Emotion Detection using EEG Signals

by

Mohammad Adnan Hossain
18101262

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
June 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Mohammad Adnan Hossain

18101262

Approval

The thesis/project titled “Emotion Detection using EEG Signals” submitted by

1. Mohammad Adnan Hossain (18101262)

Of Spring, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on June 6, 2021.

Examining Committee:

Supervisor:
(Member)

Mostafijur Rahman Akhond
Lecturer
Department of Computer Science and Engineering
Jessore University of Science and Technology

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Emotions play a vital role in how people feel, think and act which makes it worthwhile for analyzing human behavior. As the patterns of emotions and their reflections differ from person to person, their study needs to be based on methods that are effective regardless of the diverse domain of the population. Hence, the analysis of physiological signals in detecting and extracting human emotions is gaining significance. To support this, resources and standards are being developed simultaneously. In this paper, we propose a pre-processing method along with some feature extractions and a model for emotion detection using EEG Signals based on DEAP dataset, a current benchmark for Emotion Classification research. For the pre-processing of data, prominent channels which contribute most to the classification are selected based on the role of the prefrontal cortex in emotion regulation and conscious experience and as for feature extractions, wavelet energy, wavelet entropy, and standard deviation are used. DNN (Deep Neural Network), SVM (Support Vector Machine), and KNN (K-Nearest Neighbour) are considered as the proposed model to detect emotions on a quadrant, HAHV (High Arousal and High Valence) or HALV (High Arousal and Low Valence) or LAHV (Low Arousal and High Valence) or LALV (Low Arousal and Low Valence). The approach we used yielded a maximum accuracy of 64%, 64%, and 70% for valence, arousal, and dominance respectively.

Keywords: Emotion detection; EEG signal; wavelet energy; wavelet entropy; Deep Neural Network; Support Vector Machine; K-Nearest Neighbour

Dedication

Dedicated to our loved ones for all their inspiration and support.

Acknowledgement

Firstly, all praise to the Great Allah for whom my thesis has been completed without any major interruption.
Secondly, to my co-supervisor Dr. Mohammad Zavid Parvez sir for his kind support and advice in our work and to my supervisor Mr. Mostafijur Rahman Akhond sir for continuously guiding me throughout.
And finally to my parents without their throughout support it may not be possible. With their kind support and prayer I am now on the verge of my graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

γ Gamma

BCI Brain Computer Interface

CNN Convolutional Neural Network

DBN Deep Belief Network

$DEAP$ Dataset for Emotion Analysis using Physiological Signals

DNN Deep Neural Network

DWT Discrete Wavelet Transformation

EEG Electroencephalogram

EMG Electromyography

EOG Electrooculogram

FD Fractional Dimension

FFT Fast Fourier Transformation

GSR Galvanic Skin Response

$HAHV$ High Arousal High Valence

$HALV$ High Arousal Low Valence

HOC Higher Order Crossings

KNN K-Nearest Neighbour

$LAHV$ Low Arousal High Valence

$LALV$ Low Arousal Low Valence

$LSTM$ Long Short-Term Memory

ML Machine Learning

MLP Multiplayer Perception Network

$mRMR$ Maximum Relevance Minimum Redundancy

$NREM$ Non-rapid eye movement

OLS Ordinary Least Squares

PCA Principal Component Analysis

PPG Photoplethysmogram

PSD Power Spectral Density

RBF Radial Basis Function Network

RBF Radial Basis Function

$ReLU$ Rectified Linear Unit

RF Random Forest

$RMSProp$ Root Mean Square Prop

SD Standard Deviation

SGD Stochastic Gradient Descent

STA Statistical Features

SVC Support Vector Classifiers

SVM Support Vector Machine

Chapter 1

Introduction

Emotions are humans' mental states which are strongly correlated with nervous systems. Emotions define human feelings, thoughts and behaviors also depend on them. Previously, their study was limited to philosophers and psychologists. However, it's becoming very much popular with statisticians, mathematicians, as well as computer scientists, nowadays. There are different reasons for this. To begin with, a new branch of computer science known as Brain-Computer Interface (BCI) is gaining traction, with the goal of translating brain activity into a digital form that can be used as a command for a computer [1]. Secondly, Emotions are no longer a vague and highly subjective topic as there exists a numerical model (Figure 1.1) for classifying and distinguishing human emotions [2].

Thirdly, standards for emotion measurement, measuring devices, and effective datasets all are available today. What's more, there are rigorous mathematical models for analyzing a large dataset of emotion [3] and very fast computational devices for running those models in feasible time. The electroencephalogram (EEG) is a scalp-based recording of the brain's electrical activity. The waveforms observed represent cortical electrical activity, which is very feeble and measured in microvolts. There are mainly 5 types of signal frequencies in the human EEG waves and they are gamma (30-100 Hz), beta (12-30 Hz), alpha (7.5-12 Hz), theta (4-7.5 Hz), the and delta (0.1-4 Hz) (Figure 1.2).

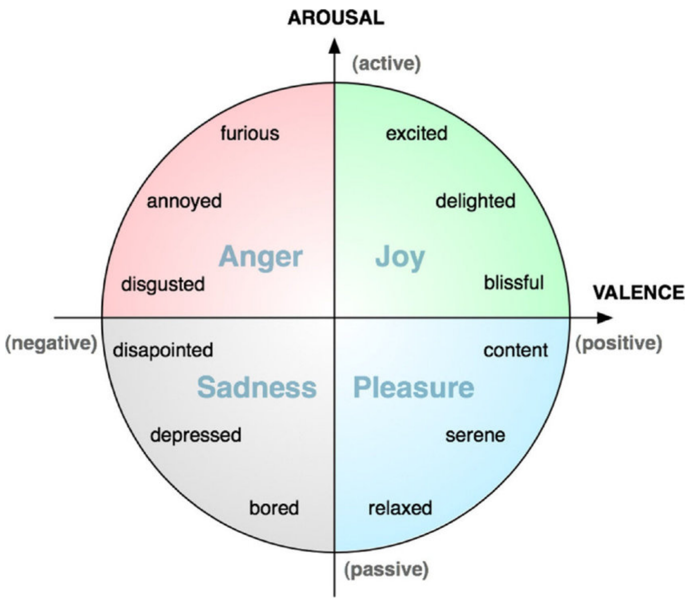


Figure 1.1: Distribution of emotions in a 2D plane

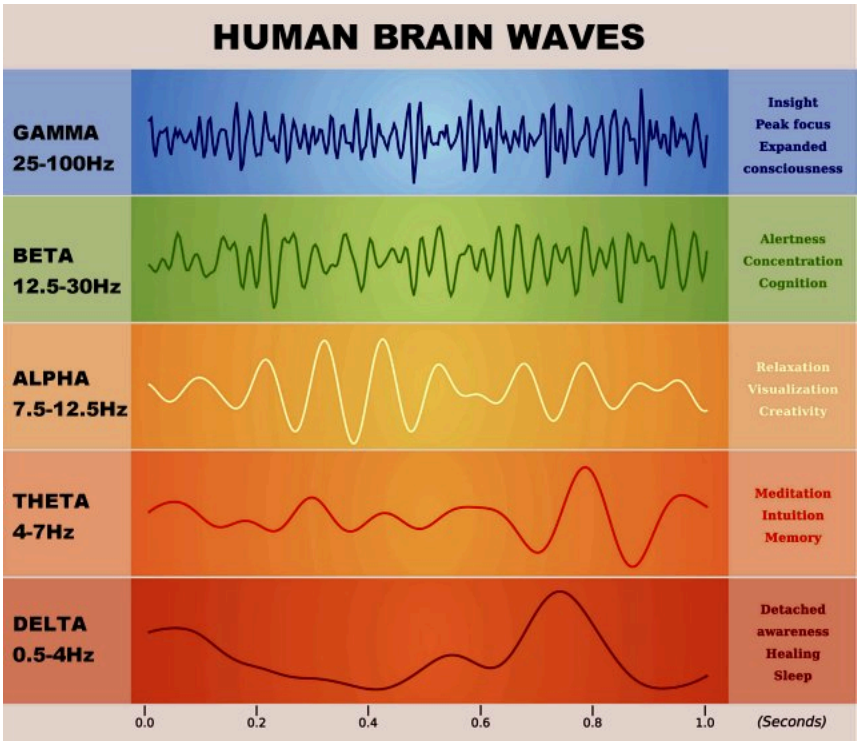


Figure 1.2: Major signal frequencies

When the information from multiple sections of the brain simultaneously is processed, gamma brainwaves are the quickest measuring EEG brain waves. They are related with higher mental activities. Beta brainwaves are linked to normal alertness or awareness, for example, active dialog or decision-making, problem-solving, task-oriented learning and beta brainwaves are easier to detect when actively thinking. Alpha brain waves are very easily detectable and were the first to be discovered, they are detectable when in a relaxed state, in light meditation, or creative super-learning mode. Theta brain waves are emitted when in deep meditation, light sleep, or daydreaming state, the source for theta activities are thought to be the frontal parts of the brain. Delta brainwaves are the slowest of all brainwaves, and they are most powerful while we are in restorative sleep, either dreamless or unconscious. NREM (non-rapid eye movement) sleep also produces the waves. Human emotion is a complex phenomenon and the generation mechanism behind it is unclear, it has been found that no matter what the culture some basic emotions are universal [4], and models like [2], Plutchik's wheel of emotions [5], and a tree of emotions [6] give us in-depth details of human emotions. Russel's [2] 2D model (Figure 1.3) portrays emotions on coordinates which denote the degree of valance (the positive or negative quality of emotion) and arousal (how responsive or energetic the subject is), it has been widely used as a reference for emotion analysis models concurrently DEAP [7] (Database for Emotional Analysis using Physiological Signals) dataset is used alongside for EEG signals [8, 9].

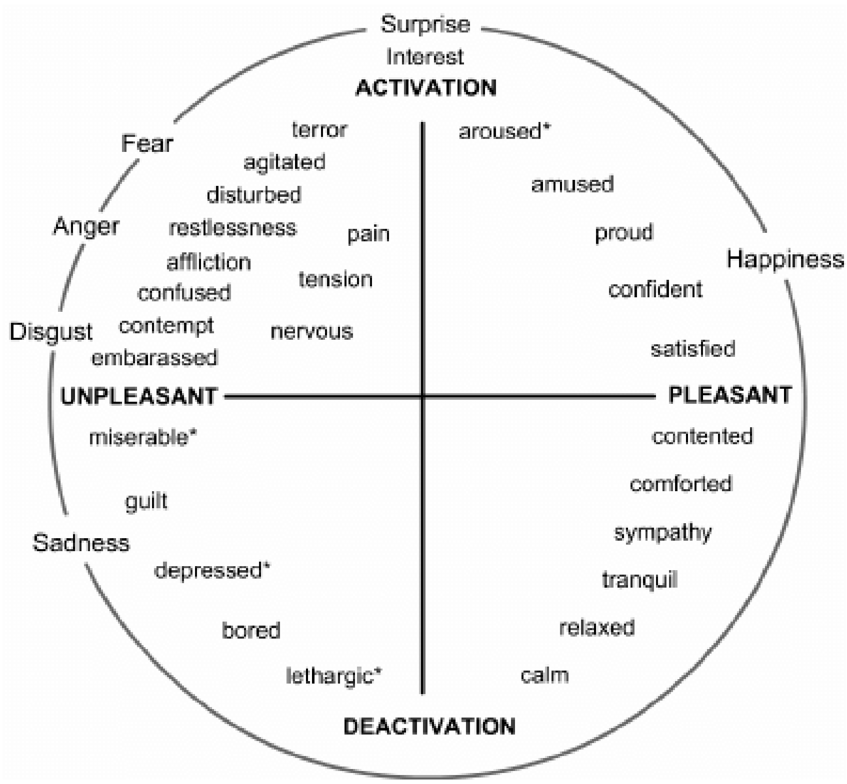


Figure 1.3: Circumplex Model of Emotions

1.1 Research Problem

Working with EEG signals can get cumbersome as a huge amount of data is considered and if a signal has a high frequency then the amount of data can increase exponentially. This causes predicaments as huge amounts of inputs are pushed into the models for training which takes enormous computing power and time. The DEAP [7] dataset is of no exception as even the preprocessed files contain a massive 322560 readings for each participant. Even after batching the 8064 readings per channel into 10 batches of approximately 807 readings each in [9], to reduce dimensionality, from each batch the mean, median, maximum, minimum, standard deviation, variance, range, skewness, and kurtosis values were extracted. After much trimming and extraction of 9 features from every 807 readings the input layer still had 4040 readings as inputs. Feature extraction for such immense data becomes challenging and complex as well.

Another predicament faced while working with such datasets is the amount of noise accumulated after the extraction of data points and portrayed on a time series line plot. Usually, these noises are filtered and then proceeded for further speculations [10], as sudden bursts of specific brainwave frequencies are to be recorded within a specific time window which is achievable using a DWT (Discrete Wavelet Transformation) as FFT (Fast Fourier Transformation) suffer from large noise sensitivity.

The author has described extensive techniques of extraction of features such as time domain (STA, HOC, FD), frequency domain (PSD), time frequency domain features in [8] (DWT, Multi-Electrode Features). For example, characteristics like Power, Mean, Standard deviation, 1st-difference, Normalized first difference, 2nd difference, and normalized second difference were employed in the Statistical Features section. Keeping in mind that this is only one of the many feature extraction methods conveyed here. And hence a large amount of these feature extractions methods are followed as it plays a very important role in emotion analysis. Concluding that such complex and lengthy procedures will be problematic to work on when such huge data is in play.

1.2 Research Objectives

The goal of this research is to ease down the complexity and make feature extraction and model training less complicated and time-consuming. As it was seen some of the channels did not play a vital role at all and the dataset consisted of extra data which could be trimmed and ignored, consequently achieving a method and a model which will not only simplify EEG extraction but will also produce better accuracy due to removal of insignificant and trivial data. Optimizing models become a vital role for any type of emotion analysis practice hence it is imminent that the optimization is reflected upon the amount of data being used and it only gets harder and cumbersome when the data increases. The number of hidden layers, the number of neurons in hidden layers, batch size, and activation functions all depend on the input layer in DNN, and thus it only gets easier as the data density decreases. In KNN, the value of K is highly dependent on the number of input values, and in SVM, the hyperline layout is also dependent on the number of values. The goal is also to use several models as using only DNN previously did not yield a good result hence by reducing data we are hoping to be able to produce better accuracy with the other models which will also require a deep insight and understanding of the models themselves. We have gone through some model optimization and selection works and implementations which are mentioned in Chapter 2 and hope to be able to apply the same to ours.

1.3 Thesis Orientation

The consequent chapters of the paper have been covered in the following order. Chapter 2 discusses similar work in the field and existing methodologies and gives an intensive analysis of the topics preceding data and information associated with our work. Chapter 3 presents the proposed model in detail. Chapter 4 give the test results and the interrelated discussions. At last, Chapter 5 closes and outlines the the thesis along with future plans.

Chapter 2

Literature Review

The release of the DEAP dataset has brought a paradigm shift in the community of affect computation researchers, as before that, most of the research was focused on facial expression or speech analysis. DEAP dataset makes the EEG signal popular for use in emotion research. As the study of EEG signals is relatively new, scientists and researchers are trying to find a way to produce a fathomable result. The study of EEG signals is very important, not only in the human-computer interface or video recommendation systems but also in medicine as most EEGs are done to diagnose and monitor seizure disorders. What's also fascinating is the link of EEG analysis with mental health where further research is being carried out for automated mental state detection [11]. There are several research works based on different mathematical models.

2.1 DEAP dataset

DEAP (A Database for Emotion Analysis using Physiological Signals) dataset, which is the current benchmark for effect computation [7], is a dataset of physiological signals of 32 subjects labeled with the data about emotions. Subjects were stimulated by 40 different 60 seconds long audio-visual stimulus (music video) and the emotional reaction was recorded as all the subjects gave a rating for each music video in terms of five parameters (Valence, arousal, dominance, liking, and familiarity). The physiological signals consist of EEG signals from 32 different channels and other peripheral signals like GSR, EMG, blood pressure, etc. There were frontal face videos of 22 subjects as well. Valence and arousal ratings were converted to binary form using a threshold of value 5 on a scale between 1 to 9. As a result, there were left only 4 emotional states; HAHV, HALV, LAHV, and LALV (L and H represent low and high, A and V represent Arousal and Valence) instead of 81 coordinates in the Arousal-Valence graph (Figure 2.1). EEG signal was recorded using 32 active BioSemi AgCl electrodes (placed according to the international 10-20 system). 8

Peripheral physiological Signals were also recorded. The signals were sampled at 512 Hz. The final resulting .bdf files were large hence a pre-processed downsampled (128 Hz) .mat files were utilized as this format is useful for testing classification or regression technique without explicitly having to process all the data. The format contained the information for 32 participants and each participant had 2 corresponding arrays, the first array has a dimension of 40 x 40 x 8064 (video/trial x channel x data), and the second array has a dimension of 40 x 4 (video/trial x label).

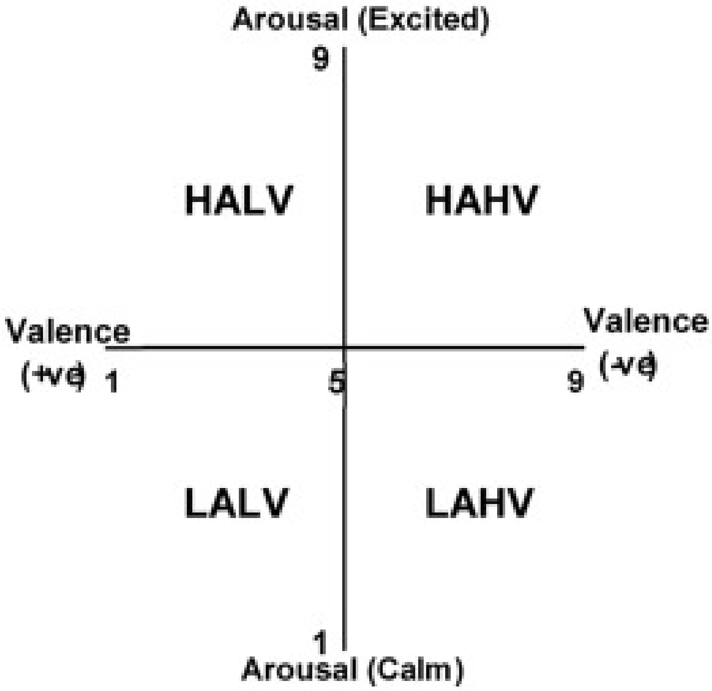


Figure 2.1: Classifying emotions in quadrants from ratings, horizontal axes represent valence ratings and vertical axes represent arousal ratings

2.2 EEG signals in emotion detection

Liu and Sorina [12] used a few extraction features but mainly focused on the Higuchi Fractal Dimension (FD) feature combined with statistical and Higher-Order Crossing (HOC) features which were extracted from four channels (based on the Fisher Discriminant Ratio channel selection method), namely FC5, F4, F7 and AF3, the Support Vector Machine (SVM) was used as the classifier. The classification was carried on using 2 datasets, the DEAP database, and some visual experimental data. For the DEAP database, a total of 8 emotions (happy, surprised, satisfied, protected, angry, frightened, unconcerned, and sad) were recognized, when the features selected were HOC, statistical, and Higuchi FD Spectrum the accuracy for 8 numbers of emotions was 54.58%. HOC,6 statistical, and FD features altogether gave an accuracy of 53.7% whilst only using 6 statistical and HOC features to classify gave 50.36% and 32.6% respectively. It is important to note that the mean accuracies of the classifications did increase as the number of emotions recognized decreased (from 8 to 2). The mean accuracies of the visual experimental data were significantly higher than that of the DEAP database, following the same feature selection tests as before the accuracies were in the order 58.68%, 56.6%, 56.94%, and 35.42%, which should be taken into account is that the highest number of emotions recognized here was 6 (from 6 to 2).

Wang and Shang [13] followed an unsupervised learning method and their model based on Deep Belief Network (DBN), which not only predicted emotions but also were able to extract features from raw signal data without the intervention of any hand-engineered feature extraction methods, to make it more scalable, the objective was to create a system that could automatically extract essential elements from raw physiological data with little human involvement. Amongst the 32 channels and 8 peripheral nervous system channels (a total of 40 channels) from the DEAP dataset, only 4 peripheral channels were considered, two Electrooculogram (EOG) and two Electromyography (EMG) channels. The EOG channels record eye movements, the two EMG channels included a zEMG which captures laughs or smiles, and a tEMG to reflects possible head movements. Finally, using the 4 channels and passing them through the DBN classifier the accuracies of arousal, valence, and liking obtained was 60.9%, 51.7%, and 68.4%, respectively. However, some drawbacks of the system were a long time for training and the selection of only 4 peripheral channels from DEAP.

Choi and Kim [11] used an artificial recurrent neural network-based model using Long Short-Term Memory (LSTM) to automatically extract features from raw data just like Wang and Shang [13] and classify the data accordingly. But unlike [13] they classified based on only arousal and valence. They used 10 channels as input data mainly the channels from the frontal lobe (Fp1, Fp2, F3, F4), temporal lobes (T7, T8), occipital lobes (P3, P4), and as reference a Photoplethysmogram (PPG) signal and a Galvanic Skin Response (GSR) signal. The hyperparameters of the model were modified at least 30 times but the top 4 settings with the best results were selected accordingly, despite the portrayal of the top 4 settings, the best result for both the arousal and valence were achieved in setting 4. For arousal, the setting consisted of 2 layers, 4 nodes, 128 batch size, 30 epochs, and 1, 1, 1 for learning rate, momentum, and forget gate bias, accordingly, with an accuracy of 74.65%. For valence, the setting consisted of 2 layers, 2 nodes, 128 batch size, 30 epochs, and 0.0001, 0.1, 1 for learning rate, momentum, and forget gate bias, accordingly, with an accuracy of 78%.

[14] proposed a simple probabilistic classifier based on Bayesian theorem and supervised learning for classifying emotions, for feature extraction Fast Fourier Transformation (FFT) was used, and for feature selection, Pearson correlation coefficient was applied. The classification was carried out on two-level classes (High and Low) and three-level classes (High, Medium, and Low) for both arousal and valence. The accuracies for the two-level class for valence and arousal were 70.9% and 70.1%, accordingly and the accuracies for the three-level class for valence and arousal were 55.4% and 55.2%, accordingly. [8] uses two models K Nearest Neighbor and Random Forest, but what's interesting in their paper is the feature extraction methods where they extracted a total of 12 features altogether and ran their models on each feature. They found that when the mRMR feature selection method was used for the dimension reduction of the feature vectors, the highest accuracy was achieved, with KNN producing 68.4% and 66.3% accuracy for arousal and valence, respectively, and RF producing 71.2% and 69.9% accuracy for arousal and valence, respectively.

The discrete wavelet transformation (DWT) for feature extraction was implemented in [10] which extracted the data as wavelet coefficients, after the decomposed signals were obtained there were a total of 5 different decomposed signals with each having its own frequency range. The following are the statistical characteristics that were utilized to depict the time-frequency distribution of EEG signals:

- The wavelet coefficients' maximum in each sub-band.
- The wavelet coefficients' minimum in each sub-band.
- Mean of the wavelet coefficients in each sub-band.
- The standard deviation of the wavelet coefficients in each sub-band.

The models implemented in this paper were Multiplayer Perception Network (MLP) and Radial Basis Function Network (RBF). The MLP consisted of 20 input nodes with two hidden layers of 24 and 14 nodes respectively along with two outputs (epileptic or not-epileptic), with 2000 epochs training. The 20 inputs were obtained by multiplying the four features with the number of wavelet decomposition (D1-D4 & A4). The model showed 97% accuracy on the classification for this two-class problem. The RBF network used the OLS (Ordinary Least Squares) algorithm, which consisted of 20 inputs and the result was a classification accuracy of 98%. Using RBF rather than MLP can speed up the training and make it extremely fast due to the OLS algorithm.

[9] provided both Deep Neural Network and Convolutional Neural Network models. The 8064 readings (of the DEAP) have been split into 10 batches of about 807 readings per channel for the feature extraction. Approximately 90 values were derived for each batch from the mean, average, maximum, minimum, defaults, variances, range, skewness and turnover numbers. The net mean, median, magnitude, minimum, standard variance, variance, range, skewedness and kurtosis data were added to the previously acquired 90 data, totaling 101 values per channel for the complete 8064 lectures together with the experiment, and participant numbers. For their Deep Neural Network (DNN) model, they used 4 fully connected dense neural networks, Rectified Linear Units (ReLu) as activation functions, Softmax for nonlinear activation functions, a dropout probability of 0.25 for the input layer, and 0.5 for the hidden layer, batch-size 310 and an epoch of 250. To speed up the learning rates RMSProp [15] was used and a learning rate of 0.00001 with a gradient direction of 0.9 was implemented in the RMSProp learning. 3 hidden layers had 5000, 500, and 1000 neurons respectively, where the first input layer takes 4040 readings, and this model achieved 75.78% and 73.13% accuracy for valence and arousal, respectively. However, their Convolutional Neural Network (CNN) achieved a fascinating 81.406% accuracy for valence and 73.36% accuracy for arousal detection, where they converted the DEAP data into 2D image format. The first Convolution layer accepts a 2D array as input and utilizes 100 filters, a 3*3 size kernel, and the activation functions 'TanHyperbolic' and 'Relu' for valence and arousal categorization, respectively. For both valence and arousal categorization, the second layer included 100 filters, a 3*3 size kernel, and 'TanHyperbolic' as the activation function. The Max-Pooling layer has a max-pooling of 2 dimension across 2x2 blocks, with the dropout of the layer outputs 0.25. The layer subsequently supplies the outputs with a totally linked neural dense stratum of 128 output dimensions. TThe dense layer has the activation function 'TanHyperbolic,' the probability of a drop-out of 0.5. The finished, completely connected neural dense layer had dimension of 2 or 3, which employed 'Softplus' as the activation function, depending on the number of output classes. In the Stochastic gradient descent (SGD), Categoric Cross Entropy has been employed as a loss function and an optimizer with a learning rate of 0.00001 and 0.001 respectively in terms of value and arousal, and with a gradient momentum of 0.9. 250 epochs were used in the experiment, and the model was trained in batches of 50 experiments each. To our understanding, the CNN model utilized in [9] has the highest accuracy of any of the other models.

Chapter 3

Proposed Methodology

3.1 Input data

The data for input was collected from the DEAP [7] dataset which consists of two parts. The first section presents the results of an online self-evaluation in which 14-16 participants scored 120 one-minute music video snippets on arousal, valence, and dominance. The second section includes assessments of participants, physiological measurements and facial registrations of an experiment where 32 people saw 40 of the music videos above. EEG and physiological data have been captured and the videos have also been assessed by each individual. Frontal video has also been captured for 22 individuals. The official data set contains every individual assessment, Youtube links of the music videos used, all the evaluators presented the videos, the answers given to the survey participants in advance of the experiment and the experiment's front-face video recordings 1-22 for the participants and the original unprocessed physiological data in BioSemi .bdf format. The physiological data recorded from the DEAP experiment in Matlab Format has been utilized to simplify and efficiently (data downsampled to 128Hz, elimination, filtering, segmentation, and so on). We have 2 arrays shown in Table 3.1 for each of the 32 individuals.

For each of the 32 subjects, the dataset comprises 40 experiments. For each of the 40 tests, the label array comprises a rating of valence, excitement and dominance. For each of the 40 studies, for each of the 32 subjects, there are 8064 physiologic/EEG signal data for each of 40 separate channels. Although there are 40 channels included in the dataset (8 peripheral physiological signals included), only 32 were considered as these were the EEG signals required for the experiment.

Table 3.1: DEAP dataset representation for each subject

Array Name	Array Shape	Array Contents
data	40 x 40 x 8064	video/trial x channel x data
labels	40 x 4	video/trial x label (valence, arousal, dominance, liking)

3.2 Data pre-processing

The total readings at hand for each experiment is 258048 (40 x 8064) and hence such huge readings need to be pre-processed before moving forward. The first step is to remove the 3 seconds baseline added to the signal, this is observed when the multiplying 63 with 128Hz (obtained from the pre-processed .mat files), and hence after removal of the extra 3 seconds it can be observed that the data is reduced to 7680 (60 x 128).

The data is then plotted on Welch's Periodogram (Figure 3.1) to observe the relationship between the brainwaves and the labels of valence and arousal provided by the subjects. In Figure 3.1 it can be seen that the waveform is partitioned by colors, for delta waves (0.5Hz - 4Hz) sky blue color was implemented, for theta waves (4Hz - 8Hz) red, for alpha waves (8Hz - 12Hz) yellow, for beta waves (12Hz - 30Hz) brown and finally for gamma waves (30Hz - 50Hz) black.

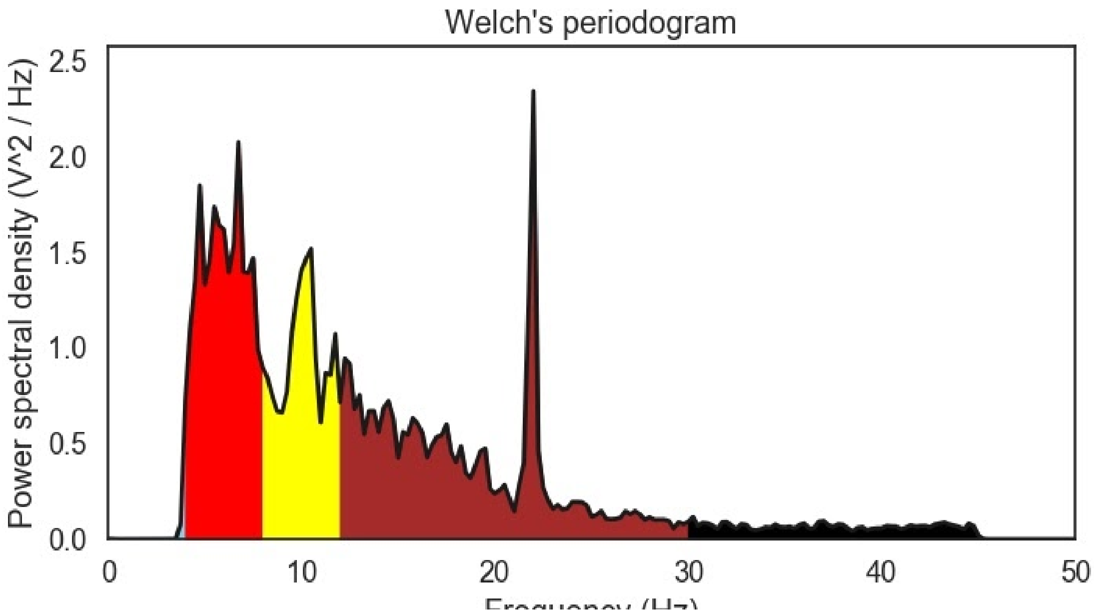


Figure 3.1: Welch's periodogram of subject 1 > trial 2 > channel 1

The x-axis portrays the Frequency in Hertz and the y-axis depicts the power spectral density (in V^2/Hz). During analysis subject-1's data readings were used from several trials mainly trial 4, 10, 22, and 20, in trial 4 where the valence was low and arousal was high (LVHA) the theta activity was observed to be high, in trial 10 the valence was low and the arousal was low as well (LVLA) and majority of the channels showed that the alpha activity was high, in trial 22 where the valence was high and arousal was low (HVLA) again a high alpha activity was detected, finally in trial 20 where valence and arousal given as high (HVHA) a relatively high theta activity was detected hence it was concluded that theta along with alpha activities were highly associated with arousal and valence. To confirm the claim the above experiment was again carried out for subject-6 as well.

3.2.1 The 10-20 system

There are predefined places for electrodes at the skull to make repeatable installations. The 10/20 system is one of such electrode placements or mounts. This system is regarded as an opportunity to determine the appropriate location of an electrode to reproduce the outcomes of the experiment. The name of the system comes from the approach used to locate the exact position of the electrode. Head size is a measurement variable. The technique therefore employs distances of many fixed places on the head in percentages. The beginning points are the nasion, the dent at the top of the nose and the inion at the rear. From the Nasion into the inion is an imagined vertical line and from the left ear lobe to the right is a horizontal line. The theoretical circle is drawn around the head, therefore 10 in the name, from 10% above the nasion and anion along the vertical line. The additional electrodes, which are indicated by the 20, are positioned at 20% inter-electrode distance. The Fz is 20% up from the nasion ring, while the top of the skull is marked with the Cz another 20%. Similarly, Pz is located on the vertical line. The horizontal marking of C3, T3, C4, and T4 is positioned in the same way. The electrodes on the imaginary circle are also 20% apart, while T3 and T4 are on the horizontal line. The rest of the electrodies are set equal to the horizontal lines of the front and parietal electrode between the vertical lines and the circle [16, 17].
Refer to Figure 3.2 for a visual depiction to make this written explanation less abstract. Please notice that this is not a regular top image of the skull that shows the position in the circle on the outside edge and doesn't display the temporal lobe. The places are extended in the outer circle of the nasion-inion. The cortical regions may also be seen as a broad indication: the top half are the front lobe, the parietal lobe directly underneath the front lobe, the temporal regions at the side and the occipital cortex at the bottom.

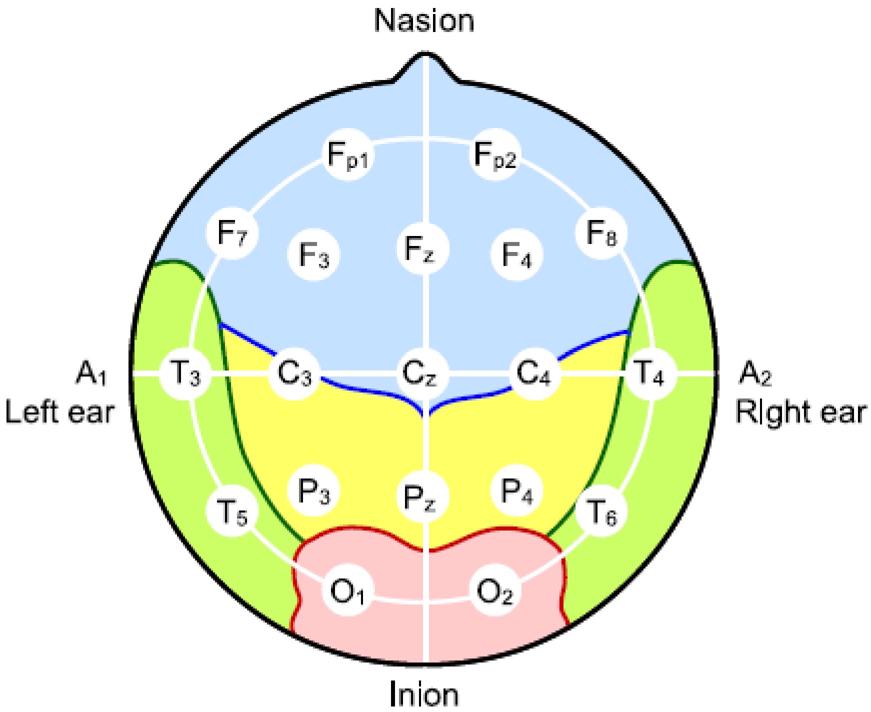


Figure 3.2: 10-20 system of electrode placement

There were 32 electrodes in the 10-20 system in [7] (Figure 3.3) and data were recorded in two distinct places. Participants 1-22 in Twente and 23-32 in Geneva have been documented. There are some small modifications in the format due to a different hardware review. First, there is a difference in the sequence of EEG channels at both sites. Secondly, for each site, the GSR measure is in another format. Hence when considering channels 9-20 for each subject the channel names might vary, for channel 9-20 the channel names in Twente are CP1, CP5, P7, P3, Pz, PO3, O1, Oz, O2, PO4, P4, and P8 whereas in Geneva they are CP5, CP1, P3, P7, PO3, O1, Oz, Pz, Fp2, AF4, Fz and F4 accordingly. However, the EEG channels have been reorganized in the preprocessed data to comply with the Geneva ruling.

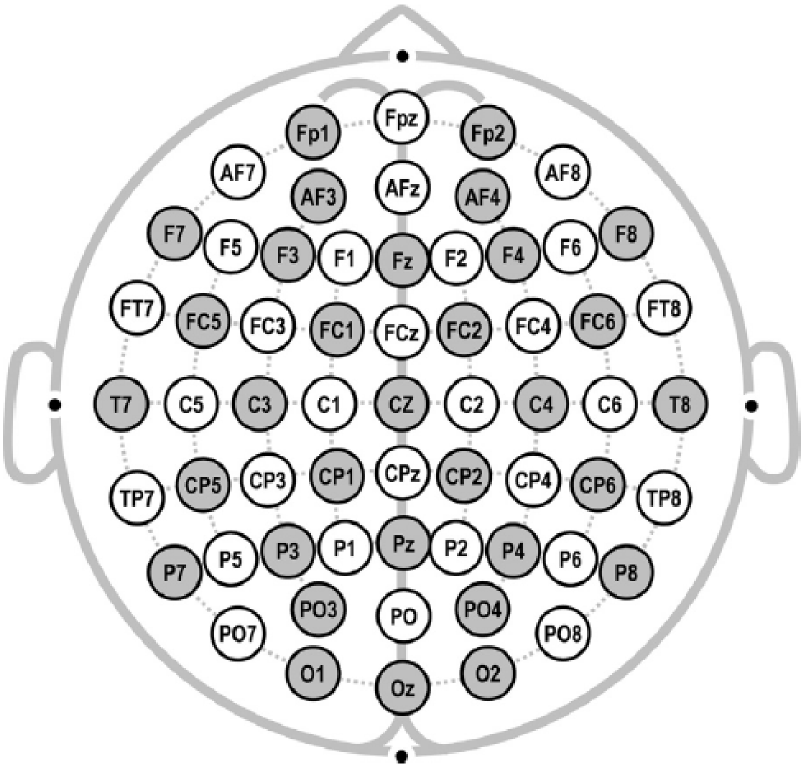


Figure 3.3: International 10 20 system for 32 electrodes (Gray circles)

3.2.2 Channel Selection

Psychological study has indicated that during the experience of negatives or good emotions, the left front lobe and right front lobe appear to have some activity. The left front lobe is especially activated while the person has a good effect and the right front lobe exhibits the same activity during a negative reaction [18]. In addition, significant research conducted in [19] demonstrate the involvement of the prefrontal cortex in emotion control and conscious experimentation. There are a total of 11 channels in the 10-20 system (for 32 electrodes) (Figure 3.3) related to the frontal lobe, namely Fp1, F3, F7, FC5, Fp2, Fz, F4, F8, FC6, FC2, and FC1 from which a total of four channels Fp1, Fp2, F3, and F4 (channel number 1, 3, 17, 20) were selected based on the studies carried out in [19, 18]. The channel selection were also based on the research carried out in the previous section where the data were plotted on a Welch's periodogram (Figure 3.1) where the Fp1 (channel number 1) showed a high activity theta activity in general when the arousal was high to support this claim a study in [20] shows why theta and arousal activity is highly correlated. It seemed that whenever a high arousal and high valence activity took place the frontal lobe of the brain showed a higher activity of theta hence the correlation between theta brainwave and arousal and valence activity were deduced.

3.3 Feature Extraction

Previously, before channel selection, we used DWT to extract wavelet energy as well as wavelet entropy from the data, standard deviation of the data was also considered. Hence, 3 features per channel were extracted giving a total of 3 x 32 x 32 (3072) features per trial.

3.3.1 Previous Feature Extraction

In the previous feature extraction procedure the data was reduced by taking a window size of 10 seconds thus 10 total segments. After which DWT were applied to the segments on each of the channel and the wavelet coefficients were extracted, the wavelet energy and wavelet entropy is estimated as follows at each sub-band (scale) l [21, 22]:

$$E_l = \sum_{n=1}^{2^{S-1}-1} \left|C_X\left(l,n\right)\right|^2 \ , N = 2^S, 1 < l < S \tag{3.1}$$

$$E_l = \sum_{n=1}^{2^{s-1}-1} \left|C_x\left(l,n\right)\right|^2 \ exp\left(\left|C_X\left(l,n\right)\right|^2\right), N = 2^S, 1 < l < S \tag{3.2}$$

where C_X(l, n) is the discrete wavelet translation at the l th scale and S is the number of scales.

The standard deviation of each channel was calculated as followed:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \tag{3.3}$$

where x_i represents each datapoint within the signal, μ represents the mean of the datapoints, and N represents the total number of datapoints.

3.3.2 Feature Extraction and Signal Refining after Channel Selection

However, the above feature extraction was discarded as the accuracy achieved was not high and as channel selection was not carried out. After selecting the four frontal channels (Fp1, Fp2, F3, and F4), first 3 seconds were excluded and then the remaining 7680 data were down-sampled to 256 data which was then fed into both the KNN and SVM classifiers.

Principal Component Analysis (PCA)

For the second feature extraction procedure Principal Component Analysis (PCA) was introduced, the Principal Component Analysis (PCA) is a linear reduction approach that may be used by projecting information from a high dimension space into a smaller sub-space. It aims to keep the vital parts of the data with the most variation and delete the non-essential sections with the least variation. PCA clusters similar data points based on feature correlation between them without any supervisions as it is an unsupervised dimensionality (feature) reduction technique. The principal components value was given 20 and 3 to project 7680-dimensional EEG data to 20-dimensional principal components for KNN model and 3-dimensional principal components for SVM model respectively which helped make training the models faster and even slightly increased the accuracy of arousal for SVM.

3.4 Emotion Classification

The emotions are separately classified for valence, arousal, and dominance. One hot encoding is applied to all of the three labels. For all the three labels value above 4.5 is encoded to 1 and below 4.5 is considered to be 0. It is to be noted that our previous experiment, which excluded channel selection and down sampling of the data, did not classify dominance. It was later added to the current models and classification to classify emotion on a 3-dimensional plane (Figure 3.4) which gives us the opportunity to explore more emotional experiences.

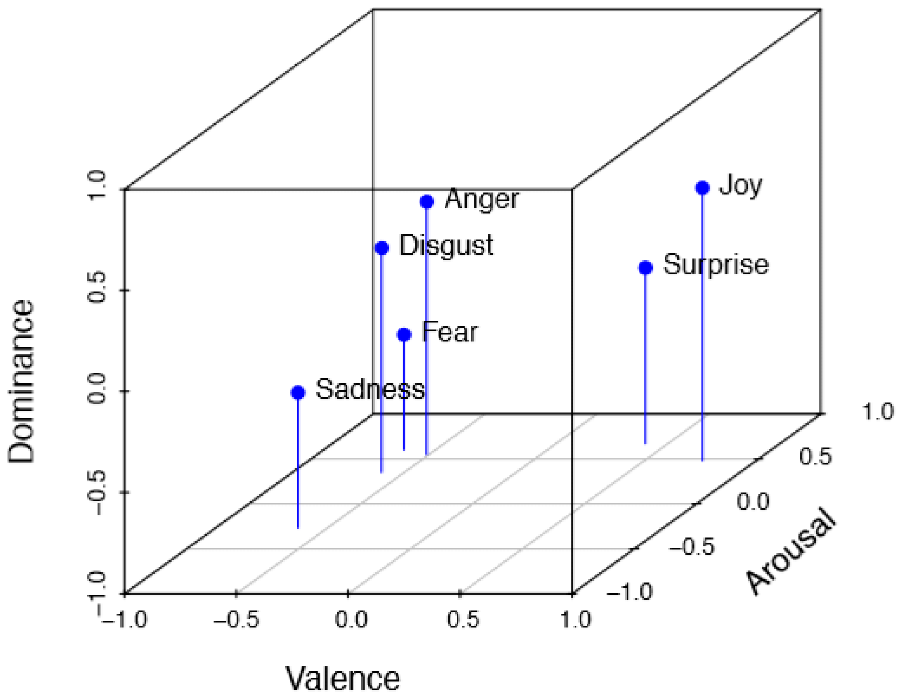


Figure 3.4: The emotion space covered by the Valence-Arousal-Dominance (VAD) model

3.4.1 Prior Models

Previously (as mentioned above), before channel selection, we used Discrete Wavelet Transformation (DWT) over the .mat files for getting all samples' Energy, Entropy, and Standard Deviation mapped with subjects' number, video number, channel number along with Subjects rating in terms of valence and arousal. Valence and arousal ratings were converted to binary form using threshold of value 5 in a scale between 1 to 9. We also implemented 2 models onto the dataset and all of them are discussed below in detail.

Support Vector Machine

Our SVM model had a higher degree kernel to yield a more flexible decision boundary and the decision function used was 'ovo' to train a classifier for each 2-pair class combination. It was trained with one of the 32 test subjects' data with X_train being the energy, entropy, Standard deviation (SD) and y_train:- arousal/valence (either high i.e. 2 or low i.e. 1), total number of input values fed to the model were 320 which were classified as either high/low valence/arousal. We ran 2 models, one for valence and the other for arousal. After training our model we saved it in a .sav file for later accuracy/prediction measurement. We tested the accuracy of the model later by loading the .sav file, after randomly testing using 5 subjects' data our accuracy output resulted in 58.5-59% and 56.3-56.5% for arousal and valence respectively.

Deep Neural Network

The DNN model we designed has an input layer with 3 neurons and 3 hidden layers. For the following layer, the output of one layer is the input. The input layer uses its three neurons to read signals like energy, entropy and standard deviation. layer 2 has 500 neurons, layer 3 has 400 neurons and layer 4 has 25 neurons. All neurons have ReLu (Rectified Linear Units) as the activation function. The output layer consists of 2 neurons using softmax as the activation function to determine whether the value is of high/low valence/arousal. Figure 3.5 demonstrates this neural network. The optimum number of neurons was selected through trial and error methods. We have trained our model in 310 sets and 200 epoch to ensure a complete process of learning. The model was running against multiple training data extracted from one of the 32 participants. The saved .sav was afterwards imported and the data of several individuals was utilized to assess the model's accuracy and to identify the effect of 58.8–58.92% and 56.5–56.7% correspondingly for arousal and valence.

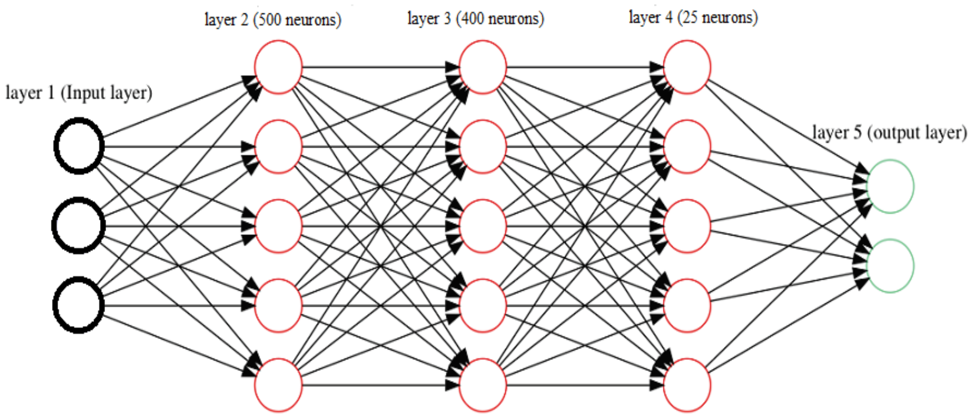


Figure 3.5: Deep Neural Network Architecture

3.4.2 Proposed Models

After the channel selection process two ML models were used, KNN and SVM. The test size was 20% leaving the training data with 80%, the training and testing data were both standardised between -1 and 1 and then fed into the models.

K-Nearest Neighbor

The optimal K value for the KNN was determined using K-fold cross-validation for valence, arousal, and dominance. For the cross-validation process the dataset was divided into 4 segments (i.e. k = 4) and a max K (for KNN) value for valence, arousal, and dominance obtained were 37, 39, and 32 respectively (Figures 3.6, 3.7, and 3.8).

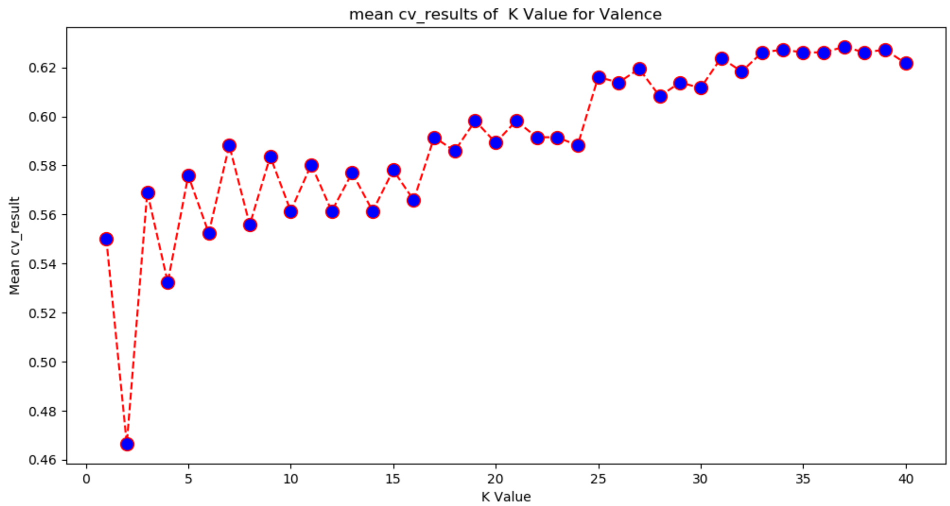


Figure 3.6: Cross validation for valence

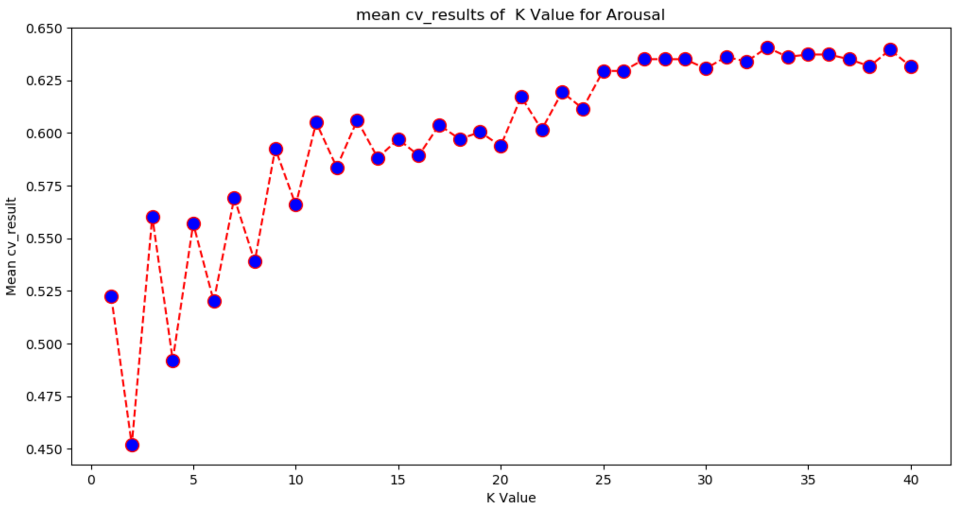


Figure 3.7: Cross validation for arousal

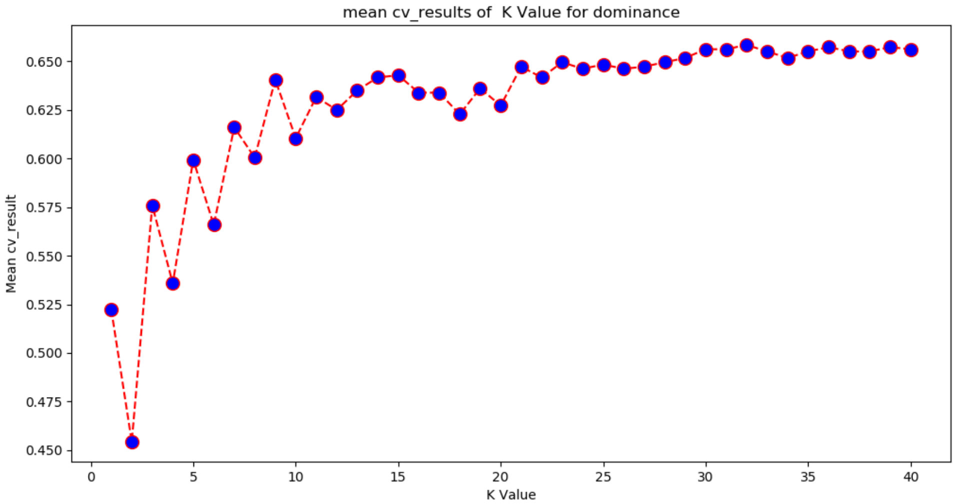


Figure 3.8: Cross validation for dominance

For the PCA data the optimal value of K was determined by square rooting the total data (i.e. 20) which was approximately 5 and the rest of the procedure was same as before. The model was used to classify whether the valence/arousal/dominance was high/low. Giving us the chance to classify emotion on a 3-dimensional plot rather than 2. All the training data were taken from the refined and downsampled dataset (prior to PCA implementation) of all the four frontal channels.

Support Vector Machine

The SVM used The Radial (RBF) Kernel to deal with the overlapping data, the radial kernel finds Support Vector Classifiers (SVC) in infinite dimensions. The Radial Basis Function is estimated as follows:

$$f(a,b) = e^{-\gamma(a-b)^2} \tag{3.4}$$

where a and b are two datapoints and γ (gamma) scales the squared distance between them $(a-b)^2$.
The principal component value was selected as 3, as the change in component value from 20 to 3 increased the overall accuracy of the model and decreased the model training time immensely as well. As for rest of the procedure for the input data, the same steps were implemented like the previous model.

Chapter 4

Results and Discussions

To get to the final result, various tactics were used. The result requires extensive investigation, with the purpose of extracting features in the most efficient manner possible. We offer a method for classifying human emotion using the DEAP dataset, a benchmark for emotion research, to convert EEG signals into time domain frequency analysis and downsample the data as well as apply PCA to the raw readings and finally use KNN, SVM, and DNN to identify emotion in this study. The accuracy of the models is determined using the following equation:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{4.1}$$

where True Positive stands TP, True Negative TN, False Positive FP, False Negative, and False Negative FN.

The dataset is divided into two arrays namely data and label (Table 3.1). The label is an array of size 32 x 4 where it contains the valence, arousal, dominance, and liking of all the 32 subjects. From which valence, arousal, and dominance were extracted and was processed using the one-hot-encoding technique.

To begin with the DEAP dataset was refined for both the procedures, before and after channel selection, by excluding the first 3 seconds of the data hence decreasing the data from 8064 to 7680. For the first procedure, prior to channel selection, each data had been divided into segments of 6 seconds (thus 10 total segments). Then on the divided segments DWT was carried out to extract 3 features. Using these features the average accuracy obtained for DNN was Valence-58.86% and Arousal-56.6% and for SVM was Valence-56.40% and Arousal-58.75%. The accuracy obtained were not that great hence a new method was considered.

For the second method, to reduce dimensionality of the data, channel selection was introduced. Researches indicate that the frontal lobe of the brain plays a vital role in emotional experiences [19, 18] based on which channels Fp1, Fp2, F3, and F4 were selected. Secondly, the data were downsampled from 7680 to 256 which was then fed to both the KNN and SVM. KNN gave an accuracy of 64.45%, 63.28%, and 70.31% for valence, arousal, and dominance respectively. SVM gave an accuracy of 64.45%, 64.84%, and 70.31% for valence, arousal, and dominance respectively. Later on pca was applied to the raw 7680 data, the components value for KNN and SVM used was 20 and 3 respectively, thus reducing the dimensionality of data further. After which KNN gave 57.03%, 52.73%, and 61.72% for valence, arousal, and dominance respectively, SVM gave 64.84%, 64.84%, and 70.31% for valence, arousal, and dominance respectively. In Table 4.1 we are comparing all the models used along with their respectable accuracies.

Table 4.1: Accuracy of all the models along with feature selection type

Model	**Feature**	**Valence Accuracy**	**Arousal Accuracy**	**Dominance Accuracy**
DNN	DWT (no channel selection)	58.86%	56.6%	-
SVM	DWT (no channel selection)	56.40%	58.75%	-
KNN	Downsampled (with channel selection)	64.45%	63.28%	70.31%
SVM	Downsampled (with channel selection)	64.45%	64.84%	70.31%
KNN	PCA (with channel selection)	57.03%	52.73%	61.72%
SVM	PCA (with channel selection)	64.84%	64.84%	70.31%

From Table 4.1 it can be seen the best accuracy was obtained when PCA was used along with SVM. Even after reducing the dimensionality of the data from 7680 to 3, it is fascinating how the accuracy still held up.
Refer to figures 4.1, 4.2, and 4.3 for a detailed visual comparison between all the models along with the feature extraction methods and all the three labels namely Valence, Arousal, and Dominance respectively.

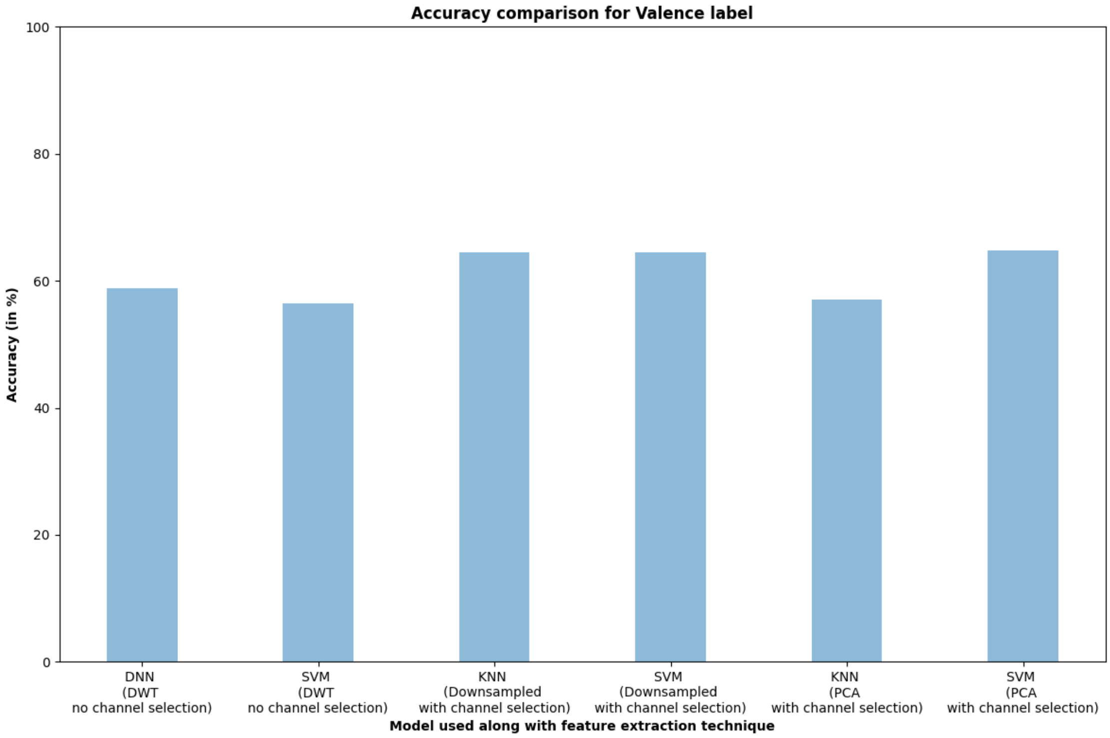


Figure 4.1: Model accuracy comparison of Valence

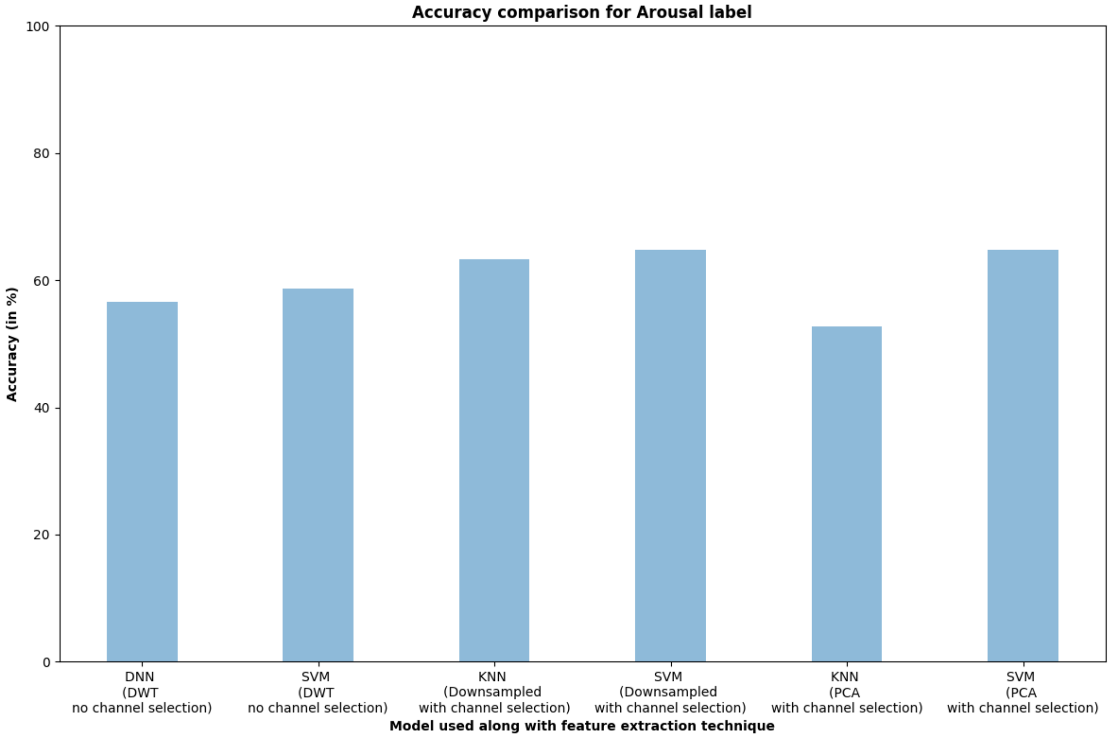


Figure 4.2: Model accuracy comparison of Arousal

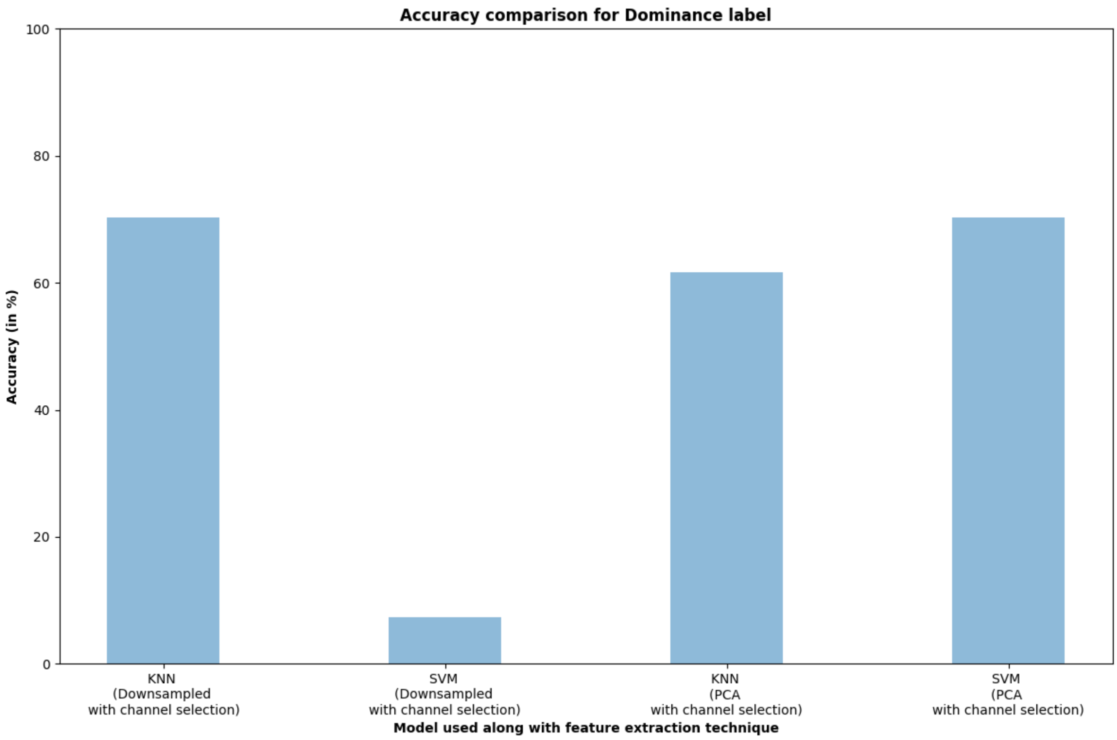
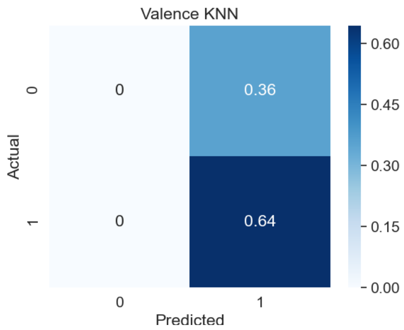
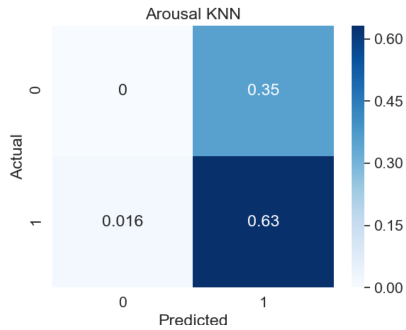


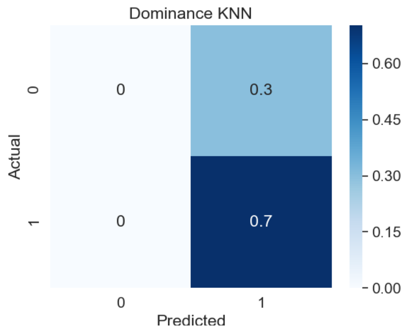
Figure 4.3: Model accuracy comparison of Dominance



(a) Confusion matrix for valence

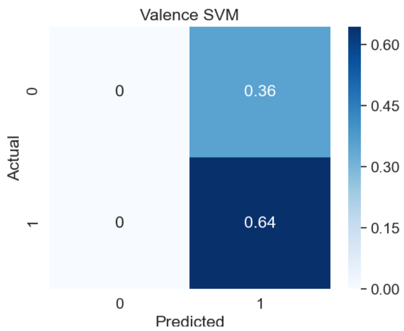


(b) Confusion matrix for arousal

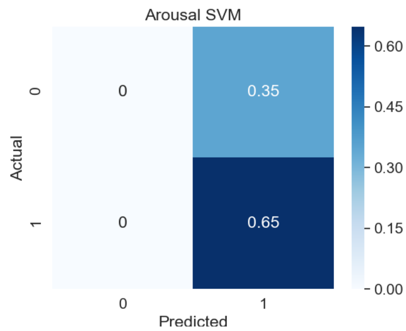


(c) Confusion matrix for dominance

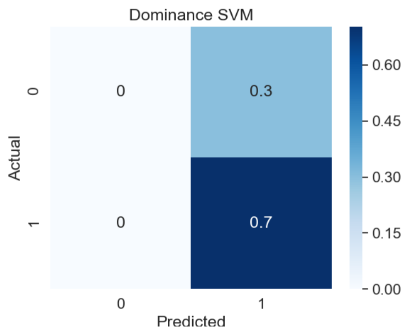
Figure 4.4: Confusion matrices for KNN



(a) Confusion matrix for valence

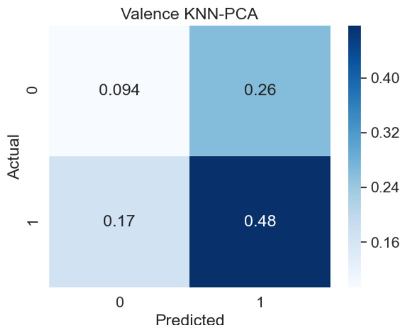


(b) Confusion matrix for arousal

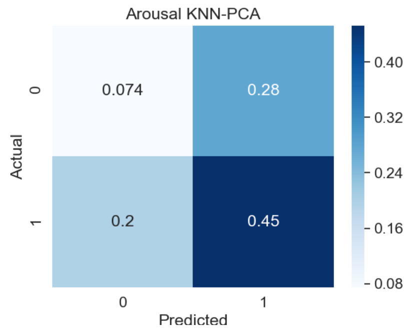


(c) Confusion matrix for dominance

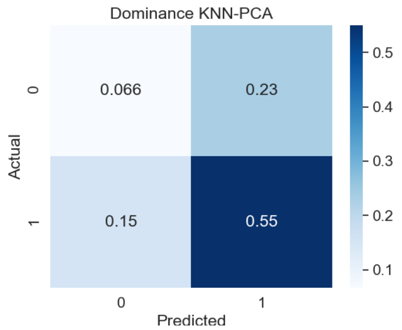
Figure 4.5: Confusion matrices for SVM



(a) Confusion matrix for valence

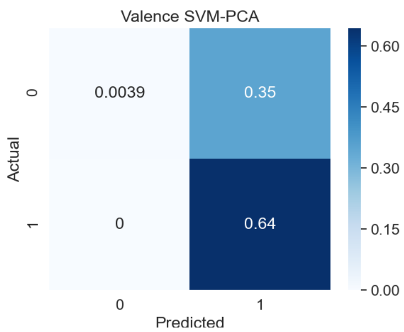


(b) Confusion matrix for arousal

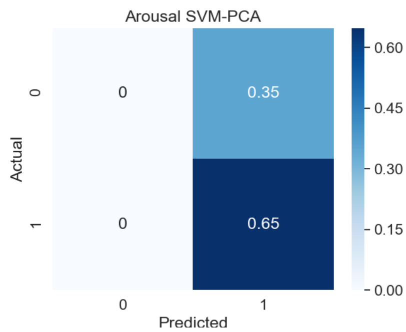


(c) Confusion matrix for dominance

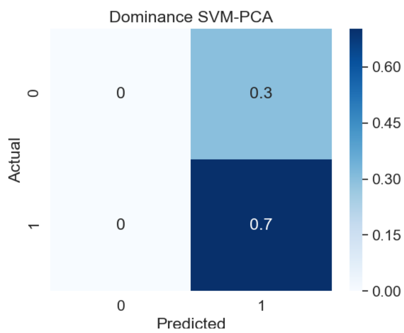
Figure 4.6: Confusion matrices for KNN with PCA



(a) Confusion matrix for valence



(b) Confusion matrix for arousal



(c) Confusion matrix for dominance

Figure 4.7: Confusion matrices for SVM with PCA

Figures 4.4, 4.5, 4.6, and 4.7 depicts the confusion matrices of the models along with certain feature extraction methods which were used to classify the three labels namely Valence, Arousal, and Dominance respectively.

Chapter 5

Conclusion

All our procedures indicate that high accuracy in emotion detection can be achieved using simpler models with fewer data points. Moreover, our procedures also indicated that majority of the emotions are related to the pre-frontal lobe of the brain consequently helping us consider channels referring to the frontal area. Channel selection immensely helps us by reducing dimensionality and the number of channels required for classification.

However, we used down-sampled data and many of the features are trimmed and a lot of relevant data is lost for reducing complexity. Escaping this and taking all the relevant data into account and excluding the redundancy more effectively will increase the accuracy in every model, we hope. Besides this fact, till now these models can only detect the emotional states in a binary manner (High/Low) not the continuous spectrum of emotion as increasing the emotional state to a higher order decreases accuracy immensely. However, other optimised models can be developed for achieving that higher goal. In that case, models will have the caliber for pinpointing a person's emotional condition at the cost of high computational power. Reducing the redundancy and making these models efficient will be a real challenge then.

Bibliography

[1] Jerry J Shih, Dean J Krusienski, and Jonathan R Wolpaw. "Brain-computer interfaces in medicine". In: *Mayo Clinic Proceedings*. Vol. 87. 3. Elsevier. 2012, pp. 268–279.

[2] James A Russell. "A circumplex model of affect." In: *Journal of personality and social psychology* 39.6 (1980), p. 1161.

[3] Gregory P Strauss et al. "Mathematically modeling emotion regulation abnormalities during psychotic experiences in schizophrenia". In: *Clinical Psychological Science* 7.2 (2019), pp. 216–233.

[4] Paul Ekman et al. "Universals and cultural differences in the judgments of facial expressions of emotion." In: *Journal of personality and social psychology* 53.4 (1987), p. 712.

[5] Robert Plutchik. "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice". In: *American scientist* 89.4 (2001), pp. 344–350.

[6] Phillip Shaver et al. "Emotion knowledge: further exploration of a prototype approach." In: *Journal of personality and social psychology* 52.6 (1987), p. 1061.

[7] Sander Koelstra et al. "Deap: A database for emotion analysis; using physiological signals". In: *IEEE transactions on affective computing* 3.1 (2011), pp. 18–31.

[8] Jingxin Liu et al. "Emotion detection from EEG recordings". In: *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. IEEE. 2016, pp. 1722–1727.

[9] Samarth Tripathi et al. "Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset." In: *Twenty-ninth IAAI conference*. 2017.

[10] Pari Jahankhani, Vassilis Kodogiannis, and Kenneth Revett. "EEG signal classification using wavelet feature extraction and neural networks". In: *IEEE John Vincent Atanasoff 2006 International Symposium on Modern Computing (JVA'06)*. IEEE. 2006, pp. 120–124.

[11] Eun Jeong Choi and Dong Keun Kim. "Arousal and valence classification model based on long short-term memory and deap data for mental healthcare management". In: *Healthcare informatics research* 24.4 (2018), p. 309.

[12] Yisi Liu and Olga Sourina. "EEG-based subject-dependent emotion recognition algorithm using fractal dimension". In: *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE. 2014, pp. 3166–3171.

[13] Dan Wang and Yi Shang. "Modeling physiological data with deep belief networks". In: *International journal of information and education technology (IJIET)* 3.5 (2013), p. 505.

[14] Hyun Joong Yoon and Seong Youb Chung. "EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm". In: *Computers in biology and medicine* 43.12 (2013), pp. 2230–2237.

[15] Tijmen Tieleman and Geoffrey Hinton. "Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude". In: *COURSERA: Neural networks for machine learning* 4.2 (2012), pp. 26–31.

[16] Herbert Jasper. "Report of the committee on methods of clinical examination in electroencephalography". In: *Electroencephalogr Clin Neurophysiol* 10 (1958), pp. 370–375.

[17] Manzoor Khazi, Atul Kumar, and MJ Vidya. "Analysis of EEG using 10: 20 electrode system". In: *International Journal of Innovative Research in Science, Engineering and Technology* 1.2 (2012), pp. 185–191.

[18] Christopher Niemic. "Studies of Emotion: A Theoretical and Empirical Review of Psychophysiological Studies of Emotion." In: (2004).

[19] Richard J Davidson, Daren C Jackson, and Ned H Kalin. "Emotion, plasticity, context, and regulation: perspectives from affective neuroscience." In: *Psychological bulletin* 126.6 (2000), p. 890.

[20] Ian R Kemp and Birger R Kaada. "The relation of hippocampal theta activity to arousal, attentive behaviour and somato-motor movements in unrestrained cats". In: *Brain research* 95.2-3 (1975), pp. 323–342.

[21] Panagiotis C Petrantonakis and Leontios J Hadjileontiadis. "Emotion recognition from EEG using higher order crossings". In: *IEEE Transactions on information Technology in Biomedicine* 14.2 (2009), pp. 186–197.

[22] Mouhannad Ali et al. "A novel EEG-based emotion recognition approach for e-healthcare applications". In: *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. 2016, pp. 162–164.