A Machine Learning Approach to EEG-based Prediction of Human Affective States Using Recursive Feature Elimination Method

Didar Dadebayev[1*], *Goh* Wei Wei[1], *Tan* Ee Xion[2]

[1]School of Computer Science and Engineering, Taylor's University, Subang Jaya, Malaysia
[2]Life Sciences, School of Pharmacy, International Medical University, Bukit Jalil, Malaysia

Abstract. Emotion recognition, as a branch of affective computing, has attracted great attention in the last decades as it can enable more natural brain-computer interface systems. Electroencephalography (EEG) has proven to be an effective modality for emotion recognition, with which user affective states can be tracked and recorded, especially for primitive emotional events such as arousal and valence. Although brain signals have been shown to correlate with emotional states, the effectiveness of proposed models is somewhat limited. The challenge is improving accuracy, while appropriate extraction of valuable features might be a key to success. This study proposes a framework based on incorporating fractal dimension features and recursive feature elimination approach to enhance the accuracy of EEG-based emotion recognition. The fractal dimension and spectrum-based features to be extracted and used for more accurate emotional state recognition. Recursive Feature Elimination will be used as a feature selection method, whereas the classification of emotions will be performed by the Support Vector Machine (SVM) algorithm. The proposed framework will be tested with a widely used public database, and results are expected to demonstrate higher accuracy and robustness compared to other studies. The contributions of this study are primarily about the improvement of the EEG-based emotion classification accuracy. There is a potential restriction of how generic the results can be as different EEG dataset might yield different results for the same framework. Therefore, experimenting with different EEG dataset and testing alternative feature selection schemes can be very interesting for future work.

1 introduction

Emotion takes an important part in human daily communication and behavior and is generally associated with an individual's mood and character. Despite its significant role in our daily life, all scientific information about the mechanism and process of emotional phenomena is yet to be fully explored. As technology has a part in almost all the areas of our daily life it would be meaningful to create a machine with a human-like intellect that can execute the needs of the user or can react depending on the surroundings. This would

---

* Corresponding author: didar2929@gmail.com

reveal new means for Brain-Computer Interface (BCI) technology and can be useful for various real-world situations. The outcomes of a successful emotion recognition research might be applied in the medical field, such as diagnosis of depression or its treatments. A machine with an ability to understand the human brain state can evaluate the effectiveness of a treatment or act as an auxiliary tool in determining a rigorous diagnosis. For people with motor impairments, it can enable some of their motor functions bypassing the neuromuscular system. A device with such abilities can save lives. In the engineering field, emotions can serve as metrics of humanity for smooth human-computer interaction (HCI) and the development of artificial intelligence. Research on EEG-based technologies might be challenging, but the importance and significance of potential applications give it a strong research value.

Even though there are no direct metrics to detect emotions, there are various indirect measures available to reveal emotional states. These measures can be obtained from a wide range of responses, including facial expression, muscle movement, and more recently evolved brain signals collected by neuroimaging techniques, such as Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). EEG was discovered at the beginning of the twentieth century, and only in recent times with the progress of artificial intelligence EEG became a vital tool in medicine, and research on EEG-based affective computing started prospering.

Reasonable data is the basis for machine learning. The process of EEG signal collection is troublesome and slow, depending on the number of channels and subjects involved. This implies an enormous time spent on the experimental process, including equipment setup, electrodes placement, and software setting. Considering the high cost of human resources and time, it might be challenging to acquire data for large-scale samples. Whereas, fewer scale samples might lead to overfitting of data and eventually yield misleading results. On the other hand, publicly available EEG databases are practical and solid, but the data acquisition and labeling might be different and not suitable for all models. Hence, one of the questions addressed in this study is how to construct an effective EEG-based emotion recognition framework using a public database. Another important aspect is the extraction of valuable features from the EEG data. There are various features tested and validated in previous EEG-based emotion recognition studies [1]. Frequency domain features are commonly extracted from EEG signals. Another proven approach is calculating geometric complexity features from the brain signals, such as fractal dimension values [2-3]. This work will follow the hypothesis that emotions can be detected from the EEG signals as fractal dimension values change.

The selection of the most valuable features is an important part of EEG-based emotion recognition research. To evaluate high-dimensional EEG features, the Recursive Feature Elimination (RFE) algorithm fused with the Support Vector Machine (SVM) algorithm will be implemented. RFE is a wrapper-based feature selection technique that adapts to a model and eliminates weak features until the specified number of features is achieved. The RFE has shown to be effective in ranking EEG features and selecting the most valuable of them by eliminating the EEG features that are not generic for all cases and developing a set of solid EEG features that are stable among all training and testing subjects [4].

This study will only cover the use of the Database for Emotion Analysis using Physiological signals, or simply DEAP, which was collectively harvested from several institutions [5]. This dataset, which is renowned in emotion recognition research, comprises EEG recordings from participants who were exposed to pre-arranged music videos and were asked to evaluate their level of arousal, valence, dominance, and liking. The DEAP dataset will be used to train and test the model presented in this work. However, there is a potential restriction of how generic the outcomes can be, as different EEG datasets might yield different results for the same classification model. There was no EEG signal

collection experiment carried out for this study, only the design of a framework including various features, classifiers, and pre-recorded brain activities to be presented for precise classification of human emotional states.

2 Background

The following subsections describe the main aspects of human emotions, including the definition of affective computing and BCI, models for emotion representation, and the structure of EEG signals.

2.1 Affective computing

For a long time, the study of human emotions has been hearth of much controversy among researchers. Some argued that it is not feasible to empirically study human emotional states as it is impossible to directly record the internal feelings of an individual [6]. Psychologists claimed that true inner brain states should be referred to as a black box, however, with the development of cognitive science, EEG turned into a powerful tool for tracking the neural activity inside the black box to reveal such phenomena as memory, cognitive load, and perception [7]. Previously researchers heatedly disputed whether emotions emerge within the cognition, therefore studies on affective states have been mostly skirted by scientists till the end of the twentieth century [8], and withal studies of emotions were thought to be very complex.

According to [9], affective computing can be described as "computing that relates to, arises from, or deliberately influences emotions" and it provides a vast range of applications for BCI. A BCI is a direct communication bridge between a brain and an external device, where nervous system activity is recorded and converted into an artificial output that can improve or even replace natural nervous system output [10]. Affective BCI is the subtype of the original BCI technology that aims to create a machine able to catch affective states from brain signals. An example of affective BCI is a prosthetic machine that attempts to recognize the emotional state of a user with a communication disorder [11]. Another example is the treatment of mental disorders by supporting a human's ability to change their mood through neurofeedback [12]. It is also applied in the entertainment field, such as adaptive games that follow a player's emotions to modify the gameplay [13]. Affective BCI is commonly used as an active link between emotions or as a passive emotion sensor to inform the device about a certain affective state. Affective computing has great potential, therefore it is an exciting research field for scientists and practitioners.

2.1 Models of emotion

The generation of emotions activates different brain areas and it is important to know how to classify them. Although there are different kinds of emotion classification models proposed by various researchers, no consensus on the general model has been reached.

The discrete emotion models sort out emotional reactions based on a small group of universal discrete emotions. They are called universal as they are common in different cultures and some are applicable for primates and other mammals, hence are naturally cultivated [6]. The discrete emotion models are frequently applied in emotion recognition systems where the computer predicts several emotional states, however, these models are not capable of fully reflecting the complexity of emotions.

The dimensional emotion models try to address the complexity of emotions by claiming that there are different dimensions of affective states. In this viewpoint, emotions are

classified by a set of scales, where each scale is a dimension of a multi-dimensional system. A dimension can be continuous or discrete, having a minimum and a maximum value. A certain emotion can be described based on the values on each scale or a point in a multi-dimensional expanse [14]. In this way, researchers can focus on emotion recognition tasks without concern about what category of emotion to follow. One of the commonly used dimensional models is Russel's circumplex model [15], where emotions are classified by two dimensions, the valence scale varying from pleasant to unpleasant and the arousal scale varying from active to inactive. For more comprehensive specifications of emotions, the three-dimensional Pleasure-Arousal-Dominance model was proposed by Mehrabian [16], where the dominance scale varies from dominant to submissive.

2.3 Electroencephalography (EEG)

EEG is an electrophysiological monitoring technique to collect information about the electrical activity of the human brain. The signals are formed by spontaneous and rhythmic impulses of the neurons towards certain events. In neuroscience and psychology, it is suggested that affective brain states and human behavior can be depicted by EEG signals [17]. Moreover, it is capable of revealing even subtle changes in the mental states of an individual. Thus, EEG is becoming a more popular modality for emotion recognition studies. EEG signals are delicate, making the process of acquisition quite difficult because they can be easily contaminated by other physiological signals. The majority of these interferences contain Electromyography (EMG), Electrocardiography (ECG), and Electrooculography (EOG) signals. Hence, raw EEG signals must undergo pre-processing procedures. On top of that, EEG signals are strongly non-linear and chaotic. Any physiological activity inside the human brain happens intuitively and directly affects signal structure. Since time and space factors from the brain activity are reflected continuously, EEG signals obtain a non-linear geometrical structure [18].

In EEG all electrodes are placed on the scalp according to the international 10-20 system, which was presented by the American Electroencephalography Society [19]. The numbers 10 and 20 indicate the distance between any adjacent electrodes, which is 10% or 20% of the whole length from front to rear side or left to the right side of the human cranium. Fig. 1 presents a 10-20 electrodes placement system.
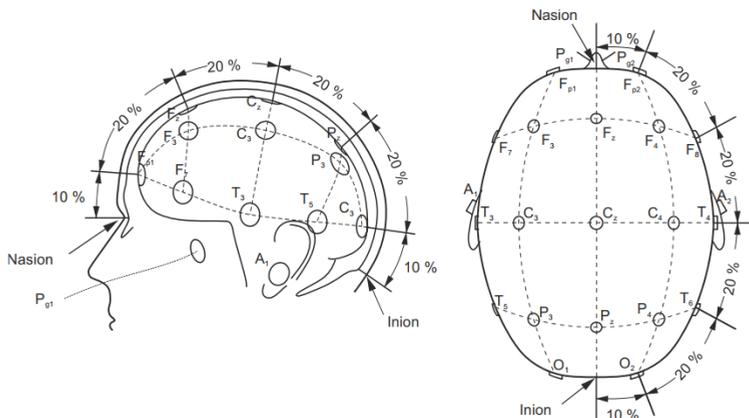

Fig. 1. 10-20 placement system [19]

EEG signals are generally characterized in terms of rhythmic activity which comes from human biological phenomena. These rhythms may be categorized into bands of frequency, namely delta (0.5 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 13 Hz), beta (13 – 30 Hz), and gamma (above 30 Hz). Once these frequency bands are obtained, the EEG signals are converted into feature vectors which are further used for the machine learning classification stage. The process of decomposing the EEG signal into its underlying sub-frequency bands is depicted in Fig. 2.
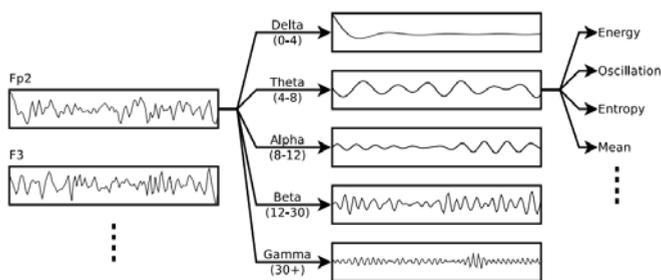

Fig. 2. Decomposition of every EEG electrode into its sub-frequency bands.

Researchers generally use 32, 64, or 128 channels electrode caps to collect EEG data. The extraction of one type of EEG feature for the corresponding electrode will lead to a large set of features. Subsequent extraction of different features will multiply this number. However, more features mean more data to be preprocessed, which affects the overall process of emotion classification. Thus, it is mandatory to reduce the number of EEG features and select those which possess information highly related to emotions. This can be achieved by applying feature reduction techniques. RFE is a wrapper-based feature reduction technique that adapts to a model and removes insignificant features until the conditioned amount of features remains. It is comparable to backward feature reduction. It begins with a group of features that are used to train a machine learning model and each feature is classified based on its significance, and the least significant are eliminated. This cycle is iterated recursively with the rest of the features. RFE is more resistant to overfitting compared to other feature selection techniques. SVM algorithm is frequently used as a learning model for RFE, and this fusion is known as SVM-RFE. Feature coefficients are calculated to verify the significance of each feature and rank them where values that are close to zero are labeled as less valuable [20].

3 DEAP Dataset

One of the main factors in the emotion recognition research field is the availability of adequate data for training and testing of the proposed model. In a study by [5], a dataset for emotion prediction studies was created, namely DEAP. This dataset comprises EEG, ECG, EMG, EOG, galvanic skin response, skin temperature, blood pressure, and respiration amplitude data of 32 participants while watching 40 one-minute-long excerpts of music videos. On top of these physiological signals, the face video recordings of 22 participants are provided. A total of 1280 trials are provided for furthers investigations. Upon completion of each trial, participants were asked to evaluate their level of arousal, valence, dominance, and liking, with a scale ranging from 1 to 9. EEG equipment that was used for the recording of the DEAP dataset has 32 electrodes, 4 located at the center, and the

remaining 28 are on the lateral sides of the brain cortex, forming 14 symmetrical pairs of electrodes.

A multimodal DEAP dataset offers not only raw physiological data but also a preprocessed version, where the EEG data were down-sampled to 128 Hz, and EOG and noise artifacts were removed by a bandpass frequency filter in a range of 4-45 Hz. The data were then averaged to a common reference. The preprocessed DEAP dataset is commonly used for a quick trial of different machine learning algorithms and it contains 32 separate files. Each file includes the physiological recordings of one volunteer watching 40 short music video excerpts, resulting in 40 sessions. For each session, a total of 40 channels have been used, 32 EEG channels, and 8 peripheral channels. Each channel provides 63 seconds of signal, where the first 3 seconds are excluded pre-session baseline and the rest 60 seconds are the actual session signals [5]. For this study, the preprocessed version will also be employed. As claimed by the originators of the DEAP dataset, it is well-suited for researchers to quickly test their proposed algorithms. Therefore, in our study, we utilized this dataset to examine our model. More information about the DEAP dataset is available in [5].

4 Related Works

The originators of the DEAP dataset, [5], conducted a classification of the collected ratings of arousal, valence, and liking using spectral power features extracted for each channel using Welch's technique. Altogether 216 features were extracted and Fisher's linear discriminant was applied for feature selection. A Gaussian Naive Bayes classifier achieved the accuracy results of 62% for arousal, 57.6% for valence, and 55.4% for liking.

In a study by [3] presented new affective EEG datasets, using audio and visual stimuli from IADS and IAPS. A consumer-grade EEG device, namely Emotiv, was used for data acquisition with 14 participants involved, and their feedback for valence, arousal, and dominance levels was collected. This study focused on the correlation between fractal dimension features and participants' reported ratings. Additionally, statistical and Higher-Order Crossing (HOC) features were extracted. An SVM-based model was tested for both the newly created database and the DEAP database. Results suggest that the fusion of all three types of features yields the best accuracy of 67.9% for recognition of 4 emotions. The accuracy was found to be consistent between the new databases and the DEAP database.

A study by [21] followed the fuzzy rules reported in previous emotion recognition literature [22], eventually presenting an algorithm for the classification of positive and negative emotions from EEG data. Statistical features were extracted from a pre-processed alpha band, including mean value, standard deviation, and means of the absolute values for the first and second differences. On top of that, the signal power of the alpha band was calculated. The study also presented an oscillation feature, obtained from all local maxima and minima of the signal. For feature reduction, Fisher's discriminant analysis was applied, revealing that oscillation and signal power values have a greater discrimination ratio compared to other features. Each fuzzy rule involved two inputs and one output, namely, certain feature values for an electrode pair as input and valence as an output. The obtained accuracy by fuzzy logic classifier was 62.62% with 10-fold cross-validation applied. Results were compared to SVM and Gaussian Naive Bayes classifiers, revealing that fuzzy logic-based algorithm outperforms.

In [23] a new strategy for emotion recognition based on Dual-Tree Complex Wavelet Packet Transform (DT-CWPT) energy features was suggested. Additionally, the difference between the energy features of all symmetrical pairs was extracted. For feature selection, Singular Value Decomposition (SVD), QR factorization with column pivoting (QRcp), and F-Ratio methods were applied. Then, the remaining features were classified with SVM.

Results were validated using the leave-one-out scheme by accuracy and F1-score values. The final results were 65.3% and 66.9% for valence and arousal, respectively.

A study by [24] presented a new emotion prediction model based on the Hierarchical Bayesian Network (HBN) that processes generic and individual characteristics of emotions for each subject singly during training and neglecting identity during testing. Spectral Power features for five frequency bands together with the difference between 14 symmetrical pairs were extracted. Principal Component Analysis (PCA) was employed for the feature selection stage. The model showed an accuracy of 58.4% and 58.0% for arousal and valence, respectively.

In [25] an EEG-based emotion recognition model was also verified using the DEAP dataset. However, they used only 14 EEG channels for feature extraction and the features extracted include fractal dimension, spectral power, and Hjorth parameters. Additionally, statistical features, namely standard deviation, kurtosis, and standard deviation were obtained. Feature selection was performed by Minimum Redundancy Maximum Relevance (mRMR), whereas SVM was trained as a classifier to recognize two-level arousal and valence states of participants. The proposed model achieved an accuracy of 73.14% for valence and 73.06% for arousal state recognition.

In another study by [26], the DEAP dataset was used to explore the correlation between bispectrum features and human emotions based on the valence-arousal emotion model with high or low levels. Basic bandpass filtering was used to remove unwanted artifacts. The backward sequential feature selection technique was applied to extracted spectral power features. The classification model based on SVM achieved an accuracy of 64.84%and 61.17%, for arousal and valence, respectively.

In [27] the ordinal pattern analysis, otherwise known as motifs, was utilized for better valence and arousal recognition. Motifs can identify iterating structures in the time domain and are noise persistent. Connectivity, asymmetry, and graph-theory features were extracted from the motifs of DEAP dataset. Analysis of Variance (ANOVA), Minimum Redundancy Maximum Relevance (mRMR), and RFE were used for feature selection, whereas SVM was employed for the classification stage. The highest accuracy achieved was 60.1% and 55.98% for valence and arousal, respectively.

A study by [28] conducted a cross-subject emotion prediction by proposing a novel multiple transferable RFE technique, or simply M-TRFE for both binary and multiclass classification. For binary classification, the least square support vector machine (LSSVM) yielded 68.98% and 64.94% accuracy, for valence and arousal, respectively, whereas for multiclass classification 65.13% was observed. The model was tested using the DEAP dataset and demonstrated improved results.

5 Methodology

The process of recognizing human emotions through EEG signals starts with the preprocessing stage, where various noises and artifacts are eliminated. Then, various feature vectors are obtained from EEG signals in the feature extraction stage. Then, extracted features undergo feature selection/reduction process to reduce feature size and facilitate the further process of emotion classification. Lastly, the selected features are fed into a machine learning algorithm to estimate and predict human emotional states [5]. Different studies propose different feature extraction and machine learning models depending on certain research objectives. Fig. 3 presents a general process of EEG-based emotion recognition research.

Obtaining DEAP Dataset	Signal Preprocessing	Feature Extraction	Feature Selection	Emotion Classification

Fig. 3. Process flow of EEG-based emotion recognition study.

To evaluate the effectiveness of the proposed framework, it must be validated in a real-life experiment. The DEAP dataset provided by [5] will undergo several steps, namely preprocessing of EEG signals, feature extraction from these signals, evaluating and selecting features by feature reduction method, and finally applying machine learning classification algorithm. The improved accuracy results will signify the effectiveness of the proposed model. The obtained results will be compared with the previous emotion recognition studies that have employed the same DEAP dataset.

5.1 Signal pre-processing

MATLAB (R2020a) signal processing toolbox will be utilized for building finite impulse response (FIR) filters in the preprocessing stage, as well as for the feature extraction process. A Hamming window with a filter order of 127 observations will be used to split EEG data into five frequency rhythms.

5.2 Feature extraction

The Power Spectral Density (PSD), also called spectral power, is usually calculated for different frequency bands and used to indicate the extent of the signal power over these bands. It is a frequency domain feature and is widely used for the studies of EEG-based emotion estimation. The most renowned technique to extract PSD features is called Welch's method [29], which aims to estimate the average power for each frequency by applying Fourier transforms, windowing, and averaging of small signal windows. Welch's method increases the accuracy of the typical periodogram, as the EEG signal is always time-varying, such as if you take 30 seconds of EEG data, it is a low chance that the signal will appear as an output of perfect sines. More likely, the spectral density of the EEG varies with time, continuously modified by the brain activity under the scalp. The challenge is that to obtain a true PSD data, a typical periodogram seeks a static EEG spectral content, which is not changing overtime period, which is never the case. Thus, spectral data is generally biased with high variance. Welch's method helps to significantly reduce this variance, by averaging the spectral content collected over short window segments. However, this is achieved at the cost of a reduced frequency resolution.

The oscillation feature was first presented in a study by [30], which helps to understand how signal power influences the activation and deactivation of specific brain areas. It was proposed that signal power, by means of high and low peaks, is highly associated with oscillation and activation/deactivation of certain brain areas. Therefore, the link between oscillation features and signal power features implies that both arise from the same emotional brain activity. The oscillation feature can be computed by dividing the total number of signal samples by the total number of the identified local maxima and local minima points of the signal. The pseudocode of the oscillation feature is presented below:

Get the signal $x(t)$, with $t = 1, 2, \ldots N$ samples
Set local minima, $L_{min} = 0$
Set local maxima, $L_{max} = 0$

$FOR\ t = 1\ to\ N - 2$
$IF\ x(t) > x(t+1)\ AND\ x(t+2) > x(t+1)$
$THEN\ L_{min} = L_{min} + 1$
$IF\ x(t) < x(t+1)\ AND\ x(t+2) < x(t+1)$
$THEN\ L_{max} = L_{max} + 1$
$END\ FOR$

Once we obtain L_{min} and L_{max} values, the oscillation features can be calculated as by the following equation:

$$O = N/(L_{min} + L_{max}) \tag{1}$$

The brain has a complex structure, thus EEG signals are generally non-linear and chaotic. The fractal dimension is a measurement of the geometrical complexity of a self-similar figure. It involves mathematical computations used for the extraction of the time-series features. The fractal dimension can help interpret the spatio-temporal dynamics of the EEG data. For instance, the fractal dimension values of EEG signals tend to be lower during the seizure activity in contrast to the normal brain state [31]. The EEG signals with similar power spectra might have different fractal dimension values, hence it can be deduced that fractal values can provide complementary information about brain activity [31]. Different studies highlighted the effectiveness of fractal dimension values for the EEG time-series feature extraction process. In [32] it was suggested that the fractal dimension as EEG features facilitate recognition of positive and negative emotional states. Fractal dimension features were also proven to be effective for EEG-based attention and concentration detection [33]. In a study by [34], fractal dimension features applied for emotional state recognition when listening to music. The traditional method of determining the dimension is to examine the scaling correlation between units of scale and the number of these units needed to measure shape, as shown in the following equation:

$$N \propto \varepsilon^{-D} \tag{2}$$

Here, N signifies the quantity of the newly scaled units needed to measure particular shapes, ε signifies the change in the unit of measurement when it increases or decreases and D denotes the fractal dimension. For a normal line, decrement in a unit of measurement by ε would imply that $N = 1/\varepsilon$, so that $D = 1$. Notably, higher resolution of a fractal line would lead to a higher level of geometrical complexity, meaning that more than $1/\varepsilon$ units would be needed when the unit of measurement is decreased by ε.

In this study, the Higuchi algorithm [35] is adopted for the fractal dimension values computation. The algorithm showed better accuracy results compared to other fractal dimension algorithms as presented in [36]. A detailed description of the Higuchi algorithm is given below.

Let $X(1), X(2), X(3) \dots X(N)$ be a finite list of time series samples. Then, a new time series can be constructed as:

$$X_t^m = X(m), X(m+t), \dots, X\left(m + \left[\frac{N-m}{t}\right] * t\right) \quad (m = 1, 2, \dots, t) \tag{3}$$

Where m is the starting time and t is the interval time [35]. The length of the curve is further defined by:

$$L_m(t) = (\sum_{i=1}^{\left[\frac{N-m}{t}\right]} |X(m + i*t) - X(m + (i-1)t|) * \frac{N-1}{\left[\frac{N-m}{t}\right]*t}$$ (4)

In the above expression, the term $\frac{N-1}{\left[\frac{N-m}{t}\right]*t}$ is referred to as a normalization factor for the length of the subset time series. For every time interval t, the mean value of the curve length through all values of m is estimated, to compute the value of $L(t)$ as (likewise $N \propto \varepsilon^{-D}$ mentioned before):

$$L(t) \propto t^{-D}$$ (5)

At this point, fractal dimension value *D* can be obtained by estimating the *L(t)* for different points of t plotted versus t and drawing a line to the points. The final value of D would be the magnitude of the slope of this line [35].

$$D = \frac{\ln L(t)}{-\ln t}$$ (6)

In a study by [3], the Higuchi algorithm was implemented and validated on the standard mono fractal signal generated by the Weierstrass function [37] where the theoretical FD values were known in advance. This was done to calculate one FD value per finite set of time series samples and Fig. 4 presents the computation of fractal dimension values of the Weierstrass function signal with different window sizes.
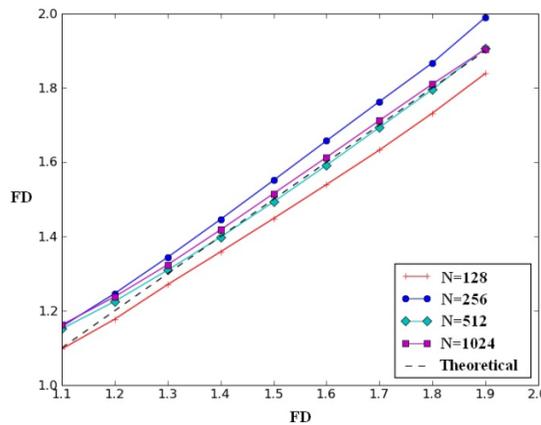

Fig. 4. Fractal dimension values of the Weierstrass function signal [3].

Results indicate that the window size of 512 samples yields fractal dimension values almost identical to theoretical. Therefore, in this study, the same window size of 512 samples is selected. In a study by [3], the t_{max} values ranging from 8 to 64 were tested to calculate the fractal dimension values. Having N = 512, the value of 32 for t_{max} has shown the least error rate as illustrated in Fig. 5.
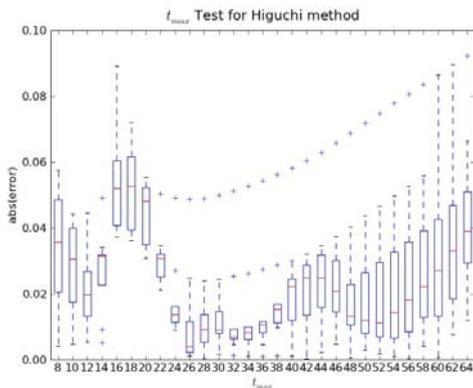
Fig. 5. Error values for various t_{max} for N = 512 [3].

The following features will be extracted in this study:

Table 1. Summary of extracted features.

Feature type	Number of features
Spectral Power features	160 (32 × 5)
Oscillation features	160 (32 × 5)
Fractal Dimension features	32
Total	352

5.3 Feature reduction

RFE is a type of a backward reduction of the features [38]. This method starts constructing a model on the entire set of features and obtains an importance value for each feature. The least important features are then eliminated, the model is reconstructed, and importance values are calculated again. The researcher would indicate the number of feature subsets to be analyzed together with the size of each subset. Thus, the size of a subset becomes a tuning parameter for RFE. The subset size that greatly improves the model performance is used to rank the remaining features.

RFE will be employed in this study to reduce the number of extracted features. SVM algorithm will be adopted as a learning method for RFE and features will be selected according to their significance in the coefficient matrix. Features that have zero or close to zero coefficient value will be marked as insignificant for SVM, whereas values more than zero, both positive and negative, will be considered as significant. Moreover, coefficient values will be further powered by two, so that values far from zero will have a more evident significance. One feature will be excluded per each iteration of RFE until reaching 25% of the whole number of features. Besides the RFE method, PCA will be evaluated for comparison. Feature reduction will be executed by Python's Scikit-learn toolkit [39], which features various feature reduction, classification, regression, and clustering algorithms, as well as validation options.

5.3 Emotion classification

A Support Vector Machine (SVM) is a type of classifier that utilizes a discriminant hyperplane to detect different classes. SVM is a powerful classifier. It locates low

dimension features into a higher dimension projection applying kernel functions which are effective for the inseverable cases [40]. When drawing a classification boundary to separate between two groups, one can notice that there are many possible ways of drawing this boundary. The challenge is how to decide which boundary is the best for your classification problem. One way of looking at it is you can take nearby data points and you can measure the distance that line to the data point. As illustrated in Fig. 6, this distance is known as margin. A line with a higher margin is better and that is what SVM tries to do. It will try to maximize the margin between the nearby data points and the line itself. These nearby data points are called support vectors and the boundary line is called hyperplane, which is a plane in an n-dimensional projection that separates different classes. Fig. 6 presents the optimal hyperplane selection by the SVM algorithm.
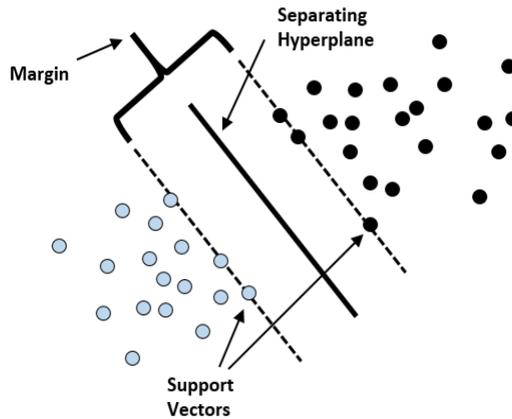

Fig. 6. The hyperplane estimation by SVM classifier.

In the SVM algorithm, the kernel function is responsible for getting the discriminant hyperplane. There are various types of kernel functions available. For instance, a linear kernel is employed for linear SVM to estimate a discriminant linear hyperplane. Other types of kernel functions can be applied to get a non-linear discriminant hyperplane, such as the Radial Basis Function (RBF) kernel and such a classifier is called SVM-RBF [41]. In this study, the polynomial kernel is adopted as presented in [42]:

$$K(x \cdot z) = (gamma \cdot x^T \cdot z + coef)^d \tag{7}$$

In the equation above, $x, z \in R^n, gamma$, and $coef$ are the kernel parameters, whereas T is the transpose operator and d indicates the polynomial kernel order. The excellence of the SVM classifier over other linear discrimination algorithms is that the hyperplane used is the one that increases the margins between different classes which enables the generalization capabilities [41]. A more detailed description of the SVM classifier is available in [43].

6 Conclusion and Future Directions

Emotions can comprehensively reflect human inner feelings and thoughts, thus they take a substantial part in human behavior and communication. Emotion estimation is a highly prospective area for BCI research as it can bring beneficial applications that can improve human daily lives. Once the development of EEG-based emotion recognition technology will turn into practical use, a bunch of key problems in various research fields will be

eliminated, such as currently known AI will gain a human-like intelligence with real emotions. The technology can also be used for emotional disorder treatment, BCI for the disabled, adaptive learning, adaptive games, and many more. Various models have been introduced lately for human emotion recognition. This experimental study will present a framework for improving emotion recognition based on EEG brain signals using a wrapper-based feature reduction method. The framework will be tested on a benchmark dataset DEAP which involved 32 participants. Three emotional scales are targeted for recognition, namely valence, arousal, and dominance. SVM-RFE will be used to reduce the number of extracted features by nearly 75%, and the remaining features will be further used for the classification step. Superiority and robustness of the proposed framework are expected, considering the stability of RFE performance and its resistance to overfitting. The efficiency of RFE will be compared with PCA feature reduction method. For the sake of robustness of the results, a 10-fold cross-validation scheme will be adopted, and mean values of the results based on accuracy and F1-score calculations will be reported. Additionally, the obtained results will be compared with previous EEG-based emotion recognition models that used the same DEAP dataset. Only those studies that split emotional scales into two levels (high/low) will be included in comparison with the proposed framework. The contributions of this study are primarily about the improvement of the EEG-based emotion classification accuracy.

Emotions are believed to be greatly individualized [44]. Hence, the majority of EEG-based emotion recognition studies have been conducted in a subject-wise way, by following the emotional labels from each participant to build subject-based models. However, this strategy limits its general usage, as the retraining and building separate subject-based models might be time-consuming and expensive. Building a system that can be readily applied for massive utilization remains a challenge for EEG-based affective computing.

As an extension of this experimental study, some other wrapper-based feature reduction methods, such as the Boruta algorithm or swarm optimization algorithm can be explored. Further investigations can focus on the evaluation of the advanced techniques of DLN, DNN, and CNN as a classifier. Additionally, the effect of age and gender for EEG-based emotion recognition can be studied in detail. Lastly, the utilization of other physiological datasets, such as SJTU Emotion EEG Datasets (SEED) [45] and MAHNOB-HCI [46] can be considered to validate the effectiveness and robustness of the proposed framework.

References

1. A. Al-Nafjan, M. Hosny, Y. Al-Ohali, A. Al-Wabil, Appl. Sci. (2017)
2. M.M. Javaid, M.A. Yousaf, Q.Z. Sheikh, M.M. Awais, S.Saleem, M. Khalid, Lec. Notes Comp. Sci. (2015).
3. Y. Liu, O. Sourina, Lec. Notes Comp. Sci. (2014)
4. Z. Yin, Z. Fei, C. Yang, A. Chen, IECON Proceed. (2016).
5. S. Koelstra, IEEE Trans. Affect. Comp. (2012).
6. J. Panksepp, *Affective neuroscience: the foundations of human and animal*. (1998).
7. S. Paradiso, Am. J. Psyc. (1998)
8. J.E. LeDoux, Annu. Rev. Neurosci. (2000)
9. E. L. Van Den Broek, Lec. Notes Comp. Sci. (2012).
10. J. R. Wolpaw, *Handbook of Clinical Neurology* (2013).
11. R. Corive, IEEE Signal Process. Mag. (2001).
12. D. C. Hammond, *Child and Adolescent Psychiatric Clinics of North America*. (2005)
13. D. Plass-Oude. *Brain-Computer Interfacing and Games* (2010).
14. K. R. Scherer, *Soc. Sci. Inf.* (2005)
15. J.A. Russell, J. Pers. Soc. Psychol. (1980).

16. A. Mehrabian, Curr. Psychol. (1996)
17. P. Ackermann, C. Kohlschein, J. Á. Bitsch, K. Wehrle, S. Jeschke, IEEE 18th Int. Conf. e-Health Net. Appl. Ser. (2016)
18. D. Garrett, D.A. Peterson, C.W. Anderson, M.H. Thaut, IEEE Trans. Neural Syst. Rehabil. Eng. (2003)
19. L. F. Nicolas-Alonso, J. Gomez-Gil, Sensors. (2012)
20. S. Maldonado, R. Weber, Inf. Sci. (2009)
21. J.W. Matiko, S.P. Beeby, J. Tudor, ICASSP. (2014).
22. A. Schmidt, J. Laurel, Cogn. Emot. (2001).
23. S.N. Daimi, G. Saha, Expert Syst. Appl. (2014)
24. Z. Gao, S. Wang, ICMR. (2015).
25. J. Atkinson, D. Campos, Expert Syst. Appl. (2016)
26. N. Kumar, K. Khaund, S. M. Hazarika, Procedia Comput. Sci. **84**, 31 (2016).
27. A. Tiwari, T. H. Falk, Comput. Intell. Neurosci. (2019)
28. J. Cai, W. Chen, Z. Yin, Symmetry. **11**, (2019).
29. P. D. Welch, IEEE Trans. Audio Electroacoust. (1967).
30. J. W. Matiko, S. P. Beeby, J. Tudor, ICASSP. 4389 (2014).
31. J. P. Pijn, J. Van Neerven, A. Noest, F. H. Lopes da Silva, Electroencephalogr. Clin. Neurophysiol. (1991).
32. L. I. Aftanas, N. V. Lotova, V. I. Koshkarov, V. P. Makhnev, Y. N. Mordvintsev, S. A. Popov. Int. J. Psychophysiol. (1998).
33. Q. Wang, O. Sourina, M. K. Nguyen, Int. Conf. Cyberworlds. (2010)
34. O. Sourina, V. V. Kulish, A. Sourin, IFMBE (2009)
35. T. Higuchi, Phys. D Nonlinear Phenom. (1988)
36. Q. Wang, O. Sourina, M.K. Nguyen, Vis. Comp. (2011).
37. P. Maragos, F. K. Sun, IEEE Trans. Signal Process. (1993).
38. I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Mach. Learn. (2002).
39. "scikit-learn: machine learning in Python — scikit-learn 0.23.2 documentation." [Online]. Available: https://scikit-learn.org/stable/. [Accessed: 21-Sep-2020].
40. W. S. Noble, Nat. Biotechnol. (2006).
41. C. J. C. Burges, Data Min. Knowl. Discov. (1998)
42. C. Chang, C.J. Lin, LIBSVM (2001).
43. N. Cristianini, J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. (2000).
44. S. Hamann and T. Canli, Cur. Opin. Neurobio. (2004).
45. "SEED Dataset." [Online]. Available: http://bcmi.sjtu.edu.cn/home/seed/. [Accessed: 21-Sep-2020].
46. "HCI Tagging Database - Home." [Online]. Available: https://mahnob-db.eu/hci-tagging/. [Accessed: 21-Sep-2020].