EEG based Emotion Prediction based on Audio Visual Stimulation using En-FS and SEn-G model

Sricheta Parui,* *Graduate Student Member, IEEE* *Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, India. *School of Computer Engineering, Kalinga Institute of Industrial Technology, India
sricheta.paruifcs@kiit.ac.in

***Abstract*—The detection of brain impulses and emotions has been a research topic of late. To define emotion as a psychological variable, various machine learning algorithms were applied but there are still some challenges with high dimensional data. To date, much research has been done on this topic to classify a maximum number of emotions and find clarity on the behaviors of the brain signal with certain stimuli. In this paper, we have thoroughly studied the data and experimented with various techniques to find out the best one and according to that we have designed the model to classify a maximum number of emotions in a short time. We have extracted various domains of features including time, frequency, wavelet, etc., and applied the Ensemble Feature Selection method to find out the best subset of features. This technique includes correlation technique, information gain measurement, and finally recursive feature elimination method to find the optimized feature set. For the classification part, various method has been reviewed and then a stacked ensemble generalization model has been adopted with respect to bagging and boosting results of state-of-the-art machine learning techniques. The results show that the ensemble approach of feature optimization (En-FS) combined with the stacked ensemble generalized model (SEn-G) classification performs better. The suggested technique has been tested using the DEAP Dataset, and the experimental findings support the efficacy of the strategy while also being compared to state-of-the-art approaches.**

***Index Terms*—Ensemble Feature Selection Stacked Ensemble Generalized Model EEG Recursive Feature Elimination**

I. Introduction

Emotion is a dynamic behavioral process requiring several stages of neuronal and chemical processing [1]. Here in our research, we are dealing only with human emotion. Due to the competition for being a survivor in these rapid changes of lifestyle, human emotion gets affected and it becomes the driver of any cognitive processing system. These biological aspects influence many factors, such as learning, communication, relationship, health issues, etc. Emotional responses play a very significant role in much human-computer interaction (HCI) systems. It is necessary to identify and recognize the emotions in computer systems that people communicate with, and to facilitate contact between humans and machines [2]. Emotion recognition has many applications, for example, knowing the mental states of a person, we can adapt a user interface and thus providing an effective man-machine interaction [3], evaluating the performance of an operator carrying out a task [4], human reliability analysis [5], etc. Many researchers are researching detecting and classifying various types of emotion by classifiers using signal processing, machine learning [6]–[9], deep learning [10]–[12] technologies. Earlier, interest in the identification of emotions from various modalities like text [13], speech [14], facial expression [15], gesture [13], [16], etc. has grown and recently drawn attention with the growth of brain-computer-interfaces (BCIs), which has came up with the term effective BCI (BCI) [17]. In recent years, EEG-based emotion recognition research has gained great interest from a broad range of interdisciplinary fields, from neuroscience to technology, including fundamental studies on emotion theory and application to affective brain-computer interactions.

Aim of the work: The aim of these studies is to improve the ability of BCI systems to recognize, process, and react to affective states of users using physiology [18]. For the following purposes, the need for a computer-intelligent approach to emotion classification is required:

- While EEG is a reasonably reliable measure and a simple interface, it suffers from the non-stationary properties of the signal. The extraction of temporal correlations of spontaneous EEG signals is therefore a major issue for the recognition of emotions from EEG signals [19].
- Due to this non-stationary property, the EEG signal is difficult to classify. Therefore, the extraction and selection of features become an integral method before classification [20].
- Because the EEG signal is very poor, the recognition of emotions involves more channels of acquisition equipment in the data signal acquisition phase, which is not conducive to realistic usage [21].
- The Human brain is a nonlinear complex machine, and EEG signals are difficult to interpret by using conventional time-frequency extraction and interpretation approaches [22].
- In EEG's emotional recognition, it is not widely accepted which features are most suitable, and there are only a few words that equate various features with each other [17].

A. Related Works and Objectives

Researchers attempted many ways to classify the emotion. A machine-learning algorithm was used for classification. Dan Nie et al. [23] and Xiao-Wei Wang et al. [24] used a vector supporting system to interpret the EEG signal when viewing films. Yuan-Pin Lin et al. [25] have also used SVM to classify four emotion states when the subjects are listening to music and obtained an averaged classification accuracy of 82.29% ± 3.06% across 26 subjects. Zhang et al. [26] proposed

TABLE I: Extracted feature from pre-processed EEG signal

Feature Type	Features
Statistical Features	Mean Max Value Min Value Median Standard deviation Kurtosis
Derivative Features	the First differential mean First differential max First differential min the Second differential mean Second differential max Second differential min Hjorth ability Hjorth mobility Hjorth complexity
Time Domain Features	Line length Mean of the vertex to vertex amplitudes Variance of the vertex to vertex amplitudes Mean of the vertex to vertex times The variance of the vertex to vertex times Mean of the vertex to vertex slope The variance of the vertex to vertex slope Zero crossings (Number of zero-crossing in a signal)
Frequency Domain Features	FFT delta [0.1 - 4 Hz] FFT theta [4 - 8 Hz] FFT alpha [8 - 13 Hz] FFT beta [13 - 30 Hz] FFT gama [30 - 40 Hz] FFT whole [0.1 - 40 Hz] FFT delta-theta ratio [delta/theta FFT delta-alpha ratio [delta/alpha] FFT theta-alpha ratio [theta/alpha] FFT delta-theta-alpha ratio [(delta+theta)/alpha]
Wavelet Domain Features	Min wavelet value Max wavelet value Mean wavelet value Median wavelet value STD wavelet value Skewness wavelet value Kurtosis wavelet value Wavelet energy Wavelet entropy First differential wavelet mean First differentiation wavelet max Second differential wavelet mean Second differential wavelet max Wavelet energy percentage Wavelet zero-crossing Wavelet coefficient of variation Wavelet total energy
Cross Correlation Features	Max ABS XCORR (Maximum positive amplitude of auto-correlation or cross-correlation function) Mean ABS XCORR (Mean value auto-correlation or cross-correlation function)

an emotional understanding framework that categorized user states into two emotional statements with 73 percent accuracy during image viewing using an SVM classifier. Apart from SVM, other approaches [24] [27] to machine learning was used to identify a person's emotional state from EEG signals. Liu et al. [28] introduced an EEG-based emotion detection algorithm based on real-time samples and tested it on trial databases and the DEAP comparison database. Aravind E Vijayan et al. suggested a method for extracting epochs from data using statistical methods such as Shannon Entropy and higher-order auto-regressive processes that match the derived characteristics. The model is used to classify human emotions by feeding them into a multi-class Vector Support Machine [29].Chanel et al. [30] found that an entertaining assessment of emotion could be achieved with a cumulative accuracy of 58% for three levels of emotion using the Naïve Bayes classifier. Ko et al. [31] showed the usefulness of using EEG relative power shifts and the Bayesian network to forecast the emotional states of the consumer.

- With the increasing interest in brain-computer interaction (BCI), EEGs (electroencephalograms) from users have also been analyzed. It is still unclear whether the EEG simply demonstrates a physiological response, or also gives insight into the emotion as to how it is mentally experienced.
- This field of research is still relatively new, and much still needs to be done to improve the existing components of the BCI, but also to identify new possibilities. The existing work, considers some emotion to get the best accuracy or satisfied with a specific limit of accuracy for all the emotions.
- None of them have considered the whole signal is coming from all 32 channels. Some work concentrates only on valence and arousal or just valence only.
- A major drawback of the above methods is that, to date, only a handful of features have been compared in each sample. In comparison, most studies depend on a separate but typically small data set.
- The asymmetry of brain regions was investigated for the recognition of emotions. These techniques, however, only examined the interaction between symmetrical electrodes in the brain and did not attach all the electrodes [22].
- In past years so many machine learning algorithm has been used to classify EEG signal for emotion recognition such as KNN [32], SVM [33], ANN [34] etc. Standard machine learning models cannot, however, achieve the high-level abstract features of the EEG [22]
- And also their work has tested only one or two classifiers. They have not verified whether their algorithm is giving better accuracy compared to all state-of-art classifiers, which may make their searching incomplete.

The recognizability of various emotions depends on how effectively it is possible to map the EEG features to the emotion representation selected. The current auditory brain-computer interface study concludes that it is better to train with visual input to increase or decrease the sensorimotor rhythm amplitude than with auditory feedback [35]. This is not linked to the recognition of feelings, although it is noted in the debate that a less evolved sense of hearing can be present in healthy individuals without eye problems [36]. Visual stimuli may be easier to recognize from brain impulses than audio stimuli since the visual sense is more developed. Following this logic, a combined attempt to evoke an emotion from both visual and auditory input can offer the optimal atmosphere for recognition of emotions [35]. In this paper the DEAP dataset [37] has been used which is available online. In their experiment, audio-visual stimuli have been used for emotion elicitation. Unlike other methods, we suggested a feature-driven classification model in which we categorized a broad variety of emotions driven on valence, arousal, liking, and dominance, taking the data from all 32 channels into consideration. And with our novel method of selection of features (i.e. Correlation matrix, ranking features by importance, and then recursive feature elimination method) and stacked ensemble generalized model

where so many classifiers are involved, we get better accuracy with all four scales.

Contributions

There is already an understanding of the current state of EEG-based emotion detection, and it is time to move back to the primary objectives of this study for its application within human media engagement. In this study, our major contributions are:

- Proposing a hybrid ensemble feature selection technique to select the optimizes feature set for emotion recognition.
- Selecting the optimized feature set considering the signals coming recorded by all the channels to classify the emotion in 4 scales.
- Proposing a Stacked Ensemble Generalized classification model to ensure a maximum number of emotion recognition in minimum time and computational cost.

B. Motivation of the work

The main reason for this model is the use of Deap data to attain improved classifications rates. Using a stacked ensemble generalization model, a new, exact EEG signal classification method is proposed to achieve that goal. Many model averaging methods have in nature been created, in which several sub models equally contribute to a composite forecast. Variable patterns are utilized to extract relevant features and the local feature generator/extractor is in the literature. The usage of effective feature generation and selection methods should be a highly accurate pattern recognition framework. To generate low, medium and high-level features for improved performance, a multi-level function generation should be used. Multilevel function generation is mostly due to the size of the function vector. A suitable function selector in this stage should be employed.Two primary aims lie in the feature selector. This increases classification capabilities and reduces the classifier's execution time.

C. Novelty and Contribution

In the field of EEG-based emotion detection, much study has been conducted, but few definitive results have been derived. The recognizably of multiple emotions depends on how well EEG characteristics can be mapped to the emotion representation selected. The novelty of this work is it finds the best features for each emotion, and also we have applied all the standard classifiers for emotion recognition. To reduce overfitting, running time and for better accuracy selection of proper features is very important. In this work, we have followed three steps to select the right features for classification. Three steps involve three types of the feature selection method. The proposed feature selection and classification method are tested on DEAP data-set, and better accuracy has been achieved to recognize emotion based on valence, arousal, liking, and dominance compared to recent works. Also, another point of concern is channel selection. There is a reason why we have considered all 32 channels for this work. Because to develop all the possible emotions the whole brain is responsible. Some part is responsible for happiness and a totally different part is responsible for calmness or sadness or there may be some parts that are common for two different emotions but not the whole part. So if someone wants to classify a maximum number of emotions s/he should consider all of the channels. In this paper, we have classified emotions in four scales. Recently the use of brain signals to classify emotions gets attention as it is a challenging job, and it needs to be accurate. People used machine learning, as well as a deep learning approach to do that.

II. Proposed Methodology

A computational intelligence approach is proposed to provide better and reliable classifications of emotion from EEG signals. An overview of the proposed method is shown in Fig. 1. The major computational steps in the proposed approach are data preparation, feature Extraction, feature optimization, and classification. The various steps in the approach are stated below:

A. Data Preparation

These steps involve resizing noise removal and filtering. We need the data window to get the corresponding frequencies, peak values, and wavelet coefficients of the signal segments. It takes a moment to grasp the emotion against the stimuli. The EEG processing time is therefore a little longer than the previous recognition time [38]. But EEG signals are typically divided by windows into sections or fragments to get accurate results.

B. Data Segmentation

The findings of previous studies suggest that smaller $(1-4s)$ windows perform better than larger windows $(5-8s)$ [39]. So epochs of 3 secs (i.e. 384-time stamp) have been made for our research segmentation, and there are $(60/3)$ for each subject and each music video. Therefore there is $20 \times 32 \times 40$for all subjects and for all music videos (i.e. 20480 epochs).

C. Feature Extraction

And for corresponding to these 25600 epochs we have created 52 features for 32 subjects which includes time domain, frequency domain, and wavelet features. Because it is very tough to work with such high dimensional data which contains a thousand features. The method also reduced the work's complexity and improved the classifier's efficiency. The generated features are listed below (Table 1).

Let D is a data matrix of size $n \times m \times e$, where n is the no of channels, m is the n is $(segmentsize \times samplingfrequency)$ and e is the no of epochs.

D. Feature Elimination and Selection

After extracting the features the data matrix becomes D' and suppose k no of features are extracted. So now the size of the data matrix becomes $n \times k \times e$ .where $k < m$. Each column of the data matrix represents one feature (Fig. 2).
We used a feature optimization method which has three steps as follows:

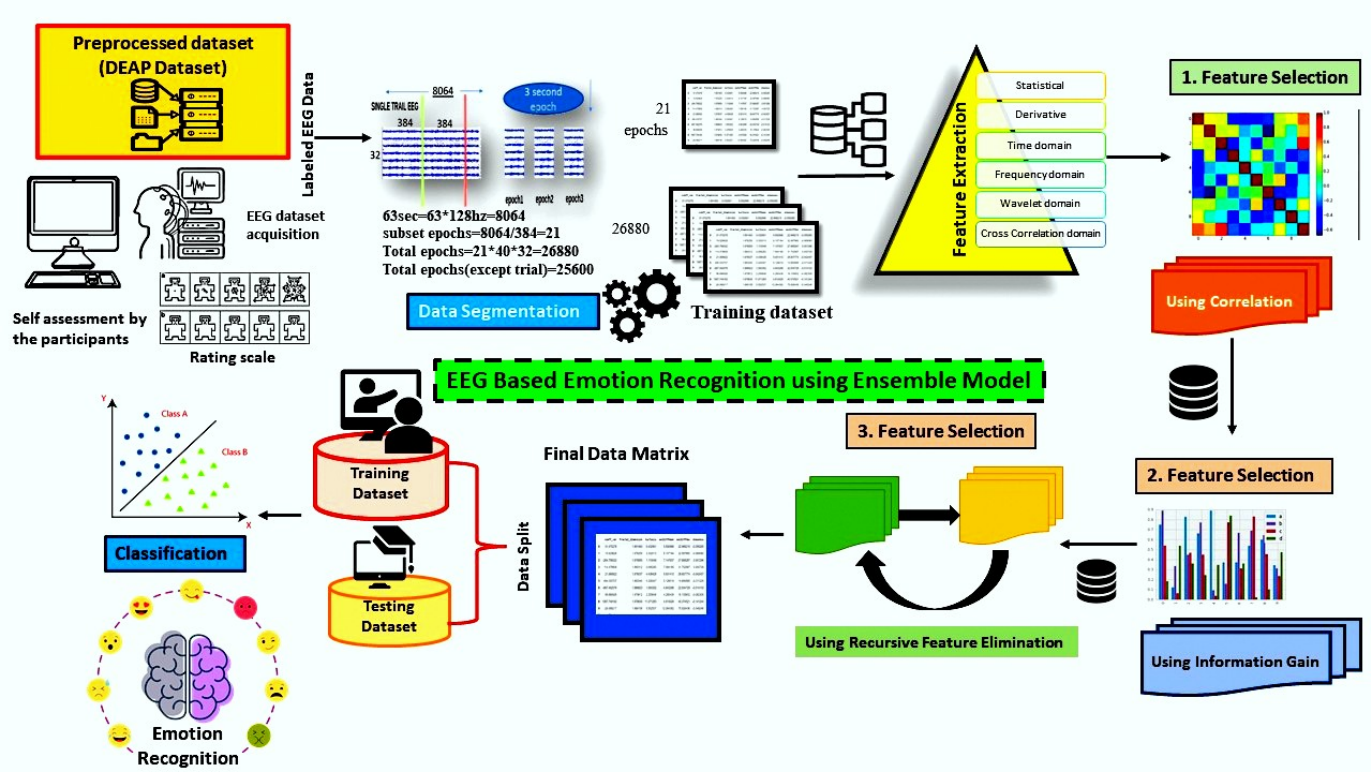


Fig. 1: Overview of the method

Step1. Eliminating redundant features:: These data features are strongly correlated with one another. Many strongly associated attributes should be eliminated and once they are eliminated, other approaches usually work better. There are some tools available to disable certain associated items.

$$B = \begin{bmatrix} 1 & b_{12} & b_{13} & ... & r_{1p} \\ b_{21} & 1 & b_{23} & ... & b_{2p} \\ b_{31} & b_{32} & 1 & ... & b_{3p} \\ . & . & . & ... & . \\ b_{p1} & b_{p2} & b_{p3} & ... & 1 \end{bmatrix} \quad (1)$$

where $r_{ab} = \frac{\sum_{i=1}^{n}(a_i-\bar{a})(b_i-\bar{b})}{\sqrt{\sum_{i=n}^{n}(a_i-\bar{a})^2}\sqrt{\sum_{i=n}^{n}(b_i-\bar{b})^2}}$ is the Pearson correlation coefficient between variable a and b. We can calculate the correlation matrix such as,

$$B = \frac{1}{n} D_s^{'} D_s \quad (2)$$

where $D_s = CDS^{-1}$ with
$C = I_n - n^{-1}1_n 1_n^{'}$ denoting a centering matrix
$S = diag(s_1, ..., S_p)$ denoting a diagonal scaling matrix.
We excluded features with an actual correlation of 0.90 or greater. As a result of strong overlap with other features 9 features are excluded. Let l no of features are left after removing the highly correlated features. Now the column size of the data matrix is changed to $n \times k'$ where $k' < k$. Let now the data matrix is D'' (Fig. 2).

Step 2. Selecting the features according to Ranking:: Creating a model may predict the value of features from the results. Decision trees have a mechanism built in to report variable significance. To measure value it measures information gain for all apps and scales it in 0-1. The benefit of knowledge informs us which properties of the function vectors are more important. We need to determine Entropy and Average entropy from the data to quantify the gain in

Algorithm 1: Eliminate Redundant Features

Input: Data matrix D' with $n \times k \times e$ features
Output: Reduced data matrix D'' with $n \times k' \times e$
Step 1.1: Calculate correlation matrix
$B \leftarrow$ calculateCorrelationMatrix(D') {Using Pearson correlation}
Step 1.2: Exclude highly correlated features for $i \leftarrow 1$ ***to*** n **do**
end
$j \leftarrow 1$ **to** n **if** $B[i][j] \geq 0.90$ **then**
end
Exclude feature i
Step 1.3: Update data matrix size
$l \leftarrow$ number of features left after removal
$D'' \leftarrow$ ResizeColumn$(D', n \times k')$ $\{k' < k\}$
Output: Reduced data matrix D'' with $n \times k' \times e$ features =0

information. The definition of entropy originates from the theory of learning. The more entropy the value the more the quality of the knowledge. It's a common way of measuring impurities, so we can remove less useful features. The entropy of attributes is calculated as:

$$Entropy(K) = -\sum_{n=1}^{N} p(K,N) \times log(p(K,N)) \quad (3)$$

Where N =number of classes, $p(K,n)$= the proportion of instances in K that are assigned to nth classes. Accordingly, the information gain by a training data set S is defined as:

$$Gain(K,S) = Entropy(K) - \sum_{v \in value} \frac{|S_{K,y}|}{|S_K|} Entropy(K_v) \quad (4)$$

Where Values (S_K) = set of values of K in S, S_K is the subset of S induced by K, and S_K,v is the subset of S in which attribute K has a value of v. Apart from the decision tree, some other algorithms are also available to estimate the importance using a ROC curve analysis which can be conducted for each feature. After estimating the variable importance it is plotted and printed so that we can decide to get the best features from those ranking. The most important attributes are taken eliminating the attributes with the least importance. After eliminating the least important features suppose we took k'' no of features. So now the data matrix D''' will look like as follows where $k'' < k'$. And so the column size of the data matrix will be reduced to k".

Algorithm 2: Select Features Based on Ranking

Input: Data matrix D'' with $n \times k' \times e$ features
Output: Reduced data matrix D''' with $n \times k'' \times e$
Step 2.1: Create predictive model and calculate feature importance
$G \leftarrow$ calculateInformationGain(D'') {Using Decision Trees}
Step 2.2: Estimate information gain and entropy for each feature for $i \leftarrow 1$ ***to*** k' **do**
end
Calculate entropy $E_{\text{feature}_i}(K)$
Calculate information gain $G_{\text{feature}_i}(K,S)$
Step 2.3: Plot and print variable importance
$V \leftarrow$ plotVariableImportance(G)
Step 2.4: Eliminate least important features
$k'' \leftarrow$ number of features to keep
$D''' \leftarrow$ ResizeColumn$(D'', n \times k'')$ $\{k'' < k'\}$
Output: Reduced data matrix D''' with $n \times k'' \times e$ features =0

Step 3. Recursive feature elimination method:: The recursive elimination of features (RFE) method determines the needed amount of features by picking the weakest features repeatedly. This method consistently removes and lists a limited number of characteristics every cycle. This method is used to reduce model dependencies and collinearity. It is an automated feature selection technique accessible in the Caret R package, which is also widely used. This regularly generates a model layout, then removes characteristics of low relevance. On each iteration, a Random Forest method is utilised to evaluate the model. The algorithm is intended to evaluate the attributes of all possible subsets. In each iteration, a random forest technique is used to validate the model, and this approach is designed in such a way that all viable subsets of the features may be evaluated. Recursive feature elimination is basically a backward selection process that incorporates all predictors into the model. Each predictor is scored according to its significance to the model. It is shown that the efficiency decreases in each iteration after re-calculation for the random forest model. This move makes the efficiency boost.

Algorithm 3: Recursive Feature Elimination (RFE)

Input: Data matrix D''' with $n \times k'' \times e$ features
Output: Final data matrix D_{final} with $n \times k_{\text{final}} \times e$
Step 3.1: Initialize RFE with all features
$k_{\text{current}} \leftarrow k''$
$D_{\text{current}} \leftarrow D'''$
while $k_{current} > k_{final}$ **do**
end
Step 3.2: Train model and evaluate feature importance
$G_{\text{current}} \leftarrow$ trainAndEvaluateModel(D_{current}) {Using Random Forest}
Step 3.3: Identify and eliminate least important feature
$f_{\text{remove}} \leftarrow$ argmin(G_{current})
$D_{\text{current}} \leftarrow$ RemoveFeature$(D_{\text{current}}, f_{\text{remove}})$
$k_{\text{current}} \leftarrow k_{\text{current}} - 1$
Step 3.4: Update final data matrix if $k_{current} = k_{final}$ **then**
end
$D_{\text{final}} \leftarrow D_{\text{current}}$
Output: Final data matrix D_{final} with $n \times k_{\text{final}} \times e$ features =0

E. Classification

The threshold for valence, dominance, arousal, and liking was set at 4.5, thus we made them a binary classification task. We will now train the model utilising massive amounts of data as a training set and the remaining data points as assessment data points. As a result, we used 60% of the data points for training and 40% for testing. Instead, we performed 5 fold-cross validation [40] for seven different classification models using the boosting approach, selected the best features, and then tested their accuracy. Ensemble machine learning techniques [41], [42] use numerous learning algorithms to achieve

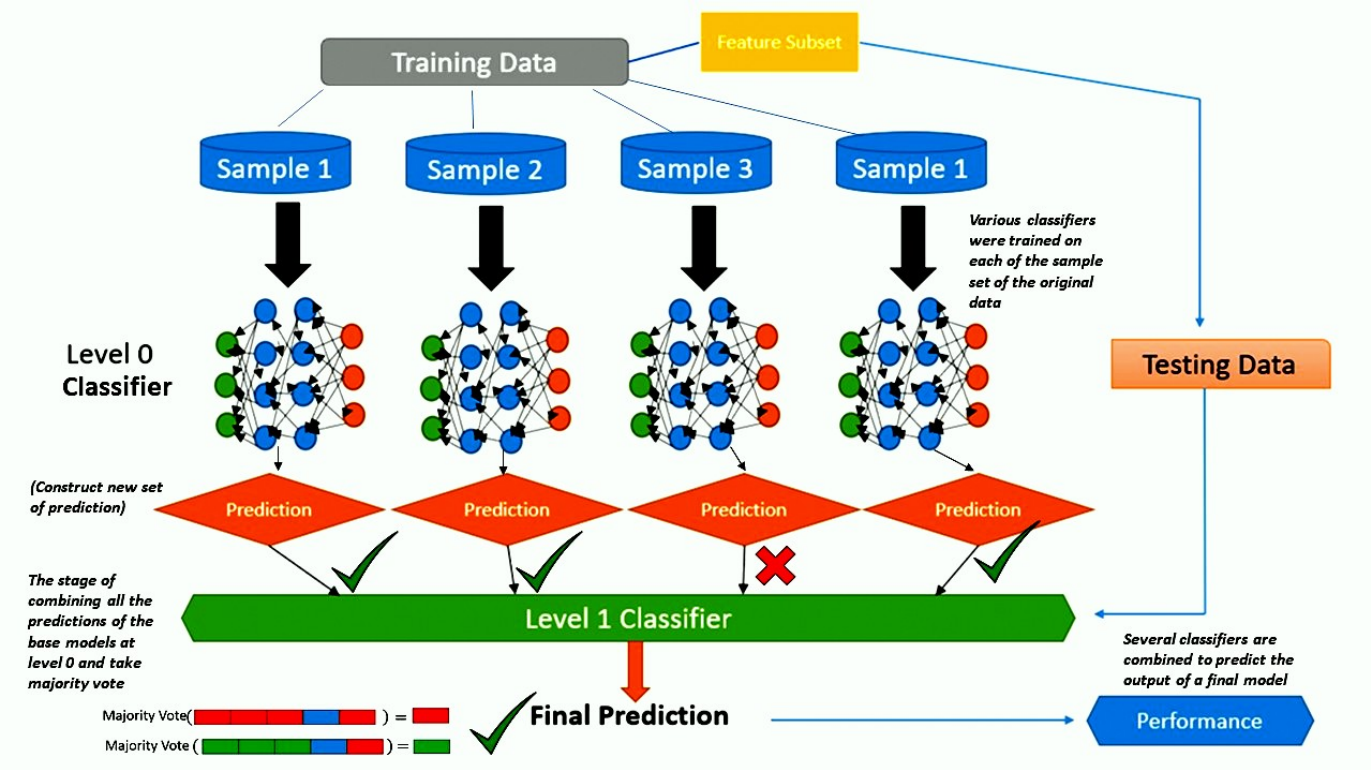


Fig. 2: Different stages of SEnGM

more prediction efficiency than any of the constituent learning algorithms could possibly achieve. Many prominent machine learning techniques nowadays are ensembles. Random Forest [43], [44] and Gradient Boosting Machine analyzing (GBM) [45], [46] are both ensemble learners. Bagging and boosting are both assembly methods that take a group of weak learners (e.g., decision trees) and combine them to generate a single, strong learner.

Stacking is an ensemble learning strategy that uses a meta-classifier or meta-regressor to merge numerous classification or regression models. The basic level models are trained using a complete training set, and the meta-model is learned using the outputs of the base level models as characteristics. The basic level also includes various learning methods, resulting in heterogeneous stacking ensembles. The following algorithm describes the stacking effect. Below is a step-by - step explanation of a basic stacked ensemble generalized model:

1) The set of trains is subdivided into 10 sections.
2) A base model (Random Forest) is fitted on 9 parts and projections for the 10th part are made. This is done for a part of a series of trains.
3) The base model (in this case, Random Forest) is then installed on the whole data collection for the train. Predictions are rendered on test set using this pattern.
4) Steps 2 to 4 are replicated with another base model (GBM, GLM), resulting in a further range of train set and test set predictions.
5) The train set predictions are used as characteristics for constructing a new model. This model is used for making final decisions about the set of test decisions.

For Stacked Ensemble Generalized Model (SEnGM): We used Random Forest, GBM (Gradient Boosting Method) and GLM (Generalized linear model) whose predictions are combined by GLM as a meta-classifier. We get an accuracy which are higher than individual classifiers.

Algorithm 4: Stacked Ensemble Generalized Model (SEnG)

Input: Training set X, labels Y, base models $M_1, M_2, \ldots, M_k$
Output: Stacked ensemble model M_{stacked}
Divide the training set X into n sections: $X_1, X_2, \ldots, X_n$
Initialize an empty list C
for $i = 1$ to n **do**
    Select $n-1$ sections as the meta-training set: $T = X \setminus X_i$
    for $j = 1$ to k **do**
        Train base model M_j on T
        Make predictions on section X_i using M_j: $P_i^{(j)} = M_j.predict(X_i)$
    end for
    Combine predictions $P_i^{(1)}, P_i^{(2)}, \ldots, P_i^{(k)}$ into a single vector: $c_i = [P_i^{(1)}, P_i^{(2)}, \ldots, P_i^{(k)}]$
    Append c_i to C
end for
Train a meta-model M_{meta} on C using labels Y
return M_{meta} =0

Algorithm 5: XGBoost

Input: Training set X, labels Y, number of iterations T, learning rate η, number of base models k
Output: Ensemble model M
Initialize ensemble model $M = 0$
Initialize residuals $r = Y$
for $t = 1$ to T **do**
  Compute gradients g and Hessians h for residuals r
  Initialize base models $M_1, M_2, \ldots, M_k$
  for $j = 1$ to k **do**
    Fit base model M_j to training set X with weights h
    Make predictions on training set X using M_j: $P^{(j)} = M_j.predict(X)$
    Update residuals $r = r - \eta \cdot P^{(j)}$
  end for
  Update ensemble model $M = M + \eta \cdot (M_1 + M_2 + \ldots + M_k)$
end for
return M =0

Fig. 3: Online self assessment tool [47]

III. Results

A. Data description

The freely accessible online emotion data set called DEAP data set is used in this analysis to measure the emotion from the physiological signal. The sampled, pre-processed, segmented data set was used in pickled Python / numpy formats. In this set of data 30 subjects $(S1 - S32)$ are shown 40 music video clips which should stimulate different types of emotions. After 20 trials the 40 video clips have been replicated for 40 trials with a brief pause. During the pre-processing period, the length of each video is 60 secs A planning time of 3 s was applied to each music video, and therefore the overall time of each video changed to 63 s. While viewing the video clip the EEG signals were captured by 32 channels that were mounted on each subject's scalp. After watching the video, the subjects were asked by an online assessment tool to rate the video clip, which contents questionnaire to underpin the state of emotion at the time.

B. Evaluation by Performance measures

The efficiency of the proposed approach has been measured with the following measure.

1) Confusion Matrix: A confusion matrix is a table often used to define a classification model's (or "classifier") output.

2) Recall and Precision: Precision-Recall is a useful measure of predictive success when the classes are extremely imbalanced. Precision is a measure of the relevance of the results in information retrieval, while recall is a measure of how many genuinely relevant results are returned. Recall (R) is defined as the number of true positives (Tp) plus the number of false negatives (Fn).

$$R = \frac{T_p}{T_p + F_n} \tag{5}$$

Precision (P) is defined as the number of true positives (Tp) over the number of true positives plus the number of false positives (Fp).

$$P = \frac{T_p}{T_p + F_p} \tag{6}$$

3) F1 Score: A statistical indicator of binary classification is called F1-score or F-score or F- measure to indicator the test accuracy. To calculate the score, we must find the test's recall (R) and precision (P) value. The F1-score rule is as follows;

$$F1 - score = \frac{P \times R}{P + R} \tag{7}$$

4) ROC Curve: Diagnosis of the receiver operating characteristics by ROC curve. This is a graphical plot that shows a binary classifier system's output calculation, as its threshold for discrimination is varied. The ROC curve is generated by the plotting at several threshold settings of the true positive rate (TPR) against the false positive rate (FPR).

$$TPR = \frac{T_p}{T_P + F_N} \tag{8}$$

$$FPR = \frac{F_p}{T_N + F_P} \tag{9}$$

C. Result 1 (Comparison with other Feature Selection Method)

The data includes strongly correlated features. We've removed characteristics with an absolute correlation of 0.90 or greater. As a result of high correlation with other features 21 features are removed. It constructs a model repeatedly, and eliminates features of low value. We measured precision on the specific subset and got full precision with 8 attributes (Table 2). For liking 11 Features are selected by calculating accuracies on all possible subsets of 15 attributes. For arousal 16 Features are selected by calculating accuracies on all possible subsets of 27 attributes. For dominance 18

Features are selected by calculating accuracies on all possible subsets of 20 attributes. And finally for valence 7 Features are selected by calculating accuracies on all possible subsets of 7 attributes. Table 2 describes the collection of features after introducing three approaches for each range of emotions. There are primarily two types of systems for choosing features. Filter and Wrapper. Wrapper methods [48] use learning to notify search by algorithm across the space of the feature subsets are mainly two types of feature selection method is available. Filter and Wrapper Method. Wrapper methods [48] search through the space of feature subsets using learning to inform search by the algorithm [49]. They calculate the learning algorithm's estimated accuracy for each function that can be added to or removed from the subset of features [49]. Therefore, an induction algorithm is "wrapped around" the feature collection, such that the operators' bias that determines the search and that of the induction algorithm interact strongly [50]. Several common examples of wrapper methods include forward feature selection, backward elimination of features, recursive elimination of features, etc.

Forward selection [51] is an iterative process by which we begin without any feature in the model. We keep adding the feature in each iteration that improves our model better before a new variable is introduced that does not improve model performance. In backward elimination [51]–[53], we begin with all the features and at each iteration, we remove the least significant feature which improves model performance. We reiterate this until no development is noticed on the removal of features.

Another popular method is The recursive Feature Elimination Method [54] . This is a greedy technique in optimization, which seeks to find the highest performing subset in apps. It constantly generates templates and sets aside the better or worst performing function at each iteration. This develops the next iteration with the traits to the left before all traits are depleted. Instead, it lists the applications according to their elimination order [52].

The biggest downside of wrapper methods over filter methods is the former 's computational cost, which comes from calling the induction algorithm for each set of features to consider. This expense has motivated some researchers to build innovative methods to speed up the assessment process [55].

Although wrapper models require optimizing a predictor as part of the selection process, filter models rely on the training data's general characteristics to select features independently of any predictor [49]. Some of the popular filter methods are Correlation, Mutual Information, Relief, etc.

Correlation is described as the linear relationship measure between two quantitative variables, such as age and body size. The correlation you may also describe is a measure of how strongly one variable depends on another. If two variables are highly correlated between themselves, they focus on providing redundant target information. Essentially, with only one of the redundant variables, we can make an accurate prediction of the goal. The second variable does not add additional information in these situations, so eliminating it can help reduce the dimensionality and also the added noise.

Features can also be selected based on Mutual information which is a measure of the two variables being mutually dependent. It tests the amount of information gained concerning one variable by analyzing the other variable. In other words, it decides how much we can learn about one aspect by knowing another — it's a little like correlation, except it's more general about the shared knowledge.

Another popular method for binary classification that gained attention is Relief Method. Kira and Rendell [56], [57] proposed the initial Relief Algorithm inspired by instance-based learning [58], [59]. Relief tests a proxy metric for each product as an independent appraisal filtering product collection mechanism which can be used to calculate the 'quality or 'relevance' value to the target group.

We have tested each aforementioned method on our data set and noticed that the features selected by a single method is not able to perform well, wherever our method of feature selection is performing way better. Table 3 summarizes the comparison of classification accuracy of our proposed classifier along with the feature selection method that we followed and the other conventional method of feature selection. The ROC curve in Fig 5 and Fig 6 shows that selection of features step by step is improving classification accuracy for all the classes rather than applying a single wrapper or filter method. In the seven feature selection algorithms, the average classification accuracy of the proposed Correlation-Mutual Information-RFE algorithm is still the highest, which also verifies the advantage of the algorithm's performance.

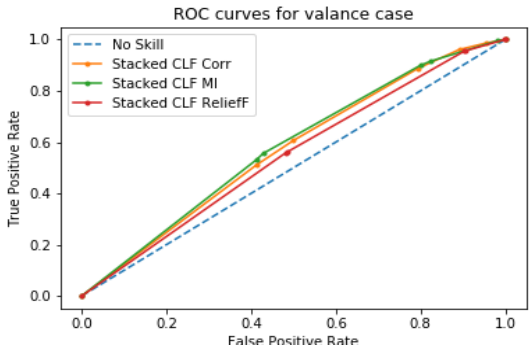
Result-2 (Comparison of performance with different number of features)

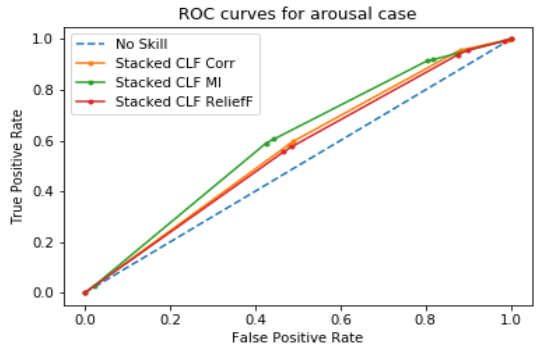
Another experiment is done with a different number of features to test how much the accuracy is improved after each step of the proposed method. Next, we'll check with 52 features. This ensures no set of features will be added and we must feed all the 52 features into the classifier. Here we have used a Random Forest classifier for the classification of emotion in four scales namely valence, arousal, liking, and dominance. Here also the threshold value is kept as 4.5.It can be proven from Table 3 that the precision is marginally increased as the number of functions reduces. Despite the reduced number of features the time to get the output is often significantly decreased. While there is no radical improvement in the classifier's accuracy with a decreased number of settings, the algorithm is fast and with 8 chosen features, the response time is cut.

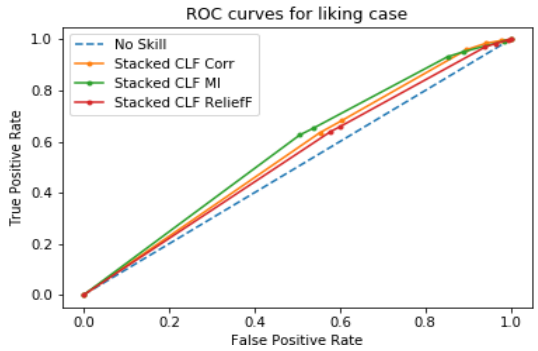
D. Result-3 (Classification Quality Comparison for different classifier)

For Valence, Arousal, Dominance, and Liking we held threshold 4.5 and made them a binary classification issue. By using 5 different classification models with boosting techniques and selecting the best features, we did 5 fold-cross-validation, and then measured the accuracy for each one.

Of all 4 classes, we implemented 7 different classification models such as Bayesian, Knn, SVM, C4.5, Decision tree, GBM, Random Forest, and XGBoost and the best accuracy







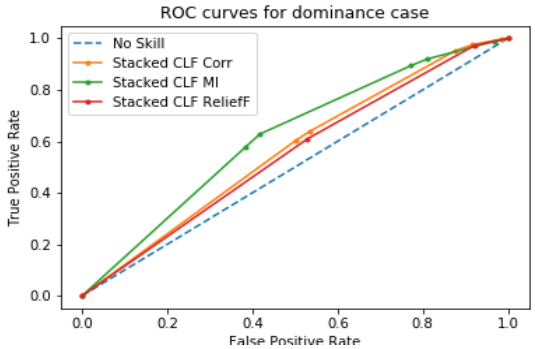
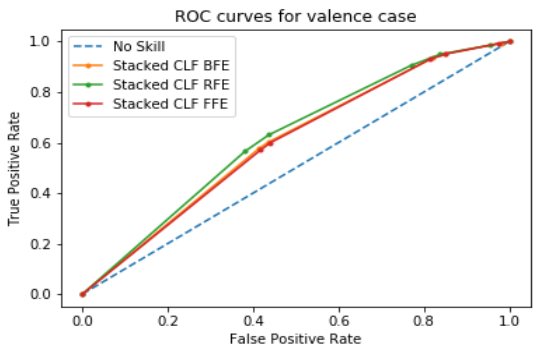
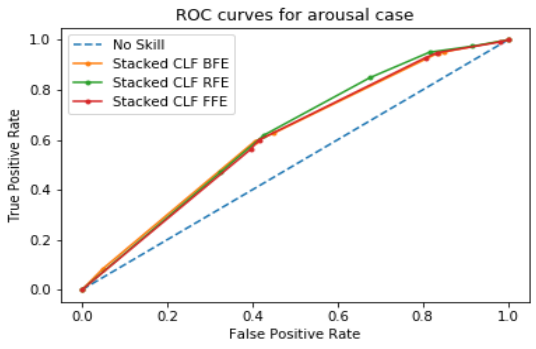
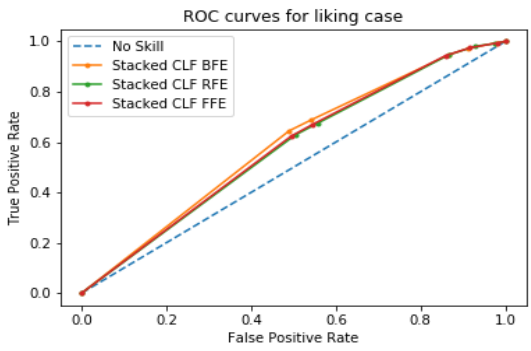


Fig. 4: ROC curve for Valence, Arousal, Liking and Dominance obtained from various types of Filter method and our proposed method. From the comparison of ROC curve it can be shown that our proposed feature selection is performing better than Correlation, Mutual Information and Relief Method.







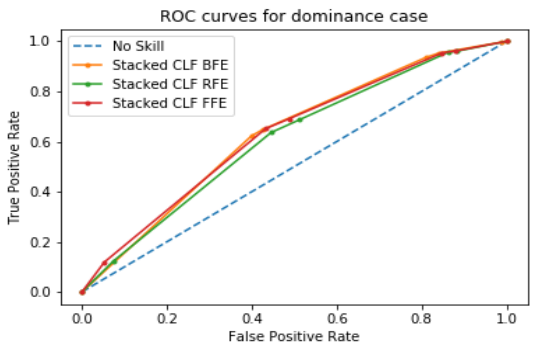


Fig. 5: ROC curve for Valence, Arousal, Liking and Dominance obtained from various types of Wrapper method and our proposed method. From the comparison of ROC curve it can be shown that our proposed feature selection is performing better than BFE, FFS and RFE method.

TABLE II: Comparison of performance metrics of the proposed method and other classifiers

Classifier	Scale	AUC	FPR	TPR	Accuracy	Gini	Recall	Precision	F1-Score
Random Forest	Valence	0.6361	0.9623	0.9917	0.7263	0.2721	0.9917	0.72786	0.83955
	Arousal	0.7045	0.8620	0.9720	0.7322	0.409	0.9720	0.73645	0.8380
	Dominance	0.7297	0.9264	0.9882	0.7438	0.4594	0.9882	0.7452	0.8497
	Liking	0.7201	0.8362	0.9675	0.7602	0.4403	0.9675	0.7690	0.8569
Decision Tree	Valence	0.56	0.9820	0.9968	0.7245	0.12	0.9968	0.7248	0.8393
	Arousal	0.6197	0.9722	0.9946	0.7166	0.2397	0.9946	0.7171	0.8333
	Dominace	06117	0.9767	0.9962	0.7362	0.2234	0.9962	0.7366	0.8469
	Liking	0.5998	0.9745	0.9953	0.7452	0.1997	0.9953	0.7462	0.8529
Gradient Boosting Method	Valence	0.6258	0.9790	0.9972	0.6258	0.251	0.9972	0.7255	0.8399
	Arousal	0.6773	0.9387	0.9896	0.7222	0.3546	0.9896	0.7232	0.8357
	Dominace	0.7008	0.9523	0.9916	0.7394	0.401	0.9916	0.7406	0.8479
	Liking	0.6851	0.9719	0.9961	0.7465	0.3703	0.9961	0.7468	0.8536
Bayesian Network	Valence	0.5315	1	1	0.7218	0.063	1	0.7218	0.8384
	Arousal	0.5402	1	1	0.7125	0.085	1	0.7125	0.8321
	Dominance	0.5335	1	1	0.7328	0.067	1	0.7328	0.8458
	Liking	0.5251	1	1	0.7421	0.0503	1	0.7421	0.8520
C4.5	Valence	0.5246	0.9320	0.9819	0.72988	0.25	0.9819	0.7345	0.8403
	Arousal	0.5632	0.8239	0.9409	0.7203	0.321	0.9409	0.7379	0.8272
	Dominance	0.5236	0.9088	0.9771	0.7437	0.569	0.9771	0.7503	0.8488
	Liking	05896	0.8535	0.9495	0.7492	0.214	0.9495	0.7699	0.8503
K Nearest Neighbour	Valence	0.6589	0.8619	0.9357	0.7224	0.012	0.9357	0.7484	0.8316
	Arousal	0.5896	0.7842	0.9356	0.7251	0.045	0.9356	0.7427	0.8281
	Dominance	0.6235	0.8195	0.9308	0.7375	0.056	0.9308	0.7659	0.8403
	Liking	0.6895	0.7814	0.9383	0.7562	0.069	0.9383	0.7814	0.8518
XGBoost	Valence	0.5795	0.8756	0.8456	0.6874	0.256	0.7658	0.7685	0.8569
	Arousal	0.6598	0.8456	0.7895	0.6523	0.365	0.7862	0.7456	0.8456
	Dominance	0.6598	0.8236	0.8569	0.6458	0.235	0.7236	0.7412	0.8652
	Liking	0.6456	0.8235	0.7965	0.6239	0.315	0.7569	0.8523	0.8463
Support Vector Machine	Valence	0.5248	0.8546	0.87596	0.6598	0.456	0.9548	0.8756	0.8756
	Arousal	0.5462	0.8463	0.7895	0.6235	0.2514	0.9263	0.8214	0.8412
	Dominance	0.6235	0.8412	0.8569	0.6548	0.124	0.8569	0.8146	0.7965
	Liking	0.5896	0.7845	0.7856	0.6215	0.365	0.9523	0.8246	0.7412
Proposed Method Stacked Ensemble Generalization Model (SEnGM)	Valence	0.7590	0.9501	0.9913	0.83300	0.012	0.9913	0.7339	0.8434
	Arousal	0.8456	0.8582	0.9782	0.8419	0.0156	0.9782	0.7433	0.8447
	Dominance	0.7896	0.9209	0.9912	0.8523	0.0189	0.9912	0.7520	0.8552
	Liking	0.7568	0.8514	0.9792	0.8642	0.0145	0.9792	0.7675	0.8605

is from the Stacked Ensemble system. The performance measures i.e the value of confusion matrix, recall, and precision, F1score, FPR, and TPR for every classifier along with our proposed method are shown in Table 2 for valence, arousal, liking and dominance. The summary of the comparison of accuracy for all of the classification techniques is described in Table 5. The ROC curves for valence, arousal, liking, and dominance from the above-mentioned classifier are shown in Fig 6.

E. Result-4 (Related Work and Comparison with our work)

To compare our work we have selected recent related work (Table 7) in which the machine learning methods are applied and tested on the same data set we have used. They tried to find the best emotion prediction approach by adjusting the number of channels, using different signal processing methods, and applying various types of ML techniques. We have therefore agreed to compare our proposed method with them to ensure that our method provides better accuracy than the most promising classifier. In Table 5 we have briefly described all the works started earlier and this suggests that with our feature selection approach and our classification methodology the model can classify the emotion into four scales i.e. valence, arousal, dominance, and liking, and also with reasonable precision given the signals coming from all the 32 channels.

IV. Research Contribution of the work

The popular techniques for emotion recognition include very few emotions by selecting signals from the channels responsible for those emotions only. But the human brain is so complex to understand when it comes to emotion. Also, the previous works had considered a few machine learning techniques to classify the emotion. Also, they have not considered all the emotions and there is no single technique is available to classify a maximum number of emotions in one go with sufficient accuracy for all of them. The proposed system can do all of the operations mentioned above. In the following, we have discussed the research contribution of the work with the reason behind doing so.

A. Why it is important to consider the signals coming from all of the channels?

Because to develop all the possible emotions the whole brain is responsible. Some part is responsible for happiness and a totally different part is responsible for calmness or sadness or there may be some parts that are common for two different emotions but not the whole part. So if someone wants to classify a maximum number of emotions s/he should consider all of the channels. In this paper, we have classified emotions in four scales. As this is a binary classification problem so for 4 scales we are getting 4*2=8 emotions classified accurately.

B. Why our feature selection method is unique and accurate?

To reduce overfitting, and running time and for better accuracy selection of proper features is very important. In this work, we have followed three steps to select the right features for classification. Three steps involve three types of feature selection methods. There are so many methods out there that give better accuracy if the highly correlated features are removed. So first we removed the highly correlated features. Next, we ranked the features by importance. The information gain is measured for each feature and based on that value we selected the best subset of features. Finally, we have applied the recursive feature elimination method. This method removes the attributes recursively. And then build a model with the rest of the attributes. To evaluate the model random forest algorithm is applied in each iteration. And thus it gets all the possible subsets of the attributes. So the combination of these three steps selects the best features to improve the performance of the classifier by getting accurate results and it is a novel idea.

C. Why the accuracy is better with stacked ensemble-based classification?

Ensemble learning improved the accuracy as it involves many classifiers and the outputs coming from all of those classifiers are combined and fed into a meta-classifier. So it is obvious that it will give much better results compared to each of them. The ensemble method also helps to reduce bias (boosting), and variance (bagging) and thus increases the stability of the classifier. Stacking adds an advantage by adding the ability to improve predictions.

V. Conclusion and Future Work

The proposed feature selection method increased classification accuracy. This experimental feature selection approach and Stacked Ensemble method provides 83.3 percent, 84.2 percent, 85.23 percent, and 86.42 percent accuracy for valence, anxiety, like, and superiority, respectively, which is better than most traditional classifiers accuracy. We hope that this analysis can be useful in the future to discover and develop new approaches for classifying human emotions that can help to discover new approach of man machine interaction. They will figure out certain types of feelings for future research. In addition, with different feelings, various important factors such as gender, age, and race should be weighed too.

In addition, considerations such as personality variations and temporal development will be recognized to make the automatic emotion detection models adaptable. Neural network research and teaching are challenging of mixed emotions. In potential research, this can be changed. The variability in the human emotional patterns must be acknowledged in the design. For different situations, the time dependency of the occurrence of emotion needs to be studied further and tested for improved results of the automated model of emotion recognition. We will aim to incorporate further BCI change in our future research which we are creating. We evaluated our concept here using just a machine learning algorithm. We should seek to gain greater precision in the future by applying

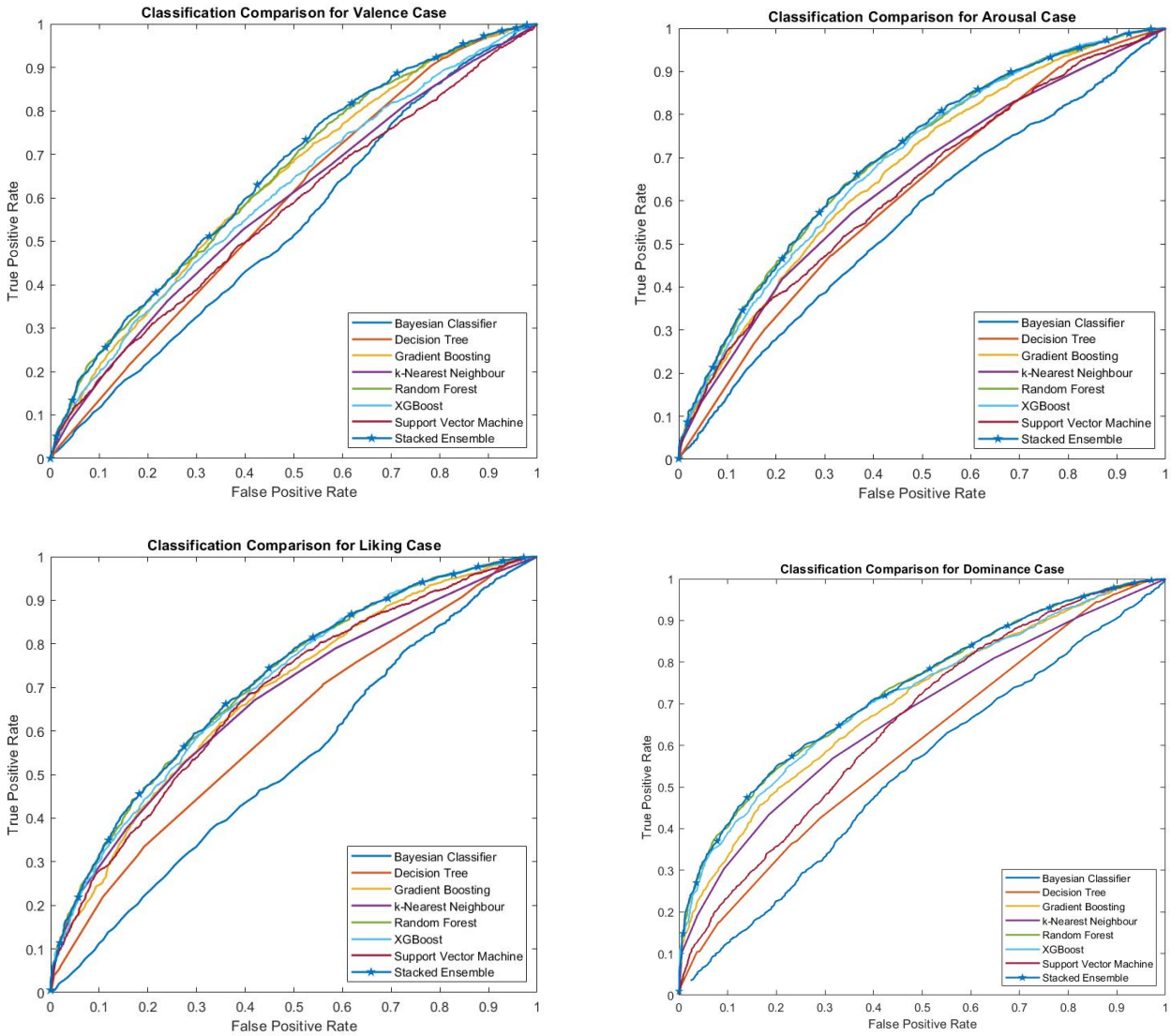


Fig. 6: ROC curve for Valence, Arousal, Liking and Dominance obtained from 7 different classifiers and our proposed method. From the comparison of ROC curve it can be shown that our Stacked ensemble model is performing better than other state of the art techniques

TABLE III: Comparison of performance with different number of features

Number of features	Classifier	Valence	Arousal	Dominance	LIking
52 (Total no of features extracted)	Random Forest	0.73214	0.7023	0.69624	0.74149
After removing redundant features	Random Forest	0.73521	0.7236	0.69568	0.74267
Once redundant features are eliminated and IG-based features selected	Random Forest	0.73987	0.72759	0.69789	0.74365
Once redundant features are eliminated and IG-based features chosen and RFE system implemented	Random Forest	0.84758	0.83254	0.80368	0.84512

a modern methodology. We will also study the Deep Neural Networks' success in detecting emotion.

REFERENCES

[1] N. P. Friedman and T. W. Robbins, "The role of prefrontal cortex in cognitive control and executive function," *Neuropsychopharmacology*, pp. 1–18, 2021.

[2] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.

[3] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, "Emotion analysis in man-machine interaction systems," in *International Workshop on Machine Learning for Multimodal Inter-*

TABLE IV: Comparison of our feature selection method with other methods

Method of Feature Selection		Valence	Arousal	Dominance	Liking
Filter Method	Correlation	0.570	0.567	0.565	0.551
	Mutual Info.	0.579	0.597	0.618	0.571
	Relief	0.548	0.556	0.549	0.535
Wrapper Method	FFS	0.598	0.609	0.631	0.578
	BFE	0.602	0.613	0.630	0.588
	RFE	0.618	0.620	0.612	0.574
Our proposed Method		*0.833*	*0.842*	*0.852*	*0.864*

TABLE V: Comparison of accuracy with the classifiers

Classifier	Valence	Arousal	Dominance	Liking
Bayesian	0.6866	0.42804	0.34421	0.71
KNN	0.7037	0.70246	0.71542	0.72453
C4.5	0.7072	0.70285	0.72234	0.73191
Decision Tree	0.7218	0.71250	0.73281	0.742187
Random Forest	0.7403	0.73561	0.74477	0.75878
GBM	0.7267	0.7255	0.7372	0.7598
XGBoost	0.6874	0.6523	0.6458	0.6239
SVM	0.6598	0.6235	0.6548	0.6215
SEn-GM	*0.8330*	*0.8420*	*0.8523*	*0.8642*

TABLE VI: Summary of feature selection after every method for all emotions

	Initial Features	Correlation	IG	RFE
Liking	52	31	15	11
Arousal	52	31	27	16
Valence	52	31	7	7
Dominance	52	31	20	18

TABLE VII: Comparison of our work with recent studies

References	No of Channels	Class	Classifer	Accuracy			
				Valence	Arousal	Dominance	Liking
I Wichakam et al. [60]	32	Binary	SVM (LOTO)	62.3%	62.9%	67.3%	NA
	10		SVM (LOSO)	64.9%	65.0%	66.8%	NA
FK Ahmed et al. [61]	32	Binary	DWPT GA-LS-SVM	48.30%	54.83%	NA	NA
			DWPT LS-SVM	46.33%	49.22%	NA	NA
			PSD+DLN (PCA+CSA)	55.07%	52.56%	NA	NA
			PSD+DLN	49.52%	49.52%	NA	NA
Koelstra S et al. [16]	10	Binary	kNN	57.6%	62%	55.4%	NA
Atkinson et al. [16]	14	Binary	SVM	73.14%	73.06%	NA	NA
A.Mart et al. [49]	18	Binary	ANN	72.87%	75.00%	NA	NA
S.Parui et al. [62]	18	Binary	XGBoost	72.67%	72.55%	73.72%	75.98%
Our Proposed Method	**32**	**Binary**	**En-FS+SEn-GM**	**83.30%**	**84.20%**	**85.23%**	**86.42%**

action. Springer, 2004, pp. 318–328.

[4] F. Bazzano, P. Montuschi, F. Lamberti, G. Paravati, S. Casola, G. Ceròn, J. Londoño, and F. Tanese, "Mental workload assessment for uav traffic control using consumer-grade bci equipment," in *International Conference on Intelligent Human Computer Interaction*. Springer, Cham, 2017, pp. 60–72.

[5] M. Strait, C. Canning, and M. Scheutz, "Limitations of nirs-based bci for realistic applications in human-computer interaction," in *BCI Meeting*, 2013, pp. 6–7.

[6] S. Mangalagowri and P. C. P. Raj, "Eeg feature extraction and classification using feed forward backpropogation algorithm for emotion detection," in *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)*. IEEE, 2016, pp. 183–187.

[7] X.-A. Fan, L.-Z. Bi, and Z.-L. Chen, "Using eeg to detect drivers' emotion with bayesian networks," in *2010 International Conference on Machine Learning and Cybernetics*, vol. 3. IEEE, 2010, pp. 1177–1181.

[8] A. Patil, C. Deshmukh, and A. Panat, "Feature extraction of eeg for emotion recognition using hjorth features and higher order crossings," in *2016 Conference on Advances in Signal Processing (CASP)*. IEEE, 2016, pp. 429–434.

[9] I. Wichakam and P. Vateekul, "An evaluation of feature extraction in eeg-based emotion prediction with support vector machines," in *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2014, pp. 106–110.

[10] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "Eeg-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2014, pp. 1–6.

[11] D. Le and E. M. Provost, "Emotion recognition from spontaneous speech using hidden markov models with deep belief networks," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*. IEEE, 2013, pp. 216–221.

[12] W.-L. Zheng, J.-Y. Zhu, Y. Peng, and B.-L. Lu, "Eeg-based emotion classification using deep belief networks," in *2014 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2014, pp. 1–6.

[13] C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin, "Emotion recognition from text using semantic labels and separable mixture models," *ACM transactions on Asian language information processing (TALIP)*, vol. 5, no. 2, pp. 165–183, 2006.

[14] O.-W. Kwon, K. Chan, J. Hao, and T.-W. Lee, "Emotion recognition by speech signals," in *Eighth European Conference on Speech Communication and Technology*, 2003.

[15] J. Edwards *et al.*, "Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review," *Clinical psychology review*, vol. 22, no. 6, pp. 789–832, 2002.

[16] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[17] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from eeg," *IEEE Transactions on Affective computing*, vol. 5, no. 3, pp. 327–339, 2014.

[18] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from eeg," *IEEE Transactions on Affective Computing*, 2017.

[19] H. Yang, J. Han, and K. Min, "A multi-column cnn model for emotion recognition from eeg signals," *Sensors*, vol. 19, no. 21, p. 4736, 2019.

[20] H. He, Y. Tan, J. Ying, and W. Zhang, "Strengthen eeg-based emotion recognition using firefly integrated optimization algorithm," *Applied Soft Computing*, p. 106426, 2020.

[21] Y. Liu and G. Fu, "Emotion recognition by deeply learned multi-channel textual and eeg features," *Future Generation Computer Systems*.

[22] Q. Zhong, Y. Zhu, D. Cai, L. Xiao, and H. Zhang, "Electroencephalogram access for emotion recognition based on a deep hybrid network," *Frontiers in Human Neuroscience*, vol. 14, 2020.

[23] D. Nie, X.-W. Wang, L.-C. Shi, and B.-L. Lu, "Eeg-based emotion recognition during watching movies," in *2011 5th International IEEE/EMBS Conference on Neural Engineering*. IEEE, 2011, pp. 667–670.

[24] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.

[25] Y.-P. Lin *et al.*, "Eeg-based emotion recognition in music listening," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1798–1806, 2010.

[26] Q. Zhang and M. Lee, "Emotion development system by interacting with human eeg and natural scene understanding," *Cognitive Systems Research*, vol. 14, no. 1, pp. 37–49, 2012.

[27] A. T. Sohaib *et al.*, "Evaluating classifiers for emotion recognition using eeg," in *International conference on augmented cognition*. Springer, 2013, pp. 492–501.

[28] Y. Liu and O. Sourina, "Real-time subject-dependent eeg-based emotion recognition algorithm," in *Transactions on Computational Science XXIII*. Springer, 2014, pp. 199–223.

[29] A. E. Vijayan, D. Sen, and A. Sudheer, "Eeg-based emotion recognition using statistical measures and auto-regressive modeling," in *2015 IEEE International Conference on Computational Intelligence & Communication Technology*. IEEE, 2015, pp. 587–591.

[30] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using eeg's and peripheral physiological signals," in *International workshop on multimedia content representation, classification and security*. Springer, 2006, pp. 530–537.

[31] K.-E. Ko, H.-C. Yang, and K.-B. Sim, "Emotion recognition using eeg signals with relative power values and bayesian network," *International Journal of Control, Automation and Systems*, vol. 7, no. 5, p. 865, 2009.

[32] M. Murugappan, "Human emotion classification using wavelet transform and knn," in *2011 International Conference on Pattern Analysis and Intelligence Robotics*, vol. 1. IEEE, 2011, pp. 148–153.

[33] Candra *et al.*, "Investigation of window size in classification of eeg-emotion signal with wavelet entropy and support vector machine," in *2015 37th Annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 7250–7253.

[34] A. Q.-X. Ang, Y. Q. Yeong, and W. Wee, "Emotion classification from eeg signals using time-frequency-dwt features and ann," *Journal of Computer and Communications*, vol. 5, no. 3, pp. 75–79, 2017.

[35] D. O. Bos *et al.*, "Eeg-based emotion recognition," *The Influence of Visual and Auditory Stimuli*, vol. 56, no. 3, pp. 1–17, 2006.

[36] F. Nijboer, A. Furdea, I. Gunst, J. Mellinger, D. J. McFarland, N. Birbaumer, and A. Kübler, "An auditory brain–computer interface (bci)," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 43–50, 2008.

[37] W. Liu, W.-L. Zheng, and B.-L. Lu, "Emotion recognition using multimodal deep learning," in *International conference on neural information processing*. Springer, 2016, pp. 521–529.

[38] M. Li, H. Xu, X. Liu, and S. Lu, "Emotion recognition from multichannel eeg signals using k-nearest neighbor classification," *Technology and Health Care*, vol. 26, no. S1, pp. 509–519, 2018.

[39] N. Thammasan, K.-i. Fukui, and M. Numao, "Application of deep belief networks in eeg-based dynamic music-emotion recognition," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 881–888.

[40] Y. Bengio and Y. Grandvalet, "No unbiased estimator of the variance of k-fold cross-validation," *Journal of machine learning research*, vol. 5, no. Sep, pp. 1089–1105, 2004.

[41] C. Zhang and Y. Ma, *Ensemble machine learning: methods and applications*. Springer, 2012.

[42] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*. Springer, 2000, pp. 1–15.

[43] A. Liaw, M. Wiener *et al.*, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[44] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.

[45] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 308–324, 2015.

[46] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[47] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap dataset (2012)," *URL http://www. eecs. qmul. ac. uk/mmv/datasets/deap/index. html. BibliographyA*, vol. 6.

[48] C.-H. Yang, L.-Y. Chuang, C. H. Yang *et al.*, "Ig-ga: a hybrid filter/wrapper method for feature selection of microarray data," *Journal of Medical and Biological Engineering*, vol. 30, no. 1, pp. 23–28, 2010.

[49] S. Das, "Filters, wrappers and a boosting-based hybrid for feature selection," in *Icml*, vol. 1, 2001, pp. 74–81.

[50] D. Koller and M. Sahami, "Toward optimal feature selection," Stanford InfoLab, Tech. Rep., 1996.

[51] R. Kohavi, G. H. John *et al.*, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.

[52] S. Kaushik, "Introduction to feature selection methods with an example (or how to select the right variables?)," *Analytics Vidhya*, 2016.

[53] R. Kohavi and D. Sommerfield, "Feature subset selection using the wrapper method: Overfitting and dynamic search space topology." in *KDD*, 1995, pp. 192–197.

[54] K. Yan and D. Zhang, "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," *Sensors and Actuators B: Chemical*, vol. 212, pp. 353–363, 2015.

[55] P. Langley *et al.*, "Selection of relevant features in machine learning," in *Proceedings of the AAAI Fall symposium on relevance*, vol. 184, 1994, pp. 245–271.

[56] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Machine Learning Proceedings 1992*. Elsevier, 1992, pp. 249–256.

[57] K. Kira, L. A. Rendell *et al.*, "The feature selection problem: Traditional methods and a new algorithm," in *Aaai*, vol. 2, 1992, pp. 129–134.

[58] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine learning*, vol. 6, no. 1, pp. 37–66, 1991.

[59] D. R. Wilson and T. R. Martinez, "Reduction techniques for instance-based learning algorithms," *Machine learning*, vol. 38, no. 3, pp. 257–286, 2000.

[60] I. Wichakam and P. Vateekul, "An evaluation of feature extraction in eeg-based emotion prediction with support vector machines," in *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE, 2014, pp. 106–110.

[61] F. K. Ahmad, A. Y. A. Al-Qammaz, and Y. Yusof, "Optimization of least squares support vector machine technique using genetic algorithm for electroencephalogram multi-dimensional signals," *Jurnal Teknologi*, vol. 78, no. 5-10, 2016.

[62] S. Parui *et al.*, "Emotion recognition from eeg signal using xgboost algorithm," in *2019 IEEE 16th India Council International Conference (INDICON)*. IEEE, 2019, pp. 1–4.