Electroencephalogram (EEG) brainwave signal-based emotion recognition using extreme gradient boosting algorithm

Article · November 2020

CITATIONS
2

READS
299

4 authors, including:

Taki Hasan Rafi
Hallym University
15 PUBLICATIONS **17** CITATIONS
SEE PROFILE

Faisal Farhan
Ahsanullah University of Science & Tech
5 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Md.Ziaul Hoque
Ahsanullah University of Science & Tech
3 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Annals of Engineering

http://www.imperialopen.com/index.php/AOE

ImperialOpen
peer-review journals

ORIGINAL RESEARCH – Biomedical Engineering

Electroencephalogram (EEG) brainwave signal-based emotion recognition using extreme gradient boosting algorithm

Taki Hasan Rafi*, Faisal Farhan, Md. Ziaul Hoque, Farhan Mohd Quayyum

Department of Electrical and Electronic Engineering, Ahsanullah University of Science and Technology, 141 & 142, Love Road, Dhaka 1208, Bangladesh.

RECEIVED
09 August 2020

REVISED
03 November 2020

ACCEPTED
04 November 2020

PUBLISHED
27 November 2020

DOI: To be assigned.

* Corresponding author.

Emails:

takihassanrafi@gmail.com (T.H.)
faisalfarhaaan@gmail.com (F.F.)
ziaul1769@gmail.com (Z.H.)
mdfarhan96@gmail.com (F.Q.)

Abstract

Emotion recognition based on electroencephalogram has been a lucrative task nowadays. Different procedures are executed to improve the computational power of emotion recognition frameworks utilizing electroencephalogram (EEG). There are several computational strategies that have performed incredibly well in emotion recognition task. An automated recognition system is typically limited to few emotions. So more effective and accurate emotion classification is the primary objective for Artificial Intelligence researchers. This paper aims to classify three emotion recognition conditions, such as positive, negative, and neutral. To reduce the over-fitting dispute, principal component analysis (PCA) is applied to extricate the most significant parts of input features. Besides, covariate move adjustment of the essential parts is executed to limit the non-stationary impact of EEG signals. Several machine learning models have been deployed to classify this problem. In this case, traditional algorithms such as logistic regression, linear support vector machine, random forest, artificial neural network, long short term memory (LSTM) and extreme gradient boosted (XGBoost) classifier have been enforced on the EEG brainwave signal dataset. The XGBoost has performed outstandingly in terms of accuracy and less time complexity.

Keywords: EEG; PCA; XGBoost; Emotion recognition; LSTM.

1. Introduction

Brain Computer Interface (BCI) revolutionized both the way we communicate with machines and the way we communicate with each other. Over the years, researchers utilized several technologies for building Brain-Computer Interfaces (Atkinson & Campos, 2016). Through the advancement of Machine Learning, Deep Learning and Artificial Intelligence technology, Brain-Computer Interface made promising development in biomedical engineering research fields. Humans can modulate their brain waves and control machines. This has become possible due to BCI (Jirayucharoensak et al., 2014). Different techniques are used in BCI technology to interface directly with the human brain for controlling external devices like computer games or other assistive devices (Henshaw et al., 2019). It also has several medical applications. One of such applications of BCI is automatic emotion recognition. Emotions basically refers to the mental state of a person linked with his/her feelings, thoughts which plays a vital role in expressing and decision making (Garg et al., 2019). Several researches have been done over the years based on human speech, their body gesture, facial expressions for emotion recognition. However, physiological signals generated from the body to detect emotions emerged as the most efficient way for emotion analysis (Garg et al., 2019). These physiological signals render more advantages than physical data because they are spontaneous, highly involuntary in nature and insusceptible to social masking of emotions (Nakisa et al., 2018, Garg et al., 2019). Human emotions can be predicted from many physiological signals like electrocardiogram

(ECG), electro dermal activity (EDA), blood volume pulse (BVP), electroencephalogram (EEG) and many more. Among all these, EEG signals have received much attention in the last decade and research are still ongoing using EEG signals to detect emotions (Bilucaglia et al., 2020).

An electroencephalogram (EEG) test is used for assessing the electrical activity in the human brain. It can also detect abnormalities in brain waves. Brain cells usually interact with each other through electrical impulses. Brain electric potentials are produced as ionic current flows in between the brain neurons. EEG signals are extracted by recording the oscillation of these signals which are measured using electrodes implanted on the scalp. All of these extracted signals refer to brain activity waves in the electrode placed regions. Based on the frequency and object conditions EEG signals are classified to five waves namely alpha, beta, delta, theta, and gamma waves (Jirayucharoensak et al., 2014). We can identify the state of human emotions from the frequency bandwidth of these waves. Alpha waves range from 8-13Hz referring that human mental state is in relax condition. Beta waves ranges from 14-30Hz and shows up when human is stressed or thinking something. Delta waves usually have frequency lower than 3.5Hz indicating that the human is sleeping. Theta wave ranges from 4-8Hz and it indicates that the human is sleepy (Hamada et al., 2018). And gamma waves range from 26-70 Hz. Neurophysiology and Psychology research has proved that EEG can identify several brain electrical activity and brain functional state. Activity and information regarding the human emotional state can also be extracted through EEG which are very much affiliated to the activity of the cerebral cortex (Chen et al., 2019). Emotion classification using EEG achieved the most appreciation as it provides strong objectivity, high time resolution, high accuracy classification, satisfactory spatial resolution as well as cheap and favourable acquisition using EEG portable collection devices (Zeng et al., 2019). EEG signals are mostly used to automatically recognize different emotions as it has been found out that there lies a strong correlation between these signals and emotions such as sadness, anger, surprise, happiness. The reason why there lies a strong correlation between several emotions and EEG signals is because the signals are extracted straight from the Central Nervous System (CNS) that captures several features about internal emotion states (Nakisa et al., 2018). Automatic emotion recognition techniques have provided promising success after affiliating with EEG based emotion classification. Research on EEG-based emotion recognition algorithm is also rising in numbers with time. Although efficiency regarding these algorithms has some boundaries (Jirayucharoensak et al., 2014). However, many Machine Learning and Deep Learning models are being introduced to increase performance and to avoid the complexity of feature extraction, feature selection and other complex features thus providing high real-time performance (Chen et al., 2019). Fig. 1 illustrated an intuition of EEG-brainwave recording.

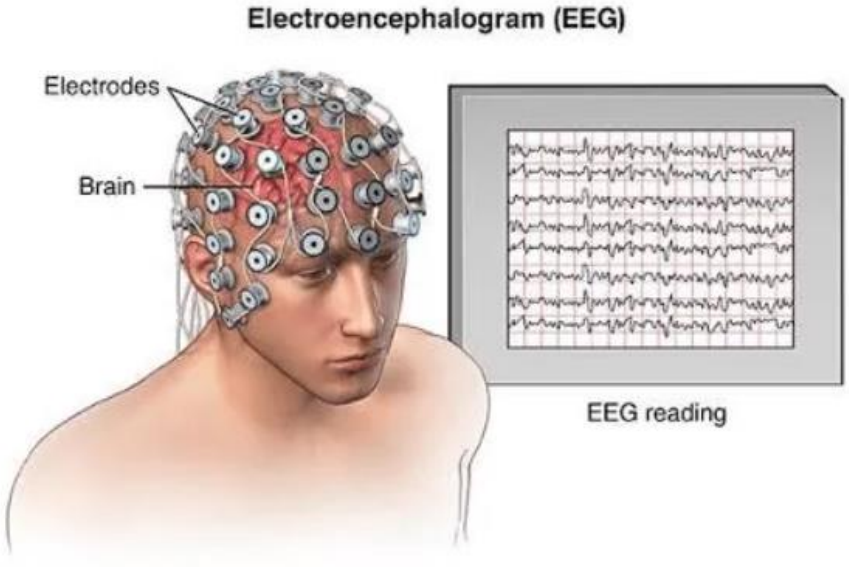

Fig. 1: EEG-brainwave recording using electrodes.

Several advanced machine learning modules have been applied recently for emotion recognition and they have achieved empirical success. Good classification results in emotion recognition has been obtained using some traditional machine learning techniques (Nakisa et al., 2018, Chen et al., 2019). Some machine learning models, Logistic Regression, Support Vector Machine, Random Forest, XGBoost and many more can be fruitfully applied in signal classification problems. XGBoost achieved much popularity among the researchers because of its execution speed and model performance. Though all the existing machine learning models have achieved success in emotion recognition sector, there still exist some limitations such as lacking learning inheriting characteristic of training samples, severe over-fitting or poor

accuracy in classification (Zeng et al., 2019). Contrarily Deep Learning models have achieved great success and proved to be more efficient in solving complex tasks. Some deep learning models like Convolution Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network are being widely used in video tracking, speech sequence modelling, image classification. They are more feasible as they are capable to learn the complex features using their strong end to end ability of self-learning (Chen et al., 2019). Convolution neural network provides significant advantages over traditional Machine learning modules and are widely being applied in EEG emotion recognition field (Zeng et al., 2019). Recurrent Neural Network (RNN) also provides great accuracy in human emotion recognition by modelling temporal sequential data. RNN, which is a class of artificial neural networks with cyclic connections and several hidden layers allow to generate outputs by assessing the previous hidden states and current inputs in the network (Jirayucharoensak et al., 2014). However due its vanishing and exploding gradient problems we cannot get efficient output in case of long-term temporal intervals. Hence, Long Short-Term Memory (LSTM) network which is a unique type of RNN is introduced that has the ability to learn long temporal sequences and solve vanishing and exploding gradient problems (Nakisa et al., 2018).

In this paper, we aim to use Extreme Gradient Boosting (XGBoost) algorithm on top of PCA to detect emotion using EEG signals. We study different machine learning models and deep learning models like logistic regression, support vector machine, random forest, XGBoost, LSTM and evaluate which model is most efficient to accurately predict the emotional categories of EEG signals. We prefer Extreme Gradient Boosting (XGBoost) algorithm for emotion recognition because it can perform well in classification and regression problems using additive learning algorithm.

2. Related works

Recognition of emotion is one of the pivotal problems in Human-Computer Interaction (HCI). This is considered as a challenging task in some respects. There are many methods which has been implied to magnify the rigor of human emotion identification through EEG. There have been many approaches to invade the space of human emotions. Thus, the research area of this kind of works are huge but comes up with less accuracy and perfection. Groundbreaking contribution has been made by different kind of researchers in this type of works. In this section, we will be going through some of the related research works, which are very productive and fruitful to our current work.

Several authors have proposed an EEG-supported emotion identification approach. Here, the utilization of the 3-Dimensional Convolutional Neural Networks (3D-CNN) is examined for utilizing various types of EEG dataset for human emotion identification. This identification approach was offered to find out the spatiotemporal qualities to establish the secular reliance among the EEG signals. As this method needs 3D signal inputs, a way that was showed the EEG signals onto a 3-dimensional type from various route signals has been established. The frame sample from various channel created 2D spatial matrix. After that to create the 3D volume input, they had to attach the time dimension by sequencing the frames together. DEAP data was used to show the productiveness of the suggested way. As the easily available public EEG datasets have less data, the data development stage was offered to enlarge the number of samples per subject and they added some noise signals too (Salama et al., 2018). There are some authors which suggest that a crucial part about emotion identification is automatic emotion identification. A high developed learning algorithm which represents high-level abstraction is required for detecting emotion from non-stationary signals. This report also suggested the use of a deep learning network (DLN) to find out the unspecified quality association among the signal inputs which is important for the learning task. They used 4 experiment setups to validate their topic. At the beginning, they applied the emotion identification by utilizing a deep learning method which includes the amount of hundred concealed nodes in every layer (DLN-100). Then they lessen the number of hidden nodes to fifty (DLN-50) for seeing the consequences of concealed node size in the DLN (Jirayucharoensak et al., 2014). After that, they utilized the PCA to be reduced the incompleteness issue of the DLN. 50most crucial components from the starting 230 input quality was pulled out by the PCA. These were put onto the DLN with 50 concealed nodes in every layer. Lastly, to solve the problem of non-stationary in EEG signals, covariate shift adaptation (CSA) method was used to amplify the efficiency of the emotion identification system. Input features was normalized by the system. This was operated for every input quality independently. Overall, the precision of SVM classifier, EEG focused emotion wise classification with DEAP dataset came out as low. But there were some limitations and the most crucial one among them was EEG-focused emotion identification algorithm and the problem was of subject differentiations of EEG signals. (Jirayucharoensak et al., 2014). There exists a considerable body of research on LSTM, RNN which is very effective for emotion classification. However, Hyper-parameter values vary the performance of these sequence classifier and the most important thing is to embrace a systematic way to get the ideal values. Considering this issue, they proposed an up-to-date structure to automatically improved LSTM hyper-parameters utilizing differential evolution (DE) (Nakisa et al., 2018).

A new framework was presented which is focused on the utilization of a DE method to improve LSTM hyper-parameters in the circumstances of emotion variations. The production was judged and set side by side with other shape-of-the-art hyper parameter enhanced methods over some new dataset which was gathered utilizing light weight physiological signals, that measures EEG & BVP signals. This production evaluation was focused on four-quadrant dimensional emotions. These are the High Arousal Positive emotions (HA-P), Low Arousal-Positive emotions (LA-P), High Arousal-Negative emotions (HA-N), and Low Arousal- Negative emotions (LA-N). The experimented outcome showed that blending of EEG & BVP signals gave the excessive presentation for varying four-quadrant dimensional emotions. (Nakisa et al., 2018). Overtime an extensive paper had been established to the target of upgrading the legitimacy of emotional identification throughout automated study of emotional quality in space and temporary dimensions of electroencephalogram, an EEG emotional quality study and categorization technique which used deep CNN, was put forward focusing on temporary qualities, frequential qualities, and to the top of all the signals inside DEAP data set (Chen et al., 2019). The judgement came out as the CNN modules which need no quality structuring obtained the greatest identification representation on temporary and frequency integrated qualities in both dimensions of arousal and valence. It is also 3.58% excessive than the representation of typical BT classifier in valence dimension and in arousal dimension 3.29% higher. They also did a variation of EEG which is a primary focus of their finding. The typical machine learning & deep learning modules for EEG-focused human emotion variation was established, and their categorization representations were varied on the EEG signals from DEAP data set (Chen et al., 2019). A recent study proposed a GRU network dealing with mortal emotion variation from the electroencephalogram (EEG) signals which are continuous. Base of the module reflects the hierarchical base of EEG signals. This sequence focuses the offering of key test subject and epochs according to its emotional diversity. Through DEAP data-set transdisciplinary emotion variation test was executed to judge the module performance. The test outcomes showed that in both dimensions, their proposed model out values the best deep baseline LSTM model by 4.2% and 4.6% respectively. They made transdisciplinary emotion variation tests on higher scale of EEG sequences in birth type of dimension to estimate the structure of their module. The outcomes showed that their H-ATT-BGRU model obtained the greatest variation of precision of 69.3% on the 0.5-s segmented EEG sequences. On the 1-s segmented EEG sequences in both the dimension, their module also obtained the variation of precision of 67.9% and 66.5% (Chen et al., 2019).

There was also a successful establishment about DL techniques which have been utilized all over. It was used also mainly in speech and image identification. To make an exact DL module precisely and systematic variation of electroencephalogram (EEG) signals are tough, mainly as EEG signals are distinguished by two different subjects or classify from a single subject. In this research work they improved and offered a SincNet-based and SincNet-R focused classifier, which has 3 convolutional layers & 3 DNN layers. SincNet-R was used to manage a tri- categorization of EEG signals. Moreover, the categorization precision among SincNet, SVM, SincNet R, LSTM and CNN was distinguished, the whole SincNet-R were analyzed too. The test outcomes showed that they also offered SincNet-R classifier is the better one than the other classifiers and also While structuring the quality extraction of 62-route EEG signals, SincNet-R could converge faster than SincNet (Zeng et al., 2019).

Amin et al., (2017) have suggested on "pattern recognition" perspective which differentiates EEG signals captured throughout various cognitive prerequisites. K-nearest neighbors (KNN), Support Vector Machine (SVM), Multi-layer Perceptron (MLP), and Naïve Bayes (NB) are mainly being utilized for classifiers. From the results of the experiments it can be seen that a 99.11% accuracy can be achieved via SVM classifier, a class of low frequencies that has a range from 0 to 3.90 Hz, for the approximation of coefficients (A5). Also, it can be seen that an accuracy of 98.57 and 98.39% can be achieved, for the elaborated coefficients, using SVM and KNN respectively; and including the ones from the lower sub-band range (from a range of 3.90-7.81 Hz) can be attained for the elaborated co-efficient (D5). At some instance for the A5 and D5 coefficients the precision rates can be compared at 97.11-89.63% and 91.60-81.07% using the classifiers MLP and NB respectively. The authors have demonstrated a way through which the EEG signals can be implemented for the tasks of cognitive and classification in a pattern recognition-based approach. The experimental results of the authors agreed with the proposed scheme. From these outcomes, it can be suggested that there is a distinction among the normal and abnormal EEG patterns which could be an important potential for the method's application. There are some authors who have suggested that a unique type, moreover a merged type, of LSTM model can be build which can detect emotions from fellow humans. For the model's applicability and accuracy, the authors have used the most precision dataset there is present for the recognition of humans' emotion and it can be cross checked with the DEAP dataset (Lin & Jung, 2017). Further, the authors have inspected about the possibility of using Recurrent Neural networks to categorize emotions using EEG signals. Further, the authors have inspected about the emotions that is picked up by the EEG Signals can be categorized by the Recurrent Neural Network (RNN). The investigation of their proposed merged LSTM model giveaways a great precision of 84.89%, 83.85%, 84.37% and 80.72% on binary classification of emotions. Also, it can be justified that a strong relation

can be drawn between the EEG signals of classifiers, usually being in set, and the way it differs from traditional learning style. Another esteemed author has experimented with SVM and K- Nearest Neighbor (KNN) implementation to detect different kinds of emotions of EEG based brain signal. They have experimented these methods on 5 different male subjects and analyzed their results respectively (Mehmood & Lee, 2015). And according to their findings it visualizes that for each subject processing and classifying the data is quite insignificant. Consequently, the outcome of the four emotions and its classification mainly varies in testing contents as it is the sole reason for the reliance of emotions. Finally, they have concluded that even though the shortcoming of powerful physiological manifestation which is related with emotion to associate at the cortical level of the brain, their proposed method of feature extraction depicts that there is the possibility of emotion being recollected. However, this analogy precisely treated for EEG signal processing that is mainly targeted on different techniques of categorization and extraction related to feature. Also, their outcomes have displayed that KNN is far more superior over SVM in all selected test subjects.

As we all know this far that brain signals can be categorized into 5 waves such as alpha, beta, delta, gamma and theta which can further be broken down into their respective frequencies to report the amount of attention, alertness, character and external stimuli (Rahmad et al., 2018). Here, also the authors have mainly focused on the EEG signal which can be categorized to identify various movement pattern. The authors have added together all the five different brain waves with genetic algorithms which can be further categorized with the help of Logistic Regression, Linear Discriminant Analysis, K-Neighbors Classifier, Decision Tree, Naïve Bayes Gaussian, and Support Vector Machine. They have found out that from the six methods above the best precision is 56% using the Logistic Regression. Finally, the authors have concluded that the research can be progressed by moreover condensing the unit of time required to pick up signals and altering the main values for the waves, of different types, to have more precise results. Another interesting topic have been suggested by other authors which instigated a system which can identify different emotion states is solely based on electroencephalogram (EEG) signals. The authors have conducted the experiments which uses the principle of movie elicitation for obtaining EEG signals of the test subject to categorize four emotional conditions, happy, calm, unhappy, and horror (Wang et al., 2011). After the EEG signals are pre-processed, the authors have examined numerous features of EEG to construct a system that can identify these emotion states. They have calculated performance of the classification, k-nearest neighbor (KNN) algorithm, multilayer perceptron and aided vector machines are being utilized as classifiers. Their experimental findings suggested that a mean test precision of 66.51% for categorizing emotion states (4) prerequisites can be achieved by operating frequency domain features and support vector machines. The detailed study of these authors have ended by their experimental results which suggested that it is attainable to seek out four emotional conditions, happy , calm, horror and unhappy, during binging movie , and a mean test precision of 66.51% is achieved. In addition, their experimental findings visualize that EEG signals which are situated at the specified lobes were extra enlightened about the different emotions (Zheng et al., 2014). We can also find out that authors have introduced advanced deep learning models to categorize two different emotional states (positive and negative) from the necessary data. The authors have instructed a deep belief network (DBN), with differential entropy features, takes input from multichannel EEG signals. Different emotional switching is extra dependable by the high precision of the hidden Markov model (HMM). They have also differentiated the execution of the deep models to KNN, SVM and Graph regularized Extreme Learning Machine (GELM). While doing the experiments the mean precision of DBN-HMM, DBN, GELM, SVM, and KNN are as follow 87.62%, 86.91%, 85.67%, 84.08%, and 69.66%, accordingly. Their experimental findings have shown that the DBN and DBN-HMM models enhance the precision of EEG-based emotion categorization in differentiation with the highest tier methods. This paper mainly focused on deep learning (DL) identifications from EEG data to build a sustainable model for the identification of the different emotional states as mentioned above. The design was being tested on 62 channels EEG signals for identifying the positive and negative emotional conditions. The dataset was obtained from different test subjects. Also, deep learning and tuning with Long Short-Term Memory (LSTM) are also being experimented by a group of researchers and they have also used Adaptive Boosting of the two types of models is experimented for each query. Three experiments were anticipated by the authors utilizing classifiers of different types (1) attention state categorization (2) emotional sentiment categorization (3) a test where the subject has think a number and the main goal is to identify that. Their findings visualize that an Adaptive Boosted LSTM can obtain a precision of 84.44%, 97.06%, and 9.94% on the attentional, emotional, and number datasets, accordingly by the researchers. The researchers have finally concluded that the outcome of the experiment was obtained within their expected limit. In their results, the approach by DEVO has led to a more stable, resource-light model that completely follow that to an extremely resource heavy deep learning model, losing a fragile amount of precision (Bird et al., 2019).

Considering all the references, we tried several machine learning classifiers on the EEG-brainwave signal dataset. Where, extreme gradient boosting algorithm has performed better than other models.

3. Methods and data

Typically, a computerized emotion recognition procedure can be completed utilizing at least one of various modalities: face, discourse, body motions, and the EEG signals. Utilizing the EEG signals, investigates center around taking care of the issue of relationship in time between feelings. Essentially, feelings keep going for short or significant stretch of time, not one minute. In this way, the connection between feeling sections in time is profoundly successful for improving recognition veracity. In this experiment, there are several traditional machine learning techniques have been used. Rest of section we discuss the methods.

To conduct this research, we utilized a publicly available electroencephalogram (EEG) brainwave-based emotion recognition dataset. The dataset has been directly collected from (Bird et al., 2019). There are three classes in this dataset such as positive, negative, and neutral. From Fig, 2, we can see the counts of each class. Here, 723 data are neural class, where 709 and 711 data are negative and positive classes, respectively.

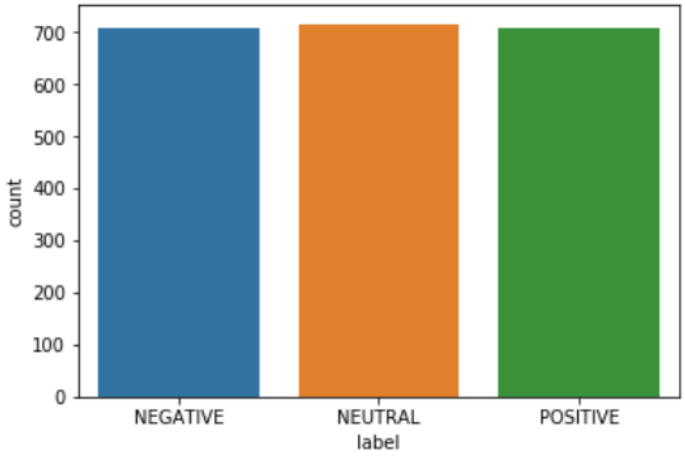

Fig. 2: EEG-brainwave dataset of three classes (Negative, Neutral and Positive).

3.1 Principal component analysis

Principal component analysis (PCA) is a pervasive method for data analysis, but out does not depend of probability model (Das, 2013). PCA can change unique low-level factors to a higher dimensional space and in this way lessen the quantity of required factors. Linear factors get merged in that case. It is a multivariate strategy that examines a set of data wherein perceptions are depicted by an internally connected quantitative ward factors (Abdi & Williams, 2010). There are some main features of principal component analysis. It can extract key information from the set of data; It can squeeze the dataset size by discarding all unimportant information; It also can analyze the pattern of the dataset by punishing several reduction techniques. So that it is easy to analyze to get further intuitions.

To find the principle component, it can be attained by singular value decomposition (SVD) of a dataset A. It can be written as $A = P\nabla Q_T$, factors of matrix can be written as Eq. 1,

$$F = P\nabla \tag{1}$$

The Q matrix determines the linear combination, which is responsible for calculation the factors of matrix. It also can be said as projection matrix due to its multiplication with A by Q computes the projection of the principle components, Eq. 2.

$$F = P\nabla = P\nabla Q^T Q = AQ \tag{2}$$

Geometrically, these components can be represented by original axis rotation. Let, A resembles three variables of a word (B), then it can be represented as three orthogonal factors. In the Eq. 2, Q can be denoted as cosine direction. And Q matrix is loading matrix. Here, A can be explained as the product of score matrix with the loading matrix, Eq. 3.

$$A = FQ^T \tag{3}$$

Where, X is the bilinear decomposition (Abdi & Williams, 2010).

In our experiment, there are three classes such as negative, neutral, and positive labels. Principal component analysis has been performed. Fig. 3 represents the principal components verses class distribution in this context.

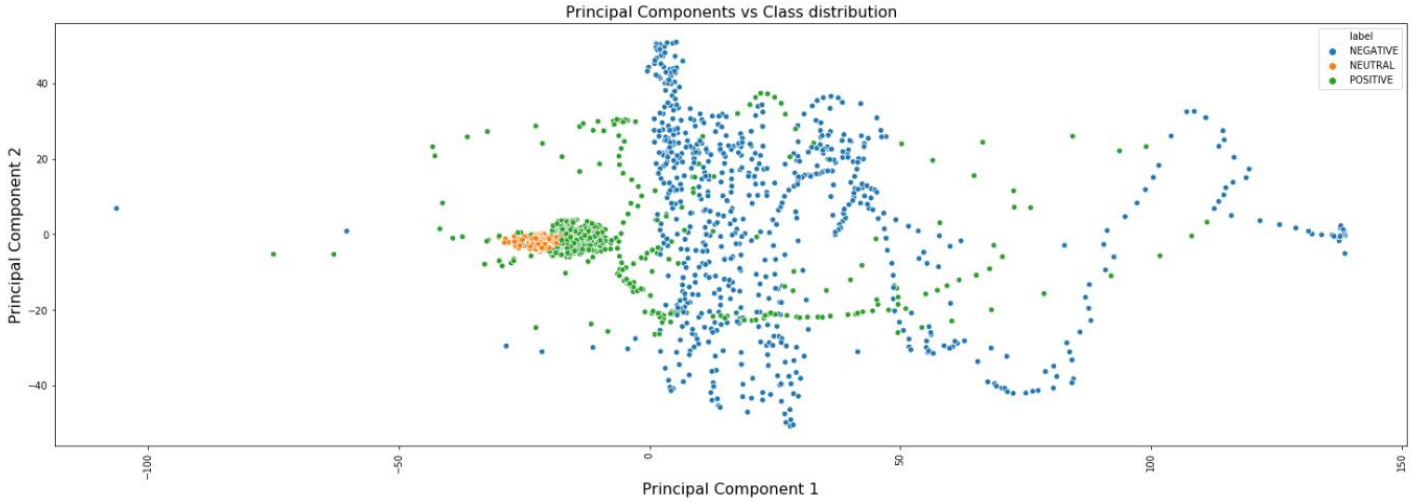

Fig. 3: Principal components vs class distribution of three classes.

3.2 Machine learning techniques

Machine Learning and Deep Learning model have achieved tremendous popularity due to their high real time performance over the years. Many machine learning modules are prevalent in research for prediction, classification, regression and solving other complicated problems. Some of the trending real-world applications of machine learning are: Image Recognition, Speech Recognition, Medical Diagnosis, Stock Market trading etc. Machine Learning has also shown promising success in Neuroscience. Recent research has shown that Machine Learning and Deep Learning techniques acquired much improvement in human emotion recognition accuracy. In this paper we are going to apply some of the widely used machine and deep learning techniques like Logistic Regression, Support Vector Machine, Random Forest, ANN Classifier, XGBoost and LSTM to evaluate which model is most efficient to accurately predict the emotional categories of EEG signals.

3.2.1 Logistic regression classifier

This algorithm is called the appropriate model of regression analysis. It only applies for the dependent and binary variables. There are several regression analyses but among them this algorithm can be called the predictive analysis. One of its usage is to narrate data. Moreover, to describe the relationship between different types of variable it has its contribution. The linear regression can also feature the biological science domain (Zhu et al., 2019).

Logistic regression is (Alkan et al., 2005) a statistical modeling technique which is widely used. Hence, in this modeling sequence technique, the probability P1, belongs to the dichotomous outcome procedure which is connected to the bunch of variables in the form of Eq. 4.

$$\log(P_1) = \ln\left(\frac{P_1}{1-P_1}\right) = \beta_0 + \beta_1 x_1 + \ldots\ldots + \beta_n x_n \quad (4)$$

This equation represents β_0 as the intercept and $\beta_1, \beta_2, \ldots\ldots, \beta_n$ are the coefficients which are interrelated with the illustrative variables $x_1, x_2, \ldots\ldots, x_n$. These variables which works as input are the mean of the wavelet coefficients of EEG signals. Two values are limited for the dichotomous variable such as yes/no, on/off or 1/0. Basically, it just represents the happening and non-happening of some event. The illustrative variables can be different types. These can be dichotomous or discrete or continuous. There is a regression technique called ordinary linear regression (OLD) which is focused on least square method. The outcome of this technique comes as dichotomous but ends up to trivial results. From the previous equation (1), the response variable also the built-in logarithm of odds proportion is the portrayal of the proportion among the chance of occurrence as per it will occur or not. Logistic regression keeps an eye on the changes of the odd logarithm of response variable where ordinary logistic regression does not. Ordinary logistic regression directly calculates the response variable itself. As, odd logarithm is linearly connected to the illustrative variables, the reverted correlation among the reaction and illustrative variables are not linear. The probability of happening of an occurrence is non-linear because

outcome of the illustrative variable is non-linear. Thus, the previous equation can be derived such as Eq. 5 (Alkan et al., 2005):

$$P_1(x) = \frac{1}{1+ e^{-\log(P_1(x))}} \quad (5)$$

As, dissimilar to ordinary linear regression, logistic regression will compel the probability values among 0 to 1. Often the maximum likelihood estimation (MLE) is used to calculate the coefficients $\beta_1, \beta_2, \ldots\ldots, \beta_n$ in the logistic regression equation (Subasi & Erçelebi, 2005). This method is different from ordinary logistic regression for estimating the coefficients. Ordinary logistic regression method searches to lessen the total of squared distances of evert data point at the regression line (Alkan et al., 2005). But the MLE method looks for to maximize log likelihood. Unlike ordinary logistic regression, the MLE method is repatriation and begins with a first arbitrary estimate of the coefficient of the regression equation. It proceeds to know the change of direction and magnitude in the coefficient which increases the likelihood function. The residuals are examined, and new estimate is calculated with a better function after the initial function is determined. This procedure keeps going until reaching some convergence criterion (Alkan et al., 2005).

3.2.2 Linear support vector machine classifier

Artificial intelligence considers machine learning as a subfield and it is connected to the establishment of various kinds of techniques and methods to make the computer learn. If we want to put it simply, the establishment of algorithms which makes the machine learn and conduct some tasks and activities. In many ways it is overlapped with statistics. Since the beginning there have been many methodologies and techniques which were established for machine to learn different kind of tasks. So, it is an extensively used machine among all the machine learning algorithms. This algorithm was very popular at the time of its development in 1990s and it continued this high performance with a little bit of tuning. The main goal of this machine is to look out for a hyper-plane in N-dimensional space that means N- numbers of features that clearly categorize data points.

All the SVMs are constructed on establishments of computational learning-based theory. As the precision and capability to cope up with a big amount of number of predictors. They have strict focus in biomedical domain (Subasi & Gursoy, 2010). A major part of the other classifiers distinct classes utilizing hyper-planes. It gets rid of the classes, utilizing by plane that is flat, inside the predictor of space. These have widened the idea of hyper-plane separation. The name of the methods came from the idea of support vectors. It is very real to think that decision boundary of the location gets impacted. By eradicating it big change on its location can be obtained. But there is a problem that occurs with the usage of a kernel function. The problem is to create linear boundaries from end to end of nonlinear transformations. It also occurs same with the sector of mapping of the predictors. Intelligent element of the technique is that which looks out for a hyper-plane which would be in the predictor space. It is also known in feature space as the products of dots and vector input. In the higher dimensional space dot product is utilized to calculate the distances among the vectors. In the feature space as the dot product with a different approach the kernel function is being utilized. The distinguished by two classes can be executed perfectly by a complex curve. It occurs in the original boundary of the predictor space. To make different the two classes perfectly it is not possible for the best liner separator. On the contrary, the real values of the predictor can be implanted into a more perfect feature space. Thus, with the help of linear decision boundary these can be compared perfectly. As the consequences, there remains an issue of looking out for the perfect transformations. The selection of kernel functions parameter values is another main problem. The fundamental support vector classifier is almost same to same as the perceptron. The replication process has been put to an end when the test subjects are classified correctly in perceptron learning.

The support vector classifier takes one definite solution: the classifier which distinguishes the classes with maximum margin. It is named as the width of the longest 'tube' which does not contain samples that can be extract around the decision boundary (see Fig. 4). It is proved that this definite solution has the best generalization capability (Subasi & Gursoy, 2010).

This classifier is blessed with many upper hands. The parameters of this classifier use standard optimization software without extra hurdle. Furthermore, it gives a very competitive performance with other methods. A limitation is that complexity with the order of the dimension of the samples. Also, the order of the number of samples. For large sample sizes NS › 1000 general quadratic programming software fails (Subasi & Gursoy, 2010). It needs to be added that this is a constructive way to avoid the issue of local minimum. It also helps to avoid the over-fitting problems which can be found in the conventional detection methods. Moreover, with the help of kernel function, the dimensionality can be avoided. (Wang et al., 2012).

Let,

We take the sample set of training to be T.

So, Eq.6.

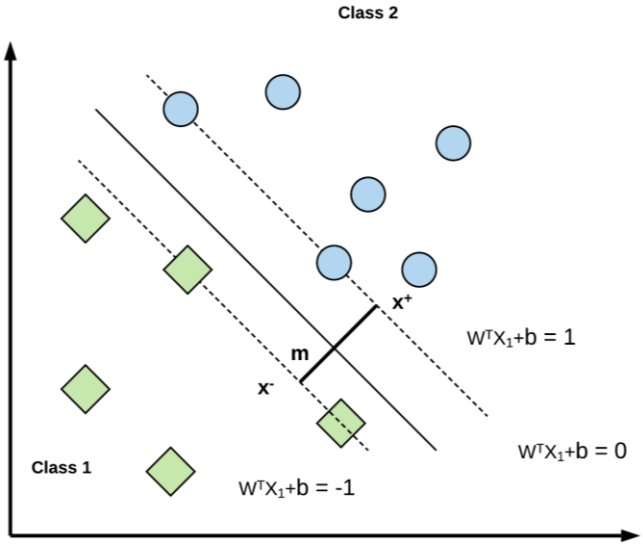

Fig. 4: Linear support vector machine classifier.

$$T = \{(x_1, y_1) \ldots\ldots (x_m, y_m)\} \in (X \times Y)^m \quad (6)$$

Here, $x_i \in R^n, y_i \in \{1, -1\}$ and $i = 1 \ldots\ldots\ldots m$

There is an optimization problem arises if we introduce the maximum interval method which is Eq.7.

$$\min \frac{1}{2} \sum_{i.j=1}^{m} y_i y_j a_i a_j k(x_i, x_j) - \sum_{j=1}^{m} a_j = \min \frac{1}{2} a^T Q a - e^T a \quad (7)$$

Here, $C \geq a_i \geq 0, i = 1 \ldots, m$ and $\sum_{i=1}^{m} y_i a_i = 0$, as $Q_{i.j} = y_i y_j k(x_i, x_j)$ is the kernel matrix. C is labeled as the regulation parameter and$= (1,1 \ldots\ldots 1)^T$. There are several kernel functions which is polynomial and typical. For example, $k_{ij} = [(x_i^T x_j) + 1]^d$ is the polynomial and $k_{i.j} = \frac{\exp(-||x_i - x_j||)}{2}$. We get the optimization resolution after resolving the issue of optimization as: $a^* = (a_1^* \ldots\ldots a_m^*)^T$.

a^* is chosen and $0 < a_j^* < C$. After that, $f(x) = sgn(\sum_{i=1}^{m} a_j^* y_i k(x, x_i) + b^*)$ is calculated as the optimal decision function.

High-generalization ability is provided by the SVM. The dependency of perfect kernel function for a particular issue is on the data. Till now no suitable technique was invented for selecting the kernel function (Wang et al., 2012). Furthermore, if we compare the artificial neural network model and SVM with a lot of local minima, support vector machine gives a distinctive result because of the convex nature of optimality problem (Fan et al., 2018). The approximated function of support vector machine algorithm is Eq. 8.

$$f(x) = \omega\varphi(x) + b \quad (8)$$

Here, $\omega(x)$ is denoted as the higher dimensional feature space which is the resultant of vector input x. ω and b are denoted as weight vector & threshold accordingly. The estimation can be made by the Eq. 9 function:

$$R(C) = C \frac{1}{n} \sum_{i=1}^{n} L(d_i, y_i) + \frac{1}{2} ||\omega||^2 \quad (9)$$

Here C is denoted as the parameter of penalty of the error which is $C \frac{1}{n} \sum_{i=1}^{n} L(d_i, y_i)$. dl is denoted as the desired value with the amount of observation n. is denoted as the imperial error where the function L_ε can be obtained as Eq. 10.

$$L_\varepsilon(d, y) = |d - y| - \varepsilon |d - y| \geq \varepsilon or\ 0\ otherwise \quad (10)$$

Where the regularization term is $\frac{1}{2}||\omega||^2$ and ε is denoted as the tube size. Lagrange multipliers and from optimality constraints the approximated function is expressed as explicit form Eq. 11.

$$f\left(x, a_i, a_i^*\right) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x, x_i) + b \quad (11)$$

Here r=the kernel function is k(x,x_i). The RBF non linear function kernel function has better perfomance of estimating H compared to ither kernel function, Eq. 12 (Fan et al., 2018).

$$k_{rbf}(x, x_i) = \exp[\frac{-(x-x_i)^2}{2\sigma^2}] \quad (12)$$

3.2.3 Random forest classifier

Deep Learning (DL) and Machine Learning (ML) with their different kinds of neural networks, in their field, such as Recurrent Neural Network (RNN), Artificial Neural Network (ANN), Long Term Short Memory (LSTM) and many others have achieved vast amount of acknowledgements by groups of researchers and scientists in their respective fields. Also, one of the popular algorithms-based classifiers used by esteemed researchers and scientists is Random Forest Classifier. Random Forest Classifier is a cluster of numerous decision-making trees and outputs of it is the mean prediction of each separate trees which is run by a method for classifications known as ensemble-learning (Edla et al., 2018). This very algorithm was firstly founded by Breiman. This classifier produces forests in casual number of trees because it is seen that higher accuracy of the result is achieved when the number of decision trees is more (Subudhi et al., 2020). In Fig. 5 each individual tree can be categorized as a standalone output of its categorization and classifier where it is all selected by the decision-making trees. Fig. 5 shows us the construction of the random forest.

Usually, the leaf is considered for the extension at every point of development of the tree. The large sets of data dataset are unanimously broken down into two categories that resembles a plain structure (Zhang et al., 2017). The algorithm and how the margin function blend together with each other are described as Eq. 13:

$$mg(X, Y) - \alpha\vartheta_k I(h_k(X) = Y) - \max_{j \neq Y} a\vartheta_k I(h_k(X) = j) \quad (13)$$

Here, I (.) represents the indicator function, $h_1(x)$, $h_2(x)$.....$h_k(x)$ are the ensemble of classifiers and X,Y are just random vectors.

The error of the forest classifier is represented as Eq. 14.

$$PE *= P_{X,Y}(mg(x, Y) < 0) \quad (14)$$

Here,$h_k(X) = h(X, \theta_k)$ is assumed in Random Forest.

The equation for the Random Forest for the margin function is represented as Eq. 15:

$$mr(X, Y) = P_\theta(h(X, \theta) = Y) - \max_{j \neq Y} P_\theta(h(X), \theta) = j) \quad (15)$$

And finally, we can measure the strength of the classifiers by Eq. 16.

$$s = E_{X,Y} mr(X, Y) \quad (16)$$

As disclosed, the forest error rate mainly depends on two key features (Zhang et al., 2017):

It can be said that if the correlation between any two trees is more than the error rate will be quite high and vice-versa.

The strength of each individual trees in the forest is inversely proportional to the error rate which means that as the strength of each tress is increased, the forest error rate falls gradually.

The researchers have used viable method to analyze the algorithm mathematically to accomplish more efficiency in the process.

Finally, we can conclude that Random Forest algorithm is a very viable method to construct upon multiple trees and combines them to a single tree to have the maximum prediction accuracy.

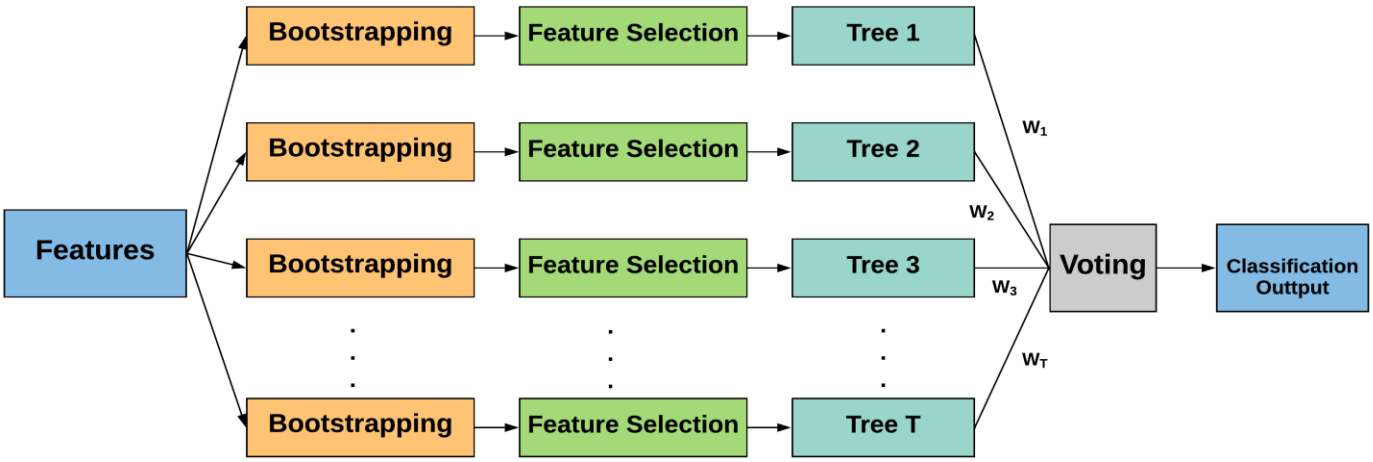

Fig. 5: Random forest classifier structure.

3.2.4 Artificial neural network classifier

Artificial Neural Network is a special branch of Neural Network where the biological neural networks and its structure is solely constructed based on computational model. Artificial Neural Network (ANN) is widely used by researchers and scientists in the field of biomedicine, engineering, analysis of brain waves etc. All interconnected artificial networks constitute to a neural network. Artificial neural network can be categorized into three simple layers (i) Input Layer (ii) Hidden Layer (iii) Output Layer (Bhatikar et al., 2005). This is displayed in Fig. 6:

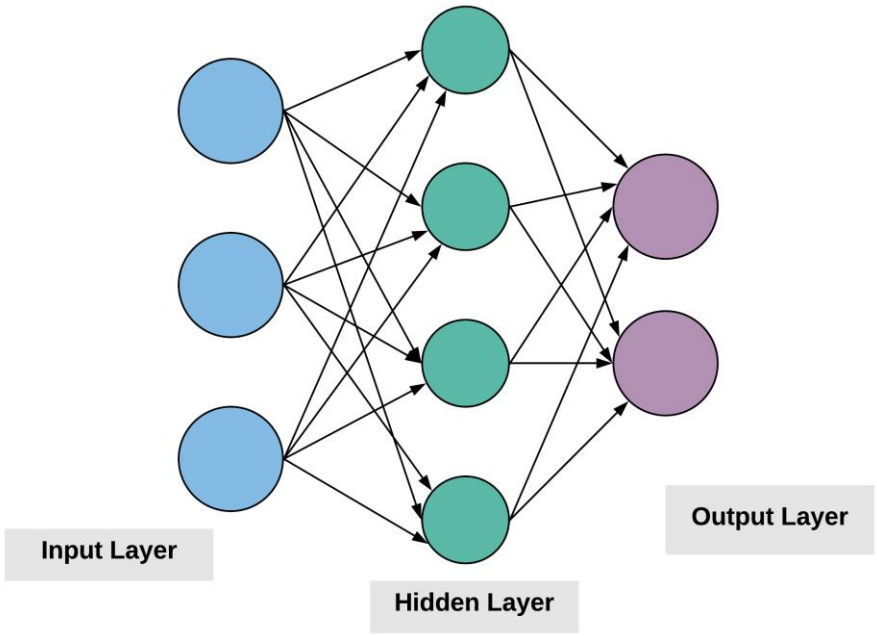

Fig. 6: Artificial neural network (ANN) three primary layers.

Artificial Neural Network mainly constitutes as a range of mathematical models that mimics attributes that are being observed under the systems of processing which contains the biological information, importantly the characteristics of 'emergent learning' whereby the system builds, to act as a certain stimulus response mapping task, for itself. The stimulus, in calculative terms, is accounted for the insert to an Artificial Neural Network (ANN), and the quantitative feedback eventually produces the output. A magnificent characteristic of all Artificial Neural Network (ANN) models is that they contain interlinked processing nodes which constitutes to a network, usually known as neurons, and usually calculate solution based on simple mathematics. At the instance, the neurons are interlinked to produce an ANN, a random neuron gets its feedback from the neurons providing into it, from which it forms a result. In the next step the output is then

circulated back to the neurons that it has been provided into. The inter-links are then being plodded and regulated by interactions that happened between the inter-neuronal. The characteristics of 'emergent learning' is mainly obtained by the Artificial Neural Network (ANN) and other specific ways. During the learning phase, weights biases must undergo adjustments until the learning process has been done. A settled projection is represented by Artificial Neural Network (ANN), from the input and output, and the characteristics of universal function approximation is shown by the projection with the aid of disguised neuron. The entanglement of this observation is that an appropriately structured ANN can roughly guess any function limited to a close and domain which is confined. The job of doing the approximation can be done by an algorithm called back-propagation which is performed on a unanimous state (Bhatikar et al., 2005).

Another study has founded that ANN can be used in the study of the input units, in the pattern recognition, where input units refers to the feature vector and the output units refers to the class of pattern which has to be categorized (Rai et al., 2013). To broadly understand how it works a visual representation is illustrated in Fig. 7.

It is seen that the input vector has been fed into the input layer and the corresponding element consisting of individual unit is its output. Net activation can be defined as an outline where hidden unit are being calculated with the weighted sums of the hidden unit's input. Basically, what happens at the hidden units is that an inner multiplication is done between the inputs and the weighted vector (Bhatikar et al., 2005). Here Eq. 16 shows the transpose value of the input and Eq. 17 shows the transpose value of weight.

$$\text{If, } x = [x_1 * x_2 \dots \dots . x_n]^T \tag{16}$$

$$w = [w_1 * w_2 \dots . . w_n]^T \tag{17}$$

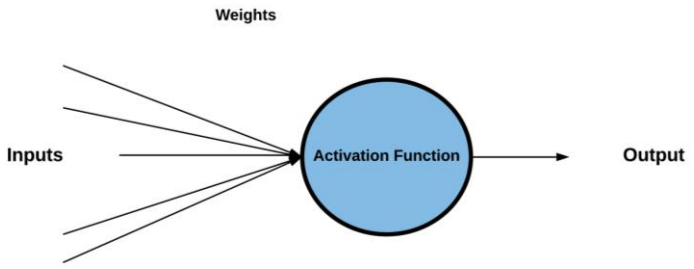

Fig. 7: ANN basic working diagram.

Here x_n and w_n represent the n^{th} term of x and w respectively.

So, finally net function can be expressed as Eq. 18.

$$\text{net}_j = \sum_{i=1}^{d} x_i w_{ij} \tag{18}$$

Here, x_i represents the function of input vector and w_i represents the elements of the weight. Individual hidden units produce a function (non-linear) that belongs to the activation and it generally constitutes to an output, Eq. 19.

$$y_j = f(\text{net}_j) \tag{19}$$

So, it is seen that Artificial Neural Network (ANN) plays a vital role for different aspects of recognition involved in Deep Learning and Machine Learning. High precision of accuracy and efficiency can be achieved by doing the required experiments if the ANN is trained and tested more with the same pattern (Rai et al., 2013).

3.2.5 Long Short-Term Memory (LSTM)

Deep learning models like Convolution Neural Network (CNN), Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM) network showed outstanding achievements recently and are being widely used in video tracking, speech sequence modelling, image classification, emotion recognition and other fields due to their capability of self-learning and providing more accurate results. Recurrent Neural Network (RNN) has been applied to solve numerous problems namely speech recognition, image classification, language modelling (Hu et al., 2020). However, RNN has some drawbacks due to its vanishing and exploding gradient problems hence sometimes efficient outputs cannot be obtained in case of long-term temporal intervals. Hence, Long Short-Term Memory (LSTM) network which is a unique version of RNN is introduced that can learn long temporal sequences and solve vanishing and exploding gradient problems (Nakisa et al., 2018).

In 1997 presented an advanced type of RNN known as Long Short Term Memory (LSTM) network which has an improved structure than RNN and is capable of learning long-temporal sequences (Nguyen-Le et al., 2020). The LSTM

network consists of three gates: i) input gate ii) output gate iii) forget gate. All three of these gates are added in the hidden layer of the network and thus it can resolve the problem of processing long temporal sequence of data (Pei et al., 2019). Therefore, LSTMs are specially designed to refrain from the long temporal sequencing problem. Structure of LSTM network is illustrated in Fig. 8.

First task of LSTM network is determining which information must be kept and which information should be forgotten from the cell state. This is determined by a sigmoid layer also known as "forget gate layer". The sigmoid layer evaluates the current input $\mathbf{x_t}$ and latest output $\mathbf{h_{t-1}}$ and then outputs a number that lies between **0** to **1** for each number in cell state $\mathbf{C_{t-1}}$ (last module LSTM memory). The output **1** refers that the data must be retained and the output **0** refers that the data must be eliminated. Thus, we get a forget function, Eq. 20.

$$f_t = \sigma(w_f.[h_{t-1},x_t] + b_f) \quad (20)$$

Here w_f refers to weight matrices of the input layer and b_f is the corresponding bias vector.

The activation function σ (sigmoid) is characterized as $\sigma(x) = 1/(1+e^{-x})$.

The next step is to determine what new data should be accumulated in the new cell state $\mathbf{c_t}$. This cell state has two parts. A sigmoid layer named as "input gate layer" determines which values should be updated. This is done in the first part. In the later part a **tanh** layer appoints a new value of array $\mathbf{c'_t}$. This new array value can be saved in the state. Both parts are then combined in Eq. 21 and Eq. 22 respectively, and the state is updated (Pei et al., 2019).

$$i_t = \sigma(w_i.[h_{t-1},x_t] + b_i \quad (21)$$

$$c'_t = \tanh(w_c.[h_{t-1},x_t] + b_c) \quad (22)$$

Here $\mathbf{w_i}$ and $\mathbf{w_c}$ are the weight matrices of the input layer, b_i and b_c are the corresponding bias vectors. The activation function **tanh** is characterized as $tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$.

After that, the old cell state $\mathbf{c_{t-1}}$ is upgraded to new memory $\mathbf{c_t}$ (Eq. 23). Upgrading is done by multiplying $\mathbf{i_t}$ with $\mathbf{c'_t}$ and then adding the term with the product of f_t and $\mathbf{c_{t-1}}$.

$$c_t = f_t * c_{t-1} + i_t * c'_t \quad (23)$$

Lastly the output result is determined based on the cell state. It is executed following a few steps. Firstly, a sigmoid layer determines the specific part of cell state that should be returned as output. Then that cell state is assessed by the **tanh** function. The output of sigmoid gate is then multiplied with this updated cell state to obtain the filtered output. Eq. 24-26 describe this whole stage:

$$o_t = \sigma(w_o.[h_{t-1},x_t] + b_o) \quad (24)$$

$$h_t = o_t * \tanh(C_t) \quad (25)$$

The output predicted value can be expressed as follows:

$$y_t = \sigma(w_y.h_t + b_y) \quad (26)$$

Here $\mathbf{w_o}$ and $\mathbf{w_y}$ are the weight matrices of the input layer, $\mathbf{b_o}$ and $\mathbf{b_y}$ are the corresponding bias vectors.

Apart from the traditional LSTM network some other variant LSTM networks are also being introduced by researchers for better optimization and to reduce the training time. A widely known LSTM variant is "peephole connections". Another variation of LSTM network is to use input gates and coupled forget gates. Rather than individually deciding which data to forget and which new information should be retained, this variant makes the decision together. Gated Recurrent Unit (GRU) is another LSTM variant introduced in 2014 (Hu et al., 2020). Unlike traditional LSTM structure that has three gates, GRU consists of two gates namely update gate and reset gate. This results in easy training, simple structure and fast weight gradient descent. Although it sometimes results in unstable training procedure (Pei et al., 2019). Some variants work better than others on certain tasks. However, be it traditional or different variant LSTM networks, all serve the same purpose that is to work efficiently in case of long-term dependencies.

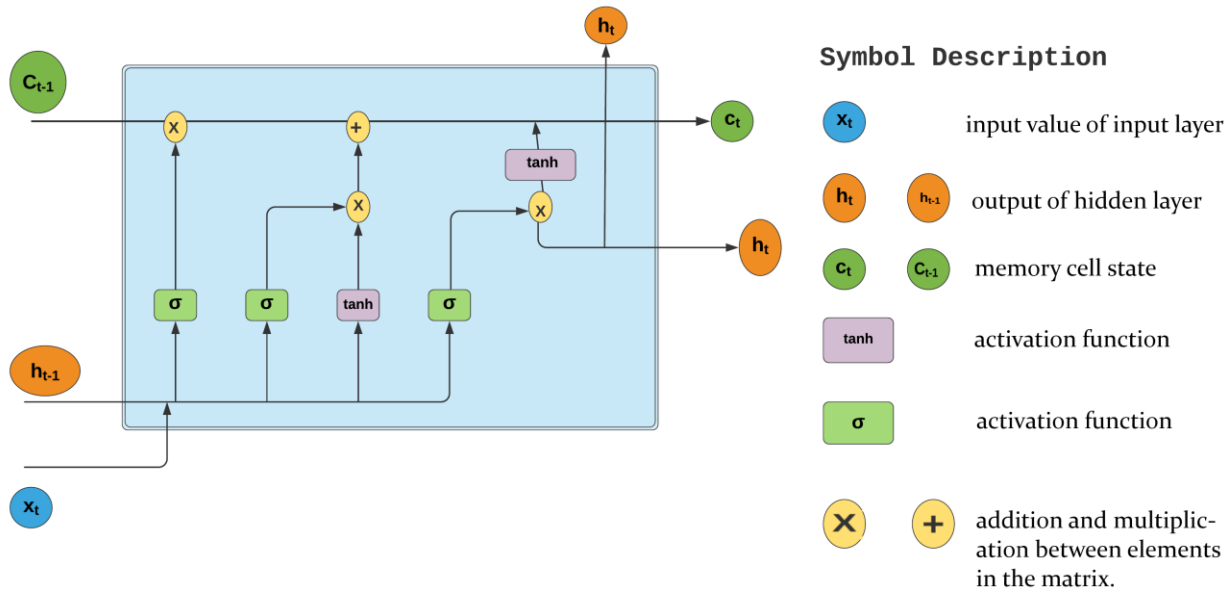

Fig. 8: Long short term memory (LSTM) network structure.

3.2.6 Proposed-Extreme Gradient Boosting Algorithm

Extreme Gradient Boosting (XGBoost) is arguably one of the most powerful machine learning algorithms out there today. The extreme gradient boosting algorithm resolved the limitations of speed and accuracy while computing data. This algorithm also requires less time and training for prediction. XGBoost also performs better than several statistical models (Fan et al., 2018). Along with optimizing computer resources, XGBoost also prevents over-fitting of data. This algorithm combines weak base classifier and turns it to a stronger base classifier with the help of additive training strategies allowing better results particularly in classification and regression problems (Gui et al., 2020). XGBoost is listed as one of the best machine learning algorithms because of its execution speed and model performance. It is fast contrast to other gradient boosting implementations and gives accurate result in classification and regression predictive modelling problems (Fan et al., 2018). The model possesses some great algorithm features like Sparse Aware (to automatically handle missing data values), Block Structure (for parallelizing tree construction) and Continued Training (to boost a prevailing fitted model on a new data). This model mainly uses gradient boosting decision tree algorithm, the structure of extreme gradient boosting (XGBoost) is illustrated in Fig. 9.

XGBoost provides a strong base classifier from a weak classifier through the process of additive learning. To implement this process some steps are followed. At the beginning, the first learner gets fitted to the input data. Then another model is fitted to the remaining for dealing with the limitations of weak learner. This goes through a repetitive process until it reaches the stopping criterion. Finally, a prediction of the model is achieved by finding the aggregation of prediction of each learner. This whole process can be characterized by Eq. 27 (Shi et al., 2019):

$$f_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = f_i^{(t-1)} + f_t(x_i) \tag{27}$$

Here x_i refers to the input variable, $f_t(x_i)$ refers to the learner step at t and $f_i^{(t)}$, $f_i^{(t-1)}$ are the predictions during the steps t and t-1.

An analytic expression is developed by the XGBoost algorithm to retain the computational pace and to prevent the overfitting issue. The expression is as Eq. 28:

$$Obj^{(t)} = \sum_{k=1}^{n} l(\bar{y}_i, y_i) + \sum_{k=1}^{t} \Omega(f_i) \tag{28}$$

Here, l refers to the loss function, n refers to the number of observation. Ω Is known as regularization term and can be defined as Eq. 29.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \| \omega \|^2 \quad (29)$$

Here, λ refers to the regularization parameter, ω refers to the vector of scores in leaves and γ represents the minimum loss required for further partitioning the leaf node. More advanced computational techniques and expressions can be known from Chen and Guestrin (Fan et al., 2018).

XGBoost has provided outstanding results over the years due its well performance in classification and regression problems using additive learning method (Shi et al., 2019). Efficient and accurate result can also be expected in emotion recognition using this popular machine learning algorithm.

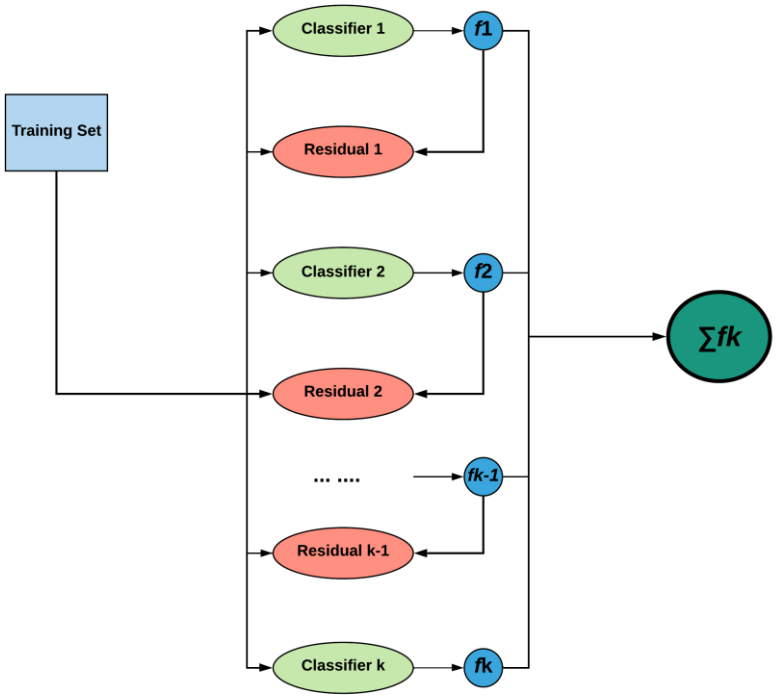

Fig. 9: Extreme Gradient Boosting (XGBoost) structure.

4. Experimental results

The main objective of this experiment is to investigate how well the machine learning classifier as well as ANN and LSTM classifiers performed EEG-brainwave signal-based emotion recognition. In this section, we discuss the machine learning classifiers performances on the brainwave dataset.

There are several techniques has been used in this experiment. These classifiers are logistic regression, random forest, linear-support vector machine, artificial neural network, RNN-LSTM and our proposed extreme gradient boosting (XGBoost). We evaluated our models by accuracy matrix (Eq. 30). Where TP, TN, FP and FN are true positive, true negative, false positive and false negative respectively. The dataset split in 80% for training and 20% for testing.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (30)$$

In this purpose to examine the further effectiveness of the classifiers, we take the CPU time, system delay and total run time of each classifier in consideration. From the experimental outcome, the classification accuracy of extreme gradient boosting classifier has better accuracy than other classifiers. We have taken all principal components in consideration to ensure best output from every classifier. Besides, artificial neural network (ANN) and long short-term memory (LSTM) are very competitive with machine learning models. Although ANN has an accuracy of 97.43%, but it consumes a tremendous execute time. Since, to accelerate the training time of the ANN, we can misuse some parallelism between two softmax classifiers. Which should not be taken for consideration in industrial deploy purpose. On the other hand, many literatures have shown that long short term memory (LSTM)- based recognition system performs very well in many cases. But in our case, LSTM has extremely good accuracy, but it has taken a good amount of execution time. For training the

LSTM model, we set epoch to 100. 120 epochs have been tested to ensure the best output from LSTM model. Table 1 presented the parameter values of LSTM model. We have done Bootstrap Aggregation for random forest model to ensure better result in terms of accuracy and time complexity. However, it performs 2nd best after our proposed Extreme gradient boosting algorithm (XGBoost). In Table 2, two principle component-based performance of logistic regression is presented.

Here, the CPU time indicates the processing time of every classifier. In this case, random forest classifier has the lowest CPU time of 5.21 sec. The system delay measures the time to start the execution time. Furthermore, we combined the CPU time and system delay to compute total run time. A conflicting fact that, although our proposed classifier extreme gradient boosting (XGBoost)+PCA has the highest accuracy than other classifiers, but it has slightly higher total run time than random forest classifier. But we take accuracy matrix more precisely to consider the best model, because in industrial purpose, the system needs more accurate results in large scale datasets. Since, logistic regression classifier also has a high run time. We declined ANN and LSTM based emotion recognition due to their extremely higher run time, which is 100-150 times higher than our proposed PCA+XGBoost classifier. Table 3 demonstrates the performance analysis of every classifiers. The accuracies are taken into 5 times run time average.

Table 1: Parameter values for LSTM Model.

Layer (Type)	Output Shape	Parameters
Lstm (LSTM)	(None, 1, 64)	668928
dropout (Dropout)	(None, 1, 64)	0
lstm_1 (LSTM	(None, 32)	12416
dropout_1 (Dropout)	(None, 32)	0
dense (Dense)	(None, 3)	99

Table 2: Logistic regression performance taking two principle components.

Algorithm	Accuracy	CPU Time	System Delay	Total Run Time
Logistic Regression	76.51%	11.26 Sec	2.34	13.6

Table 3: Performance analysis of every classifier in brainwave-based emotion recognition dataset.

Algorithm	Accuracy	CPU Time	System Delay	Total Run Time
Logistic Regression	93.31%	3 min 29 sec	2.9 sec	3 min 24 sec
Random Forest	97.96%	5.21 sec	123 ms	5.43 sec
Linear SVM	97.46%	2 min 12 sec	2.92 sec	2 min 15 sec
Artificial NN	97.43%	18 min 31 sec	7 min 45 sec	26 min 16 sec
LSTM	96.92%	7 min 26 sec	3 min 24 sec	10 min 50 sec
Proposed XGBoost	99.44%	6.19 sec	221 ms	6.40 sec

A comparison has been done with recent developed models with the proposed model of this experiment. Table 4 shows the comparison between the proposed XGBoost models performance with other developed models.

Above all discussion, we can conclude that, extreme gradient boosting classifier has performed incredibly well in terms of accuracy. But it has slightly higher run time than random forest. But the run time difference between XGBoost and random forest can be negligible. So extreme gradient boosting classifier is our proposed model for this EEG brainwave signal-based emotion recognition.

5. Conclusion

Electroencephalogram (EEG) brainwave signal-based emotion recognition is now one of the promising fields for the researchers. Many works have been done so far on the brainwave signal-based emotion recognition and neural

complexity of human brain. In regarding to this, we proposed a machine learning algorithm, extreme gradient boosting (XGBoost) for an efficient three class emotion recognition procedure. Key findings of this study can be summarized as:

- 2 principal component based logistic regression classifier has 76.51% accuracy with 13.6 sec total run time, whereas taking all 10-principal component-based classifier has 93.31% accuracy with slightly higher run time of 14.6 sec.
- On the other hand, random forest classifier has 97.33% accuracy with 5.43 sec run time, which is extremely well and the less run time than any other models.
- ANN has an accuracy of 97.43% but it took the highest run time of 26 min 16 sec; and linear-support vector machine classifier has 97.46% accuracy with 2 min 15 sec run-time.
- Improved RNN-LSTM has an accuracy of 96.92%, where it has 10 min 50 sec run-time.
- Finally, our proposed classifier, PCA+ extreme gradient boosting (XGboost) has achieved the higher accuracy of 99.44% which has 6.40 sec run-time.

Above all, extreme gradient boosting (XGBoost) and Random Forest have a very close run-time difference. Where random forest has slightly better run-time than XGBoost. But in terms of accuracy, XGBoost has the best accuracy. Therefore, including all findings and considerations, we recommend PCA+XGBoost for further utilization, because it has near 100% accuracy, but the difference with random forest is negligible in our observation.

Table 4: A comparison between proposed model and other recent models.

Authors	Algorithms	Accuracy	Dataset
Salama et al., (2018)	3D-CNN	87.44%+88.49%	DEAP Dataset
Jirayucharoensak et al., (2014)	DLN+(PCA+CSA)	53.42%+52.05%	DEAP Dataset
Chen et al., (2019)	D-CNN	85.57%	---
Chen et al., (2019)	Bidirectional GRU	69.3%	DEAP Dataset
Atkinson & Campos, (2016)	mRMR+Kernel Classfier	60.7%+73.06%	DEAP Dataset
Zeng et al., (2019)	SincNet-R	95.22%	Customized Dataset
Bilucaglia et al., (2020)	SVM+Temporal Features	63.8%	---
Šťastný et al., (2003)	Hidden Markov Model	80%	---
Amin et al., (2017)	K-NN	93.33%	RAPM Dataset
Garg et al., (2019)	Merged-LSTM	84.89%+83.85%+84.37%+80.72%	DEAP Dataset
Rahmad et al., (2018)	SVM	48%	Customized Dataset
Wang et al., (2011)	SVM	66.51%	Customized Dataset
Wang et al., (2014)	SVM+PCA	91.77%	Customized Dataset
Bird et al., (2019)	Adaboosted LSTM	84.44%+97.06%	Customized Dataset
Proposed	**PCA+XGBoost**	**99.44%**	**Kaggle Brainwave**

Authors Contribution Conceptualization: T.H., Methodology: T.H., F.F. Experiment-T.H., Software: T.H., Writing - Original Draft: T.H., F.F., Z.H., F.Q., Writing - Review & Editing: T.H., F.F., Z.H., F.Q., Visualization: T.H., F.F., Z.H., F.Q., Supervision: T.H.

Acknowledgement The authors want to thank the editor and the anonymous reviewers as their comments helped us to improve the quality of the paper.

Funding This research received no specific fund from any organization.

Conflicts of Interest The authors declare no conflict of interest.

Disclaimer The content and claims presented in this journal are of the authors only and do not necessarily represent the views of the editors and publisher.

Abbreviation and Nomenclature

PCA	Principal Component Analysis	LSTM	Long-Short Term Memory
EEG	Electroencephalogram	RF	Random Forest
LR	Logistic Regression	LSVM	Linear Support Vector Machine
ANN	Artificial Neural Network	XGBoost	Extreme Gradient Boosting

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 433–459. https://doi.org/10.1002/wics.101

Ahmed, R., Temko, A., Marnane, W., Lightbody, G., & Boylan, G. (2016). Grading hypoxic-ischemic encephalopathy severity in neonatal EEG using GMM supervectors and the support vector machine. Clinical Neurophysiology, 127(1), 297–309. https://doi.org/10.1016/j.clinph.2015.05.024

Alkan, A., Koklukaya, E., & Subasi, A. (2005). Automatic seizure detection in EEG using logistic regression and artificial neural network. Journal of Neuroscience Methods, 148(2), 167–176. https://doi.org/10.1016/j.jneumeth.2005.04.009

Amin, H. U., Mumtaz, W., Subhani, A. R., Saad, M. N. M., & Malik, A. S. (2017). Classification of EEG signals based on pattern recognition approach. Frontiers in Computational Neuroscience, 11(November), 1–12. https://doi.org/10.3389/fncom.2017.00103

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. https://doi.org/10.1016/j.eswa.2015.10.049

Bhatikar, S. R., DeGroff, C., & Mahajan, R. L. (2005). A classifier based on the artificial neural network approach for cardiologic auscultation in pediatrics. Artificial Intelligence in Medicine, 33(3), 251–260. https://doi.org/10.1016/j.artmed.2004.07.008

Bilucaglia, M., Duma, G. M., Mento, G., Semenzato, L., & Tressoldi, P. (2020). Applying machine learning EEG signal classification to emotion-related brain anticipatory activity. F1000Research, 9, 173. https://doi.org/10.12688/f1000research.22202.1

Bird, J. J., Faria, D. R., Manso, L. J., Ekárt, A., & Buckingham, C. D. (2019). A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction. Complexity, 2019(August). https://doi.org/10.1155/2019/431654

Chen, J. X., Jiang, D. M., & Zhang, Y. N. (2019). A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification. IEEE Access, 7, 118530–118540. https://doi.org/10.1109/access.2019.2936817

Chen, J. X., Zhang, P. W., Mao, Z. J., Huang, Y. F., Jiang, D. M., & Zhang, Y. N. (2019). Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks. IEEE Access, 7, 44317–44328. https://doi.org/10.1109/ACCESS.2019.2908285

Edla, D. R., Mangalorekar, K., Dhavalikar, G., & Dodia, S. (2018). Classification of EEG data for human mental state analysis using Random Forest Classifier. Procedia Computer Science, 132(Iccids), 1523–1532. https://doi.org/10.1016/j.procs.2018.05.116

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., & Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. Energy Conversion and Management, 164(January), 102–111. https://doi.org/10.1016/j.enconman.2018.02.087

Garg, A., Kapoor, A., Bedi, A. K., & Sunkaria, R. K. (2019). Merged LSTM Model for emotion classification using EEG signals. 2019 International Conference on Data Science and Engineering, ICDSE 2019, 139–143. https://doi.org/10.1109/ICDSE47409.2019.8971484

George, S. T., Subathra, M. S. P., Sairamya, N. J., Susmitha, L., & Joel Premkumar, M. (2020). Classification of epileptic EEG signals using PSO based artificial neural network and tunable-Q wavelet transform. Biocybernetics and Biomedical Engineering, 40(2), 709–728. https://doi.org/10.1016/j.bbe.2020.02.001

Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., & Wierstra, D. (2015). DRAW: A recurrent neural network for image generation. 32nd International Conference on Machine Learning, ICML 2015, 2, 1462–1471.

Gui, K., Che, H., Zeng, Z., Wang, Y., Zhai, S., Wang, Z., Luo, M., Zhang, L., Liao, T., Zhao, H., Li, L., Zheng, Y., & Zhang, X. (2020). Construction of a virtual PM2.5 observation network in China based on high-density surface meteorological observations using the Extreme Gradient Boosting model. Environment International, 141(April), 105801. https://doi.org/10.1016/j.envint.2020.105801

Hamada, M., Zaidan, B. B., & Zaidan, A. A. (2018). A Systematic Review for Human EEG Brain Signals Based Emotion Classification, Feature Extraction, Brain Condition, Group Comparison. Journal of Medical Systems, 42(9). https://doi.org/10.1007/s10916-018-1020-8

Henshaw, J., Liu, W., & Romano, D. M. (2019). EEG Signal Processing: Feature extraction, selection and classification methods. EEG Signal Processing: Feature Extraction, Selection and Classification Methods, February. https://doi.org/10.1049/pbhe016e

Hochreiter, S. (1997). Long Short-Term Memory. 1780, 1735–1780.

Hu, F., Zhu, Y., Liu, J., & Li, L. (2020). An efficient Long Short-Term Memory model based on Laplacian Eigenmap in artificial neural networks. Applied Soft Computing Journal, 91, 106218. https://doi.org/10.1016/j.asoc.2020.106218

Jirayucharoensak, S., Pan-Ngum, S., & Israsena, P. (2014). EEG-Based Emotion Recognition Using Deep Learning Network with Principal Component Based Covariate Shift Adaptation. Scientific World Journal, 2014. https://doi.org/10.1155/2014/627892

Kumar, J. S., & Bhuvaneswari, P. (2012). Analysis of electroencephalography (EEG) signals and its categorization - A study. Procedia Engineering, 38, 2525–2536. https://doi.org/10.1016/j.proeng.2012.06.298

Lin, Y. P., & Jung, T. P. (2017). Improving EEG-based emotion classification using conditional transfer learning. Frontiers in Human Neuroscience, 11(June), 1–11. https://doi.org/10.3389/fnhum.2017.00334

Mehmood, R. M., & Lee, H. J. (2015). Emotion classification of EEG brain signal using SVM and KNN. 2015 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2015. https://doi.org/10.1109/ICMEW.2015.7169786

Nakisa, B., Rastgoo, M. N., Rakotonirainy, A., Maire, F., & Chandran, V. (2018). Long short term memory hyperparameter optimization for a neural network based emotion recognition framework. IEEE Access, 6, 49325–49338. https://doi.org/10.1109/ACCESS.2018.2868361

Nguyen-Le, D. H., Tao, Q. B., Nguyen, V. H., Abdel-Wahab, M., & Nguyen-Xuan, H. (2020). A data-driven approach based on long short-term memory and hidden Markov model for crack propagation prediction. Engineering Fracture Mechanics, 235, 107085. https://doi.org/10.1016/j.engfracmech.2020.107085

Pei, S., Qin, H., Zhang, Z., Yao, L., Wang, Y., Wang, C., Liu, Y., Jiang, Z., Zhou, J., & Yi, T. (2019). Wind speed prediction method based on Empirical Wavelet Transform and New Cell Update Long Short-Term Memory network. Energy Conversion and Management, 196(April), 779–792. https://doi.org/10.1016/j.enconman.2019.06.041

Rahmad, C., Ariyanto, R., & Yunianto, D. R. (2018). Brain signal classification using Genetic Algorithm for right-left motion pattern. International Journal of Advanced Computer Science and Applications, 9(11), 247–251. https://doi.org/10.14569/ijacsa.2018.091134

Rai, H. M., Trivedi, A., & Shukla, S. (2013). ECG signal processing for abnormalities detection using multi-resolution wavelet transform and Artificial Neural Network classifier. Measurement: Journal of the International Measurement Confederation, 46(9), 3238–3246. https://doi.org/10.1016/j.measurement.2013.05.021

Salama, E. S., El-khoribi, R. A., Shoman, M. E., & Shalaby, M. A. W. (2018). EEG-Based Emotion Recognition using 3D Convolutional Neural Networks. 9(8), 329–337.

Shi, H., Wang, H., Huang, Y., Zhao, L., Qin, C., & Liu, C. (2019). A hierarchical method based on weighted extreme gradient boosting in ECG heartbeat classification. Computer Methods and Programs in Biomedicine, 171, 1–10. https://doi.org/10.1016/j.cmpb.2019.02.005

Šťastný, J., Sovka, P., & Stančák, A. (2003). EEG signal classification: Introduction to the problem. Radioengineering, 12(3), 51–55.

Subasi, A., & Erçelebi, E. (2005). Classification of EEG signals using neural network and logistic regression. Computer Methods and Programs in Biomedicine, 78(2), 87–99. https://doi.org/10.1016/j.cmpb.2004.10.009

Subasi, A., & Gursoy, M. I. (2010). EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications, 37(12), 8659–8666. https://doi.org/10.1016/j.eswa.2010.06.065

Subudhi, A., Dash, M., & Sabut, S. (2020). Automated segmentation and classification of brain stroke using expectation-maximization and random forest classifier. Biocybernetics and Biomedical Engineering, 40(1), 277–289. https://doi.org/10.1016/j.bbe.2019.04.004

Wang, C. M., Zhang, C. M., Zou, J. Z., & Zhang, J. (2012). Performance evaluation for epileptic electroencephalogram (EEG) detection by using NeymanPearson criteria and a support vector machine. Physica A: Statistical Mechanics and Its Applications, 391(4), 1602–1609. https://doi.org/10.1016/j.physa.2011.09.010

Wang, X. W., Nie, D., & Lu, B. L. (2011). EEG-based emotion recognition using frequency domain features and support vector machines. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7062 LNCS(PART 1), 734–743. https://doi.org/10.1007/978-3-642-24955-6_87

Wang, X. W., Nie, D., & Lu, B. L. (2014). Emotional state classification from EEG data using machine learning approach. Neurocomputing, 129, 94–106. https://doi.org/10.1016/j.neucom.2013.06.046

Zeng, H., Wu, Z., Zhang, J., Yang, C., Zhang, H., Dai, G., & Kong, W. (2019). EEG emotion classification using an improved sincnet-based deep learning model. Brain Sciences, 9(11). https://doi.org/10.3390/brainsci9110326

Zhang, T., Chen, W., & Li, M. (2017). AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier. Biomedical Signal Processing and Control, 31, 550–559. https://doi.org/10.1016/j.bspc.2016.10.001

Zheng, W. L., Zhu, J. Y., Peng, Y., & Lu, B. L. (2014). EEG-based emotion classification using deep belief networks. Proceedings - IEEE International Conference on Multimedia and Expo, 2014-September(Septmber). https://doi.org/10.1109/ICME.2014.6890166

Zhu, C., Idemudia, C. U., & Feng, W. (2019). Improved logistic regression model for diabetes prediction by integrating PCA and K-means techniques. Informatics in Medicine Unlocked, 17(January), 100179. https://doi.org/10.1016/j.imu.2019.100179