ERNN: A biologically inspired feed-forward neural network to discriminate emotion from EEG signal

Reza Khosrowabadi, Student Member, IEEE, Chai Quek, Senior Member, IEEE, Kai Keng Ang, Member, IEEE, Abdul Wahab, Member, IEEE

Abstract— Emotions play an important role in human cognition, perception, decision making and interaction. This paper presents a six-layer biologically inspired feed-forward neural network to discriminate human emotions from EEG. The neural network comprises a shift register memory after spectral filtering for the input layer, and estimation of coherence between each pair of input signals for the hidden layer. EEG data are collected from 57 healthy participants from 8 locations while subjected to audio-visual stimuli. Discrimination of emotions from EEG is investigated based on valence and arousal levels. The accuracy of the proposed neural network is compared with various feature extraction methods and feed-forward learning algorithms. The results showed that the highest accuracy is achieved when using the proposed neural network with a type of radial basis function.

Index Terms— Affective computing, EEG-based emotion recognition, Arousal-Valence plane, functional connectivity.

I. INTRODUCTION

EMOTIONS are not only important in human creativity and intelligence, but also in human rational thinking, decision making, curiosity and human interaction [1]. Research in emotions involved multidisciplinary areas such as psychology, neuroscience and affective computing. Although emotions have been studied extensively, the underlying neural mechanisms especially in terms of expression in response to emotional stimuli or perception of them are not studied extensively [2-4]. When a person experiences an emotional episode, the cognition process involved in understanding the situation can be distracted or facilitated [5-7]. The mechanism of such interactions between emotions and cognition can be investigated using functional and computational models [8, 9]. In general, these models are designed based on theoretical and experimental works in psychology and neuroscience and follow the related underlying biological mechanism. Subsequently, these architectures are applied to discriminate emotional states. Usually, experimental studies for emotion recognition are based on face, voice and biosignals features[10-14]. However, human biosignals are relatively more consistent across cultures and nations than face or voice features [15]. Therefore, using biosignal features would yield more consistent results.

Nevertheless, emotions are psycho-physiological phenomena associated with a wide variety of subjective feeling, thus it is difficult to propose a general architecture. However, almost all healthy human have similar patterns of cognitive process and physiological stimulation at same emotional states. Therefore, corresponding biological neural network involved in emotional perception is described. Subsequently, a feed-forward neural network for discrimination of emotions is proposed in this paper that could follow the underlying biological and functional mechanism.

Certain brain regions are associated in the processing of emotional stimuli. For instance, according to modern recording studies, the thalamus, amygdala, hippocampus, basal ganglia, insular cortex and orbitofrontal cortex are involved in emotion perception [16-18]. The amygdale is engaged in implicit memory tasks and the hippocampus is usually employed in explicit memory tasks [19]. These studies show that cortical and subcortical regions are involved in emotion regulation [20, 21]. In fact, information of the external world is acquired by sensory receptor cells and transmitted through synapses in neural pathway, spinal cord, peripheral ganglion, and brainstem to the thalamus. The thalamus maps the topography of this information to sensory area of the cortex. Subsequently, the mapped information is interpreted and encoded in neocortex. The encoded information is then classified using a task specific learning structure [22] as it is described in following paragraph.

The neocortex consists of the grey matter and is organized in vertical columns of 6 horizontal layers whereby the 1st layer is the outmost layer, and the 6th layer is the innermost layer. These 6 layers are separated by a characteristic distribution of cell types and neuronal connections. The structure of the
neocortex is relatively uniform, but there are exceptions of this uniformity such as lack of the 4th layer in motor cortex [23]. Furthermore, the neocortex is divided into the frontal, parietal, occipital, and temporal lobes. Each lobe performs a specific function, such as the occipital lobe engages in visual tasks, and the temporal lobe engages in auditory tasks. After the projection of information to the 1st layer of the primary sensory lobe in the neocortex, the selected information is then transmitted to other sensory areas that ultimately elicit a response. This process includes the segregation of information at the 2nd layer, the estimation of coherence between patterns at the 3rd layer, and the updating of the memory rate (eg. speed of motion) at the 5th layer. The attention level to the external stimuli also influences the 4th layer that has direct connections to other layers. After encoding the information, selective attention determines the information to move to short term memory, and then classifies them based on their meaning, and subsequently stores them in the long term memory.

Considering the fact that memory is not localized because different memory regions are used for different mental tasks and cognitive function cannot be assigned to any specific part of the brain [24]. Hence, the interactions between the brain regions during the perception of emotional stimuli will be more interpretable if the interactions can be explained functionally.

In general, the functional interactions between brain regions can be explained using top-down or bottom-up approaches [25]. Perception of emotional stimuli involves a deeper integration so that investigation of both top-down and bottom-up approaches are required [26]. In the top-down approach, emotion is described as a product of a cognitive process that translates the emotional stimuli using appraisal theory [27]. In the bottom-up approach, emotion is explained as a response to stimuli with intrinsic or learned properties and the reinforcement of them [28]. In this study, top-down approach is computationally modeled. The bottom-up approach is also substituted with subject’s feeling. The subject’s feeling is based on his/her previous experiences and measured using Self Assessment Manikin (SAM) [29] questionnaire. Therefore, a biologically inspired model is developed considering both top-down and bottom-up approaches. It is based on a functional model called EmoCog architecture [9] shown in Fig 1. The EmoCog architecture functionally describes the interaction between the brain regions involved in emotion regulation. This functional model corresponds to the biological brain structure in the outlook of the cognitive process [9]. In fact, EEG shows direct brain responses to external stimuli. It carries a lot of information related to translation and encoding of sensory information. Therefore, this signal is acquired during a controlled paradigm to evaluate the computational model.

The remainder of this paper is structured as follows. Section II describes the proposed feed-forward neural network for discrimination of emotions from EEG. Section III clarifies the experimental protocol. Section IV presents the experimental results. Section V concludes the paper.

II. ARCHITECTURE OF PROPOSED NEURAL NETWORK

The proposed biologically inspired feed-forward neural network is shown in Fig 3.

The neural network is construed to discriminate the emotions from EEG. The process of the emotional states discrimination in each layer of the proposed neural network is described as follows.

A. Functions of each layer

This section explains the connectionist architecture of each layer presented in Fig 3. For convenience of the readers, a list of essential mathematical symbols is described in Table I.

The multi-channel EEG data are the network input and the valence/arousal level is the output.

1) 1st layer – Spectral filtering

EEG signal is often contaminated by noises and artifacts such as AC power-line interference (50 Hz in Singapore), heart beat, ocular and muscular artifacts that mainly are located in lower frequencies. A spectral filtering (Fig 2) is performed on the EEG using a band pass filter to extract the rhythmic activity from 4 to 40 Hz using

\[ X'_i = \sum_{k=1}^{n} H^{k*} E_{i}^{k-n} \]  

(1)

where \( E_{i}^{n} \) denotes the \( n \)th sample of the \( i \)th channel of the acquired EEG data and \( X \) denotes the band-pass filtered data.

The EEG time series after spectral filtering \( X \in \mathbb{R}^{n_{ch}} \) are then applied as input to the 2nd layer.

2) 2nd layer – A short term memory

It has been shown in other studies that emotion variations last for some time till the next emotional episode happens, and these variations are detectable using EEG [11]. Therefore, the period of emotional episode represents the use of a short term memory. Typically, EEG data for a period of 1-4 seconds is used to discriminate an emotional state [30] because EEG is assumed to remain stationary during short intervals. In the proposed neural network, a serial-in/parallel-out shift register memory is used to accumulate the filtered EEG data for a
period of 1 second using a rectangular window. The optimum length of EEG data is selected using a genetic algorithm (GA).

The spectral filtered EEG $X$ is presented to a rectangular window $f_w$ to produce $W$. The rectangular window function is calculated using

$$f_w(n) = \begin{cases} 1 & s_t \leq n \leq N + s_t \\ 0 & \text{otherwise} \end{cases}$$

(2)

$$W^n_i = f_w(n)X^n_i; \quad \forall n \in \left[ s_t, N + s_t \right]$$

(3)

where $w^n_i$ denotes the $n$th sample of the $i$th channel of windowed filtered EEG and $s_t$ is the start point of window.

The $W$ is represented as input to the 3rd layer.

$$W = \left[ w_1, ..., w_{n_c} \right]$$

(4)

3) 3rd and 4th Layers- Connectivity features

Studies have shown that the cortical-subcortical interactions and interaction between different brain regions play an important role in perception of emotional stimulus [20, 31, 32]. Therefore, brain connectivity features would be very informative to investigate the relationship between emotion and cognition during the perception of emotional stimuli. In addition, the functional connectivity features can overcome the handedness issue because they are not reciprocal [33]. Therefore, the magnitude square coherence estimation (MSCE) [34] is applied to compute the functional connectivity features between brain regions from EEG signal. To compute the MSCE features, at 3rd layer of the network, the windowed time series EEG $W$ is transformed to frequency domain. The MSCE features are then computed in frequency domain with high resolution at 4th layer.

The weight between $n$th node of $i$th channel at 2nd layer and $j$th hidden node at 3rd layer is computed using

$$v_{n,i}^f = e^{-J2\pi f_i n}$$

(5)

where the $J$ denotes the imaginary unit and $e(.)$ is the exponential function.

The transfer function of $f_j$ hidden node at 3rd layer produces a response $z_i^f$ given in

$$z_i^f = \sum_{n=1}^{n_c} w_i^n v_{n,i}^f$$

(6)

After, at 4th layer the MSCE features [35] are computed using data transferred to frequency domain $Z$. The MSCE features are computed for all pairs of EEG channels in all frequencies.

The weights between hidden nodes of 3rd and 4th layers are considered to be one.
\[ V_{i,j}^2 = I, \forall i, j \in \{1, ..., n_c\} \quad (7) \]

Since the \(m^{th}\) hidden node at \(4^{th}\) layer computes the MSCE between pairs of \(z_iju\) and \(z_{j'}\), the transfer function for \(m^{th}\) hidden node at forth layer \(c_{m}^j = c_{m}^{j'}\) is computed using

\[ c_m^j = \frac{|z_m^i z_m^{j'}|^2}{(z_m^i z_m^{j'}) (z_m^{j'} z_m^{j'})} \quad (8) \]

where the \(z_m^{j'}\) denotes the complex conjugate of the \(z_m^{j}\).

However, some of these extracted features \((c_{m}^{j'})\) are irrelevant or redundant and have a negative effect on the accuracy of the classifier. In addition, structure of neural network at next layers is chosen based on the number of selected features. Consequently, the network would be very computationally extensive in case of using the huge number of features. Therefore, a number of significant features should be selected. Several supervised and unsupervised method can be applied. In this study, nonnegative sparse principal component analysis (NSPCA) is used to extract the significant features in unsupervised manner. NSPCA transforms the original features to a lower dimensional space. This transformation maximizes the variance of the transformed features using parts of the original coordinates and creates a sparse projection [36]. In particular, this method obtains a more biologically grounded original features. Consequently, the network would be very computational extensive in case of using the huge number of features. Therefore, a number of significant features should be selected. Several supervised and unsupervised method can be applied. In this study, nonnegative sparse principal component analysis (NSPCA) is used to extract the significant features in unsupervised manner. NSPCA transforms the original features to a lower dimensional space. This transformation maximizes the variance of the transformed features using parts of the original coordinates and creates a sparse projection [36]. In particular, this method obtains a more biologically grounded original features. Consequently, the network would be very computational extensive in case of using the huge number of features. Therefore, a number of significant features should be selected. Several supervised and unsupervised method can be applied. In this study, nonnegative sparse principal component analysis (NSPCA) is used to extract the significant features in unsupervised manner. NSPCA transforms the original features to a lower dimensional space. This transformation maximizes the variance of the transformed features using parts of the original coordinates and creates a sparse projection [36]. In particular, this method obtains a more biologically grounded original features. Consequently, the network would be very computational extensive in case of using the huge number of features. Therefore, a number of significant features should be selected. Several supervised and unsupervised method can be applied. In this study, nonnegative sparse principal component analysis (NSPCA) is used to extract the significant features in unsupervised manner. NSPCA transforms the original features to a lower dimensional space. This transformation maximizes the variance of the transformed features using parts of the original coordinates and creates a sparse projection [36].

Initially, the extracted features \((c_{m}^{j'})\) are centered by subtracting off the mean. After, nonnegative principal components of the centered features are calculated. Finally, significant number of features \(n_{out}\) are selected. The \(n_{out}\) is a constant number and it is one of the network parameter that is selected using optimization method in this study. Size of original features extracted for a subject is \(n_{f} \times n_{f}^j\) as shown in Table I.

![Feature Ranking (NSPCA)](image)

Feature Ranking

The indexes of most significant features are applied to trigger the outputs of the \(4^{th}\) layer as shown in Fig 4 and Fig 3 using

\[ I = \{0, 1\}^{n_{out}}, I = f_s(c_{m}^{j'}, n_{out}) \quad (9) \]

where the \(f_s(.)\) denotes the feature ranking function and \(I\) denotes the indexes of activated hidden nodes at \(4^{th}\) layer. The active nodes are presented in orange at Fig 3. These selected features are the input of \(5^{th}\) layer and computed using

\[ C_m^j = I_m^j C_{m}^{j'} | I_m^j | = 0 \quad (10) \]

After, the most significant features are classified using a 2-layer, radial basis function type learning algorithm.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X[n]</td>
<td>(\in \mathbb{R}^{n \times 1}), (n)-channel time series EEG after spectral filtering; (n \in [1 n_j])</td>
</tr>
<tr>
<td>W[n]</td>
<td>(\in \mathbb{R}^{n \times 1}), (n)-channel accumulated EEG using a rectangular window (f_w); (n \in [1 N])</td>
</tr>
<tr>
<td>C[f]</td>
<td>(\in \mathbb{R}^{n \times 1}), Furrier transform of W[n]; (f \in [-F_s/2, F_s/2])</td>
</tr>
</tbody>
</table>

4) \(5^{th}\) and \(6^{th}\) layers - The classification stage

After choosing the most significant connectivity patterns between the brain regions, these patterns are correlated to emotional states in a feed-forward manner at layers \(5^{th}\) and \(6^{th}\).
According to EmoCog architecture, a one pass learning algorithm is implemented using a radial basis function type network. The RBF network is presented in two-layer structure. The activated hidden nodes at layer 4th \((C_n')\) are the input layer of this classifier. Each hidden unit at layer 5th implements a radial activated function. Following at layer 6th, the output unit implements a hard limit function on the weighted sum of 5th layer’s hidden units. The transfer functions of hidden nodes at layer 5th are calculated using

\[
\phi(V_m,Y,C,b_m') = g_l\left(b_m'\left\| C - V_m' \right\| \right) \tag{11}
\]

where \(C\) denotes the selected features, the \(b_m'\) denotes the bias of \(m\)th hidden node at 5th layer which is calculated using equation (12) and \(g_l(.)\) is also defined by equation (13).

\[
b_m' = \frac{\sqrt{\ln(2)}}{\sigma \sqrt{n}} \tag{12}
\]

\[
g_l(X) = e^{-x^T \cdot X} \tag{13}
\]

where \(\sigma\) denotes the spread of radial basis functions, \(b_m(.)\) is natural logarithm function and \(e(.)\) denotes the exponential function. 

The output unit function at layer 6th is also calculated using

\[
\hat{Y} = \sum_{n=1}^{n_h} V_n^4 \phi(V_n^3,C,b_n') + b' \tag{14}
\]

where \(\hat{Y}\) denotes the estimated classes labels and the \(b'\) denotes the bias of output unit.

In addition, the output of our binary classifier \(\hat{Y}\) is assigned to its class label using a hard-threshold (step-function) using

\[
Y' = g_2(\hat{Y},\theta) = \begin{cases} 1 & \hat{Y} \leq \theta \\ 0 & \hat{Y} > \theta \end{cases} \tag{15}
\]

where the \(\theta\) is calculated from training data using

\[
\begin{align*}
  u_1 &= \max(\hat{Y}) \\
  u_2 &= \min(\hat{Y}) \\
  \theta &= \frac{u_1 + u_2}{2} \tag{16}
\end{align*}
\]

Initially the 5th layer doesn’t have any nodes and the hidden nodes of 5th layer are added self-adoptively by orthogonal least square (OLS) algorithm [38]. The procedure is started by computing the errors associated with input vectors using

\[
e = \left( (P' \cdot Y')' \right) / \left( (\sum Y' \cdot Y')' (\sum P' \cdot P')' \right) \tag{17}
\]

where \(P = \phi(C',C,b_n')\).

Subsequently, the input vector with greatest associated error is detected and a node is added to 5th layer with weights equal to this vector. The network parameters at 5th and 6th layers are updated using equations (18), (20) and (21) respectively.

\[
V_n^4 = [V_n^3, C_n'] \tag{18}
\]

\[
\beta = Y \left[ \phi(V_n^3', C, b_n'); \text{ones}(1, N_n) \right] \tag{19}
\]

\[
b_n' = \beta(N_n + 1) \tag{20}
\]

where \(Y\) denotes the desired output of training set, and \(C_n'\) is transpose of the selected vector of \(C\) with greatest associated error. Then, the actual error of network is calculated using mean square normalized error. The actual error of network is then compared with defined goal; if the goal has not been reached, another node is added. This process is continued until the sum-squared of actual error falls below the defined goal error or the number of hidden layer nodes at layer 5th reaches to maximum defined number \(n_h\).

B. Learning and testing process

The learning process as shown in Fig 5 consists of three stages:

1) Computing the parameters of neural network in layer one, two, three and four in unsupervised manner (computing the MSCE features).
2) Selecting of active hidden nodes in 4th layer using an unsupervised method (NSPCA).
3) Computing the network parameters for 5th and 6th layers in a supervised manner (classification of labeled data).

The binary classes are configured using

\[
\begin{align*}
  Y_e \in \mathbb{N}^{n_h}, Y_e \in \{0,1\} \\
  Y_a \in \mathbb{N}^{n_h}, Y_a \in \{0,1\} \tag{22}
\end{align*}
\]

where \(Y_e\) denotes the valence group labels and \(n_h\) is the total number of subjects in this group. Similarly, \(Y_a\) denotes the arousal group labels and \(n_h\) is the number of subjects in this category.

In testing phase, stage 1 and 2 are repeated. The selected features are then classified using parameters calculated in learning phase.

However, this network is sensitive to value of \(\sigma\) and \(n_h\). Furthermore, radial basis networks even when designed efficiently tend to have many times more neurons than other comparable feed-forward networks in the hidden layer [39]. These parameters should be tuned properly to lead a high level of accuracy. Nevertheless, network can converge to an
optimum accuracy rates by applying a proper value for $\sigma$ and large enough value for $n_h$. In addition, the RBF network is fast and can be directly implemented in the network [40]. Therefore, other feed-forward learning methods are also applied such as extreme learning machine (ELM) [41], general regression neural network (GRNN) [42], k-nearest neighbor method (KNN) [43], Naive bayesian (NB) [44], support vector machine (SVM) [45]. The network accuracy using all mentioned methods is shown in Table II. The results confirm that the RBF network work better than other possible networks. The accuracy of network is also compared with higher order crossing [46] and discrete wavelet transform [47] which are the two existing feature extraction methods for emotion recognition from EEG.

C. Defining the emotional states- class label

Emotion theories and researches have suggested a number of basic emotions although there is no coherent definition [7, 48-51]. Basic emotions are defined as the emotions that are common across cultures and selected by nature because of their high survival factors [51]. Commonly accepted basic emotions include: happy, sad, fear, anger, surprise and disgust, and complex emotions such as shame and disappointment are a combination of these basic emotions [49]. Emotions can also be measured by two axes of valence and arousal plane [49, 52]. Valence measures unpleasant to pleasant, and arousal measures calm to excited levels. Basic emotions can then be mapped onto the valence–arousal space. However, different subjects may feel differently while they are exposed to similar emotional stimuli. Therefore, the emotional responses of subjects have to be ascertained using questionnaires. This task is performed using the SAM in this study. The SAM is a non-verbal pictorial assessment technique that directly measures the valence, arousal, and dominance levels associated with a person’s affective reaction to a wide variety of stimuli [53].

The proposed neural network is applied to discriminate the changes of the cognitive process in response to emotional stimuli. These changes are interpreted from the changes in EEG and mapped to subjects’ SAM responses in terms of valence and arousal. The process of emotion discrimination from EEG using the proposed neural network is shown in Fig 6 (a). Noted that the arousal and the valence levels are discriminated simultaneously using a parallel structure as is shown in Fig 6(b).

III. EXPERIMENTAL DESIGN

The performance of the proposed neural network is investigated using EEG data collected from healthy subjects. The experimental design in the collection of EEG data is described in this section.

A. Emotion Elicitation

Studies have shown that elicitor (subject elicited vs. event elicited), setting (controlled condition in the lab vs. real world), focus (expression vs. perception) and subject awareness (open recording vs. hidden recording) are factors that can influence the emotion elicitation results [13]. Subject elicited category refers to the instruction given to the subject to express a specific emotion (for example to mimic the facial expression of happiness), or recalling past emotional episodes. Event elicited category refers to use of some images, sounds, video clips or any emotionally evocative stimuli. The International Affective Picture (IAPS) [29], International Affective Digitized Sound System (IADS) [54], Bernard Bouchard’s synthesized musical clips [55] and Gross and Levenson’s movie clips [56] are used to elicit emotional response in this study. Although touch, smell and taste are also known to influence human emotion, these are less studied in the literature [57] and thus are not used in this study.
A combination of arousing pictures from IAPS (1022, 1026, 1052, 1110, 1200, 1220, 1920, 1932, 2030, 2040, 2071, 2141, 2165, 2205, 2276, 2312, 2340, 2345, 2530, 2700, 2900, 3530, 3571, 5720, 5779, 5780, 5800, 5811, 5814, 5836, 6010, 6821, 6834, 6836, 6883, 6940, 8120, 8350, 8461, 8497, 9220, 9530) and synthesized musical excerpts belonging to Bernard Bouchard are used to elicit emotional responses from the subjects. These data are generally accompanied by affective evaluations from experts or average judgments of several people. However, the emotional feeling or perception from a stimulus differs from subject to subject based on the subject’s experience.

Therefore, even though predefined evaluation labels are available, the SAM questionnaire is used in this study to rate the subjects’ emotions. An example of the SAM questionnaire is shown in Fig 7.

B. Experimental Protocol

The duration of emotion elicited can be categorized into three categories: full blown emotions that last from seconds to minutes, moods that last from minutes to hours, and emotional disorders that last for years or an entire lifetime [10]. Ideally, an emotion recognition system should be able to discriminate the emotional states from the EEG as fast as possible [58]. Hence, this study focuses on full blown emotions whereby the emotional stimuli are presented for one minute in a counterbalanced and random order. The protocol of the experiment is shown in Fig 8.

The data are collected with subjects seated in a comfortable chair in a registration room whereby the experimental procedure is explained to them. The subjects are then asked to fill in a handedness questionnaire [59]. The EEG is recorded using the BIMEC from Brainmarker BV. The BIMEC has one reference channel plus eight EEG channels with a sampling rate of 250 Hz. The impedance of the Ag/AgCl electrodes is kept below 10 kΩ. Considering the cerebral lateralization during emotional perception [60], the eight Ag/AgCl electrodes are attached bilaterally on the subjects’ scalps using the 10/20 system of electrode placement as shown in Fig 6(a) where the Cz is the reference channel [61]. EEG data are collected for a 6-min period of time that comprised of 1 min eyes-closed, 1 min eyes-open, and 1 min for 4 emotional stimuli. The eyes-closed and eyes-open resting states are applied to bring all subjects back to a similar mental state. The subjects are exposed to 4 emotional stimuli in different arousal and valence levels. The visual stimuli are displayed on a 19 inch monitor positioned 1 meter far from the participant’s eyes and the audio stimuli are played by speakers with a constant output power. The categories of emotional stimuli are presented randomly such that each stimulus category is seen one time for every subject.

C. Subjects

EEG data were collected from 57 healthy subjects (age: 17-33, 9 women and 48 men).

The valence and arousal levels measured from SAM questionnaire are used for labeling the EEG data. The valence and arousal levels provide a dynamic representation of the emotional states. The valence-arousal plane provides a dynamic representation of emotional states. The valence and arousal levels are evaluated separately. The boundaries between different classes are determined from the subjects’ answers to the SAM questionnaire. Negative emotions are labeled when Valence ≤ 3 and positive emotions are labeled when Valence ≥ 7. Calm emotions are labeled when Arousal ≤ 3 and excited emotions are labeled when Arousal ≥ 7. For example, the subjects with SAM responses if Valence ≤ 3 are labeled as negative whereas subjects with Valence ≥ 7 are labeled as positive.

IV. Experimental Results and discussion

It was explained in previous sections that the emotional states are recognized based on valence and arousal levels in this study. The negative and positive states from valence dimension and calm and highly exited states from arousal dimension are investigated (Class1 and Class3 in ). The EEG signals are acquired during watching audio-visual stimuli through an explained paradigm. The subjects’ emotional states are scored using SAM questionnaire. The collected EEG signals are then labeled with SAM responses. Subsequently, accuracy of the proposed biological inspired network is checked using this data. A single-trial EEG data of 1 second is used to test the proposed neural network and classification accuracy of the network for valence/arousal level identification is computed. A 5-fold cross validation method is then applied to validate the results. Nevertheless, performance of proposed network is evaluated using other trials with size of 1 second as well. Average and standard deviation of classification accuracies calculated for all trials using the proposed network is presented in Table II.

The cross validation method assesses how the results generated by the methods will generalize to an independent dataset. In 5-fold cross validation method, 1/5 of observation from the original sample set is used as the testing data, and the remaining observations are used as the training data. The partitions are generated using the “cvpartition” function in the MATLAB Bioinformatics toolbox. The population size of arousal and valence categories are shown in Fig 9 where the classes’ populations are \( n_v = 175(c_{v1} = 80, c_{v2} = 95) \), and \( n_a = 119(c_{a1} = 60, c_{a2} = 59) \).

![Fig 9. Population of classes in each group](image-url)
The input of neural network is the normalized EEG data in range of 0 to 1. The normalization is done to remove the DC component of EEG signal and it is done for each subject separately. Every individual element of each subject (EEG sample points of each channel) is divided by square root of summation of square of all its elements (EEG sample points of all channels acquired at the same time) as is given in

\[ x_i(n) = \frac{x(n)}{\sqrt{\sum_{i=1}^{n} x_i^2(n)}}, \quad x_i(n) \in [0,1] \quad (23) \]

where \( x_i(n) \) and \( x_i(n) \) denote the \( n \)th element of \( i \)th EEG channel before and after normalizing, respectively. After, the normalized EEG is processed in proposed neural network as explained in section II. The output of network is valence/arousal level. The results shown in Table II presents classification accuracy for arousal and valence recognition.

Table II. Inter-subject classification accuracy for EEG-based arousal and valence recognition

<table>
<thead>
<tr>
<th>Feature selection, Classification method</th>
<th>Parameters</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERNN</td>
<td>( n_o = 2n_a ) or ( 2n_a ), ( \sigma = 3.28 ) layer=10%</td>
<td>Arousal: 70.83, Valence: 71.43</td>
</tr>
<tr>
<td>MSCE-KNN</td>
<td>( n_o = 5 )</td>
<td>Arousal: 62.50, Valence: 62.86</td>
</tr>
<tr>
<td>MSCE-ELM(Sig)</td>
<td>Noise at 5th layer=10%</td>
<td>Arousal: 65.22, Valence: 65.71</td>
</tr>
<tr>
<td>MSCE-SVM</td>
<td>kernel rbf, ( \sigma = 6 )</td>
<td>Arousal: 66.67, Valence: 68.57</td>
</tr>
<tr>
<td>MSCE-GRNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 56.52, Valence: 57.14</td>
</tr>
<tr>
<td>MSCE-NB</td>
<td>Noise at 5th layer=10%</td>
<td>Arousal: 65.22, Valence: 68.57</td>
</tr>
<tr>
<td>Hj [62]-KNN</td>
<td>kernel rbf, ( \sigma = 6 )</td>
<td>Arousal: 45.83, Valence: 48.57</td>
</tr>
<tr>
<td>Hj-ELM(Sig)</td>
<td>( n_o = 24 ), ( n_a = 5 )</td>
<td>Arousal: 54.17, Valence: 51.43</td>
</tr>
<tr>
<td>Hj-SVM</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 47.83, Valence: 54.29</td>
</tr>
<tr>
<td>Hj-GRNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 45.83, Valence: 55.43</td>
</tr>
<tr>
<td>Hj-NB</td>
<td>( n_o = 24 ), ( n_a = 5 )</td>
<td>Arousal: 47.83, Valence: 51.43</td>
</tr>
<tr>
<td>FBM-KNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 47.83, Valence: 51.43</td>
</tr>
<tr>
<td>FBM-ELM(Sig)</td>
<td>Noise at 5th layer=10%</td>
<td>Arousal: 52.17, Valence: 51.43</td>
</tr>
<tr>
<td>FBM-SVM</td>
<td>kernel rbf, ( \sigma = 6 )</td>
<td>Arousal: 54.17, Valence: 54.29</td>
</tr>
<tr>
<td>FBM-GRNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 52.17, Valence: 51.43</td>
</tr>
<tr>
<td>FBM-NB</td>
<td>Noise at 5th layer=10%</td>
<td>Arousal: 47.83, Valence: 54.29</td>
</tr>
<tr>
<td>GMM STFT[63]-KNN</td>
<td>kernel rbf, ( \sigma = 6 )</td>
<td>Arousal: 52.17, Valence: 51.43</td>
</tr>
<tr>
<td>GMM STFT-ELM(Sig)</td>
<td>( n_o = 2 ), ( n_a = 5 )</td>
<td>Arousal: 47.83, Valence: 54.29</td>
</tr>
<tr>
<td>GMM STFT-SVM</td>
<td>Noise at 5th layer=10%</td>
<td>Arousal: 52.17, Valence: 57.14</td>
</tr>
<tr>
<td>GMM STFT-GRNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 47.83, Valence: 51.43</td>
</tr>
<tr>
<td>GMM STFT-NB</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 52.17, Valence: 51.43</td>
</tr>
<tr>
<td>HOC [46]-KNN</td>
<td>Noise at 8-30Hz, 4-40Hz</td>
<td>Arousal: 54.17, Valence: 54.29</td>
</tr>
<tr>
<td>HOC-KNN</td>
<td>( n_o = 5 )</td>
<td>Arousal: 58.33, Valence: 57.14</td>
</tr>
<tr>
<td>HOC-ELM(Sig)</td>
<td>( n_o = 5 )</td>
<td>Arousal: 54.17, Valence: 51.43</td>
</tr>
<tr>
<td>HOC-SVM</td>
<td>Noise at 8-10Hz, 4-40Hz</td>
<td>Arousal: 56.52, Valence: 57.14</td>
</tr>
<tr>
<td>HOC-GRNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 54.17, Valence: 54.29</td>
</tr>
<tr>
<td>HOC-NB</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 56.52, Valence: 57.14</td>
</tr>
<tr>
<td>W.C [47]-KNN</td>
<td>Noise at 8-68Hz</td>
<td>Arousal: 45.83, Valence: 47.83</td>
</tr>
<tr>
<td>W.C-KNN</td>
<td>Noise at 8-40Hz</td>
<td>Arousal: 54.17, Valence: 54.29</td>
</tr>
<tr>
<td>W.C-ELM(Sig)</td>
<td>Noise at 8-10Hz</td>
<td>Arousal: 42.86, Valence: 47.83</td>
</tr>
<tr>
<td>W.C-SVM</td>
<td>kernel rbf, ( \sigma = 6 )</td>
<td>Arousal: 56.52, Valence: 57.14</td>
</tr>
<tr>
<td>W.C-GRNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 45.71, Valence: 51.43</td>
</tr>
<tr>
<td>W.C-NB</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 54.17, Valence: 54.29</td>
</tr>
<tr>
<td>MI-KNN</td>
<td>Noise at 8-5Hz</td>
<td>Arousal: 45.83, Valence: 45.71</td>
</tr>
<tr>
<td>MI-ELM(Sig)</td>
<td>Noise at 8-10Hz</td>
<td>Arousal: 47.83, Valence: 48.57</td>
</tr>
<tr>
<td>MI-SVM</td>
<td>kernel rbf, ( \sigma = 6 )</td>
<td>Arousal: 52.17, Valence: 48.57</td>
</tr>
<tr>
<td>MI-GRNN</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 47.83, Valence: 51.43</td>
</tr>
<tr>
<td>MI-NB</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 52.17, Valence: 51.43</td>
</tr>
<tr>
<td>ERNN (All trials)</td>
<td>( \sigma = 0.8 )</td>
<td>Arousal: 65.85 ± 9.73, Valence: 67.3 ± 9.4</td>
</tr>
</tbody>
</table>

The proposed neural network has 4 variable parameters including \( N, n_{out}, n_h \) and \( \sigma \). The maximum number of hidden neurons is set to \( n_h = 2n_a \) for arousal level recognition and \( 2n_a \) for valence level recognition. Since the network with optimum parameters is appreciated and considering the fixed topology of the network, a straightforward approach is to apply a GA as an optimization tool to compute the optimum \( n_{out}, \sigma \) and \( N \).

The GA is an optimization algorithm, which is invented based on genetics and evolution. Usually, the initial population of individuals is generated randomly. After, the fitness function which is a measure of improvement of approximation is calculated for each individual. Then, crossover and mutation are performed on the selected individuals to create a new individual that replaces the worst members of the population offspring. These procedures are continued until the end-condition is satisfied [64].

In this study, the classification accuracy of network is considered as fitness function with population initial range of [250 1000], [1 100], and [0.01 10] for \( N, n_{out} \) and \( \sigma \) respectively. The fraction rate for crossover is 0.8 in a scattered format with applying an adaptive feasible method as the mutation function. Population size is 20 and maximum generation is 100. The programming is done using global optimization toolbox in Matlab. According to GA results, the \( N, n_{out} \) and \( \sigma \) are set to 256, 12, 3.28 respectively. The classification accuracies shown in Table II are computed using the optimum parameters.

In addition, for comparison reasons, different feature extraction techniques from EEG data for emotion recognition are implemented. The extracted features are then classified using different feed-forward methods including NB, KNN, and ELM with sigmoid output function. The results are shown in Table II, where the FBM denotes fractional brownian motion [65-68], Hj explains the Hjorth parameters including activity, mobility, and complexity [62]. Furthermore, two recent applied feature extraction techniques including the wavelet transform and zero order crossing are also implemented. The HOC presents the higher order zero crossing method that has been used by petrantonasik et al [46] and the W_C denotes the wavelet based features that has been applied by Murgappan et al [47] for EEG based emotion recognition. Murgappan et al have employed the discrete wavelet transform for 5 scales using the daubechies 4th order orthonormal bases and the extracted wavelet coefficients at the \([1,2,3,4,5]\)th scales that correspond to the alpha, beta, gamma band to estimate the wavelet energy based features called recoursing energy (REE) efficiency, its modified form namely logarithmic REE (LREE) and absolute logarithmic REE (ALREE). However, to have a fair comparison, both methods are tested either in the frequency ranges that have been selected their papers [8-30 Hz] and [4-68 Hz] or in our selected frequency range [4-40 Hz].

On the other hand, the ERNN computes the MSCE features between different brain regions at different frequencies ranged in [4,40 Hz]. Subsequently, it is very important to know what are the most effective EEG channels and the most effective frequencies for valence/arousal level recognition from EEG data. This task is performed automatically by doing the feature...
The results shown in Table IV also indicate that during perception of arousal level of emotional stimuli, more amount of changes will happen at functional connectivity between right central region (C4) and right frontal (F4) and parietal regions (P3) at lower EEG frequencies (alpha band). Furthermore, changes in arousal level of stimuli causes more changes in functional connectivity between right parietal (P3) and right central regions (C4), and right temporal (T8) and left frontal (F3) regions at upper EEG frequencies (beta and gamma bands).

The demonstrated results in Table III and Table IV are graphically illustrated in Fig 11. The connectivity in different frequencies has been highlighted in different colors. The blue line describes the theta band, the red line denotes the alpha band, the green line indicates the beta band and black dashed line refers to the lower gamma band.

It can be concluded from the results that brain has different policies (functional connectivity patterns) for recognition of valence and arousal level of the emotion. This phenomenon would be very useful in applications such as the improvement of the brain plasticity [69, 70] or the investigation of brain aging using various types of stimuli. There are evidences for a global posterior-anterior shift in aging [71]. However, it should be noted that these results are based on 8 EEG channels and using audio-visual stimulus.

Furthermore, obtaining the network parameters using other adaptive methods such as [72-74] should be considered instead of using GA algorithm for optimizing them. However, as is shown in Table II, the mutual information doesn’t lead to a good network performance, therefore, looking for a better adaptive method is suggested in future works.

V. Conclusion

This paper presents a biological inspired feed-forward neural network discriminating emotion from EEG based on valence and arousal levels. The EEG data from the perception of emotional stimuli in healthy participants are collected. The top-down approach is formulated and bottom-up approach is bypassed using SAM answers. Subsequently, the performance of the proposed neural network for discriminating emotions is
evaluated using the EEG data and SAM responses. The results show that there are patterns of brain regions connectivity in the perception of individual emotional stimuli. These patterns are detectable by estimating the connectivity between different brain regions from the EEG data. However, these patterns vary in different subjects but common patterns can be selected at specific frequencies. Nevertheless, the feed-forward architecture is presented by considering a constant level of attention, mood and mental health for all subjects. Therefore, further assessment for understanding the impact of attention level, moods and mental disorders on perception of emotional stimuli should be done. In addition, improvement of the network for multi-class valence/arousal problem is proposed for the future works.

Acknowledgment

The authors would like to thank Mr. Michel Heijnen and Mrs. Sau Wei Tan for their support in the data collection of the experiment and Drs. Björn Crüts for his valuable advice on designing the experiment.

References


