KNN	10-Fold CV	72.88	70.08	71.25
SVM	10-Fold CV	81.83	76.23	79.15
LIB-SVM	10-Fold CV	83.61	81.68	79.86
LSTM_3 Classifier	10-Fold CV	**94.12**	**89.89**	**84.79**

The impact of the proposed deep learning-based methodology, in general, is higher as compared to traditional machine learning algorithms like MLP, KNN, SVM, and LIB-SVM. The results demonstrated that LSTM classifier has provided a significant increase in classification performance from +6.3% and up-to 20.4%. Apart from classification accuracy gains, LSTM classifier also has the ability to sustain above 92.65% specificity that quantitatively results in having very low false prediction rate.

Table 13 presents the comparison of classification accuracies of different related work studies considered with our approach. Based on the number of emotions considered, channels of EEG device, data partition method used, along with the name of the dataset used for experiments are described. Three datasets are considered for this comparison named as DEAP, SEED, and own Created dataset by the authors. Also, based on the data partition method used by the different studies simply makes the comparison easier to analyze the accuracy achieved by our proposed method and related state-of-the-art methods. In Table 13 method used by the different studies are provided with a superscript having numbers 1, 2, 3, 4, 5, 6, and 7. Where these numbers refer to the type of data partitioning method used in the approach proposed by the authors of different studies. Result for 50–50, 60–40, 70–30, 80–20, and 90–10 training–testing data partition method it is represented by 1, 2, 3, 5, and 6. For 10-fold cross-validation and 8-fold cross-validation, it is represented by 4 and 7.

The results indicate that the structure of LSTM based deep learning network classifier is significantly better in terms of classification accuracy than other approaches. The authors of these works have not indicated whether the findings are the highest accuracy attained by their classification approach or the average accuracy of several runs or in which training and testing partition the data is segregated. On our created dataset of four class of emotions proposed method to achieve maximum accuracy as 83.12%, 86.94%, 91.67%, 94.12% for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing partition respectively for 100 epochs of LSTM based deep learning classifier as shown in Table 13. For three class of emotions proposed method to achieve maximum accuracy as 81.33%, 85.41%, 89.44%, 92.66% for 50–50, 60–40, 70–30, and 10-fold cross-validation training–testing

Table 13
Comparison of classification accuracies of different studies with LSTM_3 model.

Author Year	Number of Electrodes	Number of Emotions	Method	Dataset	Classification Accuracy (%)
Koelstra et al. (2010)	32	4	SVM[d]	Created	57.00
Bastos-Filho et al. (2012)	32	2	KNN[d]	DEAP	69.50
Wijeratne and Perera (2012)	32	3	SVM[c]	DEAP	80.00
Wijeratne and Perera (2012)	32	3	SVM[e]	DEAP	75.00
Anh et al. (2012)	14	5	SVM[d]	Created	70.50
Daimi and Saha (2014)	32	4	SVM[d]	DEAP	67.75
Jirayucharoensak et al. (2014)	32	3	SVM[d]	DEAP	49.52
Liu et al. (2016)	32	4	Bimodal Deep AutoEncoder[f]	DEAP	83.25
Atkinson and Campos (2016)	32	3	GA-SVM[g]	DEAP	62.33
Atkinson and Campos (2016)	32	2	GA-SVM[g]	DEAP	73.14
Atkinson and Campos (2016)	32	5	GA-SVM[g]	DEAP	46.69
Bhatti et al. (2016)	1	4	MLP[d]	Created	78.11
Ackermann et al. (2016)	32	3	Random Forest[d]	DEAP	55.23
Yin et al. (2017)	32	2	Linear SVM[d]	DEAP	78.00
Zhuang et al. (2017)	32	2	SVM[d]	DEAP	71.99
Mert and Akan (2018)	32	4	KNN[d]	DEAP	67.00
Nakisa et al. (2018)	5	4	Probabilistic Neural Network[d]	DEAP	67.47
Song et al. (2018)	62	3	DGCNN[d]	SEED	79.95
Qing et al. (2019)	32	4	KNN + DT + Random Forest[d]	DEAP	62.63
Qing et al. (2019)	62	4	KNN + DT + Random Forest[d]	SEED	74.85
Zhong et al. (2020)	62	2	RGNN[d]	SEED	85.30
Acharya et al. (2020)	1	2	GP Tree[a]	Created	80.14
Acharya et al. (2020)	1	2	GP Tree[b]	Created	80.76
Acharya et al. (2020)	1	2	GP Tree[c]	Created	81.87
Acharya et al. (2020)	1	2	GP Tree[d]	Created	85.97
Cimtay and Ekmekcioglu (2020)	62	2	CNN[d]	SEED	86.50
Cimtay and Ekmekcioglu (2020)	62	3	CNN[d]	SEED	78.30
Cimtay and Ekmekcioglu (2020)	32	2	CNN[d]	DEAP	72.80
This study	4	4	LSTM_3[a]	Created	**83.12**
This study	4	4	LSTM_3[b]	Created	**86.94**
This study	4	4	LSTM_3[c]	Created	**91.67**
This study	4	4	LSTM_3[d]	Created	**94.12**
This study	4	4	LSTM_3[a]	DEAP	**80.66**
This study	4	4	LSTM_3[b]	DEAP	**85.42**
This study	4	4	LSTM_3[c]	DEAP	**86.22**
This study	4	4	LSTM_3[d]	DEAP	**91.38**
This study	4	3	LSTM_3[a]	SEED	**83.26**
This study	4	3	LSTM_3[b]	SEED	**86.94**
This study	4	3	LSTM_3[c]	SEED	**87.29**
This study	4	3	LSTM_3[d]	SEED	**89.34**
This study	4	3	LSTM_3[a]	Created	**81.33**
This study	4	3	LSTM_3[b]	Created	**85.41**
This study	4	3	LSTM_3[c]	Created	**89.44**
This study	4	3	LSTM_3[d]	Created	**92.66**

[a] Result for 50–50 training–testing data.
[b] Result for 60–40 training–testing data.
[c] Result for 70–30 training–testing data.
[d] Result for 10-fold cross validation.
[e] Result for 80–20 training–testing data.
[f] Result for 90–10 training–testing data.
[g] Result for 8-fold cross validation.

partition respectively for 100 epochs of LSTM based deep learning classifier as shown in Table 13.

On DEAP dataset for 50–50, 60–40, 70–30, and 10-fold cross-validation proposed LSTM based deep learning model has maximum classification accuracy as 80.66%, 85.42%, 86.22%, and 91.38%. While analyzing the results it is clear that most of the other research work has used 10-fold cross-validation method to do the splitting between training and testing of data which can be considered as a standard method to compare different approaches using the same datasets. In Table 13 LSTM based method has achieved a classification accuracy of 91.38% on DEAP dataset followed by the approach proposed by Liu et al. (2016) using Bimodal Deep Auto Encoder and achieved a classification accuracy of 83.25%.

On SEED dataset for 50–50, 60–40, 70–30, and 10-fold cross-validation our proposed LSTM based deep learning model has attained a maximum classification accuracy as 83.26%, 86.94%, 87.29%, and 89.34%. Followed by the work proposed by Cimtay and Ekmekcioglu (2020) using dense convolutional neural network on SEED dataset and achieved a classification accuracy of 86.50%.

The reason behind the supremacy of the proposed model is that LSTM's worked efficiently at extracting patterns in input feature space, where the input data spans over long sequences as we have EEG dataset. The gated architecture of LSTM's used in the paper has the ability to manipulate its memory state. Also, LSTM model is able to handle data with multiple input variables as EEG dataset has many features and also they are non-linear signals. It can be seen from the results that as the number of emotions increases from two to three or more the classification accuracy decreases as the complexity increases, but in this scenario also our proposed approach has outperformed the other methods as it has the capacity to handle data with multiple input variables and features.

One of the main issues for EEG-based recognition of emotions is whether emotion for each subject is credible and stable at varying times. Every subject was asked to take this experiment twice in intervals of 1 Week to find a solution to this problem. After this data acquisition again, the whole process is repeated to recognize emotions. The findings show that our model has attained equivalent prediction accuracy for the twice tests that were conducted in a 1-week difference per subject, even though having a difference in the time slot and the clips shown to them. Limitation of this study is that only one feature extraction technique is used in future we will use the hybrid feature extraction method and also utilize more strategies such as feature smoothing and deep network like convolutional neural network and hybrid LSTM network along with convolutional neural network to improve the performance of our framework.

These results indicate the reliability of LSTM based deep learning classifier that is able to classify four and three class of emotions (positive, negative, and neutral) dataset using brain signals.

9. Conclusion

In the duration of this work, we propose brain signals based human emotion recognition system using LSTM based deep learning model, which is analyzing the emotional conditions of participants through brain wave assessment. Our system consists of four different modules: emotion elicitation, EEG data collection, feature extraction, and emotion classification. For this, an LSTM based deep learning model classifier was used for classification after feature extraction through EMD, which is responsible for the resulting high accuracy over existing state-of-the-art methods. To validate our results, we performed multiple experiments, which are valuable in comparing our LSTM based approach with the existing methods. These results indicate a benefit in terms of accuracy and the potential to recognize several comparable, discrete emotions over the current state-of-the-art methods using brain signals for emotions recognition. As this dataset is unbalanced, so the value of recall is also explained as it defines that the classifier is either able to classify minority classes or not. Comparison of recall, precision, F1 score, and specificity shows that our method outperformed the existing literature work. This research work has also considered human behaviour analysis. It is also verified from the results that sadness and angry emotions are easiest to classify by the classifiers used on the other side independently. It is also observed that surprise genre generated approximately equal response from both the participating subjects under study. Brain signals of age group 26–35 years are recognized to provide the highest precision of emotional identification in compliance with self-reports. For gender-based responsiveness, analysis results have proven that among the considered genders, females tend to respond more as compared to males. This system uniquely uses MUSE 2 a commercially available device for reading EEG signals. In general, the findings of the suggested LSTM based deep learning methodology has confirmed to outperform the outcomes of current state-of-the-art methods and is also able to classify unbalanced datasets.

However, considering the restricting elements of these studies, it is essential to notice that only the EMD feature extraction technique is considered. There are other feature extraction techniques like Fast Fourier transformation (FFT), Ant Colony Optimization, Simulated Annealing, Genetic Algorithm, and Particle Swarm Optimization, which can be considered. Despite the promising results, this paper has identified a significant limitation that is the portability of EEG devices and also as the number of emotions recognized increases the classification accuracy decreases which are also validated in the Table 13 where comparison of classification accuracies of different state-of-the-art methods with proposed LSTM based approach is shown.

This will enable us to open up many research opportunities in a different framework. In future, the data acquisition process can be enhanced with the help of multi-channel EEG devices. The number of participants in this research work is 45 only; it would be interesting to see the behavior of the proposed classifier with more number of participants. Also, multiple classes of participants can be considered to check the response and behavior of participants. Moreover, in future, the response of participants concerning videos and audio on different age groups considering both clinical and non-clinical populations can be analyzed. However, for real-world application in our study, a four-channel MUSE 2 is used to recognize four class of negative emotions and three class of positive, negative, and neutral emotions while watching emotional clips. The results with classification accuracy for the positive, negative, and neutral class of emotions for 50–50, 60–40, and 70–30, and 10-fold cross-validation training–testing partition validate the use and feasibility of EEG signals for emotion recognition in this domain. For future research work, EEG devices classification accuracy for multi-channel portable devices along with multi-class of emotions will be carried out to check the performance of LSTM classifier as the problem becomes complex. Also, the complex architecture of LSTM based deep learning model along with variations in the architecture will be carried out.

CRediT authorship contribution statement

Aditi Sakalle: Conceptualization, Investigation, Resources, Methodology, Writing - original draft, Formal analysis, Writing - review & editing. **Pradeep Tomar:** Data curation, Formal analysis, Resources, Writing - review & editing, Methodology, Software. **Harshit Bhardwaj:** Data curation, Formal analysis, Resources. **Divya Acharya:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - original draft. **Arpit Bhardwaj:** Investigation, Project administration, Supervision, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Acharya, D., Billimoria, A., Srivastava, N., Goel, S., & Bhardwaj, A. (2020). Emotion recognition using fourier transform and genetic programming. *Applied Acoustics, 164*, Article 107260.

Acharya, D., Goel, S., Asthana, R., & Bhardwaj, A. (2020). A novel fitness function in genetic programming to handle unbalanced emotion recognition data. *Pattern Recognition Letters*.

Acharya, D., Goel, S., Bhardwaj, H., Sakalle, A., & Bhardwaj, A. (2020). A long short term memory deep learning network for the classification of negative emotions using eeg signals. In *2020 International joint conference on neural networks (IJCNN)* (pp. 1–8). IEEE.

Acharya, D., Varshney, N., Vedant, A., Saxena, Y., Tomar, P., Goel, S., & Bhardwaj, A. (2021). An enhanced fitness function to recognize unbalanced human emotions data. *Expert Systems with Applications, 166*, 114011.

Acharya, Divya, Jain, Riddhi, Panigrahi, Siba Smarak, Sahni, Rahul, Jain, Siddhi, Deshmukh, Sanika Prashant, & Bhardwaj, Arpit (2021). Multi-class Emotion Classification Using EEG Signals*, 1367. Advanced Computing. IACC 2020, Communications in Computer and Information Science* (pp. 3–16). Singapore: Springer, 978-981-16-0401-0, In this issue.

Ackermann, P., Kohlschein, C., Bitsch, J.Á., Wehrle, K., & Jeschke, S. (2016). Eeg-based automatic emotion recognition: Feature extraction, selection and classification methods. In *2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom)* (pp. 1–6). IEEE.

Al-Shargie, F., Kiguchi, M., Badruddin, N., Dass, S. C., Hani, A. F. M., & Tang, T. B. (2016). Mental stress assessment using simultaneous measurement of eeg and fnirs. *Biomedical Optics Express, 7*(10), 3882–3898.

Albanie, S., Nagrani, A., Vedaldi, A. & Zisserman, A. (2018). Emotion recognition in speech using cross-modal transfer in the wild. arXiv preprint arXiv:1808.05561.

Anh, V. H., Van, M. N., Ha, B. B., & Quyet, T. H. (2012). A real-time model based support vector machine for emotion recognition through EEG. In *2012 International conference on control, automation and information sciences (ICCAIS)* (pp. 191–196). IEEE.

Atkinson, J., & Campos, D. (2016). Improving bci-based emotion recognition by combining eeg feature selection and kernel classifiers. *Expert Systems with Applications, 47*, 35–41.

Bajaj, V., & Pachori, R. B. (2011). Classification of seizure and nonseizure eeg signals using empirical mode decomposition. *IEEE Transactions on Information Technology in Biomedicine, 16*(6), 1135–1142.

Bashivan, P., Rish, I., Yeasin, M. & Codella, N. 2015. Learning representations from eeg with deep recurrent-convolutional neural networks. arXiv preprint arXiv: 1511.06448.

Bastos-Filho, T. F., Ferreira, A., Atencio, A. C., Arjunan, S., & Kumar, D. (2012). Evaluation of feature extraction techniques in emotional state recognition. In *2012*

4th international conference on intelligent human computer interaction (IHCI) (pp. 1–6). IEEE.
Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks, 5*(2), 157–166.
Bhardwaj, A., Tiwari, A., Varma, M. V., & Krishna, M. R. (2014). Classification of eeg signals using a novel genetic programming approach. In *Proceedings of the companion publication of the 2014 annual conference on genetic and evolutionary computation* (pp. 1297–1304).
Bhardwaj, A., Tiwari, A., Varma, M. V. & Krishna, M. R. 2015. An analysis of integration of hill climbing in crossover and mutation operation for eeg signal classification. In Proceedings of the 2015 annual conference on genetic and evolutionary computation (pp. 209–216). ACM.
Bhardwaj, A., Tiwari, A., Krishna, R. & Varma, V. (2016). A novel genetic programming approach for epileptic seizure detection. Computer Methods and Programs in Biomedicine, 124, 2–18.
Bhardwaj, H., Sakalle, A., Bhardwaj, A., & Tiwari, A. (2019). Classification of electroencephalogram signal for the detection of epilepsy using innovative genetic programming. *Expert Systems, 36*(1), Article e12338.
Bhatti, A. M., Majid, M., Anwar, S. M., & Khan, B. (2016). Human emotion recognition and analysis in response to audio music using brain signals. *Computers in Human Behavior, 65*, 267–275.
Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.
Chen, T., Ju, S., Yuan, X., Elhoseny, M., Ren, F., Fan, M., & Chen, Z. (2018). Emotion recognition using empirical mode decomposition and approximation entropy. *Computers & Electrical Engineering, 72*, 383–392.
Cheng, B., & Liu, G. (2008). Emotion recognition from surface emg signal using wavelet transform and neural network. In *Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)* (pp. 1363–1366).
Cimtay, Y., & Ekmekcioglu, E. (2020). Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition. *Sensors, 20*(7), 2034.
Clifford, G., Hitchcock, C. & Dalgleish, T. 2020. Negative and positive emotional complexity in the autobiographical representations of sexual trauma survivors. Behaviour Research and Therapy, 103551.
Daimi, S. N., & Saha, G. (2014). Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications, 41*(13), 6057–6065.
Davidson, R. J. (1994). On emotion, mood, and related affective constructs. The nature of emotion: Fundamental questions (pp. 51–55).
Day, R.-F., Lin, C.-H., Huang, W.-H., & Chuang, S.-H. (2009). Effects of music tempo and task difficulty on multi-attribute decision-making: An eye-tracking approach. *Computers in Human Behavior, 25*(1), 130–143.
Deb, S., Strodl, E., & Sun, J. (2015). Academic stress, parental pressure, anxiety and mental health among indian high school students. *International Journal of Psychology and Behavioral Sciences, 5*(1), 26–34.
Diez, P. F., Mut, V., Laciar, E., Torres, A., & Avila, E. (2009). Application of the empirical mode decomposition to the extraction of features from eeg signals for mental task classification. In *2009 Annual international conference of the IEEE engineering in medicine and biology society* (pp. 2579–2582). IEEE.
Gers, F. A., Schmidhuber, J. & Cummins, F. (1999). Learning to forget: Continual prediction with lstm.
Hossain, M. S., & Muhammad, G. (2017). An emotion recognition system for mobile applications. *IEEE Access, 5*, 2281–2287.
Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. -C., Tung, C. C. & Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454 (1971), 903–995.
IMDb. Movies list. 2018.
Jirayucharoensak, S., Pan-Ngum, S., & Israsena, P. (2014). Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation. *The Scientific World Journal, 2014*.
Kartelj, A., Filipović, V., & Milutinović, V. (2012). Novel approaches to automated personality classification: Ideas and their potentials. In *2012 Proceedings of the 35th international convention MIPRO* (pp. 1017–1022). IEEE.
Khosrowabadi, R., Quek, C., Ang, K. K., Tung, S. W., & Heijnen, M. (2011). A brain-computer interface for classifying eeg correlates of chronic mental stress. In *The 2011 international joint conference on neural networks* (pp. 757–762). IEEE.
Kim, J., & André, E. (2008). Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30*(12), 2067–2083.
Kim, Y., Lee, H., & Provost, E. M. (2013). Deep learning for robust feature generation in audiovisual emotion recognition. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 3687–3691). IEEE.
Koelstra, S., Yazdani, A., Soleymani, C., Mohammad and, Lee, J. -S., Nijholt, A., Pun, T., Ebrahimi, T. & Patras, I. (2010). Single trial classification of eeg and peripheral physiological signals for recognition of emotions induced by music videos. In International conference on brain informatics (pp. 89–100).
Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing, 3*(1), 18–31.
Lai, H.-L., & Li, Y.-M. (2011). The effect of music on biochemical markers and self-perceived stress among first-line nurses: A randomized controlled crossover trial. *Journal of Advanced Nursing, 67*(11), 2414–2424.
Lalitha, S., & Tripathi, S. (2016). Emotion detection using perceptual based speech features. In *2016 IEEE annual India conference (INDICON)* (pp. 1–5). IEEE.
Lee, K. & Cho, M. (2011). Mood classfication from musical audio using user group-dependent models. In 2011 10th International conference on machine learning and applications and workshops (Vol. 2, pp. 130–135).
Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R., & Chen, J.-H. (2010). Eeg-based emotion recognition in music listening. *IEEE Transactions on Biomedical Engineering, 57*(7), 1798–1806.
Liu, W., Zheng, W.-L., & Lu, B.-L. (2016). Emotion recognition using multimodal deep learning. In *International conference on neural information processing* (pp. 521–529). Springer.
Liu, Y.-J., Yu, M., Zhao, G., Song, J., Ge, Y., & Shi, Y. (2017). Real-time movie-induced discrete emotion recognition from eeg signals. *IEEE Transactions on Affective Computing, 9*(4), 550–562.
Logemann, H. A., Lansbergen, M. M., Van Os, T. W., Böcker, K. B., & Kenemans, J. L. (2010). The effectiveness of eeg-feedback on attention, impulsivity and eeg: A sham feedback controlled study. *Neuroscience Letters, 479*(1), 49–53.
Mann, H. B. & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 50–60.
Mert, A., & Akan, A. (2018). Emotion recognition from eeg signals by using multivariate empirical mode decomposition. *Pattern Analysis and Applications, 21*(1), 81–89.
Mohammadi, Z., Frounchi, J., & Amiri, M. (2017). Wavelet-based emotion recognition system using eeg signal. *Neural Computing and Applications, 28*(8), 1985–1990.
Nakisa, B., Rastgoo, M. N., Tjondronegoro, D., & Chandran, V. (2018). Evolutionary computation algorithms for feature selection of eeg-based emotion recognition using mobile sensors. *Expert Systems with Applications, 93*, 143–155.
Ng, H. -W., Nguyen, V. D., Vonikakis, V. & Winkler, S. (2015). Deep learning for emotion recognition on small datasets using transfer learning. In Proceedings of the 2015 ACM on international conference on multimodal interaction (pp. 443–449). ACM.
Niemic, C. (2004). Studies of emotion: A theoretical and empirical review of psychophysiological studies of emotion.
Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in pet and fmri. *Neuroimage, 16*(2), 331–348.
Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 10*, 1175–1191.
Qing, C., Qiao, R., Xu, X., & Cheng, Y. (2019). Interpretable emotion recognition using eeg signals. *IEEE Access, 7*, 94160–94170.
Rao, V. P., Puwakpitiyage, C. H., Azizi, M. M., Tee, W., Murugesan, R., & Hamzah, M. (2018). Emotion recognition in e-commerce activities using eeg-based brain computer interface. In *2018 Fourth international conference on advances in computing, communication & automation (ICACCA)* (pp. 1–5). IEEE.
San-Segundo, R., Gil-Martín, M., D'Haro-Enríquez, L. F., & Pardo, J. M. (2019). Classification of epileptic eeg recordings using signal transforms and convolutional neural networks. *Computers in Biology and Medicine, 109*, 148–158.
Schuller, B., Arsic, D., Rigoll, G., Wimmer, M., & Radig, B. (2007). Audiovisual behavior modeling by combined feature spaces. In *2007 IEEE international conference on acoustics, speech and signal processing-ICASSP'07, 2* p. 733.). IEEE.
Scott, M. J., Guntuku, S. C., Huan, Y., Lin, W. & Ghinea, G. (2015). Modelling human factors in perceptual multimedia quality: On the role of personality and culture. In Proceedings of the 23rd ACM international conference on multimedia (pp. 481–490). ACM.
Song, T., Zheng, W., Song, P., & Cui, Z. (2018). Eeg emotion recognition using dynamical graph convolutional neural networks. *IEEE Transactions on Affective Computing*.
Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional state classification from eeg data using machine learning approach. *Neurocomputing, 129*, 94–106.
Wijeratne, U., & Perera, U. (2012). Intelligent emotion recognition system using electroencephalography and active shape models. In *2012 IEEE-EMBS conference on biomedical engineering and sciences* (pp. 636–641). IEEE.
Wu, Z., Huang, N. E., Long, S. R., & Peng, C.-K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences, 104*(38), 14889–14894.
Yin, Z., Wang, Y., Liu, L., Zhang, W., & Zhang, J. (2017). Cross-subject eeg feature selection for emotion recognition using transfer recursive feature elimination. *Frontiers in Neurorobotics, 11*, 19.
Zaremba, W., Sutskever, I. & Vinyals, O. (2014). Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.
Zhang, Y.-D., Yang, Z.-J., Lu, H.-M., Zhou, X.-X., Phillips, P., Liu, Q.-M., & Wang, S.-H. (2016). Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access, 4*, 8375–8385.
Zhang, Y., Liu, Y., Weninger, F., & Schuller, B. (2017). Multi-task deep neural network with shared hidden layers: Breaking down the wall between emotion representations. In *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 4990–4994). IEEE.
Zhao, L., & He, Y. (2013). Power spectrum estimation of the welch method based on imagery eeg. *Applied Mechanics and Materials, 278*, 1260–1264.
Zheng, W.-L., & Lu, B.-L. (2015). Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on Autonomous Mental Development, 7*(3), 162–175.

Zheng, W.-L., Liu, W., Lu, Y., Lu, B.-L., & Cichocki, A. (2018). Emotionmeter: A multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics, 49*(3), 1110–1122.
Zhong, P., Wang, D., & Miao, C. (2020). Eeg-based emotion recognition using regularized graph neural networks. *IEEE Transactions on Affective Computing*.
Zhuang, N., Zeng, Y., Tong, L., Zhang, C., Zhang, H., & Yan, B. (2017). Emotion recognition from eeg signals using multidimensional information in emd domain. *BioMed Research International*.