A FILTER-DOMINATING HYBRID SEQUENTIAL FORWARD SEARCH method for channel selection in EEG-BCI Based Emotion Recognition

PARAG JYOTI DAS

A FILTER-DOMINATING HYBRID SEQUENTIAL FORWARD SEARCH method for channel selection in EEG-BCI Based Emotion Recognition

Thesis submitted to the
Indian Institute of Information Technology Guwahati
for award of the degree

of

Master of Technology

by

PARAG JYOTI DAS

under the supervision of

Dr.Upasana Talukdar

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY GUWAHATI

April 2022

CERTIFICATE

This is to certify that the thesis entitled **"A Filter-Dominating Hybrid Sequential Forward Search method for Channel Selection in EEG-BCI Based Emotion Recognition"***, submitted by* **Parag Jyoti Das** *to the* **Indian Institute of Information Technology, Guwahati** *for the award of the degree of* **Master of Technology***, is a record of bona fide research work carried out by him under my supervision and guidance. The thesis, in my opinion, is worthy of consideration for the award of the degree of Master of Technology in accordance with the regulations of the Institute. To the best of my/our knowledge, the results embodied in the thesis have not been submitted to any other university or institute for the award of any other degree or diploma*

Dr.Upasana Talukdar,

Assistant Professor,

Department of Computer Science and Engineering

Indian Institute of Information Technology Guwahati

Date:
April 25,2022

DECLARATION

I certify that

a. The work contained in this thesis is original and has been done by me under the general supervision of my supervisor(s).

b. The work has not been submitted to any other Institute for any degree or diploma.

c. I have followed the guidelines provided by the Institute in writing the thesis.

d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

e. Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f. Whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

PARAG JYOTI DAS

ACKNOWLEDGMENTS

I acknowledge with deed sense of gratitude, the help and valuable guidance given to me at every stage of the work by **Dr.Upasana Talukdar** and other faculties of **Indian Institute of Information Technology, Guwahati** . Without their help I could not have produced this project.

Further, We express gratefulness to those also who have contributed their valuable inspiration and suggestion towards completion of this project .

ABSTRACT

Human brain is one altogether the most amazing organs that differentiate humans from all of the other organisms. Our brain doesn't just manage the organs, but also it can think and remember.People with disabilities have limited freedom to socialize. Literature reports that,15% of the world's population lives with some form of disability. Brain-computer interfaces (BCIs) can be played a vital role to reinstate the independence of disabled individuals.A remarkable developoment has been made and BCI research is out of laboratories obtaining wide acceptance. The fundamental idea of a BCI system is to capture the brain signals and translate them directly into the device commands, without any involvement of peripherals. Numerous brain imaging technologies are used to acquire brain signals. Some common examples are functional MRI (fMRI) or Functional resonance imaging,Electroencephalography (EEG), Magneto encephalography (MEG) etc.EEG is the preferred choice because of its low cost, portability,non-invasiveness and easy to use.Emotion recognition using EEG enables the direct evaluation of the inner state of a user.

The thesis aims to address the high dimensionality of EEG features by selecting the most relevant channels. Sequential Forward Search (SFS) is a popular feature selection method. The same concept is applied in selecting the best channels carrying the most discriminative information. The thesis mainly proposes a filter-dominating hybrid sequential forward search method with an aim to attain high efficiency in the performance of EEG-based BCIs. Experiments with this new hybrid approach have been conducted on DEAP dataset.In this dataset, EEG data were recorded from 32 healthy participants where 16 males and 16 female. The sampling frequency of the original dataset is 512Hz and the sampling frequency of the preprocessed dataset used for the study is 128Hz. During the experiment, each participants watched a one minute long music video. After each trial/video, the participants perform self-assessment of their level of emotions in terms of valence, arousal, like/dislike, and dominance. The wrapper method under consideration include support vector machine (SVM), and the filter method include the Fisher Score. We extracted the wavelet based features from the selected channels to capture the time-frequency representation followed by Genetic Algorithm Based Feature Selection. SVM is used for classification of emotional states into two classes: Low/High Valence (unpleasant to pleasant) and Low/High Arousal (bored to stimulated)

Key Words: Brain Computer Interface(BCI),Electroencephalography (EEG),Fisher Score,Support Vector Machine,Genetic Algorithm,Valence,Arousal.

List of Abbreviations

ANN	Artifcial Neural Network
BCI	Brain Computer Interface
CAR	Common Average Reference
CNS	Central nervous system
CSD	Current Source Density
CSP	Common Spatial Pattern
ECG	Electrocardiograms
EEG	Electroencephalogram
EGG	Electrogastrography
FFT	Fast Fourier Transform
fMRI	Functional magnetic resonance imaging
fNIR	Functional Near InfraRed
GMM	Gaussian Mixture Model
GUI	Graphical user interface
KNN	K Nearest Neighbor
LDA	Linear Discriminant Analysis
MEG	Magnetoencephalography
MSR	Magnetically Shielded Room
NN	Neural network
PCA	Principal Component Analysis
PSD	Power spectrum density
RMS	Root mean squared
SFFS	Sequential forward floating search
SNR	Signal to Noise Ratio
SSEP	Somatosensory Evoked Potential
STD	Standard deviation
SSVEP	Steady-state visually evoked potentials
SVM	Support Vector Machine
WPD	Wavelet Packet Decompositions
WT	Wavelet transform

List of Symbols

Symbol	Description
α	alpha band power
β	beta band power
γ	gamma band power
θ	theta band power

Contents

List of Figures

List of Tables

Chapter 1

Introduction

BCI (brain-computer interface) is a system to manage and control a device for example, computer,wheelchair or a neuroprothesis by human intention which doesn't rely on the brain's normal output pathways of peripheral nerves and muscles. [8] It provides a direct communication between the brain and the computer without muscle control. People with severe physical disabilities, such as damaged brain-stem stroke, limbs,cerebral palsy, spinal cord injury, amyotrophic lateral sclerosis (ALS) or other neuro-muscular diseases, BCI is the exclusive way to communicate. [9] Presently, several functional imaging modalities like fMRI, EEG, MEG etc, are available for research. Among these non-invasive devices like Electroencephalography (EEG) is unique and most commonly used, since it provides high temporal resolution of the measured brain signals, as well as it is very affordable, relatively convenient,safe and easy to use for both healthy users and the disabled. Over the last few years, EEG based BCI research has rapid progression to enhance the computer processing speeds as well as effective signal analysis.

1.1 Application

According to the application area, BCI systems that uses EEG signals can be classified into into different categories as shown in the Fig. 1.1

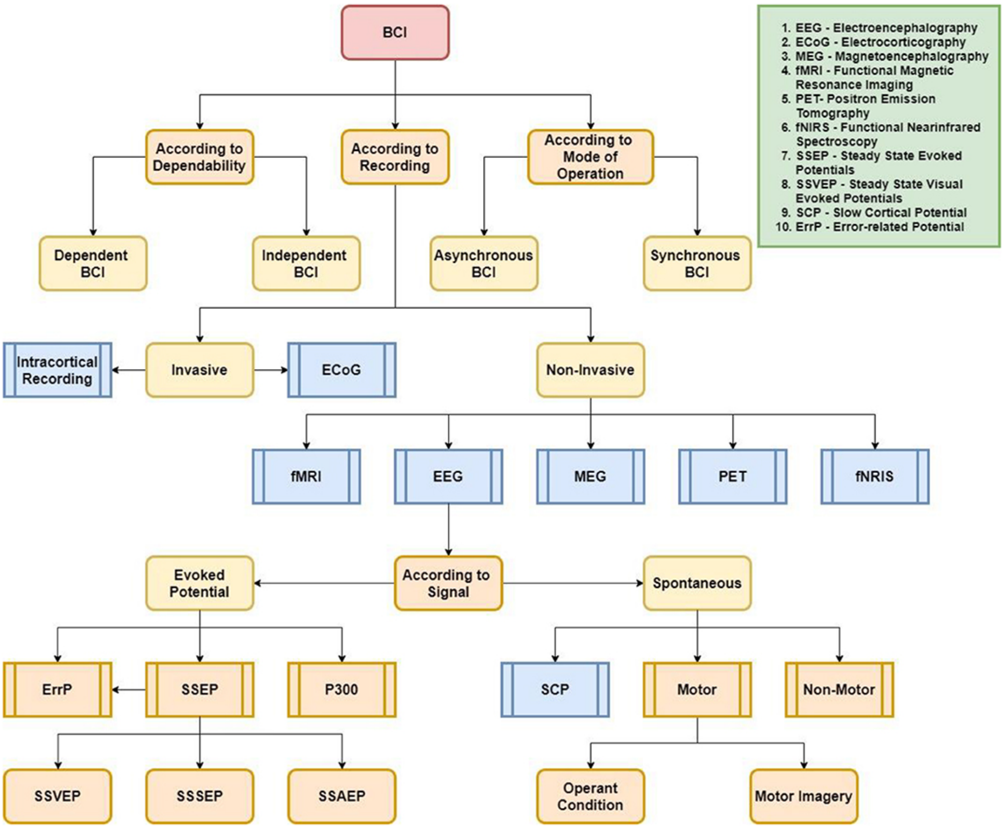


Figure 1.1: Classification of BCI systems in terms of dependability, recording method, and mode of operation. [1]

1.2 Challenges

Various issues have been observed on the EEG based BCI and accompanying changes in EEG signals.Some of the major issues are:

- Non-stationary
- Non-linearity
- High Dimensionality
- Noisy

1. Non-stationary:-One such issue is the non-stationary nature of EEG signals. EEG signal differs from one subject to another. Also, EEG signals diverge trial by trial and from day to day. This kind of non-stationary behaviour of EEG signals can be seen in both inter-session and intra-sessions. As a result, while evaluating the feature space of the training and testing session, a significant shift in the feature space may be found; which might make a trained model not optimal for the evaluation session.

2. Non-linearity:- Human brain is a highly complex nonlinear system where disorder behavior of neural ensembles can be detected. Hence EEG signals characterized by nonlinear dynamic methods rather than linear methods.

3. High Dimensionality:-Since, EEG BCI usually requires a large number of electrodes or lengthy time period of trials, that's why EEG BCI system has a high dimension of extracted features.

4. Noise and Artifacts:- EEG signals are highly noisy and hence have low SNR(Signal to Noise Ratio)

1.3 Problem Statement

In EEG BCI systems, the number of extracted features is usually very high. This is because EEG BCI usually requires a large number of electrodes or lengthy time period of trials.The focus of this project is to reduce the dimensionality of the data by removing redundant and irrelevant EEG channels. Our experiment put forwards an emotion recognition framework, SFCSER(Sequential Forward Channel Selection for Emotion Recognition) based on subject specific hybrid channel selection method. Hybrid Sequential Forward Search based on Fisher Score and Support Vector Machine is used to identify the optimal set of channels. Fisher Score is used as a Filter Method and Support Vector Machine(SVM) is used as a Wrapper Method to select the channels. Specifically, the EEG signals defined in time-frequency domain are more representative as compared to that of time domain or frequency domain. In our model, after channel selection, we extracted the wavelet based features from the selected channels to capture the time-frequency representation followed by Genetic Algorithm Based Feature Selection. SVM is used for classification of emotional states into two classes: Low/High Valence (unpleasant to pleasant) and Low/High Arousal (bored to stimulated).The proposed method is evaluated on DEAP Dataset [10], a dataset of emotion analysis using EEG, physiological and video signals. The evaluation is made based on subject dependant analysis.

Chapter 2

Literature Review

2.1 Biological Background

2.1.1 Human Brain

Human Brain is the most wonderful part of human body. Weighs of an average human brain about 1.5 kilograms and it consists of approximately 100 billion neurons,interconnected via axons and neurons. A neuron has a cell body, known as Soma, a long axon and many dendrites. [11] Neurons acquire stimulus from other interconnected neurons(about 103 to 105) through synapses. These stimulus move through axon as electrical impulses and it helps to control body movements, emotions and other aspects of body coordination.

Parts of Human Brain (according to position) The brain can be classified based on the position as:

- **Forebrain** - consisting cerebrum, thalamus, and hypothalamus
- **Midbrain** - consisting tectum and tegmentum
- **Hindbrain** - consisting cerebellum, pons and medulla

Parts of Human Brain (according to functions) Human brain is the core hub to think, control other voluntary and involuntary actions of a body and to manage all kind of sensory activities on a human body. Different regions of the brain are responsible for different specific tasks. Hence, brain can be classified into various parts based on their functions. It can be classified into :

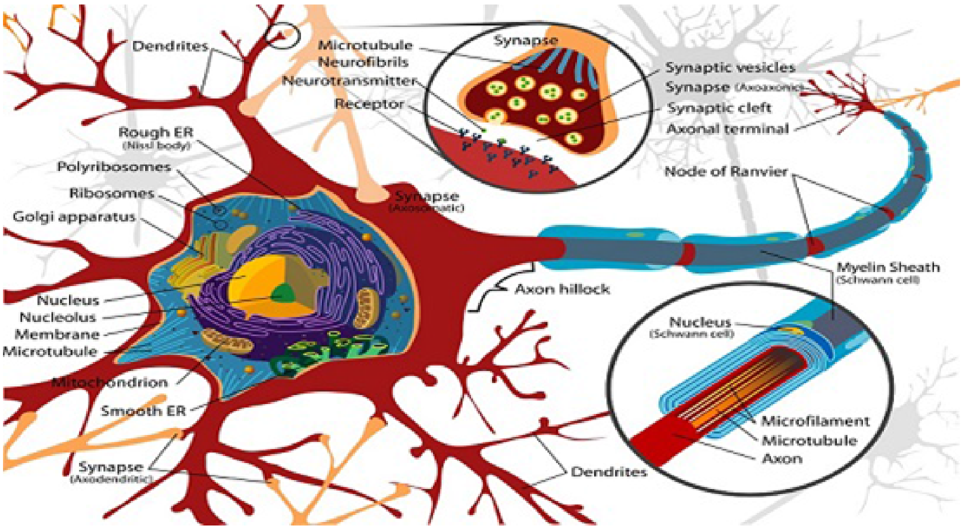


Figure 2.1: Neuron [2]

- Cerebrum
- Cerebellum
- Limbic system
- Brain stem

Cerebrum:- The cerebrum is also known as cortex, which is found only in mammals.It is the largest part of human brain. Cortex is site for complex brain functions as thought and action. Cerebral cortex has large number of folding, that increases the surface area and the number of neurons within it.

The cerebrum is divided into two halves, which are known as Left and Right hemispheres. It is connected by Corpus callosum which is a bundle of axons.The left hemisphere is associated to right part of the body and right hemisphere is associated to left parts of the body. The hemispheres are symmetrical, and each of them is divided into four sections. These are :

- **Frontal Lobe** - Associated with planning,reasoning,parts of speech, emotions,movement, and problem solving.
- **Parietal Lobe** - Associated with movement,recognition, orientation, perception of stimuli.
- **Occipital Lobe** - Associated with visual processing.

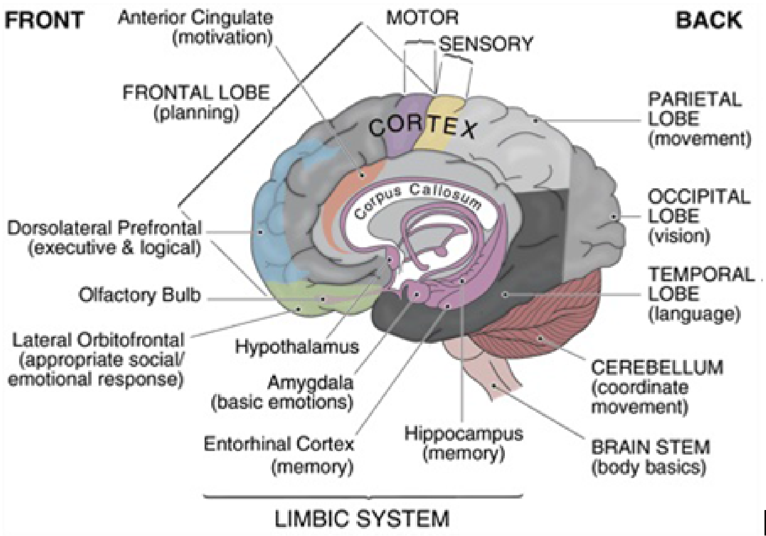


Figure 2.2: Human Brain [3]

- **Temporal Lobe** - Associated with memory, and speech,perception and recognition of auditory stimul.

Cerebellum:- The Cerebellum is also known as “Little Brain”.It is also divided into two hemispheres. It is responsible for coordination of movement and regulation, posture and balance. The cerebellum is highly folded to increase number of neurons in the area and the surface area.

Limbic System: The Limbic System is referred to as the “Emotional Brain”; which is found buried within the cerebrum. This system consists of the hypothalamus, thalamus, hippocampus and amygdala.

- **Thalamus** - It is a large mass of gray matter. It helps in motor and sensory functions. Also, it acts as router for action potentials.
- **Hypothalamus** - It controls pituitary and autonomic nervous system. For example : thirst, hunger, circadian rhythms,homeostasis,
- **Amygdala** - It is found in temporal lobe. It is associated with emotion, memory and fear
- **Hippocampus** - It is found in temporal lobe and it helps converting short term memory to permanent one.

Brain stem:- The brain stem is underneath limbic system, which is responsible for basic functions like blood pressure, breathing, heartbeat. The brain stem consists of the pons, midbrain, and medulla.

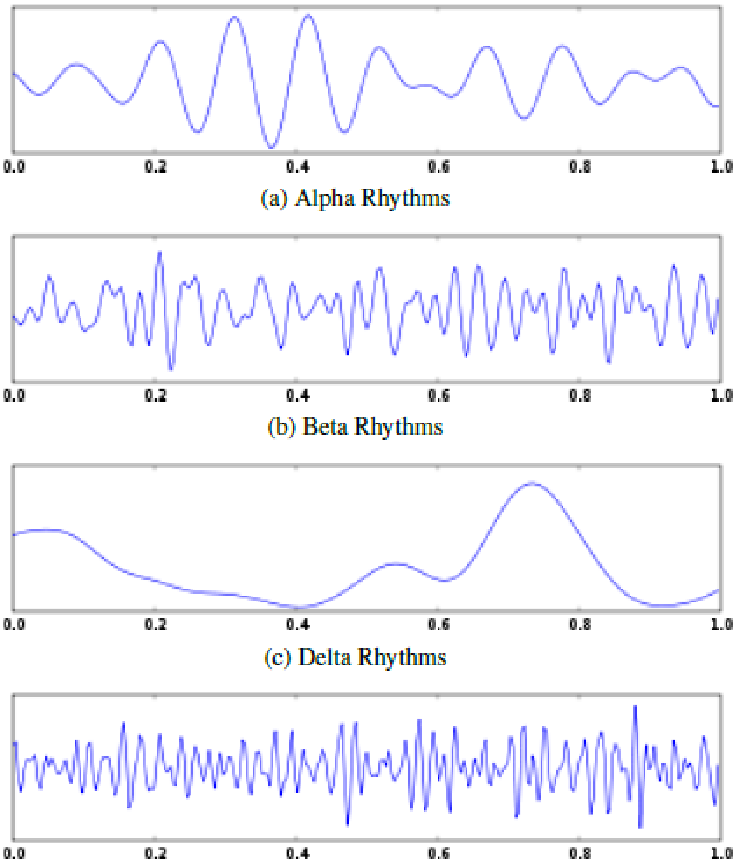
- Midbrain - It includes tegmentum and tectum and it helps with hearing,vision, eyemovement, and body-movement.
- Pons - It helps with sensory analysis,motor control.
- Medulla - It situated between spinal cord and pons.It helps with breathing and heart rate.

2.1.2 Brain Activity Patterns

The brain contains large number of neurons. When humans do some task, electric potential is developed as well as it travels across axons of neurons. Since, various parts of brain are associated to different functions, the electric signals generated due to neural activity and it widely different in terms of their amplitude, frequency,shape and the position. These are commonly termed as Brain waves or Brain rhythms. Figure 2.3 shows the different types of brain rhythms and Table 2.1 depicts the summary of different brain rhythms.

These Brain waves are classified as :

- **Delta waves** - It has frequency range from 0.5 to 3.5 Hz and they are associated with coma mental state, deep sleep.
- **Theta waves** - Derived from central, temporal and parietal parts of head. The frequency range is from 3.5 to 7.5 Hz. They are associated with stressed,thinking, and deep meditating state.
- **Alpha waves** - It is derived from occipital lobe and backside of the head. The frequency range is from 7.5 Hz to 12 Hz. They are associated with calm and relaxed states in awake humans.
- **Beta waves** - Derived from central area of the brain and front side of head.It has frequency range from 13 to 30 Hz. They are associated with high concentration level, deep thinking,and anxious state.
- **Gamma waves** - It has frequency range above 30 Hz. They are associated with simultaneous work, motor functions, and while multi-tasking.



(a) Alpha Rhythms

(b) Beta Rhythms

(c) Delta Rhythms

(d) Gamma Rhythms

Figure 2.3: Different Types of Brain Rhythms [4]

2.2 Brain-Computer Interfaces (BCIs)

2.2.1 Measuring Brain Activity

The goal of Brain Computer Interfaces (BCIs) system is to provide a direct communication between the brain and device.It enables brain to control the device or a computer by passing commands and messages. BCI system transforms the electrophysiological signals which is generated by various mental activities to device specific messages or commands. Depending on the location of the sensors, it used to record the aforementioned signals. BCI systems can be categorized into

Table 2.1: Summary of Brain Rhythms (Adapted from [4])

Brain Rhythms	Typical frequency range (Hz)	Normal amplitude (micro-volts)	Comments
Delta	0.5 - 4	<100	Dominant in infants During deep stages of adult sleep Found at central cerebrum and parietal lobes
Theta	4 - 7	<100	State of drowsiness Found at frontal, temporal and parietal regions
Alpha	8 - 13	20 - 60	Associated to alert state Found at occipital and parietal lobes
Mu	9- 11	<50	Hand Movements Found at motor and somatosensory cortex
Beta	14 - 30	<20	Also associated to hand movements
Gamma	>30	<2	Attentive state - when the subject is paying attention, response to stimulus

- **Non-invasive** - When the sensors are placed on the scalp, for example Electroencephalography, (EEG), Magnetoencephalography (MEG)
- **Semi-invasive** - When the electrodes are placed on the exposed surface of the brain, e.g. Electrocorticography (ECoG)
- **Invasive** - When micro-electrode arrays are placed directly into the cortex.

Noninvasive systems records the electrical activity and magnetic activity response due to neural activity in brain. Now, Noninvasive BCI system are reliable enough to be used as alternative means of communication outside dedicated research facilities. EEG devices are comparatively less expensive, and it is basically used for BCI research.There are some inherent challenges, like its poor spatial resolution and low signal to noise ratio tells some disadvantages of EEG. MEG gives higher spatio-temporal resolution, also it has sensitive sensors and magnetically shielding, that's why it has limited research potential compared to EEG.

Invasive techniques offers a very high temporal resolution, spatial resolution, and signal to noise ratio. The recording varies significantly from the noninvasive devices.Hence, it requires different signal processing approach. Due to the risky nature of this technique, research mainly focus on animals like monkeys and rats etc. Even though such systems have been demonstrated in humans as well.Some commercial applications of using BCIs are been created with non-invasive, as well as with invasive BCI systems.

Semi-invasive systems uses electrocorticography (ECoG) signal. As it is being located nearer to the site of neutral activity, so it provides better spatial resolution and higher signal to noise ratio. ECoG relies upon the same neurophysiologic mechanisms like EEG, and it requires similar signal processing approaches.Table 2.2 is displayed the summary of different brain imagery techniques.

Table 2.2: Summary of different brain imagery techniques (Adapted from [5])

Neuroimaging Method	Activity Measured	Measurement Type	Temporal Resolution	Spatial Resolution	Risk	Portability
EEG	Electrical	Direct	0.05 s	10 mm	Non-invasive	Portable
MEG	Magnetic	Direct	0.05 s	5 mm	Non-invasive	Non-Portable
ECoG	Electrical	Direct	0.003 s	1 mm	Invasive	Portable
fMRI	Metabolic	Indirect	1 s	1 mm	Non-invasive	Non-Portable
NIRS	Metabolic	Indirect	1 s	5 mm	Non-invasive	Non-Portable

2.2.1.1 Types of Brain Signals

Brain imagery techniques are to investigate the spatial and temporal organization of the brain.Due to the electro-chemical transmitters, electrophysiological activity of the brain occurs.Ionic current is generated by neurons while exchanging information between them. Electrophysiological activity could be measured by , electrocorticography (ECoG),electroencephalography (EEG), magnetoencephalography (MEG), and invasive electrical measurements.

The hemodynamic response of the brain helps to differentiate between activated and less activated neurons. These are called indirect methods, as they measure only the variation of local ratio of oxyhemoglobin to deoxyhemoglobin and do not directly characterize neuron activity as electrophysiological methods. The blood releases glucose to active neurons at a much higher rate than in the area of inactive neurons. Thus, the presence of glucose and oxygen results in surplus of oxyhemoglobin in the veins. This could be measured by the methods as functional magnetic resonance (fMRi) and near infrared spectroscopy (NIRS).

Electroencephalography (EEG)

EEG(Electroencephalography) measures the brain's electric potential caused by the neural activity during synaptic excitation of the dendrites. EEG measures the electric activity using the electrodes placed on the scalp. So it is a non-invasive technique in nature. EEG is the most widely spread recording method, as it is the most inexpensive non-invasive technique, which provides high temporal resolution(1ms) and is portable. But EEG are only able to measure signals of thousands of neurons, as the sensors are placed on the scalp and hence it has a poor spatial resolution.

Magnetoencephalography (MEG)

MEG (Magnetoencephalography) detects the magnetic fields resulting from the electrical currents in neurons. Magnetic fields are orthogonal to the electric signals measured by EEG. The magnetic field are less distorted, thus Magnetoencephalography provides a better spatial and temporal resolution.However, it requires typically expensive and much sensitive devices. Nonetheless, the measurements are required to be taken at magnetically shielded rooms. This procedure measures only the shallow parts of brain and is too bulky to be suitable for everyday use.

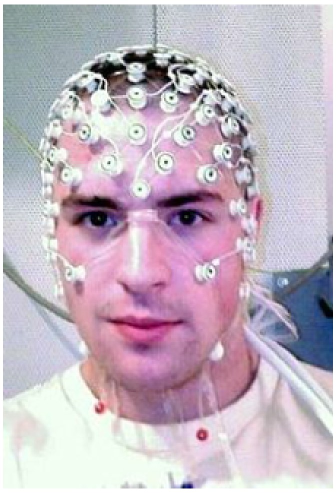
Electrocorticography (ECoG)

In ECoG(Electrocorticography) the electrodes are placed under the dura matter,and directly on the surface of the cortex without penetrating the cortex. Electrocorticography provides better spatial, temporal resolution and better signal quality. This signal has higher amplitudes and is less vulnerable to the artifacts such as eye movements and eye-blinks etc. Nevertheless, it is an semi-invasive technique, which requires risky surgery. ECoG are being primarily used in experiments with animals. Electrocorticography implants remain stable for several months and could be used to record signals. However long term stability remains unclear to date.

Functional magnetic resonance imaging (fMRI)

fMRI(Functional magnetic resonance imaging) is known as a non-invasive technique that depends on hemodynamics. Functional magnetic resonance imaging determine the blood oxygen level variations and hence it provides high spatial resolution. One of the disadvantage is, it suffers from poor time resolution,also susceptible to head motion artifacts. Functional magnetic resonance imaging like MEG are very expensive equipment and are not suitable for individual and everyday applications.

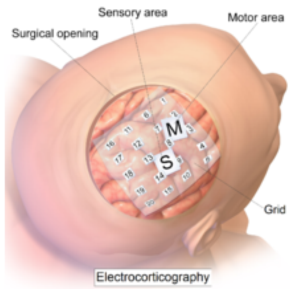
Different devices for brain imagery have shown in the figure 2.4



(a) EEG

(b) MEG

(c) fMRI



(d) ECoG

Figure 2.4: Different devices for brain imagery [5]

2.3 Electroencephalogram (EEG) based BCI

The term Electroencephalogram is originated from the concepts of :

- **Electro** - the electrical activities of the brain.
- **Encephalo** - the emission of signals from the head.

In modern techniques, the measures of electric patterns from scalp and digitizes them. The sensors measure the potentials as micro-volts, and it amplifies them before the digitalizing. Generally, the sensors are generally made up of gold or silver and it uses a conductive gel on scalp to enhance the signal to noise ratio so that we get a better result.

2.3.1 Brief History of EEG

Richard Caton (1842-1926), a physician had published his findings about electrical phenomena of the exposed cerebral hemispheres of rabbits and monkeys in the British Medical Journal in 1875.Another Polish physiologist Adolf Beck had presented an investigation of spontaneous electrical activity of the brain of rabbits and dogs in 1890.Beck experimented with the electrical brain activity of animals, placing electrodes directly on the surface of brain. His observation summarized in determination of brain waves. Napoleon Cybulski and Jelenska-Macieszyna recorded EEG for experimentally induced seizures in 1914. In 1929, German physiologist and psychiatrist, Hans Berger published 'das Elektrenkephalogramm', which marks the beginning of research on the human electroencephalogram [5].

2.3.2 The 10-20 System of Electrode Placement

For electrode placement on the scalp,the International 10-20 system, which has been standardized by the American Electroencephalographic Society is used . The system uses two reference points in head, they are:

- **Inion** - bony lump at base of skull.
- **Nasion** - at top of nose, and at level of eyes.

The electrodes are placed on the traverse and median plain, containing these points, at intervals of 10% and 20% as shown in the figure 2.5. The letters at each electrode position denotes the different brain regions such that :

- **"A"** represents ear lobes
- **"C"** represents central region
- **"P"** represents parietal region
- **"F"** represents the frontal lobe
- **"Fp"** represents the frontal polar region and
- **"O"** represents the occipital lobe

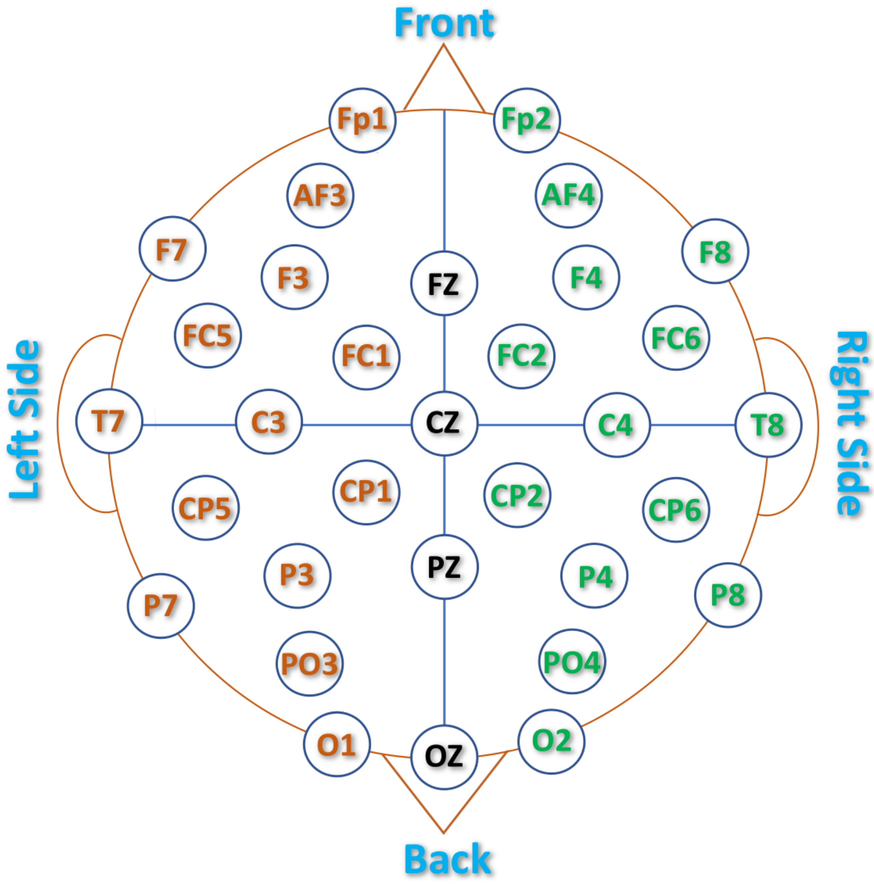


Figure 2.5: The 10-20 System of Electrode Placement(the "10" and "20" refer to the 10% or 20% inter-electrode distance) [2]

2.3.3 Feature Extraction

After completion of the signal prepossessing, next step is to extract the information to classify individual.Extracting the optimal feature vector is an important step of the study.Some of the commonly employed feature extraction methods for EEG signal are:

- **Time domain methods** - It exploits the temporal information of the signal
- **Frequency domain methods** - It exploits the frequential information embedded
- **Hybrid methods** - These are based on time-frequency domain and it exploits both temporal and frequential information.
- **Other approaches** - In addition to the aforesaid feature extraction methods, literature has

also reported other different methods as for instance phase locking, fractal dimensions, bispectrum, sparse representation are some of the examples.

1. Time domain methods To define the neurophysiological data acquired, a temporal method uses the variations of signals with time.

1.1 EEG signal amplitude The amplitude of the raw EEG signals which is captured, are prepossessed and aggregated into a single feature vector. The EEG data, captured from various channels/sensors placed on the scalp. Generally, the signals are down-sampled.Also the dimensionality gets reduced like using spatial filters during the prepossessing step, before these are aggregated as Feature Vectors.

1.2 AutoRegressive (AR) parameters In the AutoRegressive (AR) method,it models the signal at any given time, as a weighted sum of signals at previous time and some noise, i.e. mostly Gaussian white noise. Mathematically, it could be formulated as:

$$X(t) = a_1X(t1) + a_2X(t2) + ... + a_pX(tp) + E_t \tag{2.2}$$

where, X(t) is the measured signal at time t, E_t is the noise term and a_1 to a_p are the auto-regressive parameters.

2. Frequency domain methods The human brain waves or rhythms are responsible for different mental tasks. Each waves have a specific frequency, although the amplitude of the waves may change for different tasks but the rhythms are synchronized with the stimulus frequency and hence the frequential information is the key source to extract features. The main features that could be extracted are band power and power spectral density.

2.1 Power Spectral Density (PSD) Features PSD, Power Spectral Density shows the strength or distribution of power of a signal as a function of frequency domain. Generally, it is calculated by squaring the Fourier transform of the signal or by calculating auto-correlation function and then transform it. Power Spectral Density of a signal forms a very useful and meaningful feature for the study of BCI.

2.2 Band Power Features Band Power Features, provides the power distribution of a given band. Here, the preprocessed signals are band pass to a particular frequency range(depending on band like alpha, beta, gamma, theta, delta etc), squared and finally averaged over a time window. The normal distribution of the signal is being obtained by taking log transform. It has been a most useful feature for distinguishing of motor imagery classes.

3. Time-frequency representations Time-frequency method extracts the information that are in both present in the time and frequency domains. This method helps in recognizing sudden changes in the temporal domain while still analyzing frequency domain. Short Time Fourier Transform also known as STFT or wavelets are the useful for such feature extraction.

3.1 Short Time Fourier Transform (STFT) By taking Fourier transform of a signal window on a short time period gives STFT(Short Time Fourier Transform). Here,the size of the window size is fixed and it causes identical frequenctial and temporal resolutions.

Let, the signal considered to be x(n) and the window be w, the STFT of X(n,w) is given by :

$$X(n,w) = \sum_{n=-\infty}^{+\infty} x(n)w(n)e^{-jwn} \tag{2.3}$$

3.2 Wavelet transform(WT) In Wavelet Transform, the EEG signals are non-stationary i.e. the statistical property of the signals changes over the time. Hence, STFT features are useful. However the window size is fixed, and it is difficult to determine the window size. Smaller window size gives good temporal properties, while larger windows gives better frequency information. Furthermore,the Fourier Transform does not represent the finite and on-periodic signal correctly.

Wavelet transform(WT) gives best trade off between temporal resolution and frequency. The WT uses finite basis function that is also known as wavelets rather than sines and cosines functions.

4. Statistical Features The statistical features are based on the basic shape of the signals and amplitude.The statistical feature vector is being defined by the characteristic property of amplitude distribution. Example:-Min Value, Standard Deviation,Mean,Median, Max Value.

5. Entropy One of the most important feature property is Entropy.It is the measure of the complexity of the EEG signal. The spectral entropy(HS) is calculated as follows:

$$H_s(X) = \frac{-1}{logN_f} P_f(X) log_e P_f(X) \tag{2.4}$$

where, P_f is the estimated PSD for EEG segment X and N_f is the number of frequency components in the PSD estimate.

6. Hilbert Transform Hilbert transform is a linear operation, that generates H(u)(t), with same domain as the input signal u(t). The Hilbert transform derives the analytic representation of a signal u(t). An "analytic" i.e complex time signal Y(t) can be constructed from a real valued input signal y(t) as:

$$Y(t) = y(t) + jh(t) \tag{2.5}$$

where, Y(t) is the analytic signal constructed from y(t) and its Hilbert transform y(t) is the input signal h(t) is the Hilbert Transform of the input signal

2.3.4 Feature Selection

Since EEG features are usually huge in size, dimension reduction holds promise. One common way of reducing the dimension is by selecting the optimal features from the feature space. Alternatively, channel selection also plays vital role in dimension reduction. Selecting the EEG channels not only selects the optimal feature subset but also reflects the part of the brain dominated for a particular task. Different ways of feature/channel selection are described below:

1. Filter methods: Filter method majorly highlight the dependency of the features on the target class. These methods evaluate the effectiveness of the features based on a predefined criterion, independent of the classifiers.Since they do not incorporate the learning that's why Filter methods are fast .Scalability is another advantage of filter method. However, one drawback with filter method is that it may suffer from low accuracy.The mostly used criteria for filter methods are information gain, mutual information, correlation and regression coefficients.Steps include in the Filter Methods are like initialization,evaluation,subset generation,subset evaluation etc.

2. Wrapper methods: With respect to a particular classifier,in the wrapper methods,the optimal characterization condition signifies the maximal classification accuracy or minimal classification error. These methods use classification accuracy of the classifiers as the evaluation metric to identify a subset of features. Wrapper methods ensure higher classification accuracy as compared to the filter approach at the expense of high computational cost .Wei and Wang [12] proposed a methodology for channel selection during the classification of motor imagery of left hand, right hand, and foot based on a binary multi-objective particle swarm optimization algorithm. Guyon et al. [13] proposed an approach of feature selection that utilizes the internal structure and parameters of SVM. The idea is to use the weighting of the support vectors as the evaluation metric of features. Another work by Maldonado put forward a feature selection method using SVM where the method eliminates those features whose removal has less impact on the final solution.

3. Embedded methods: Embedded methods is little bit different from filter and wrapper because of the interaction of feature selection and learning process. In these methods, feature selection and learning cannot be separated.Schroder et al. [14] proposed a robust EEG channel selection algorithm across subjects in BCI systems adopting an embedded approach with a sequential search strategy for subset channel selection. Embedded methods select features in the process of learning as for instance Lasso and Elastic Net.

4. Hybrid approach: Even though the wrapper methods ensure higher classification accuracy as compared to the filter methods, Because of its high computational costs, the use of wrappers is not encouraged in case of a very high dimensional feature space . Thus, an amalgamation of filter and wrapper methods hold promise. Gan et al. [15]proposed a novel hybrid approach that amalgamates both filter and wrapper methods to identify a set of features with much less computational cost and higher classification accuracy.

5. Search methods: The problem of identifying a subset of features can also be solved using a traditional search problem. In literature, different search strategies were employed to select a subset of best features for classification. Population based heuristic search methods such as particle swarm optimization genetic algorithms, ant colony optimization have been used as feature selection algorithms.

2.3.5 Classification

The final goal of BCI systems is to recognize the cognitive intent of the user so as to convert the identified intent into a certain action. Thus, after pre-processing, extraction of features and selection of features, using classifiers to classifies the signal at a particular time point into one of its classes (cognitive tasks). Lotte et al. [16] put forward a review on the most commonly used classifiers in EEG BCI. While selecting classifiers, Lotte et al. [16] presented two main considerations. These are a. Curse of dimensionality and b. Bias-variance tradeoff :

a. The curse of dimensionality:The data must determine a class that grows exponentially with the dimensions of the feature set. Classifiers are commonly used to sample a particular class of training data at least 5-10 times as its dimension. However, this is not possible with BCI design. Extracted features are usually large in size and training data is small in size. Therefore, the curse of dimensionality is an important dilemma in the design of EEG-BCI.

b. Bias-Variance Tradeoff: There are three causes of poor classifier performance: noise, bias, and variance. Noise is system-specific and errors caused by noise cannot be completely avoided. Bias represents the difference in performance between the best classifier and the estimated classifier, and variance represents the sensitivity of the training set. Formally, classification performance depends on using mapping to find the true label of the feature vector. This assignment is learned from the training set. The classification performance obtained from the best mapping in the training set is considered the best classification.

To get the best classifier performance, you need to have low bias and variance. However, stable classifiers usually have low variance and high bias, whereas unstable classifiers have high variance and low bias. An example of a stable classifier for the is a less complex linear discriminant analysis (LDA). Small fluctuations in the training set do not affect performance and are therefore considered a stable classifier. On the other hand, in the case of unstable classifiers such as the Multilayer Perceptron (MLP), the high complexity and low variability of the training set can result in significant performance changes. Various techniques, such as a combination of classifiers and regularization, can be used to reduce the variance. However, due to the temporary nature of the EEG signal, training data from different sessions will vary significantly. Therefore, low dispersion can play a decisive role in overcoming the variability problem of the EEG-BCI system.

A simple categorization of the commonly used classification approaches is portrayed as follows:

(a) Linear classifiers: These classifier categories use linear functions to distinguish between different classes. LDA is based on hyperplane calculations to distinguish between different classes. Linear SVMs are also based on hyperplane calculations that maximize the boundaries of different classes. Build a support vector around the boundary. There are several types of kernel-based SVMs that have been fooled to implement non-linear boundaries. Naive Bayes classifiers belong to the category of stochastic classifiers based on Bayes' theorem. This guarantees independence of functionality.

(b) Artificial neural networks: Motivated by the animal's nervous system, the Artificial Neural Network (ANN) was developed. Classifiers in this category are non-linear in nature. There are many variations of ANN classifiers, including multi-layer perceptrons, neural networks with fuzzy art maps, neural networks with radial basis, deep neural networks, and neural networks with learning vector quantization.

(c) Generative models: The ultimate goal of the above classification method is to implement linear or non-linear boundaries that distinguish between different classes. Another approach to machine learning is to compute a data model that can be used for both predicting and classifying data. That is, these models can map the data distribution and classify it as a generative model. Common examples of generative models are the Gaussian mixed model and the hidden Markov model.

(d) Ensemble Learning: In machine learning, ensemble techniques are based on the fusion of a finite set of learning algorithms to achieve better classification accuracy and higher predictive performance than is possible with a single learning algorithm. Examples of ensemble learning include random forests, model buckets, and Bayesian model combinations.

2.4 EEG BCI for Emotion Recognition

2.4.1 Interpretation of Emotions

Varied definitions of emotions have been presented in the literature. Table 2.3 presents different views on basic emotional choices of researchers. The researchers agree that complex emotions are formed by superimposing basic emotions [17].

Table 2.3: Basic Emotions Selected by Different Researcher (Adapted from [7]).

Researcher	**Basic Emotion**
James [18]	Anger, Fear, Sad, Love
Ekman [19]	Anger, Fear, Sadness, Enjoyment, Disgust, Surprise
Clynes [20]	Anger, Hatred, Sadness, Happiness, Love, Romantic love, Reverence
Izard [21]	Anger, Fear, Guilt, Pain, Joy, Shame, Surprise, Interest, Disgust
Frijda [22]	Anger, Fear, Pain, Pride, Happiness, Accidental Disgust, Contempt

Human being has a diversity of researches on emotion. Emotions get change with the changes of different external stimuli. Numerous researchers presented different definitions of emotions. There is no unite conclusion for the disputation that emotions are independent or continuous. Generally, the emotional state is divided into two models; first one is known as a discrete emotional model, which contains several basic emotions. But different researchers have different opinion on the selection of basic emotions.However, researchers agree that complex emotions are made by the superposition of basic emotions. The second type is a dimensional-based emotional model. It evolves from a valence-arousal two dimensional model to a valence-arousal-like three dimensional model. Then there is higher valence-arousal-like-dominate four dimensional model. Emotion states used in this study based on the valance-arousal score level has been shown in the figure 2.6

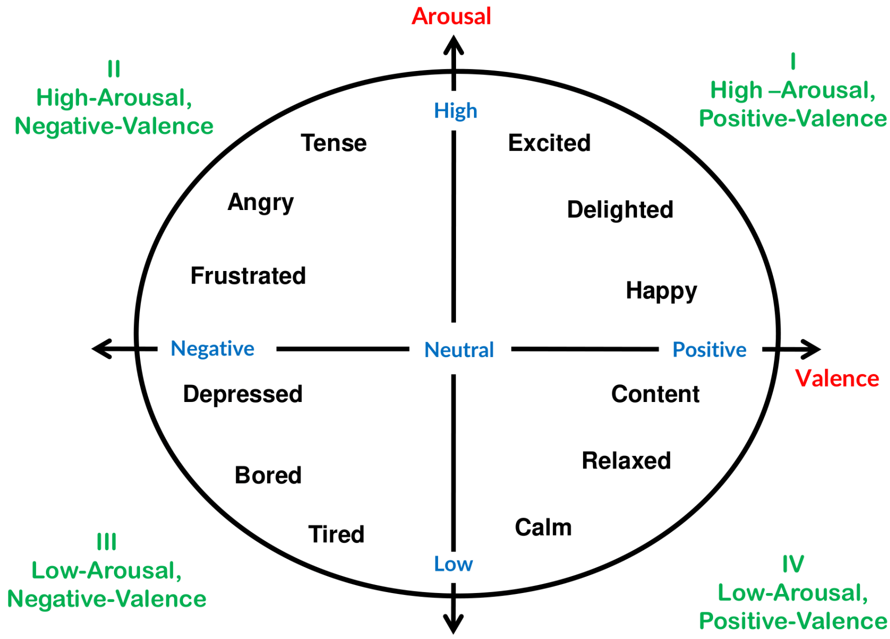


Figure 2.6: Emotion states used in this study based on the valance-arousal score level. [6]

2.4.2 Emotion Recognition

Emotion recognition using EEG signal enables the direct evaluation of the inner state of a user, which is considered an important feature in human machine interaction. Many methods for feature extraction have been already studied and explored the selection of both appropriate features.Electrode locations are usually based on neuro scientific findings. Nevertheless, suitability for emotion recognition have been tested using a small amount of distinct feature sets.Generally, on small data sets. A major drawaback is that no systematic comparison of features exists. To make Human Machine Interaction more natural and real, knowledge about the emotional state of a user is considered an important factor. Emotions are very important for correct interpretation of actions along with communication. Interest in emotion recognition from different modalities (e.g. face, posture, motion, voice) has risen in the past several decades.Recently it gained attention in the field of brain computer interfaces (BCIs), which has coined the term affective BCI (aBCI).

In EEG emotion recognition, EEG channel selection is the main challenging and most important factor affecting the performance of emotion recognition with the deepening of research. The main purpose of EEG channel selection is to select a part of electrodes i.e channels from all electrodes to reduce the computational cost and to improve the accuracy rate of emotion recognition. In the past few years, some EEG channel selection algorithms have been presented by many researchers. For instance, Rizon et al [23] presented an asymmetric ratio (AR) based channel selection method for human emotion recognition from EEG signals. The result shows in their method could reduce channels and classify the emotions effectively.Lin et. al [24] adopted the F-score index based on the ratio of between-class and within-class variances to and a set of optimal EEG channels for EEG-based emotion recognition. He et. al [25] presented a Rayleigh coefficient (RC) maximization-based genetic algorithm (GA) for channel selection in motor-imagery BCI system.They achieved the optimal subset of all channels. J Zhang et al [26] proposed the ReliefF-based method to obtain the EEG channel that has the closest relationship with the emotion, and sharply reduce the number of EEG channels without sacrificing the recognition rate. Zheng et al [27] presented a method based on deep neural networks to learn the average absolute weight distribution to select the optimal EEG channels, and obtain preferable experimental results. In recent years, a new method are proposed by Gupta et al [28].He proposed a flexible analytic wavelet transform (FAWT) based on six known channels for emotion recognition. The result has shown a significantly better performance for emotion classification as compared to the existing method. Bajaj et al [29]presented multiwavelets decomposition based features for EEG emotion classification. Their results is also performed well. Bajaj et al [29] proposed a novel method for emotion recognition using multiwavelet transform with multiclass least squares support vector machine (MC-LS-SVM). It provided classification accuracy of 84.79% for emotions. The result shows the effectiveness of the proposed method for EEG emotion recognition.Shortly, we can conclude that, in the previously presented work, researchers proposed various channel selection methods to select the optimal channel subsets. Different types of classifiers have been applied to recognize emotion, and improve the accuracy of the classification.

Chapter 3

Design of Experiment

3.1 Dataset Used

The dataset we have used for the experiment are from the DEAP database. It is espsecially used for emotion analysis using physiological signals. The sampling frequency of the original dataset is 512Hz and the sampling frequency of the preprocessed dataset is 128Hz that collected from 32 healthy participants where 16 males and 16 females. Each participants watched a one minute long music video to record their signals. After each trial/video, each participants perform self-assessment of their level of valence, arousal,like/dislike, and dominance. Thus, there are 40 trials for each subject and each trial lasts for 63s, with 3s pretrial has been included. DEAP dataset has 40 channels where 32 are EEG channels and 8 are other peripheral channels. In this experiment we will be using just only 32 EEG channels as our focus only two dimensions of emotions like valence and arousal and ignored liking and dominance. The ratings of valence and arousal ranges from 1 to 9. The valence and arousal ratings have been employed and the ratings of dominance and likelihood are ignored. The ratings of valence and arousal ranges from 1 to 9. This paper divides the ratings (1-9) with a threshold of 5 into two levels of valence and arousal states based on the analysis made by Koelstra et. al [30]. The rating in the range of 1-5 was categorized as Low valence/arousal state and rating in the range of 5-9 was categorized as High valence/arousal states. [31] Every participant's file consists of two arrays as described in the Table 3.1.EEG signals are sliced into 60s pieces with a sliding window. This dataset is available in three different file formats. These are original biosemi(.bdf) format,prepossessed matlab format(.mat) and prepossessed python format(.dat). In our experiment we are using python format.

Table 3.1: Original DEAP Dataset

Array Name	Array Shape	Array Contents
Data	40x40x8064	video/trialxchannelxdata
Labels	40x4	video/trialxlabel

Table 3.2: Derived Dataset(only for EEG Data)

Array Name	Array Shape	Array Contents
Data	40x32x7680	video/trialxchannelxdata
Labels	40x2	video/trialxlabel

3.2 Tools and Tech Used(Software)

3.2.1 Editor Used

PyCharm is a dedicated Python Integrated Development Environment (IDE) that provides a wide range of essential tools for Python developers.It tightly integrated to create a convenient environment for productive Python, web, and data science development.

Google Colab is a cloud based platform which allows to write and execute python code in the browser. It provides both CPU and GPU facility to build and develop powerful machine learning models.All the necessary libraries are already pre installed in this notebook which makes it more unique than other editor.

3.2.2 Library Used

MNE Python is an open source Python package for visualizing, exploring, and analyzing human neurophysiological data like MEG, EEG, sEEG, ECoG, The facility it provides are like Source Estimation,Machine Learning,Encoding Models,Statistics,Data Visualization etc.

Scikit Learn is the most popular and useful library for machine learning using Python. It offers a selection of efficient tools for machine learning and statistical modeling including clustering,regression,classification, and dimensionality reduction etc. This library, that is largely written using Python, is built upon SciPy NumPy and Matplotlib etc.It is a free library.

SciPy is a free and open-source Python library used for technical computing and scientific computing. It contains modules for linear algebra, optimization,interpolation,integration, special functions, signal and image processing, FFT, ODE solvers and other tasks common in science and engineering.SciPy is written in Python, also a few segments are written in C.

NumPy is a library for the Python programming language. It adds support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.It provides a high-performance multidimensional array object, and tools to work more faster than traditional python lists.It was originally created by Jim Hugunin with the help from several other developers. NumPy is free and open-source software library. NumPy is a NumFOCUS sponsored project.

Chapter 4

Methodology

The overall framework of the proposed methodology is depicted in Fig 4.1

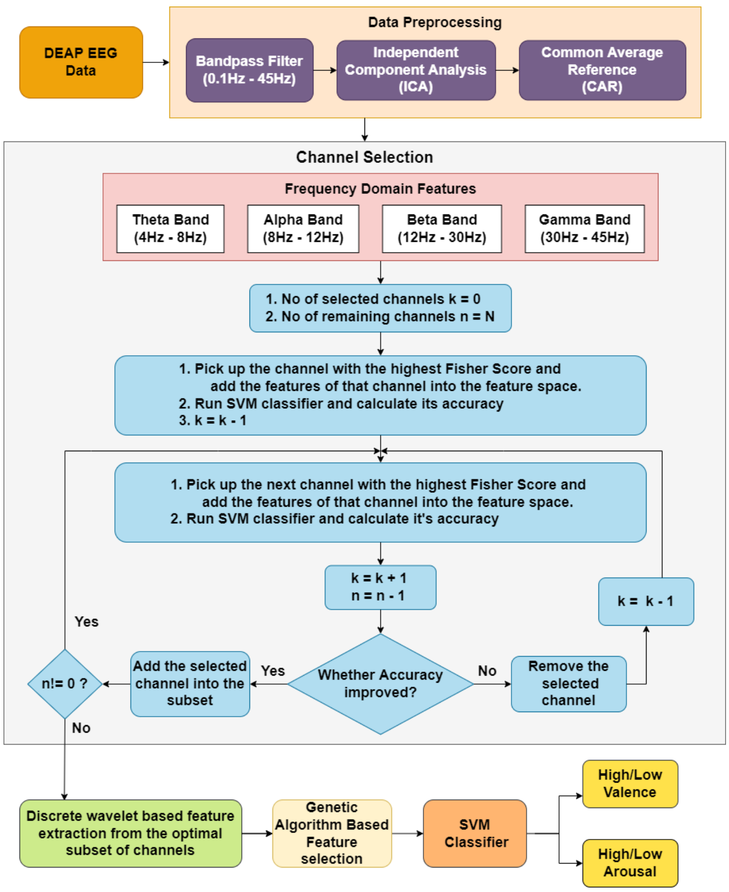


Figure 4.1: Steps in Our Emotion Recognition Framework

4.1 Preprocessing

In this experiment, we have taken only the EEG data which was processed by eliminating the three second pre-trial baseline. Different artefacts like eyeblinks has been removed using Independent Componenet Analysis(ICA).ICA can be considered as an advance technique and is used when the signals acquired are linear combination of signals from multiple sources, noises have comparable amplitudes.Next, the EEG data were filtered using a (0.1-45)Hz band-pass filter followed by Common Average Reference.In CAR, the readings of all the electrodes are averaged and the average value is obtained by subtracted from the data of each electrode. It helps to remove the common noise of all the electrodes.

4.2 Feature Extraction(Frequency Domain)

As an initial choice of feature, the power spectral density(PSD) is used. PSD is a commonly used frequency domain feature in the studies of emotion recognition from EEG. Usually, it is computed for a number of frequency bands and used as an indicator of the extent of brain activity within each of these(alpha,beta,gamma,theta,delta) bands.Welch's method is used to extract the PSD from the dataset. The EEG data that we are using, have already been down sampled to 128 Hz and low pass filtered to remove frequencies above the desired range.Welch's average periodogram method [32] has been implemented for estimating Power Spectral Density (PSD). In Welch's average periodogram method we divide the time domain signal data into successive blocks/sections to form the periodogram for each block/sections and then we take the averages of the spectra for each block/section, which leads to reduce the effects of temporarily unstable signals or noises from the signal.

Brain wave comprises of five principal frequency bands: theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz) and gamma (30-45 Hz). Since, delta waves are mainly associated with deep sleep, we are excluding it for our further analysis.

The mean power of each of the remaining frequency bands associated with each channel is taken as a statistical features which is computed using Equation 4.1.

$$Mean(\mu_x) = (\frac{1}{N}\sum_{i=1}^{N} X_i) \tag{4.1}$$

where, X_i is the spectral data obtained after the computation of PSD for a particular band and a particular channel, N is the total no of spectral data points of that particular frequency band and particular channel, and μ_x represents the mean power of the frequency band.

For each EEG electrode, we have extracted 4 frequency band's mean power. To choose the dominant electrode we have extracted 4 features (mean bands power) from 32 electrodes or channels. Hence in total per video/sample we have 32*4 = 128 features. Thus for each subject we have 40(trials)*128(features).

4.3 Channel Selection

4.3.1 Laplacian Score based fisher score

Laplacian Score [33] is a filter based unsupervised feature selection algorithm which selects a set of optimal features that can preserve locality and manifold structure of the data i.e. if two samples or instances that are close enough to each other generally they belong to the same class and to do so LS uses the nearest neighbor graph.

Given N number of data points with class labels are represented as $\{x_i, y_i\}_{i=1}^{N}$, where $y_i \in \{1, 2, \ldots, c\}$, c is the total no of class labels. Let L_r denote the Laplacian Score of the rth feature and f_{ri} denote the ith sample of the rth feature, $i = 1, 2, \ldots, m$. So, rth feature vector is defined as $\mathrm{f}_r = [\mathrm{f}_{r1}, \mathrm{f}_{r2}, \ldots, \mathrm{f}_{rm}]^T$.

First a nearest neighbor graph G is constructed in which each node denotes an instance or a sample i.e. ith node corresponds to x_i instance. For each pair of instances, say x_i and x_j, if x_i and x_j are "*close*" i.e. if instance x_i is one of the k-nearest neighbor of the instance x_j or instance x_j is one of the k-nearest neighbors of x_i, then an edge is put between ith node and jth node in the graph G. Similarity a weight matrix or affinity matrix, W is obtained using a nearest neighbor graph G.

Affinity matrix is constructed in the following way, if ith node and jth node are connected by an edge in the graph G for some constant t then, $W(i,j) = e^{-\frac{\|x_i - x_j\|^2}{t}}$ otherwise $W(i,j) = 0$.

Then, the diagonal matrix D is defined as $D = diagonal(WI)$ i.e. $D(i,i) = \sum_{j=1}^{n} W(i,j)$ where $I = [1, 1, \ldots 1]^T$ and the Laplacian matrix L is $L = D - W$. Laplacian Score of rth feature f_r is computed as:

$$\mathbf{L}_r = LaplacianScore(\mathbf{f}_r) = \frac{\widetilde{\mathbf{f}_r}^T \mathbf{L} \widetilde{\mathbf{f}_r}}{\widetilde{\mathbf{f}_r}^T \mathbf{D} \widetilde{\mathbf{f}_r}} \tag{4.2}$$

where

$$\widetilde{\mathbf{f}_r}^T = \mathbf{f}_r - \frac{\mathbf{f}_r{}^T\mathbf{DI}}{\mathbf{I}^T\mathbf{DI}}\mathbf{I}$$

Laplacian Score based fisher score: Fisher score for rth feature vector (f_r) is computed using Equation e2:

$$\mathbf{F}_r = FisherScore(\mathbf{f}_r) = \frac{\sum_{i=1}^{c} n_i(\mu_{ri} - \mu_r)^2}{\sum_{i=1}^{c} n_i \sigma_{ri}^2} \tag{4.3}$$

where, n_i denotes the number of data points in ith class, μ_r denotes mean value of rth feature vector(f_r), μ_{ri} denotes mean value of rth feature vector(f_r) for samples belongs to class i and σ_{ri}^2 denotes variance value of rth feature vector(f_r) for samples belongs to class i.

He et al. [33] stated that Fisher score is equivalent to Laplacian score with a special graph structure in which the affinity matrix is defined as equation eqw.

$$W(i,j) = \begin{cases} \frac{1}{n_l} & \text{if } y_i = y_j = l \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

Using the above affinity matrix they proved that if F_r denotes the Fisher score of the rth feature and With the weight matrix W defined as eq. eqw, then we have $L_r = \frac{1}{1+F_r}$

$$Thus, F_r = \frac{1}{L_r} - 1 \tag{4.5}$$

4.3.2 Subject Specific Hybrid Sequential Forward Channel Selection Methodology

Sequential Forward Selection (SFS) method was originally developed by Whitney [34]. In SFS, a null variable set is considered at the beginning of the procedure. Let say X_k is variable subset space and k is cardinality of the variable subset. So initially when k = 0 then $X_k = \phi$. Next, the algorithm is repeatedly updated by adding the best variables in X_k. This process is repeated till convergence criterion is satisfied. This paper employs SFS for selecting the optimum set of EEG channels. Both filter and wrapper method have been used to select the channels. Fisher score (using eq.4.5) is the filter index wheareas SVM is used as a wrapper index. Fisher score is designed to select variables whose values are more uniformly distributed for samples with the same class but more dissimilar for samples in different classes. Higher value of Fisher Score implies better separability of the features.

We have adapted the similar concept for the channel selection procedure. For a particular subject we have 40 samples and each sample has been recorded through 32 channels. Lets say, s_i denotes ith subject, v_{ij} denotes jth sample/video corresponding to ith subject and ch_{ij}^k denotes as kth channel w.r.t. jth sample/video for ith subject, where $i \in [1, 32]$, $j \in [1, 40]$ and $k \in [1, 32]$.

Now construct a channel space for ith subject and jth sample/video so that $ch_{ij} = [ch_{ij}^1, ch_{ij}^2, ch_{ij}^3, ch_{ij}^4, \ldots, ch_{ij}^{31}, ch_{ij}^{32}]$. $\forall channel \in ch_{ij}$, we have extracted 4 frequency band's (theta, alpha, beta, and gamma) mean power and we are saying it as 4 features for each channel.

Let say, from channel, ch_{ij}^k we have extracted t_{ij}^k, a_{ij}^k, b_{ij}^k and g_{ij}^k features which represents theta, alpha, beta, and gamma band's mean power respectively.

Now we will construct four 2-D matrix denoted by θ_i, α_i, β_i and γ_i in which i represents the subject no. Matrix θ_i, α_i, β_i and γ_i consists of theta, alpha, beta and gamma band's mean power from all the 32 electrodes for 40 sample/video w.r.t a particular subject. In each matrix row represents the video samples and column represents the EEG channels. So shape of θ_i matrix is (40*32) i.e. (no of video samples)*(theta band's mean power from 32 channels). More precisely,

$$\theta_i = \begin{bmatrix} t_{i1}^1 & t_{i1}^2 & t_{i1}^3 & \cdots & t_{i1}^{32} \\ t_{i2}^1 & t_{i2}^2 & t_{i2}^3 & \cdots & t_{i2}^{32} \\ t_{i3}^1 & t_{i3}^2 & t_{i3}^3 & \cdots & t_{i3}^{32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{i32}^1 & t_{i32}^2 & t_{i32}^3 & \cdots & t_{i32}^{32} \end{bmatrix}$$

Similarly,

$$\alpha_i = \begin{bmatrix} a_{i1}^1 & a_{i1}^2 & a_{i1}^3 & \cdots & a_{i1}^{32} \\ a_{i2}^1 & a_{i2}^2 & a_{i2}^3 & \cdots & a_{i2}^{32} \\ a_{i3}^1 & a_{i3}^2 & a_{i3}^3 & \cdots & a_{i3}^{32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{i32}^1 & a_{i32}^2 & a_{i32}^3 & \cdots & a_{i32}^{32} \end{bmatrix}$$

$$\beta_i = \begin{bmatrix} b_{i1}^1 & b_{i1}^2 & b_{i1}^3 & \cdots & b_{i1}^{32} \\ b_{i2}^1 & b_{i2}^2 & b_{i2}^3 & \cdots & b_{i2}^{32} \\ b_{i3}^1 & b_{i3}^2 & b_{i3}^3 & \cdots & b_{i3}^{32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{i32}^1 & b_{i32}^2 & b_{i32}^3 & \cdots & b_{i32}^{32} \end{bmatrix}$$

$$\gamma_i = \begin{bmatrix} g_{i1}^1 & g_{i1}^2 & g_{i1}^3 & \cdots & g_{i1}^{32} \\ g_{i2}^1 & g_{i2}^2 & g_{i2}^3 & \cdots & g_{i2}^{32} \\ g_{i3}^1 & g_{i3}^2 & g_{i3}^3 & \cdots & g_{i3}^{32} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_{i32}^1 & g_{i32}^2 & g_{i32}^3 & \cdots & g_{i32}^{32} \end{bmatrix}$$

Now Fisher score for θ_i, α_i, β_i and γ_i matrix is denoted by F(θ_i), F(α_i), F(β_i) and F(γ_i).

$F(\theta_i) = \begin{bmatrix} T_i^1 & T_i^2 & T_i^3 & T_i^4 & T_i^5 & \ldots & T_i^{31} & T_i^{32} \end{bmatrix}$

where, T_i^k represents the Fisher score for kth channel's theta band corresponding to ith subject.

similarly, $F(\alpha_i) = \begin{bmatrix} A_i^1 & A_i^2 & A_i^3 & A_i^4 & A_i^5 & \ldots & A_i^{31} & A_i^{32} \end{bmatrix}$

$F(\beta_i) = \begin{bmatrix} B_i^1 & B_i^2 & B_i^3 & B_i^4 & B_i^5 & \ldots & B_i^{31} & B_i^{32} \end{bmatrix}$

$F(\gamma_i) = \begin{bmatrix} H_i^1 & H_i^2 & H_i^3 & H_i^4 & H_i^5 & \ldots & H_i^{31} & H_i^{32} \end{bmatrix}$

Now we have defined the Fisher score of kth channel and for ith subject as

$$sc_k^i = \frac{T_i^k + A_i^k + B_i^k + H_i^k}{4} \tag{4.6}$$

After computing the Equation 4.6 $\forall$ k $\in$ [1, 32] we will get a vector for ith subject, which will contain all the channels Fisher score in vector form. let say that vector form is denoted by sc_i for ith subject such that,

sc$_i$ = $\begin{bmatrix} sc_1^i & sc_2^i & sc_3^i & sc_4^i & \ldots & sc_{32}^i \end{bmatrix}$

After getting the scores of different channels, first consider the channel which has the highest fisher score and check its accuracy by putting it feature vector on a classifier. Next, check accuracy by adding another feature vector of the channel which has second highest score, if the accuracy increases then select that channel, but if the accuracy decreases prune the channel and pick up the next channel. Repeat the same experiment until it reaches up-to the 32 channels. It gives an optimal subset of channels.

4.4 Algorithm

1. Compute the fisher score of each channel.

2. Pick up the channel with the highest fisher score and add the features of that channel to the feature vector.

3. Run SVM and calculate its classification accuracy.

4. Pick up next channel with the second highest (or third highest fisher score and so on) fisher score add the features of that channel to the feature vector along with the previously added features.

5. Run SVM and calculate its classification accuracy.

6. If its classification accuracy increases as compared to the previous iteration, repeat the 4 and 5, if it decreases, prune that channel and pick up the next one.

7. Continue this until all the 32 channels reached.

4.5 Feature Extraction from the Selected Optimal Set of Channels using Time-Frequency Domain(Wavelet based)

Since EEG signals changes at different instants of time or periods within a particular trial, temporal information plays a substantial role. Hence, once the most relevant channels are selected and we have extracted different types of features from Time-Frequency Domain to achieve more efficient results. We have implemented the 0.1Hz to 45Hz band-pass filter on the raw data then we have implemented ICA followed by CAR. We have taken the 4 labels of decomposition on the raw data. On the 1^{st} label of decomposition we will be having 0.1Hz to 30Hz and 30Hz to 60Hz (*Gamma Wave*). On the 2^{nd} label we will be having 0.1Hz to 15Hz and 15Hz to 30Hz (*Beta Wave*). On the 3^{rd} label we will be having 0.1Hz to 8Hz and 8Hz to 15Hz (*Alpha Wave*). On the 4^{th} label we will be having 0.1Hz to 4Hz (*Delta Wave*) and 4Hz to 8Hz (*Theta Wave*). Features extracted in this study from time-frequency domain (discrete wavelet based) for all frequency bands are listed in the Table 4.1

Table 4.1: Features extracted in this study from Time-Frequency domain

Domain	Features
Time-Frequency	Mean, Variance, Mode, Median, Skew, Standard Deviation, Kurtosis, Energy, Average Power, RMS, Shannon Entropy, Approximate Entropy, Permutation Entropy, Weighted Permutation Entropy, Hurst Exponent, Higuchi Fractal Dimension, Petrosian Fractal Dimension, Spectral, Entropy, Mean of Peak Frequency, Auto Regressive and Auto Regressive, Moving Average model parameters computed on decomposition coefficients

4.6 Genetic Algorithm based Feature Selection

Once the features are extracted, next step is to select the most relevant features. GA is employed to identify the optimum features using the methodology as described by Marjit et al. [35]. In GA, each chromosome in the individual is represented as a bit-string of 0's and 1's. In our work, the size of the individual is the no of features present in the feature space. The presence of a feature vector is represented by the following rules in the paper: If we have "0" bit in a gene position then we are not taking that feature and if we have "1" in a gene position then we are taking that feature for our classification problem.

In our experiment, feature selection process is represented as a multi-objective optimization problem, where we have used two objective functions i.e. (1) our first objective is to reduce the number of features and (2) the second one is to increment the accuracy of the classification model (SVM accuracy). We have given higher priority for classification accuracy rather than number of features selected as we can't accept a lower dimensional feature vector if it does not provide reasonable classification accuracy. The fitness function of *GA* is given in equation e4.

$$f = w_1 * f_1 + w_2 * f_2 \tag{4.7}$$

where f = Fitness of an Individual, f_1 = accuracy of the SVM model using selected features, $f_2 = (1 - \frac{|d_1|}{|d_2|})$ in which $|d_1|$ and $|d_2|$ shows the no of selected features and total no of features respectively and w_1 and w_2 are the weights assigned to accuracy and feature set dimension respectively in such a way that $w_1 + w_2 = 1$. For our problem we have taken $w_1 = 0.9$ and $w_2 = 0.1$.

The 7-way tournament selection procedure is used as our selection strategy followed crossover and mutation phases. One-point cross over is employed here. The number of generations is taken as 50, size of the population as 100, crossover and mutation probability as 0.65 and 0.05 respectively as stated by Dejong [36].

4.7 Classification

The emotional state represented by various emotional dimensions such as arousal and valence. For the classification, we divided the values into two classes based on whether the value is higher or lower than the midpoint value. For Arousal, we have two classes: high arousal (HA) and low arousal (LA). For valence, they are high valence (HV) and low valence (LV).For this binary classification problem,many methods can be used for the classification such as k-nearest neighbour (kNN), support vector machine (SVM), Naive Bayes classifier, and random forest (RF) etc .But in this experiment we have chosen SVM with polynomial kernal as it gives the best performance(based on different study) in many binary classification tasks to select best features in the classification process.In order to make the classification results convincing, we used a 10 fold cross validation split for the dataset.

Chapter 5

Experimental Result

To test the proposed methodology, initially we have taken the 32 subjects from the dataset.Dataset visualization has been shown in the Figure 5.1 for the Participant-1. Extracted the frequency domain features in the range of Theta(4-8)Hz,Alpha(8-12)Hz,Beta(12-30)Hz and Gamma(30-45)Hz using Welch's PSD Method(Fig 5.2), where dimension of each band is 32x4(i.e 32 is the no. of channels and 4 is the no. of band power for each channels).Fisher Score for each channels in the class of Valence and Arousal is shown in the Fig 5.3 and 5.4 for Subject-1. After applying the proposed methodology for Subject Dependent Analysis,the results are shown in the table 5.1,5.2,5.3 where Table 5.1 Number of selected electrodes for different classification problem are given. Similarly we selected channel subset for different classification problem in the table 5.2 and Table 5.3 depicts the Performance of proposed method based on 10-fold cross validation for Subject Dependent Analysis.Effect of Subject Dependent Channel Selection for High/Low Valence Classification is shown in the Figure 5.5 and Effect of Subject Dependent Channel Selection for High/Low Arousal Classification is shown in the Figure 5.6

5.1 Subject Dependent Analysis

First optimal set of channels has been chosen based on fisher score for each subjects. For channel selection we have used the PSD features from different frequency bands. After getting the optimal set of channels for each subject we have extracted the wavelet based features from the selected channels followed by GA based feature selection. The number of optimal channels for each subject based on the proposed methodology is presented in Table 5.1.

In figure 5.5 and 5.6 we have plotted the accuracy using optimal set of channels for each subject and as a comparison we have also added accuracy while taking all the channels. There is an improvement of accuracy as well as computational cost reduction due to channel selection. The emotion recognition accuracies For High/Low valence classification problem we got mean/average accuracy of **88.83%** , For High/Low arousal classification problem we got mean/average accuracy of **91.9%**

150 wavelet features were extracted from each of the selected channels followed by GA based feature selection and classification by SVM. The performance of the proposed methodology is illustrated in Table 5.3.

5.2 Analysis for Dominant Channels

Human Brain has divided into 5 main segments i.e. Frontal (F) Lobe, Parietal (P) Lobe, Occipital (O) Lobe, Cerebellum (C) Lobe and Temporal (T) Lobe. Although there are few segments are also there, namely Prefrontal (Fp), in between Frontal and Central Lobe Mid Frontal-Central (FC), in between Central Lobe and Parietal Lobe Mid Central-Parietal (CP) and in between Parietal Lobe and Occipital Lobe Mid Parietal-Occipital (PO). It is observed from Table 5.2 that the channels of frontal, parietal and temporal regions were repeated in case of almost all subjects with respect to the different classification (Low/High Valence, Low/High Arousal) problems. Hence, it can be concluded that the frontal, parietal and temporal regions of the brain plays inevitable role in emotion recognition as compared to the other regions. This result shows that emotion recognition involve multiple brain areas, which is expected as Yildirim et al. [37] and Pan et al. [38].

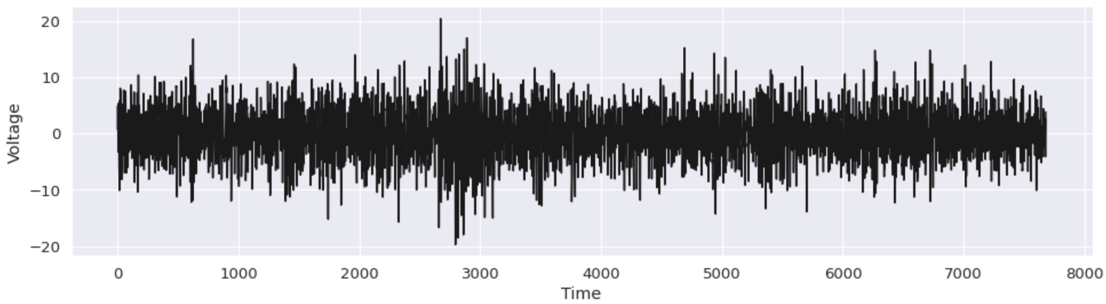
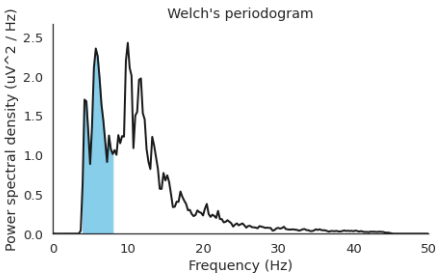
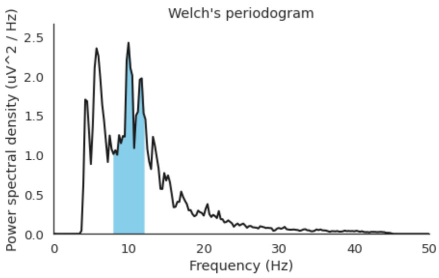


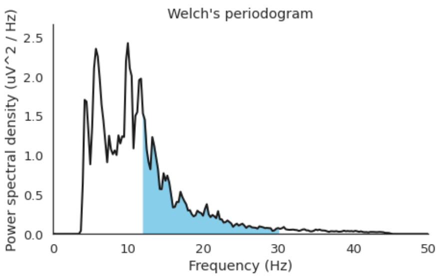
Figure 5.1: Data visualization for the Participant-1



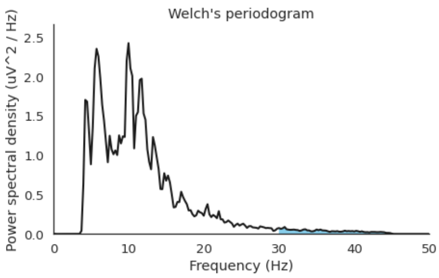
((**a**)) Theta Band



((**b**)) Alpha Band



((**c**)) Beta Band



((**d**)) Gamma Band

Figure 5.2: Feature Extraction(Frequency Domain)

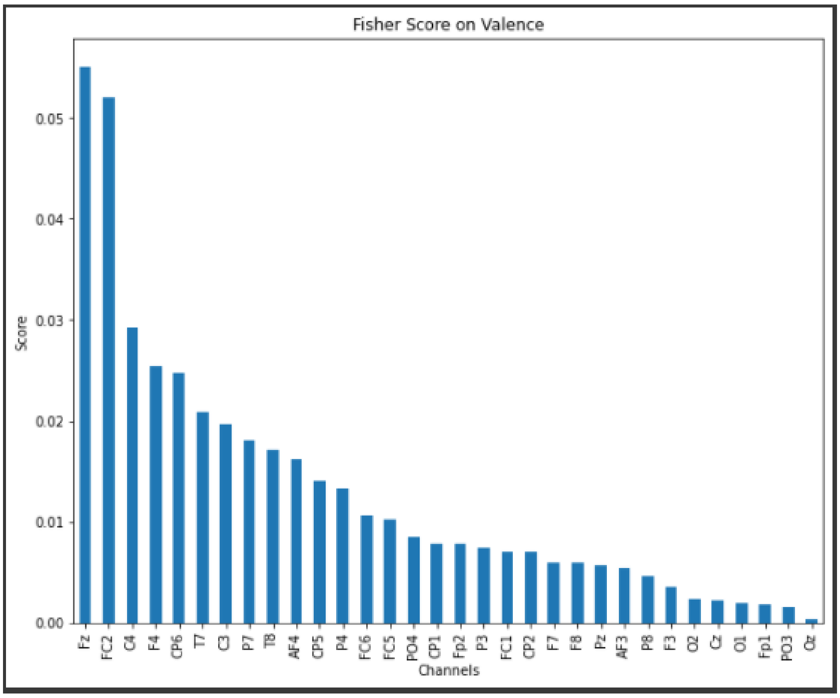


Figure 5.3: Fisher Score for Valence

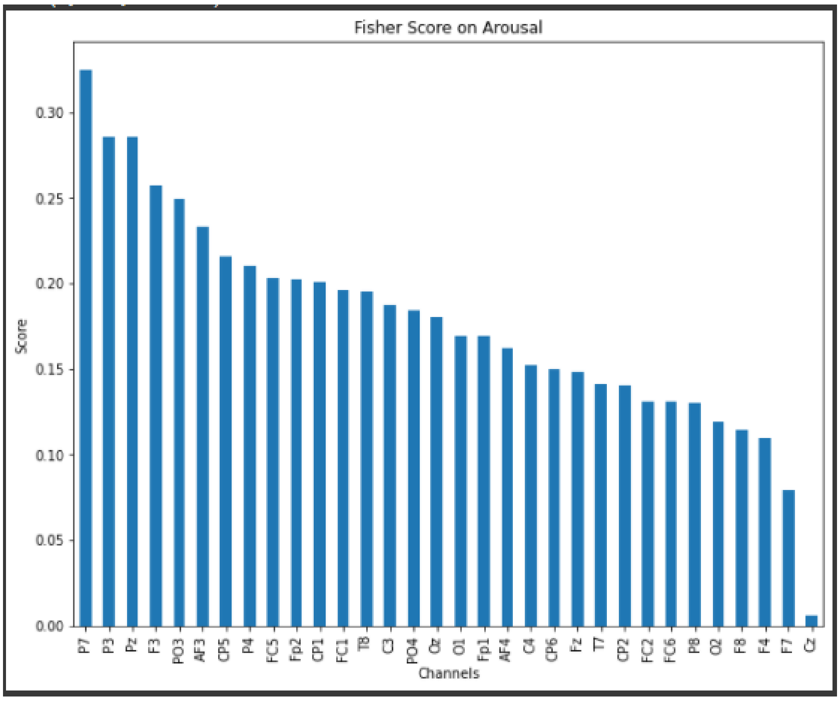


Figure 5.4: Fisher Score for Arousal

Table 5.1: No of selected electrodes for different classification problem

Subject	High/Low Valence	High/Low Arousal
s01	5	8
s02	13	2
s03	3	11
s04	3	5
s05	15	6
s06	8	12
s07	8	6
s08	16	9
s09	4	1
s10	3	10
s11	4	1
s12	3	5
s13	7	3
s14	6	10
s15	2	12
s16	24	5
s17	4	4
s18	13	6
s19	4	12
s20	1	17
s21	8	2
s22	15	7
s23	4	5
s24	3	32
s25	2	1
s26	6	7
s27	22	11
s28	12	8
s29	11	10
s30	14	24
s31	4	11
s32	13	7

Table 5.2: Selected channel subset for different classification problem

Subject	High/Low Valence	High/Low Arousal
s01	Fz, FC2, CP6, F7, FC1	P7, CP1, P3, Fz, P8, FC2, F7, C4
s02	AF3, F4, F7, CP1, O1, Fz, FC6, CP5, T7, C4, FC1, P3, Fp2	FC2, P7, AF4, FC6, AF3, F4, O1
s03	O2, Oz, FC1	AF4, FC1, O2, F3, AF3, P8, T7, C4, Pz, FC5, PO3
s04	O2, CP1, PO3	P8, CP2, P7, Pz, P3
s05	C4, Cz, CP6, P7, FC6, FC2, O2, F4, F7, F3, Fp1, AF3, PO3, T7, AF4	Fz, P7, T7, CP6, O2, Fp1
s06	P4, FC6, CP2, F4, O1, O2, Oz, Cz	C3, AF4, AF3, P7, F3, F4, CP5, T8, Oz, FC6, F7, Cz
s07	PO4, P4, T7, FC6, F4, FC5, F7, AF3	FC1, F3, O1, FC5, P3, Fz
s08	Fp1, P8, CP6, P7, AF3, O2, C4, Pz, T8, FC2, P3, AF4, PO3, Cz, CP5, Fz	T8, Fp1, FC6, Fp2, Fz, CP2, PO3, CP1, Cz
s09	AF4, Oz, C4, F4	AF4
s10	P3, CP6, FC6	FC1, Fp1, CP5, P7, T8, PO4, C3, PO3, F4, Cz
s11	P4, F4, P7, FC2	T7
s12	CP1, F4, F3	FC6, CP5, F3, Cz, FC2
s13	Cz, Fp2, Pz, AF3, FC2, P4, F7	PO3, P7, Fp2
s14	FC2, P3, F8, P8, FC5, AF4	PO3, FC2, P3, AF3, CP5, Pz, F7, T7, O2, F3
s15	Pz, AF3	FC2, FC1, Fz, P3, AF4, Cz, CP1, CP2, Fp2, AF3, CP5, Pz
s16	FC6, Fp1, Fp2, Cz, P4, P3, PO3, Oz, C4, CP2, F3, CP1, O1, P8, T8, C3, PO4, Pz, Fz, F8, FC2, CP6, F4, P7	O2, FC1, C3, C4, Cz
s17	F8, P3, PO3, CP1	P4, CP2, FC2, Fp1
s18	FC6, AF4, FC1, Pz, Oz, C3, F8, CP1, T7, Cz, PO4, Fp1, F7	AF4, FC6, FC5, PO4, P3, CP1
s19	C4, F3, CP6, FC6	CP5, P8, C4, FC6, PO4, O2, P4, CP6, P7, CP2, PO3, FC2
s20	AF4	F3, P7, CP6, F4, FC6, F8, CP5, Fz T7, FC5, AF3, O1, CP1, Cz, F7, C3, FC1
s21	AF3, P3, F4, PO4, Oz, CP2, AF4, T8	P4, Fz
s22	C3, T8, P4, PO3, Fp2, AF4, C4, Fp1 F4, AF3, CP5, FC2, FC5, FC1, Oz	O1, CP2, FC1, CP1, CP6, F8, P8
s23	T8, P4, F4, AF3	Fp1, F3, CP5, Cz, FC5
s24	FC1, Pz, PO3	Pz, O2, P8, CP2, O1, PO3, F8, FC1, P3, CP1, Cz T8, FC2, C3, FC5, Fz, CP5, F4, PO4, F7, Oz, T7 P7, AF4, Fp2, FC6, F3, P4, Fp1, AF3, C4, CP6
s25	FC2, Fp2	FC2
s26	Pz, P3, FC2, FC1, P7, T7	AF4, FC5, CP5, Oz, FC2, C3, Cz
s27	Oz, FC5, CP6, O2, C4, CP2, FC6, F3, C3, P3, CP5, P4, Cz, PO4, FC1, Pz, O1, Fz, T7, FC2, P7, Fp2	PO3, Oz, AF4, CP6, F8, Cz, FC6, O1, P8, F4, Fp2
s28	C3, CP2, F3, O1, CP1, Pz, Oz, PO4, O2, P4, P8, CP5	PO3, F4, FC6, Oz, CP2, O2, P4, CP6
s29	AF3, Cz, CP6, P4, Fp1, F8, F3, P7, FC6, F4, AF4	O2, CP6, Cz, PO4, O1, FC1, T8, P4, F4, C3
s30	PO4, Fz, Cz, AF3, FC2, Oz, P4, FC1, Fp2, O1, O2, Fp1, C3, T7	Oz, P4, CP1, P3, FC6, AF3, F3, F7, CP2, Cz, CP5, C4 FC1, P8, T8, T7, F8, C3, O2, CP6, Fp1, AF4, O1, F4
s31	CP6, CP2, Cz, CP5	FC2, C4, PO4, P4, F4, FC1, C3, CP6, AF3, F7, FC5
s32	CP2, O1, O2, CP6, CP5, PO4, FC6, P8, FC1, AF3, Cz, Oz, Fp1	Fz, Cz, P7, F4, F8, C4, P4

Table 5.3: Performance of proposed method based on 10-fold cross validation for Subject Dependent Analysis

Subject	Valence				Arousal			
	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
S-01	95	95	95	93.334	95	95	100	96.667
S-02	72.5	70.833	100	80	97.5	95	100	96.0476
S-03	97.5	100	95	96.667	85	85	100	89.238
S-04	97.5	100	95	96.667	85	91.667	86.667	86.476
S-05	85	78.333	100	85.667	80	90	76.667	78
S-06	85	83.333	100	88.81	82.5	80	95	84.238
S-07	95	96.667	97.5	96.6	87.5	90	91.667	88.476
S-08	82.5	80.833	97.5	85.267	85	78.334	100	86.476
S-09	85	89.167	88.333	88.267	87.5	95	88.333	89.333
S-10	87.5	97.5	88.333	90.743	87.5	90.833	91.667	89.238
S-11	87.5	95	85.833	87.905	97.5	100	97.5	98.571
S-12	92.5	90	97.5	91.905	95	95	100	97.143
S-13	92.5	95	90	90	97.5	97.5	100	98.571
S-14	92.5	91.667	95	91.267	87.5	88.333	96.667	91.143
S-15	92.5	95	92.5	91.905	95	100	91.667	94.667
S-16	67.5	75	85	73.97	90	100	83.333	88.333
S-17	87.5	80	100	86.333	95	94.167	100	96.571
S-18	87.5	95	85	87.333	95	96.667	97.5	96.571
S-19	92.5	95	91.667	93	92.5	91.667	100	95.143
S-20	97.5	100	97.5	98.627	90	89.1667	100	93.714
S-21	87.5	100	81.667	88	87.5	86.667	100	92.286
S-22	77.5	78.333	85.833	77.933	97.5	100	97.5	98.571
S-23	95	94.167	100	96.571	95	91.667	100	94.667
S-24	95	96.667	96.667	96	92.5	90	100	93
S-25	92.5	100	88.333	92.667	95	88.333	97.5	91.238
S-26	87.5	91.667	89.167	89.933	92.5	95	93.333	92.667
S-27	87.5	89.167	96.667	91.238	90	89.167	100	93.714
S-28	90	92.5	92.5	90.505	85	83.333	100	89.81
S-29	97.5	100	96.667	98	92.5	90	100	94
S-30	82.5	85.167	87.5	83.476	90	100	87.5	92.381
S-31	90	90.833	96.667	92.571	95	94.167	100	96.571
S-32	87.5	81.667	97.5	86.572	95	94.167	100	96.571
Total:	**88.83 $\pm$ 6.9**	**90.73 $\pm$ 8.1**	**93.31 $\pm$ 5.3**	**89.93 $\pm$ 5.7**	**91.1 $\pm$ 4.8**	**92.1 $\pm$ 5.52**	**95.4 $\pm$ 6.1**	**92.5 $\pm$ 4.6**

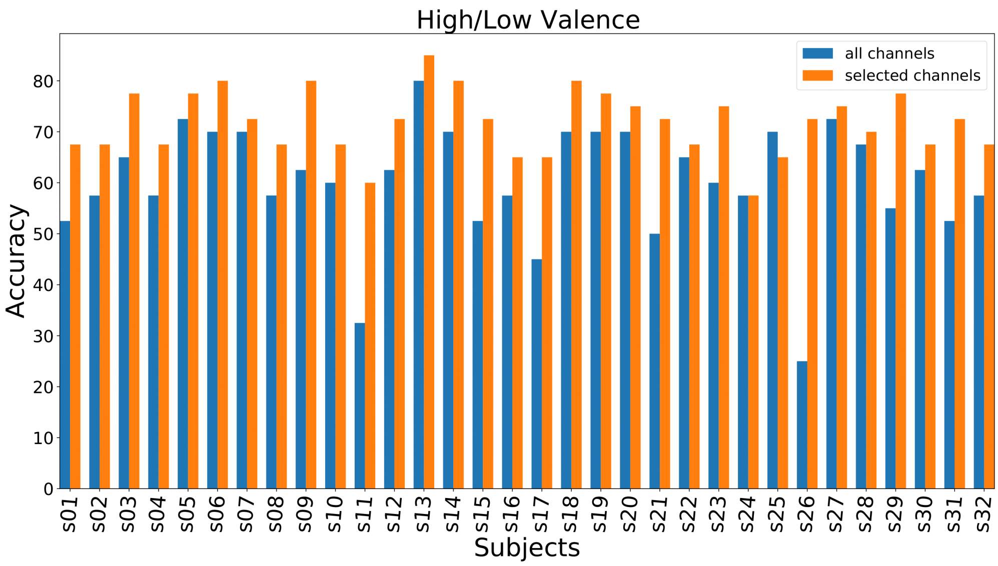


Figure 5.5: Effect of Subject Dependent Channel Selection for High/Low Valence Classification

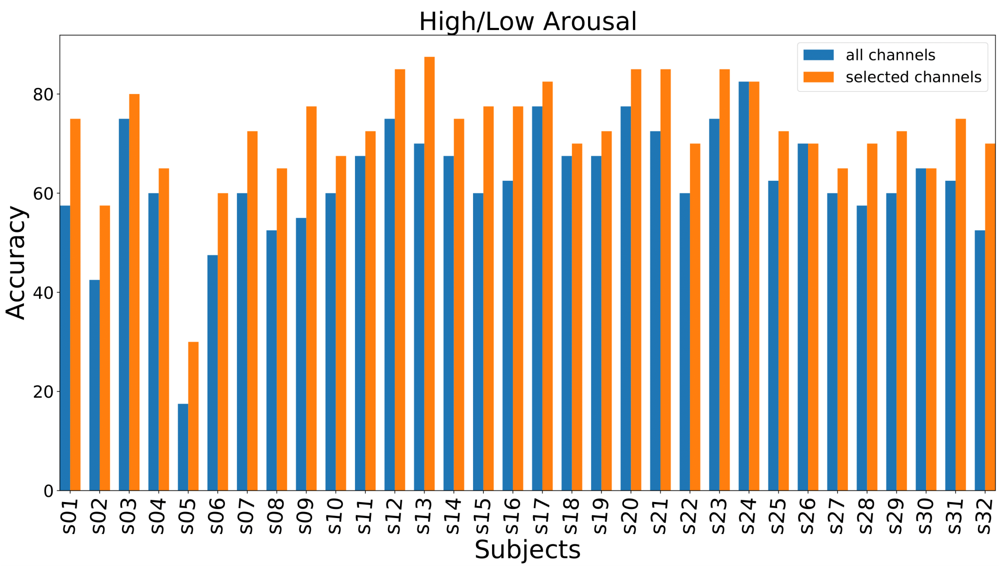


Figure 5.6: Effect of Subject Dependent Channel Selection for High/Low Arousal Classification

Chapter 6

Comparison and Discussion

Usually channel selection methods require a large number channels to participate in the emotion recognition paradigm. However, by using our proposed channel selection method, the number of channels falls down from 100% to 20-30%. The final selected channels for valence and arousal can be utilized in our daily life scene for emotion states monitoring. At the same time, our method obtains a relatively higher accuracy of 88.83% for valence and of 91.9% for arousal with selected channels, respectively. It is higher than the related work.Furthermore, comparison results with related work using EEG signals in DEAP dataset are displayed in Table 6.1.

Now the most important question is what are the advantages which make our methodology unique and better performer compared to the other existing methodology. So, we are approaching to solve the problem in a hybrid way to get a higher accuracy by amalgamating both filter and wrapper method instead of taking only one method for channel selection.Another challenge is is to extract the right features.As our task is to emotion recognition we know that to get the right features if we extract the PSD features from frequency domain then it will give us more solid results because different emotions like calm,concentration,relax are associated with different bands like alpha,beta,gamma and theta.Since the brain signals are non-stationary, time–frequency(discrete wavelet based) analysis may better present the dynamic changes in the signals, keeping mind of this advantages we have extracted discrete wavelet based features from the optimal set of selected channels. For searching the right features into a high dimensional feature space is also tedious job that's why we have adopted genetic algorithm to get the more stability for feature selection. Lastly if we see the survey of EEG-based BCI Emotion Recognition(2015-2020) we get to know that the literature's usage statistics of ICA,CAR are 26.8% and 5.0% which is better than other method like surface laplacian(0.4%), common spatial patterns(17.7%) etc. Similarly for channel selection , literature's usage statistics of filter method(fisher score) is 7.3% and wrapper method is 15.6%. Also we have used the genetic algorithm for feature selection after getting the optimal set of selected channels and GA(32.3%) has the highest literature usage statistics among all other feature selection method like mRMR(11.5%),univariate(6.3%),multivariate(6.3%) etc due its performance. Therefore from the above analysis we can conclude that, according to the literature usage statistics based on past experiments conducted by other researchers we have applied those techniques and methods by integrating them togetherly, which are performing best in terms of their effectiveness and accuracy so that it makes our experiment unique and advantageous from other experiments and we can achieve a better accuracy

The result of the experiment shows that the proposed method can select the key or relevant channels for daily-life EEG emotion recognition and ensure a relatively high accuracy.We know that, each channel carries a large amount of data. After we conducted the channel selection experiment,the number of channels reduced sharply, hence the overall data is also greatly reduced.Before the channel selection, the amount of data is large, and 32 full channels are connected to each other to form a fully connected brain network, which has a complicated structure. But after channel selection, only a few channels related to emotions is being extracted out of 32.Thus, the amount of data is greatly reduced. Then the connection relationship between channels is relatively simple and clear, indicating that the proposed method is indeed efficient. .

Table 6.1: Comparison with related work using EEG signals in DEAP dataset.

Authors	Approach	Accuracy(%)
Koelstra et al. [39]	Power Spectral Features	57.6(valence) 62.0(arousal)
Chung et al. [40]	Power spectral analysis with Bayes Classifier	66.6(valence) 66.4(arousal)
Yoon et al. [41]	FFT enhanced fetaure extraction and classification	70.9(valence) 70.1(arousal)
Zhuang et al. [42]	Spectral Power	70.9(valence) 67.1(arousal)
Candra et al. [43]	Three band wavelet entropy	65.1(valence) 64.8(arousal)
Verma et al. [44]	Multiresolution analysis and multilayer perceptron	63.5(valence) 69.6(arousal)
Wang et al. [45]	NMI-based channel selection	74.41(valence) 73.64(arousal)
Our Method	Hybrid Sequential Forward Channel Selection	88.83(valence) 91.9(arousal)

Chapter 7

Conclusion and Future Work

Our proposed methodology presents a novel channel selection strategy amalgamating filter and wrapper method where fisher score is used as filter method and support vector machine as a wrapper method to select optimal set of channels for EEG emotion recognition. Apart from that, we have used frequency and time-frequency domain features in our emotion recognition paradigm followed by genetic algorithm based feature selection. SVM is used for classification.The experiment has been carried out on DEAP database. The number of EEG channels have been reduced to 1/3 for subject dependent analysis in case of both valence and arousal using our proposed methodology.The results show the proposed method effectively improves the rate of emotion recognition while reduces the channels sharply.Finally, we can also explore the brain regions that are related to different emotions, which help us to study the relationship between specific brain regions and emotions.According to the result of Table7, it is observed that the channels of frontal regions were repeated for almost all subjects with respect to the different classification(Low/High Valence,Low/High Arousal) problems. Hence our experiment validated that the frontal region of the brain is more active compared to the other regions such as parietal,posterior,central,occipital etc as the dominance power of the frontal channels are more.

In future we plan to validate the proposed approach on other emotion recognition datasets and to test it with different combinations of features.The experiment can be carried out on subject independent analysis using multiclass classification of emotions(HVHA,LVLA,HVLA,LVHA) from the subset of selected channels using hybrid feature space. Another investigation is needed to answer concrete way in the future is whether certain types of features work better together than by themselves.

Bibliography

[1] Anwar P. P. Abdul Majeed Rabiu Muazu Musa Ahmad Fakhri Ab. Nasir Bifta Sama Bari Mamunur Rashid, Norizam Sulaiman and Sabira Khatun. Current status, challenges, and possible solutions of eeg-based brain-computer interface: A comprehensive review. In *Front. Neurorobot.*, June 2020.

[2] https://en.wikipedia.org/wiki/neuron. 2016.

[3] http://www.brainwaves.com/. 2016.

[4] V. Gandhi. Brain-computer interfacing for assistive robotics: Electroencephalograms. In *Recurrent Quantum Neural Networks, and User-centric Graphical Interfaces*. Academic Press, 2014.

[5] L. F. Nicolas-Alonso and J. Gomez-Gil. Brain computer interfaces, a review. In *Sensors*, volume 12, page 1211–1279, 2012.

[6] https://www.google.com/images/researchgate.

[7] P. Hanchuan, L. Fuhui, and D. Chris. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 27:1226–1238, 2005.

[8] D. J. McFarland G. Pfurtscheller J. R. Wolpaw, N. Birbaumer and T. M. Vaughan. Brain–computer interfaces for communication and control. In *Clinical neurophysiology*, volume 19, page 767–791, 2002.

[9] J.-J. Vidal. Toward direct brain-computer communication. In *Annual review of Biophysics and Bioengineering*, volume 2, page 157–180, 1973.

[10] https://www.eecs.qmul.ac.uk/mmv/datasets/deap/. 2012.

[11] R. P. Rao. In *Brain-computer interfacing: an introduction*. Cambridge University, 2013.

[12] T. C. E. Cheng Wei Wang, Gang Li. Channel selection in a supply chain with a multi-channel retailer: The role of channel operating costs. International Journal of Production Economics, March 2016.

[13] Stephen Barnhill Vladimir Vapnik Isabelle Guyon, Jason Weston. Gene selection for cancer classification using support vector machines. volume 46, pages 389–422. Springer, January 2002.

[14] Thilo Hinterberger Martin Bogdan N. Jeremy Hill Niels Birbaumer Wolfgang Rosenstiel Bernhard Schölkopf Michael Schröder, Thomas Navin Lal. Robust eeg channel selection across subjects for brain-computer interfaces. In *EURASIP Journal on Advances in Signal Processing volume*, volume 21, pages 289–302. Springer, November, 2005.

[15] Chun Sing Louis Tsui John Q Gan, Bashar Awwad Shiekh Hasan. A hybrid approach to feature subset selection for brain-computer interface design. Intelligent Data Engineering and Automated Learning - IDEAL 2011 - 12th International Conference, Norwich, UK,, Sept,2011.

[16] A Cichocki M Clerc M Congedo A Rakotomamonjy F Lotte, L Bougrain and F Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3), 2018.

[17] Hongpei Xu, Xingbo Wang, Weimin Li, Haibin Wang, and Qing Bi. Research on eeg channel selection method for emotion recognition. *International Conference on Robotics and Biomimetics Dali, China*, December 2019.

[18] W. James. What is an emotion? mind. 9:188–205, 1884.

[19] Paul Ekman. *Basic Emotions*, chapter 3, pages 45–60. John Wiley Sons, Ltd, 1999.

[20] M. Clynes. Sentics: The touch of emotions,. *New York, Doubleday*, 1997.

[21] C. E. Izard. Basic emotions, relations among emotions, and emotion-cognition relations. psychological review,. 99:561–566, 1992.

[22] N. H. Frijda. The emotions,. *Cambridge: Cambridge University Press*, 1986.

[23] R. Nagarajan S. Yaacob M. Rizon, M. Murugappan. Asymmetric ratio and fcm based salient channel selection for human emotion detection using eeg. WSEAS TRANSACTIONS on SIGNAL PROCESSING, January,2008.

[24] Tzyy-Ping Jung Yuan-Pin Lin. Improving eeg-based emotion classification using conditional transfer learning. Front. Hum. Neurosci., June, 2017.

[25] Yuanqing Li Lin He Youpan Hu, Daoli Li. Channel selection by rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery eeg. Neurocomputing, December, 2013.

[26] Shaokai Zhao Sanqing Hu Zhiguo Shi Jianhai Zhang, Ming Chen and Yu Cao. Relieff-based eeg sensor selection methods for emotion recognition. Sensors, September 2016.

[27] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deepneural networks. volume 7. IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT.

[28] Ram Bilas Pachori Vipin Gupta, Mayur Dahyabhai Chopda. Cross-subject emotion recognition using flexible analytic wavelet transform from eeg signals. IEEE Sensors, November 2018.

[29] Ram Bilas Pachori Varun Bajaj. Human emotion classification from eeg signals using multiwavelet transform. IEEE International Conference on Medical Biometrics, May 2014.

[30] Mohammad Soleymani Jong-Seok Lee Ashkan Yazdani Touradj Ebrahimi Thierry Pun Anton Nijhol Sander Koelstra, Christian Muhl. Deap: A database for emotion analysis ;using physiological signals. IEEE Transactions on Affective Computing, June 2011.

[31] https://www.eecs.qmul.ac.uk/mmv/datasets/deap/. 2012.

[32] P. Welch. The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967.

[33] Xiaofei He, Deng Cai, and Partha Niyogi. Laplacian score for feature selection. NIPS'05, page 507–514, Cambridge, MA, USA, 2005. MIT Press.

[34] Whitney AW.. A direct method of nonparametric measurement selection. *IEEE Trans Comput.*, 20, 1971.

[35] Shyam Marjit, Upasana Talukdar, and Shyamanta M Hazarika. Enhancing EEG-Based Emotion Recognition using MultiDomain Features and Genetic Algorithm based Feature Selection. In *9th International Conference on Pattern Recognition and Machine Intelligence*. ISI-Kolkata (submitted), 2022.

[36] Kenneth Alan De Jong. Analysis of the behavior of a class of genetic adaptive systems. *Technical report*, 1975.

[37] Esen Yildirim, Yasn Kaya, and Fatih Kiliç. A channel selection method for emotion recognition from eeg based on swarm-intelligence algorithms. *IEEE Access*, 9:109889–109902, 2021.

[38] Junhao Pan, Liying Zhan, ChuanLin Hu, Junkai Yang, Cong Wang, Li Gu, Shengqi Zhong, Yingyu Huang, Qian Wu, Xiaolin Xie, Qijin Chen, Hui Zhou, Miner Huang, and Xiang Wu. Emotion regulation and complex brain networks: Association between expressive suppression and efficiency in the fronto-parietal network and default-mode network. *Frontiers in Human Neuroscience*, 12, 2018.

[39] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3:18–31, 2011.

[40] S. Y. Chung and H. J. Yoon. Affective classification using bayesian classifier and supervised learning. *in Proc. 12th Int. Conf. Control, Automat. Syst.*, pages 1768–1771, Oct. 2012.

[41] H. J. Yoon and S. Y. Chung. Eeg-based emotion estimation using bayesian weighted-log-posterior function and perceptron convergence algorithm. *Comput. Biol. Med.*, 43(12):2230–2237, 2013.

[42] X. Zhuang, V. Rozgic, and M. Crystal. Compact unsupervised eeg response representation for emotion recognition. *in Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform.*, pages 736–739, Jun. 2014.

[43] H. Candra, M. Yuwono, R. Chai, A. Handojoseno, I. Elamvazuthi, H. T. Nguyen, and S. Su. Investigation of window size in classification of eeg-emotion signal with wavelet entropy

and support vector machine. *in Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 7250–7253, Aug. 2015.

[44] G. K. Verma and U. S. Tiwary. Affect representation and recognition in 3d continuous valence,arousal,dominance space. *Multimedia Tools Appl.*, 76:2159–2183, Jan. 2017.

[45] ZHONG-MIN WANG, SHU-YUAN HU, and HUI SONG. Channel selection method for eeg emotion recognition using normalized mutual information. *IEEE Access*, pages 736–739, Sept. 2019.