B.Sc. in Electrical and Electronic Engineering Thesis

Automatic Emotion Recognition from EEG Signal in Time-Frequency Domain with Reduced Channels Using Deep Neural Architecture

Submitted by

Student 1
1606050

Student 2
1606075

Supervised by

BUET EEE Teacher

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

May 2022

CANDIDATES' DECLARATION

This is to certify that the work presented in this thesis, titled, "Automatic Emotion Recognition from EEG Signal in Time-Frequency Domain with Reduced Channels Using Deep Neural Architecture", is the outcome of the investigation and research carried out by us under the supervision of BUET EEE Teacher.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma or other qualifications.

Student 1
1606050

Student 2
1606075

CERTIFICATION

This thesis titled, **"Automatic Emotion Recognition from EEG Signal in Time-Frequency Domain with Reduced Channels Using Deep Neural Architecture"**, submitted by the group as mentioned below has been accepted as satisfactory in partial fulfillment of the requirements for the degree B.Sc. in Electrical and Electronic Engineering in May 2022.

Submitted By:

Student 1

Student 2

Supervisor:

BUET EEE Teacher
Professor
Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology

ACKNOWLEDGEMENT

I would like to thank ...

Dhaka
May 2022

Student 1

Student 2

Contents

List of Figures

List of Tables

List of Algorithms

ABSTRACT

Emotion recognition using the electroencephalogram (EEG) has attained great attention in the arena of human-computer interaction (HCI) due to the existence of variations in neural activities at different types of emotions. In this thesis, an automatic emotion recognition scheme is adopted utilizing multi-channel EEG signal that can discriminate between different classes of emotions as well as select some significant channels which are mostly responsible in the elicitation of emotions. The main theme of the proposed method is to classify emotions with deep neural architectures with the raw EEG data or the extracted features from it. The baseline of the EEG signal is excluded from the raw data in the preprocessing stage to obtain the trial signal relevant to emotion elicitation in response to the audio-visual stimuli. The trial signal is divided into multiple segments and combining all the channel information, the obtained 2D matrix is applied to the proposed deep neural architecture. The proposed network with its channel attention mechanism offers satisfactory classification performance. As the spectral information provides salient information regarding different emotional states, the trial signal is decomposed into several sub-bands and 3D frames are formed combining the temporal and spectral information of the available channels. The frame contains significant information that eventually provides superior classification performance. In addition, EEG signals are also analyzed in the time-frequency domain where the multi-level discrete wavelet transform (DWT) coefficients or the extracted feature from the continuous wavelet transform (CWT) domain is considered. For a given channel of EEG data, each CWT coefficient from different scales is mapped into a corresponding strength-to-entropy component ratio (SECR) plane to obtain CEF2D feature matrix. In order to reduce the computational complexity, effective channels and CWT scale selection schemes are proposed based on the energy-to-entropy ratio in the CWT domain. Extensive experimentation is carried out on a publicly available emotion (DEAP) dataset and very satisfactory classification performance is obtained for valence and arousal types of emotions both in 3-class and 2-class scenarios.

Description, mathematical notation, diagram, figure

introduction - novelty

ablation study contribution of blocks hyper-parameter tuning

baseline method/benchmark

error analysis/qualitative study

methodology - clean notation

Chapter 1

Introduction

Human emotion is a complex psychophysiological phenomenon, and it is widely associated with the cognition, perception, reasoning and intelligence level of human being [1]. An individual can understand the emotional state of other persons through words, voice intonation, facial expressions and body language. In recent years, human emotion analysis has played a crucial role in analyzing mental states and cognitive functions which helps physically and mentally impaired people to express themselves using brain-computer interfaces [2]. Brain-computer interfaces allow users to monitor and control the activity of the computer using the responses of their brain and the acquired brain signals are relayed to output devices to perform desired tasks [3].

1.1 Theories of Basic Emotions

Emotions play a pivotal role in the evolution of consciousness, neurobiological development. The psychiatric and neuroscientific research on emotion analysis posit that human beings are endowed with a small set of fundamental emotions by nature. Other complex emotions are the derivatives of the basic emotions. A discrete and independent neural system subserves every emotion that is considered having evolved through their adaptive values in dealing with life tasks. Emotions can be viewed as an internal neural activity that drives the behaviour and responses depending on different external stimuli. Different affective states share a common characteristics like short duration, rapid onset, unbidden occurence, coherence among responses etc [4]. Different emotional states are the combination of physiological arousal, psychological appraisal which influence an individual to respond to different states.

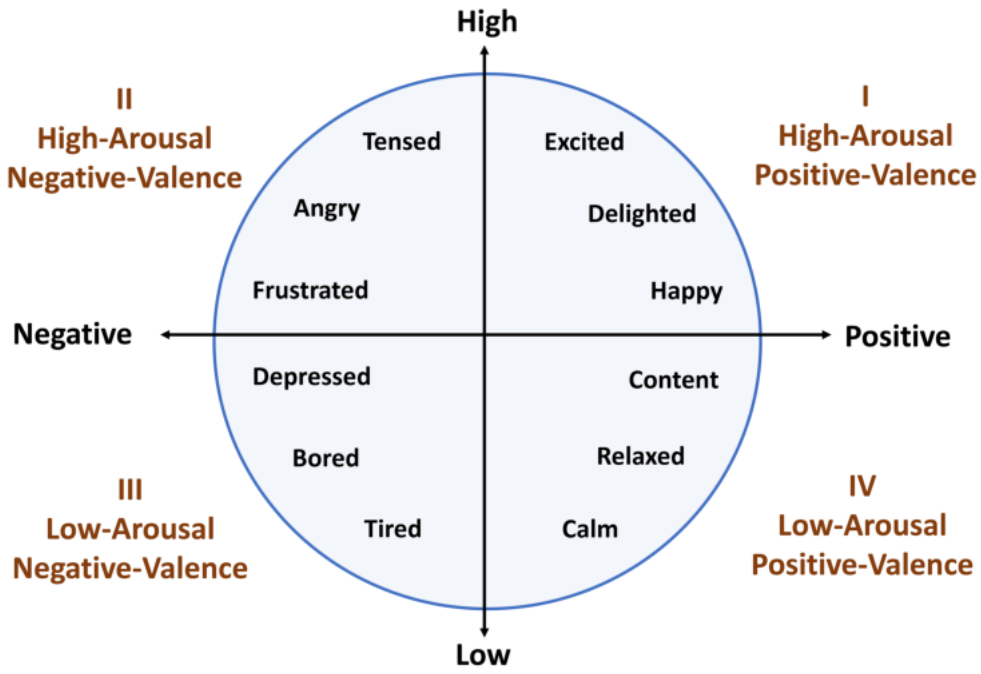


Figure 1.1: Representation of different types of emotions in two-dimensional valence-arousal space

1.2 Valence-Arousal Model

A two-dimensional circumplex valence-arousal (V-A) model is most widely used to quantify the emotional states [5]. Human emotions can be conceptualized in a two-dimensional circular space where valence represents horizontal axis and arousal represents vertical axis. In psychological terms, valence and arousal are associated with stimulus. Valence is an affective quality that refers to how an emotion is positive or negative, ranging from unpleasantness to happiness. On the other hand, arousal denotes the intensity of emotion. A hypothetical 2D valence-arousal model is illustrated in Figure 1.1.

1.3 Analysis of EEG Signal

Electroencephalography (EEG) is an efficient modality of medical imaging which helps to acquire the brain signal from different regions of the brain. EEG signal is obtained by placing electrodes on the surface of the brain. Human brain consists of millions of neurons which are playing a vital role to control the emotions, behaviour of human beings in response to different internal or external motor stimuli. Since the neurons act as information carriers, neural stimuli help to analyze the cognitive and mental states. Hence, the analysis of EEG signal in the context of emotion recognition has attained a widespread popularity among the researchers.

1.3.1 Source of EEG signal

Electroencephalogram (EEG) is a physiological method to record the potential difference between two electrodes placed at different cerebral locations of the brain. The EEG signal is associated with the neural activity where the neurons in the human cortex process the information by means of electrical signals. Large cortical pyramidal neurons in deep cortical layers play a pivotal role in the generation of EEG signal.

1.3.2 10-20 electrode system

The $10-20$ electrode placement system is an internationally recognized method that is used to describe the position and location of EEG electrodes. The $10-20$ system is based on the relationship between the location of the electrodes and the underlying area of the brain. In this system, the "10" and "20" refers to the actual distance between the adjacent electrodes which are either 10% or 20% of the total front-back or right-left distances of the skull.
In this measurement system, specific anatomical landmarks are used where the letters F, T, C, P, O denotes the frontal, temporal, central, parietal and occipital lobes respectively. "Fp" stands for "Front Polar". Even numbers (2, 4, 6, 8) refer to the right hemisphere and odd numbers (1, 3, 5, 7) refer to the left hemisphere of the brain. In addition to this, there are also (Z) sites that refers to the midline sagittal plane of the skull. The smaller the number, the closer its position to the midline.

1.3.3 Band Analysis of EEG Signal

It is crucial to investigate the frequency bands of the electroencephalogram in order to find the characteristics of the brain wave associated with different physical and mental states of the human beings. In this regard, different frequency sub-bands are prevalent in EEG signal analysis by defining different spectral thresholds. The brain signal in the order of lowest to highest frequency range is given as follows: delta(δ), theta(θ), alpha(α), beta(β) and gamma(γ).

Delta wave

Delta waves are mostly associated with deep level of sleep and relaxation. The frequency range of the delta wave is 0.5-3.5 Hz. EEG delta waves are high-amplitude brain waves and are mostly associated with deep sleep stages. It is the slowest recorded brain wave in human beings and most commonly found in young children. Along with the sleep and dreaming stages, delta bands are also prominent in cognitive processing especially in event related studies.

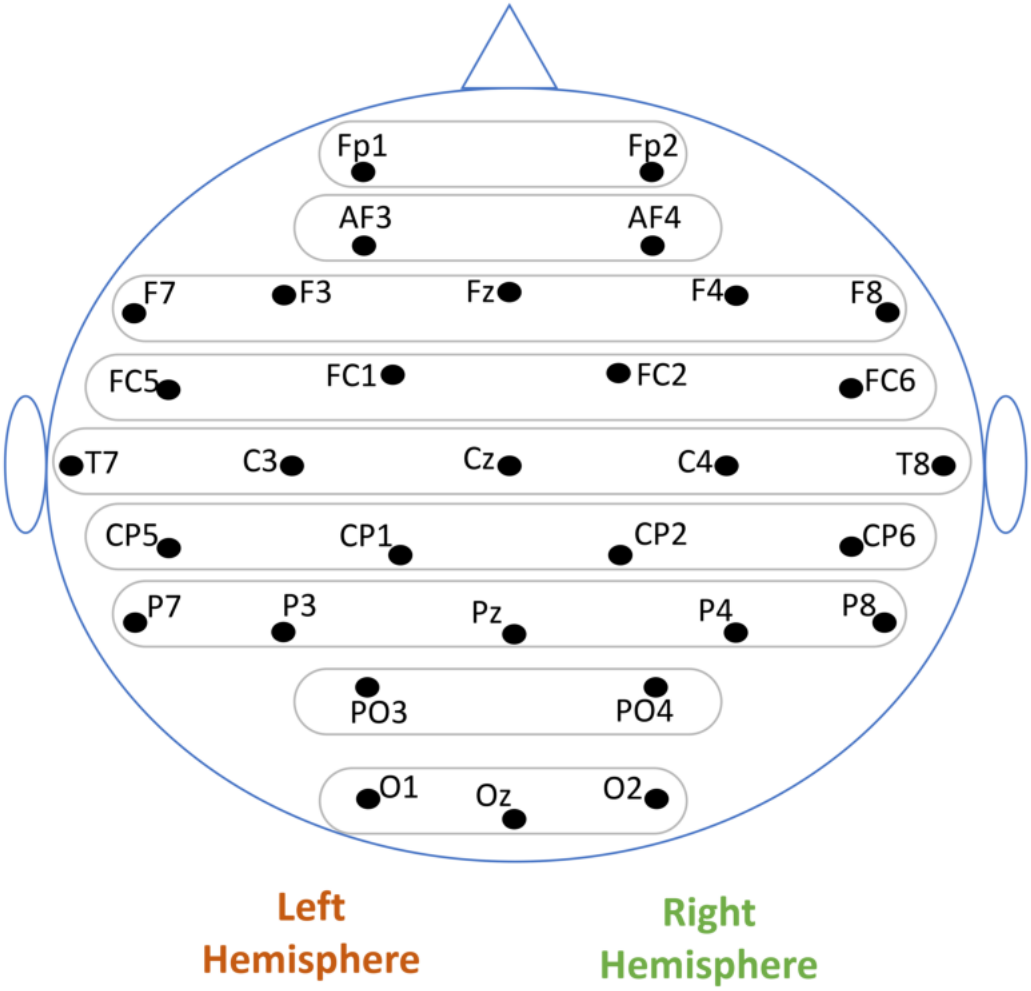


Figure 1.2: Electrode placement according to the 10-20 system

Theta wave

Theta oscillations are prevalent in various cortical structures, but mostly prominent in hippocampus. Since they are present when once in a trance or hypnotic state, these waves are also known as 'suggestible waves'. In humans, hippocampal θ rhythm is prevalent during active REM sleep. The frequency range of θ wave lies in between 3.5-7.5 Hz. In adults, theta bands are mostly observed during the transition from wakefulness to sleep. At the awaken states, θ-band power is associated with different cognitive functioning and memory related tasks.
High level theta waves are related with ADHD and impulsive activity. On the other hand, low level theta rhythm is associated with anxiety and higher stress levels.

Alpha wave

Alpha oscillation is predominantly recorded from the occipital lobes at fully awaken states or wakeful relaxation with closed eyes. The waves are reduced with drowsiness and sleep. The α rhythm plays a vital role in network coordination and communication. The frequency range of the alpha band most prominently lies in between 7.5-13 Hz. The α rhythm can reflect a lot of cognitive information of the human body. Alpha waves are known to be the 'frequency bridge' between our conscious and sub-conscious thinking.

Beta wave

Beta waves (β) are most commonly observed at our awaken state and involved in conscious thoughts and logical thinking. Prominence of beta wave results in anxiety, high arousal and stress. On the other hand, suppression of beta wave results in attention deficit hyperactivity disorder (ADHD), depression, daydreaming. In optimal condition, beta rhythm is associated with conscious thoughts, focus, memory. The frequency range of beta wave lies in between 13-30 Hz. The entire range of beta waves can be broadly classified into three groups namely low beta (13-15 Hz), mid-range beta(15-20 Hz), high beta (18-30 Hz) waves. Low beta waves are mostly associated with quiet and focused concentration. Mid-range beta waves are responsible for anxiety and performance. High beta waves have significant relations with stress, anxiety, paranoia and high energy.

Gamma wave

Gamma waves (γ) are considered to be the fastest oscillation among all brain waves (> 30 Hz). High gamma activity is involved in attention, working memory and long-term memory process. Gamma waves are found to be responsible for psychiatric disorder, hallucination and epilepsy. Along with that, the rhythm acts as a binding tool to process complex information and cognitive functions. More recently, A strong link is found between meditation and gamma waves. Different band frequency waves are illustrated in Figure 1.3.

1.4 Literature Review

In the domain of affective computing, data of different modalities are used like speech signal, facial expression, electrocardiogram (ECG), electroencephalogram (EEG), electromyogram (EMG), and galvanic skin response [6–9]. Considering the fact that neural stimulation drives all physiological activities of a human body, among different signals, EEG signal is widely used to detect emotion [10, 11], which is obtained by placing electrodes on the scalp.

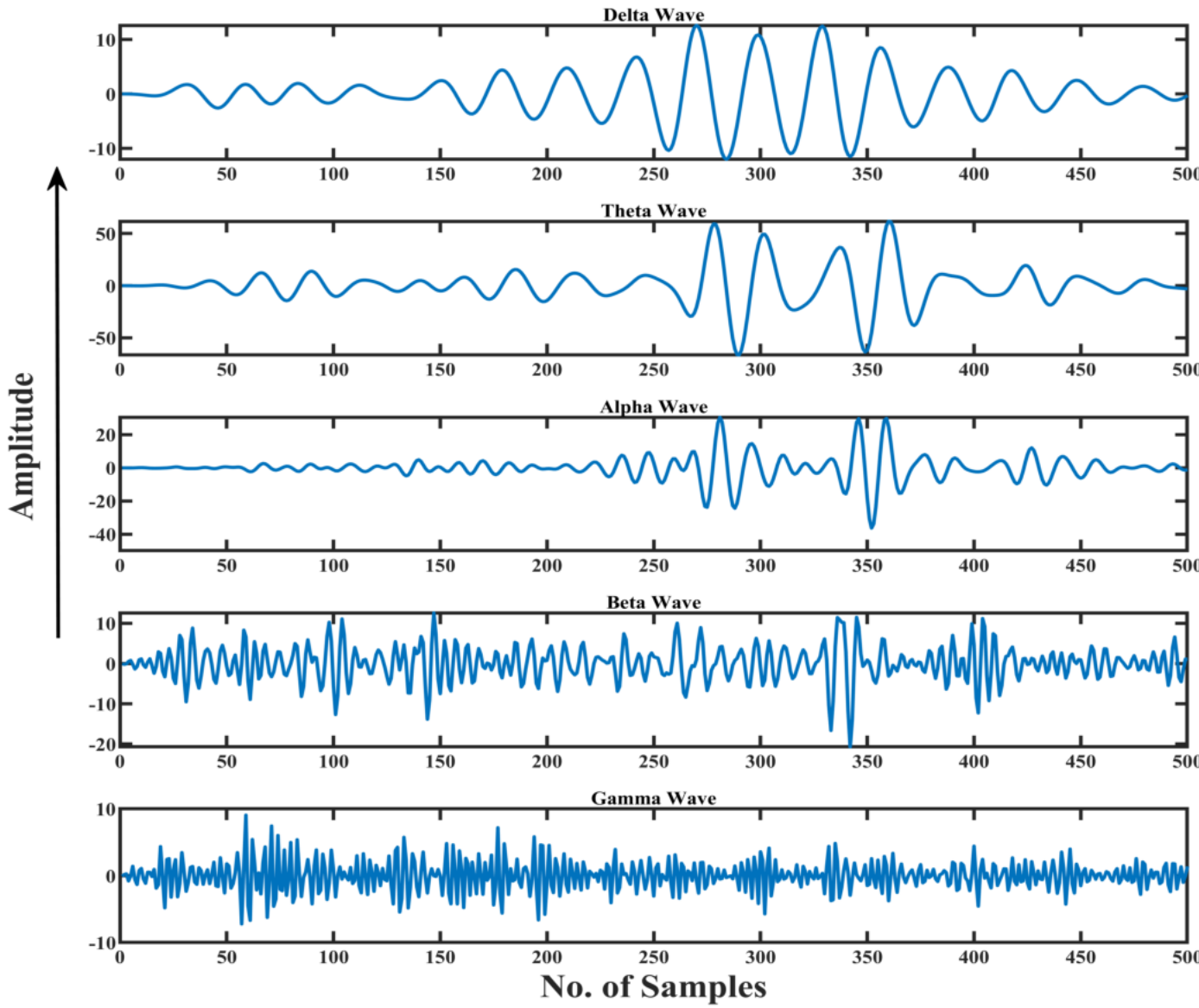


Figure 1.3: Different frequency bands of the EEG signal

Traditional machine learning-based emotion recognition methods extract hand-crafted features from the given EEG signal in time, frequency or time-frequency domain, and then utilize the features in different supervised classifiers [12–15].

In [12], different time-domain features, such as short-time energy, activity, mobility and complexity are used in support vector machine (SVM) classifier to classify emotion. In [13], discrete wavelet transform (DWT) coefficients are utilized as features in SVM classifiers. Different decomposition techniques are also applied to EEG signals prior to feature extraction. In their work, 10-channel EEG signal is used and higher frequency band (γ band) is reported to obtain better classification accuracy compared to lower frequency bands.

In [14] and [15], EEG signal is decomposed into intrinsic mode functions (IMFs) employing empirical mode decomposition (EMD), and different features are extracted from the IMFs. In [14], multivariate extension of EMD is proposed for feature extraction purposes and power ratio, power spectral density, entropy, Hjorth parameters and correlation are extracted from multichannel IMFS. In [15], features are extracted from the second-order difference plot of the IMFs. The extracted features are utilized to classify emotion with a support vector machine (SVM) and 2-hidden layer multi-layer perceptron (MLP).

In [16], an experiment is conducted on a self recorded dataset and features are extracted in time-domain, frequency-domain and time-frequency domain. The acquired features are compared employing different machine learning techniques and significant features are selected for the purpose of emotion recognition.

In [17], linear-frequency cepstral co-efficients (LFCC) feature are extracted from raw EEG signal using pre-trained ResNets and k-nearest neighbor (KNN) classifier is employed to classify emotion. In their study, two channels namely FP1 and C4 are selected based on the largest average sample entropy information and features obtained from two channels and ResNet-50s are fused to evaluate the performance of emotion recognition.

In [18], multi-channel information of EEG is exploited by forming 2D frame sequences. The frame sequences are considered to capture the spatial position relationship among different channels. In their work, a classification model is constructed using the deep forest and the constructed frames are fed as inputs to the model.

In recent years, different deep learning-based frameworks have been proposed for emotion recognition, where the EEG signal or extracted features from it are used as the input to the neural network.

In [19], EEG signal is divided into five frequency sub-bands using dual tree complex wavelet transform (DT-CWT) and features are extracted from time, frequency and non-linear analysis. The extracted features from the band-limited signals are used in a simple recurrent unit network to classify emotions. In their work, different ensembling strategies are used to improve the performance of emotion recognition.

The method proposed in [20] utilizes differential entropy feature and deep belief networks to categorize emotion. Four different profiles of 4, 6, 9 and 12 channels are selected in their study

to comare the performance with the original 62 channels. The computational time and training process of this technique are lengthy enough to apply for practical applications.
In [21], the spectral power is extracted from the frequency bands (δ, θ, α, β, γ) of the raw EEG signal and used as an input to a 1D deep neural network. Here an augmentation technique is employed to address the class imbalance problem, which strongly dictates the classification performance.
In [22], different conventional pre-trained deep learning networks (AlexNet, VGG16, ResNet50, SqueezeNet, MobileNetV2) are used to extract features from continuous wavelet transform (CWT) based scalogram images and then SVM classifier is applied to classify the emotion.
In [23], a long short-term memory (LSTM) based graph convolutional neural network (GCNN) is proposed, and differential entropy feature is extracted to classify emotion. Multiple GCNNs are constructed parallelly to extract graph domain information from a feature cube and LSTM cells are utilized to memorize the change of relationship between two channels.
The method proposed in [24] incorporates a stacked bi-directional Long Short-Term Memory (Bi-LSTM) network to classify emotion. In their work, statistical features, wavelet features and Hurst exponent are extracted from EEG data where the feature selection process is performed by the Binary Gray Wolf Optimizer.
In [25], Pearson correlation coefficient (PCC) featured images are generated, and channel correlation of EEG sub-bands namely α, β and γ are used to classify emotion with a deep neural network. The dimensions of the PCC featured images are reduced to lower the computational complexity involved in the process.
In [26], an LSTM network with channel attention autoencoder is proposed in the context of emotion recognition. The attention network is described to highlight the segments of the EEG signal that will contribute to the emotion recognition.
It is to be noted that most of the methods available in the literature consider EEG signals obtained from all the channels. However, in the case of emotion recognition, not all the channels necessarily contribute equally [27]. Hence, a deep neural network-based scheme utilizing an efficient feature which not only offers a better classification performance but also helps in reducing the number of channels is still in great demand.

1.5 Dataset Used

In this paper, a publicly avaiable widely used DEAP dataset [28] is chosen for demonstrating the experimental results. The dataset contains 32 channel EEG signals as well as other physiological signals of 8 channels (EOG, EMG, GSR) of 32 participants while watching 40 one-minute-long excerpts of music videos. After watching, the participants rated each video in terms of the labels of valence, arousal, dominance and liking using self-assessment mannequins (SAM) on a 9-point scale. In our experiment, we consider valence and arousal dimensions as

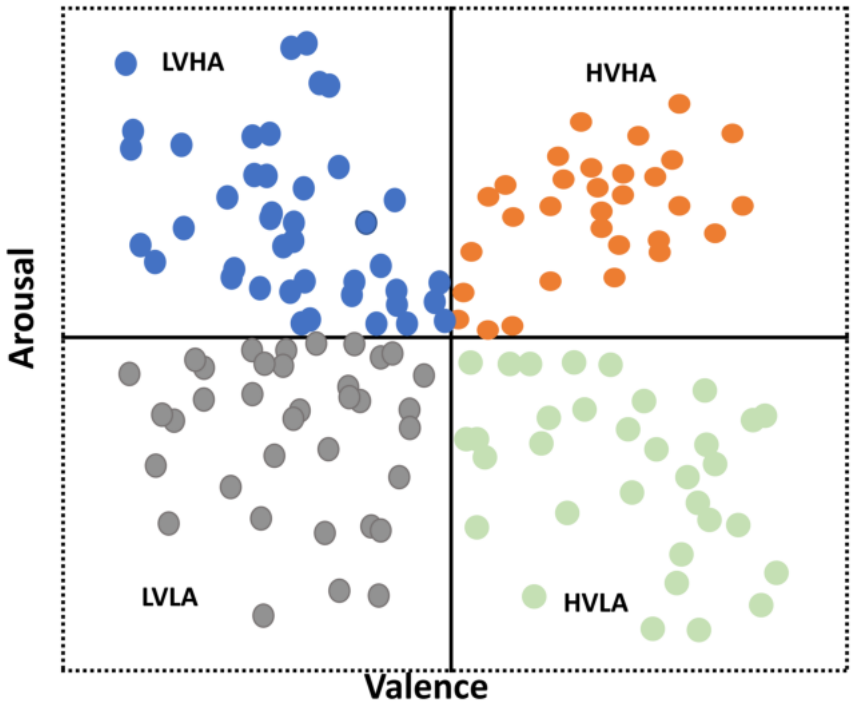


Figure 1.4: Distribution of valence and arousal scores of DEAP dataset

emotion evaluation criteria.
The original signal was recorded at a frequency of 512 Hz. In the pre-processing stage of the dataset preparation, the data were downsampled to 128 Hz and a bandpass filter with $4-45.0$ Hz frequency range was applied after removing the electrooculogram (EOG) artefacts of the EEG signal. Each subject file in DEAP dataset contains two arrays and the description is summarized in Table 1.1

Table 1.1: Summary of DEAP dataset

Array name	Array shape	Array content
Data	$40 \times 40 \times 8064$	video/trial $\times$ channel $\times$ data
Labels	40×4	video/trial $\times$ label (valence, arousal, dominance , liking)

1.6 Objectives and Scope

The objectives of the thesis are:

1. To develop different deep neural architectures and analyze the effects of attention mechanism, sequence learning in the context of emotion recognition.

2. To develop sub-frame based approach to analyze the effects of frequency domain or the extracted power in frequency domain.

3. To investigate the performance of the emotion recognition scheme in time-frequency domain by

4. To select significant channels in the context of emotion recognition.

5. To validate the effectiveness of the proposed approach using a publicly available emotion (DEAP) dataset.

1.7 Organization of the Thesis

In the first chapter, the basic theories of emotions and a two-dimensional model to describe different emotional states are described. The reason behind the analysis of EEG signal for the purpose of emotion recognition is also analyzed in the chapter. Moreover, the chapter provides the motivation and objectives of the thesis by presenting the past and present researches on emotion recognition. Furthermore, the importance of selecting the EEG channels in the context of emotion recognition is also discussed in the chapter. The rest of the thesis is organized as follows.

In chapter 2, different deep neural architectures are proposed for the purpose of emotion recognition from multi-channel raw EEG data. The baseline of the EEG data is removed to obtain more relevant information in response to different audio-visual stimuli. The baseline excluded raw signals are divided into several non-overlapping sub-frames and each sub-frame is applied to deep neural architectures to categorize different classes of emotions. Along with CNN-based deep neural architecture, long short term memory (LSTM), bidirectional long short term memory (Bi-LSTM) based networks are also discussed to analyze the temporal dependencies of the EEG signal. The significant effects of attention mechanism i.e, channel-wise attention and multi-head attention are also explored in the chapter. Detailed analysis and experimentations are carried out on a publicly available DEAP dataset.

In chapter 3, the effects of different frequency bands and the spectral power are investigated for the purpose of emotion recognition. The EEG signals are decomposed into several sub-bands and the temporal and frequency band information of all channels are considered in 3D frame. The obtained frame is inputted to a deep neural network to classify emotion. In case of spectral power-based analysis, the decomposed sub-band are divided into non overlapping sub-frames and spectral power is extracted from each window section. The feature vectors obtained from all frequency band signals are combined to a final 1D feature vector which is applied to a 1D CNN for emotion classification. Detail experimental result are presented for the same dataset.

In chapter 4, EEG signals are analyzed in time-frequency domain to classify emotions. Like

chapter 2 and 3, all transformations are applied to the baseline excluded raw EEG signal. The transformed signal frame in DWT domain or the extracted strength-to-entropy component ratio (SECR) feature from the window segment in CWT domain are applied to a deep neural network for the purpose of emotion classification. Detail experimentation are carried out considering the same dataset.

In chapter 5, a thorough analysis is performed on channel and scale selection process. In order to reduce the data storage and computational complexity associated with the CWT process, an efficient scheme is designed in CWT domain where some significant channels and a range of scales are selected based on higher energy-to-entropy ratio (EER) value. The selected channels and scales exhibit superior performance compared to other channels and range of scales.

Chapter 6 summarizes the outcomes of the thesis with some concluding remarks and possible future works.

Chapter 2

Emotion Recognition with Different Neural Architectures Using Raw EEG Signal

In this chapter, an automatic approach of emotion recognition is proposed with deep neural architecture utilizing multi-channel raw EEG data. Since EEG data contain rich spatial and temporal information, neural network maps significant feature exploiting the details from the signal. Firstly, after taking the baseline excluded raw EEG trial data, the full signal is windowed to maintain the stationarity constraint and a 2D matrix is constructed combining the information of all available channels. Finally, the 2D frame is applied to a deep neural network to categorize different classes of emotions. In order to extract the inter-channel relationship among different EEG channels, an efficient scheme of grouping of EEG channels is designed in the study.

In the first section, the test EEG trial is directly fed into a CNN-based neural architecture to extract an efficient feature vector that is subsequently applied to a dense classifier to perform the classification task. In addition to this, the contribution of channel-wise attention block is also observed for the purpose of emotion recognition.
In the subsequent part of the chapter, a long short-term memory (LSTM) based deep neural network is introduced to capture the temporal information of the EEG signal and exploit the details to classify emotions. Along with this, the effects of both channel-wise attention and multi-head self-attention blocks with the LSTM based architecture are also explored in the section.

Figure 2.1: Proposed CNN-based methodology

2.1 CNN-Based Deep Neural Network

2.1.1 Proposed Method

The significant steps involved in the proposed method are presented in Figure 2.1, which include grouping of EEG channels, pre-processing and windowing, CNN-based deep feature extraction with channel-wise attention mechanism and classification with the dense classifier. Firstly, after grouping the EEG channels in an orderly manner, the baseline of the raw EEG signal is removed. In order to maintain the stationarity constraint and increase the total number of EEG trials, the baseline excluded raw data are windowed with a proper frame length. Subsequently, channel-wise attention is applied to each sorted EEG channel to assign a specific amount of weight and then each EEG trial is employed as the input to a CNN-based neural architecture to extract deep features of the EEG signal. The extracted deep features are then fed into a dense classifier to categorize different classes of emotions. Since the features inherit characteristic information of the EEG signal, they are effective in representing the varying neural activities caused by different emotional states. In order to highlight the inter-channel relationship among different EEG channels, a thorough analysis is performed considering the spatial positions of the channels on the scalp and an efficient grouping scheme is designed. In the classification stage, both valence and arousal dimensions are classified into binary classes (high and low). In what follows, the steps involved in the proposed method are illustrated in detail.

2.1.2 Grouping of EEG Channels

The spatial location of the scalp electrodes plays a key role in the formation of EEG channels. Different regions of the brain (frontal, parietal, temporal, occipital) tend to be informative at the event of a particular activity. In this context, channel-based analysis is widely popular in the study of EEG signal [29]. In view of incorporating the channel information with respect to elicition of different emotions, an efficient grouping scheme of EEG channels is presented in the proposed study.

In the proposed method, the raw EEG signal from each channel is mapped based on the spatial information of the EEG electrodes on the scalp. In this regard, starting from the frontal region

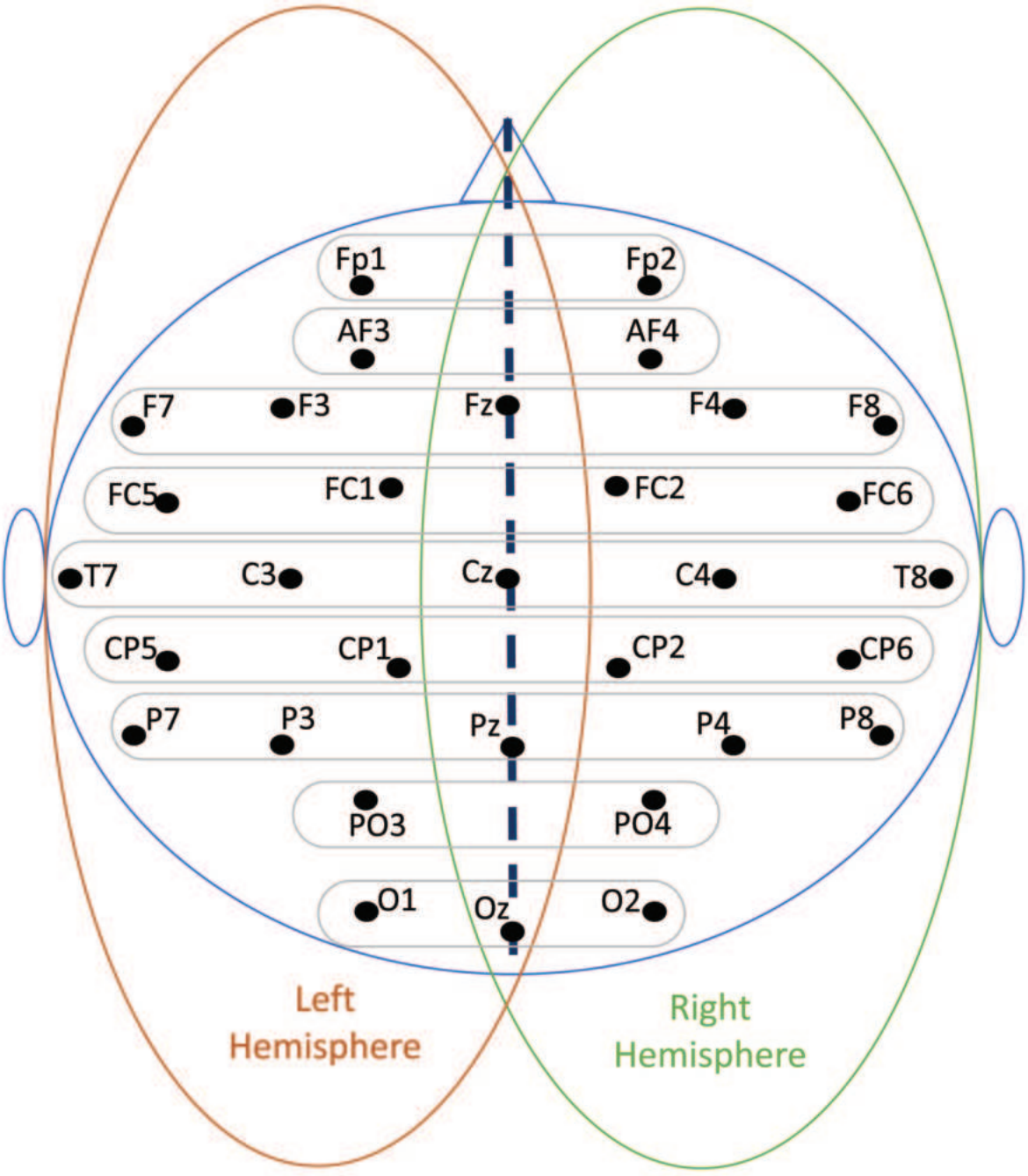


Figure 2.2: The location of left, right and midline hemispheric EEG channels by considering a mirror in the midline

to the occipital region of the hemisphere, the available left, right and midline hemispheric EEG channels are symmetrically organized considering a mirror in the midline of the hemisphere, which is illustrated in Figure 2.2. The channels on the left and right hemispheres are separated by considering the mirror method and those on the midline are included in both hemispheres. As a result, the initial channels are arranged using spatial information on the scalp.The organized channels name with the newly assigned channel number is shown in Table 2.1.

Table 2.1: Organized channels (using the proposed midline hemisphere mirror method)

Left hemispheric channels			**Right hemispheric channels**		
Channel Number		2*Channel	**Channel Number**		2*Channel
After	**Before**		**After**	**Before**	
01	01	FP1	**19**	17	FP2
02	02	AF3	**20**	18	AF4
03	19	Fz	**21**	19	Fz
04	03	F3	**22**	20	F4
05	04	F7	**23**	21	F8
06	06	FC1	**24**	23	FC2
07	05	FC5	**25**	22	FC6
08	24	Cz	**26**	24	Cz
09	07	C3	**27**	25	C4
10	08	T7	**28**	26	T8
11	10	CP1	**29**	28	CP2
12	09	CP5	**30**	27	CP6
13	16	Pz	**31**	16	Pz
14	11	P3	**32**	29	P4
15	12	P7	**33**	30	P8
16	13	PO3	**34**	31	PO4
17	15	Oz	**35**	15	Oz
18	14	O1	**36**	32	O2

2.1.3 Pre-processing and Windowing

Pre-processing stage of the proposed method includes exclusion of baseline signal. Let $X_R = [X_B, X_E] \in \mathbf{R}^{M \times N}$ be the recorded EEG signals with F Hz sampling frequency, where M is the total number of EEG channels after employing the grouping mechanism and N is the length

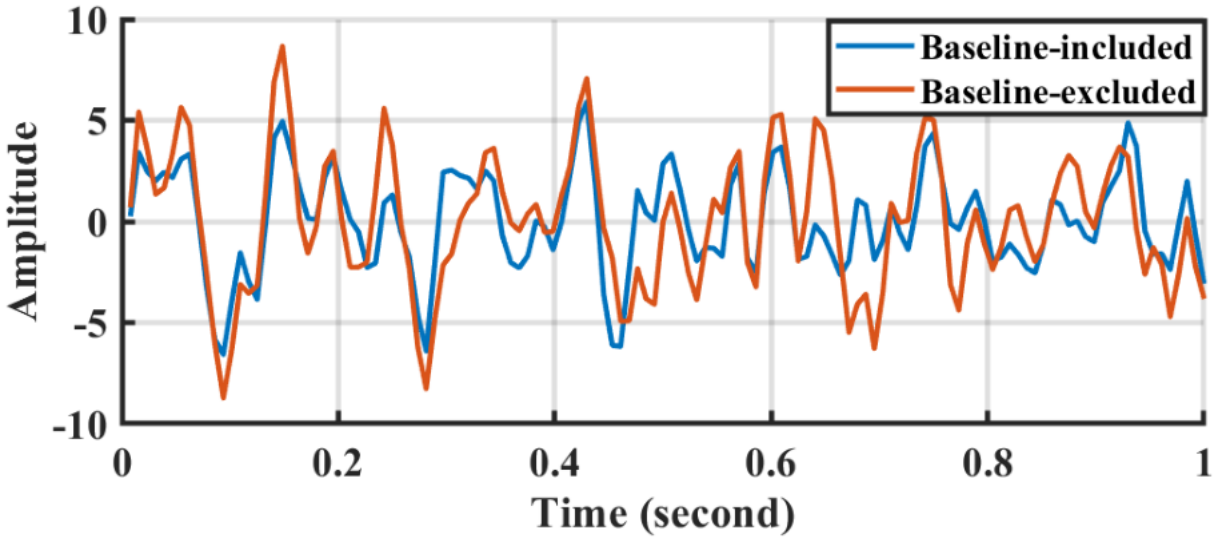


Figure 2.3: The baseline-included and the baseline-excluded raw EEG trial signal

of the EEG data. In addition, $X_B \in \mathbf{R}^{M\times L}$ indicates the baseline data where L is total number of sampling points of the baseline signal, $X_i(i = 1, 2, \ldots, \frac{L}{F}) \in \mathbf{R}^{M\times F}$ denotes the i-th second baseline signal. Let $\overline{X}_B \in \mathbf{R}^{M\times F}$ denotes the mean value of the baseline signal per-second which can be calculated as

$$\overline{X}_B = \frac{F}{L}\sum_{i=1}^{\frac{L}{F}} X_i. \tag{2.1}$$

Moreover, $X_E \in \mathbf{R}^{M\times J}$ refers to the baseline included raw EEG signal, where J denotes its length. In order to remove the baseline signal, X_E is segmented into several slices $X_j(j = 1, 2, \ldots, \frac{J}{F}) \in \mathbf{R}^{M\times F}$ with a one-second non-overlapping sliding window, and the baseline excluded segments of the raw EEG signal can be extracted as

$$X'_j = X_j - \overline{X}_B. \tag{2.2}$$

Following the extraction of the baseline-excluded one-second non-overlapping slices of the raw EEG signal, they are concatenated into a new matrix X_T, that follows the same shape as X_E. The one-second slice of the raw EEG signal with and without the baseline part is displayed in Figure 2.3.

In order to increase the total number of trials to perform the task of emotion recognition, the given baseline excluded EEG signal (X_T) is segmented into several small frames using a proper non-overlapping window. The trial signal $X_W = [\mathbf{X}_W^{1\,T}, \mathbf{X}_W^{2\,T}, \ldots, \mathbf{X}_W^{M\,T}]^T$ is obtained from X_T, where $\mathbf{X}_W^i(i = 1, 2, \ldots, M) \in \mathbf{R}^{1\times W}$ represents the trial signal at the i-th EEG channel and X_W is assigned the same emotion label as X_T. For the non-overlapping window, the total number of trials obtained from a given X_T will be $\frac{J}{W}$, where W denotes the window length.

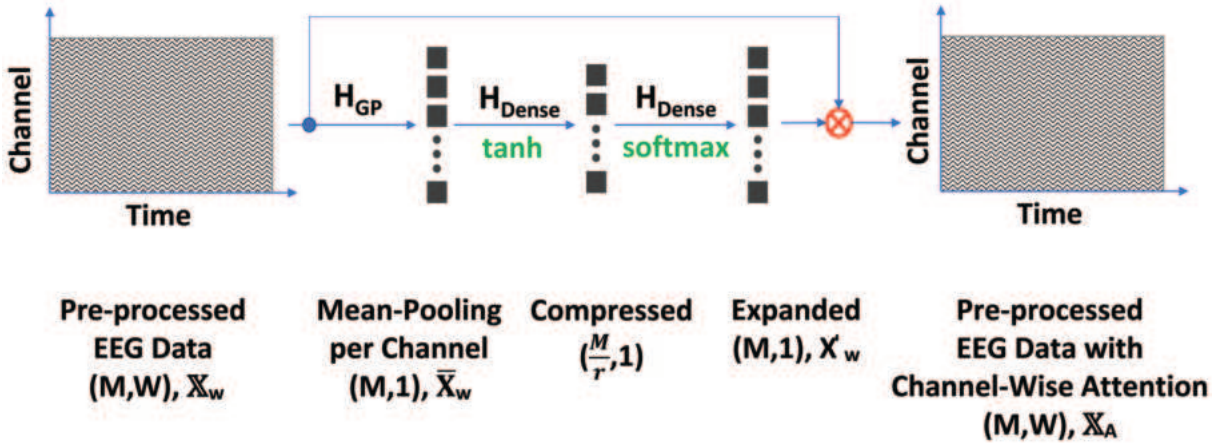


Figure 2.4: The proposed channel-wise attention mechanism

2.1.4 Channel-Wise Attention (CA)

The human attention mechanism can utilize a sequence of partial glimpses and selectively focus on the main parts to better capture a visual structure. Inspired by this nature, the spatial attention mechanism is employed to distribute the significance of EEG channels. Channel-wise attention can extract more particular information about channels with the change of the weights of different channels by exploring the information of the feature map. Therefore, the channel-wise attention mechanism can be used to exploit inter-dependencies among feature channels. In addition, it is trainable with CNNs, and it can be integrated into CNN architectures to explore the importance of the channels of EEG signals. As a result, a CNN can extract more discriminative spatial information because multichannel EEG signals contain spatial information via channels.

In order to extract the discriminative spatial information via channels, the proposed method comprises the channel-wise attention mechanism. The structure of the proposed channel-wise attention mechanism is shown in Figure 2.1.4, where the adaptive attention mechanism in a channel-wise manner in the EEG signals is implemented. The different channels may contain redundant or less relevant information and so the significance needs to be distributed artificially among the different channels of multichannel EEG signals. Consequently, the information of all channels can be considered and assigned weights adaptively to different channels based on their importance. In this framework, $X_W = [\mathbf{X}_W^{1\ T}, \mathbf{X}_W^{2\ T}, \ldots, \mathbf{X}_W^{M\ T}]^T$ represents EEG trial signal and $\mathbf{X}_W^i (j = 1, 2, \ldots, M)$ denotes the i-th channel of EEG trial, and M is the total number of channels of each trial. Firstly, mean pooling for each channel of the EEG sample is applied to obtain channel-wise statistics, which can be shown as $\overline{\mathbf{X}}_W = [\overline{X}_W^1, \overline{X}_W^2, \ldots, \overline{X}_W^M]$, and where $\overline{X}_W^i (i = 1, 2, \ldots, M)$ is the mean of the i-th channel. In this proposed mechanism, the channel-wise attention block adopts two fully-connected (FC) layers around the non-linearity, i.e., a dimensionality-reduction layer with parameter W_1 and bias terms $\mathbf{b}_1$ with reduction ratio r and tanh function as the activation function, and a dimensionality increasing layer with parameter W_2 and bias terms $\mathbf{b}_2$ and softmax function as the activation function. The FC layers

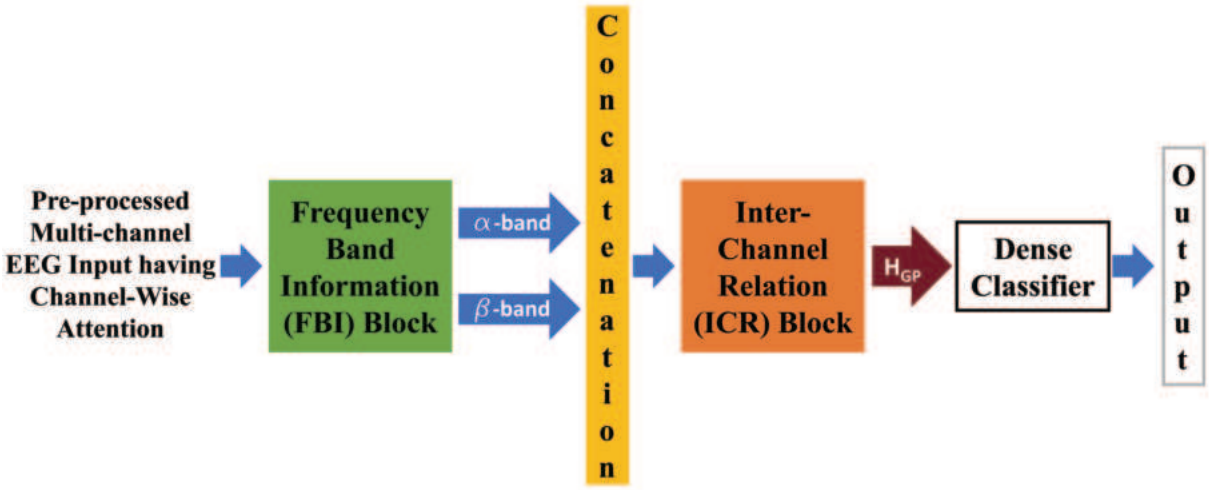


Figure 2.5: Workflow of the proposed CNN-based model with the dense classifier

of the channel-wise attention mechanism are expressed as

$$\mathbf{X}'_W = softmax(W_2 \cdot (tanh(W_1 \cdot \overline{\mathbf{X}}_W + \mathbf{b}_1) + \mathbf{b}_2), \tag{2.3}$$

where the softmax function transforms the importance of channels to probability distribution $\mathbf{X}'_W = [p_1, p_2, \ldots, p_M]$, which represents the significance of different channels. Finally, the probability is considered as the weight to recode the information of the trial EEG signal in each channel and the i-th $(i = 1, 2, \ldots, M)$ attentive channel feature is extracted. Consequently, $X_A = [\mathbf{X}_A^{1\,T}, \mathbf{X}_A^{2\,T}, \ldots, \mathbf{X}_A^{M\,T}]^T$ represents the trial EEG signal with a channel-wise attention, where $\mathbf{X}_A^i = \mathbf{X}_A^i \cdot p_i, (i = 1, 2, \ldots, M)$.

2.1.5 CNN-Based Deep Feature Extraction and Binary Classification Using the Dense Classifier

The raw EEG signal in the form of a 2D time-series data, contains the spatial and information of all channels. The brain activity changes from time to time and these temporal features can be reflected in the time dimension, whereas the activation patterns of the neurons across different functional areas due to the different locations of the brain can be extracted via the spatial dimension. Hence, in the proposed CNN-based model to extract more extensive and informative features in the EEG signal's temporal and spatial dimensions, a frequency band information (FBI) block and an inter-channel relation (ICR) block are used respectively. The proposed CNN-based model with the dense classifier is depicted in Figure 2.5.

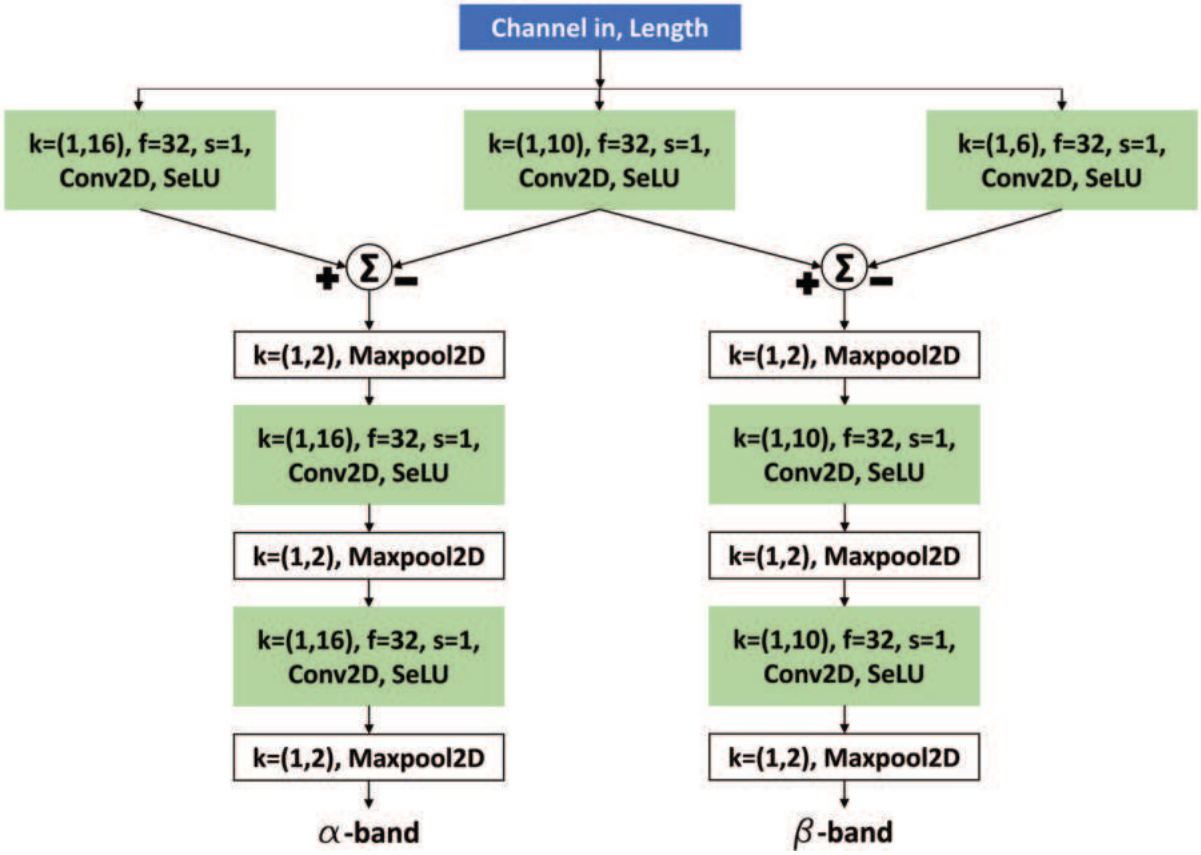


Figure 2.6: Architecture of the proposed Frequency Band Information (FBI) block

Frequency Band Information (FBI) Block

The time dimension contains five frequency bands which are delta ($0.5 - 3.5$ Hz), theta ($3.5 - 7.5$ Hz), alpha ($7.5 - 13$ Hz), beta ($13 - 30$ Hz) and gamma ($30 - 50$ Hz). According to some studies, δ and θ waves are mostly associated with deep sleep, relaxation and they are less relevant to the cognitive tasks of the human brain [30]. On the other hand, α, β and γ waves are dominant in the cases of information processing, conscious thoughts, learning and emotional tasks. As γ wave contains high-frequency bands, they are involved in the processing of multi-dimensional complex tasks [31]. Hence, the α and β waves are expected to capture different emotional states better than the γ wave. As a result, in the proposed method, the α and the β band signals are considered to extract the temporal information from the raw EEG signal. Furthermore, in the proposed CNN architecture, the FBI block is considered to pull out the dynamic temporal information where it consists of three multi-scale temporal kernels (FBI kernels), which is illustrated in Figure 2.6.

In order to extract the dynamic temporal representations in the α and the β band, we set the length of the temporal kernels as the specific ratios of sampling rate F of EEG. These ratios are defined as $a_i \in \mathbf{R}$, where the ratio coefficients a_i will become $[\frac{16}{128}, \frac{10}{128}, \frac{6}{128}]$, capturing frequency at 8 Hz to above, 12.8 Hz to above and 21.3 Hz to above as well. The 12.8 Hz to the above signal and the 21.3 Hz to the above signals are then subtracted from 8 Hz to the above

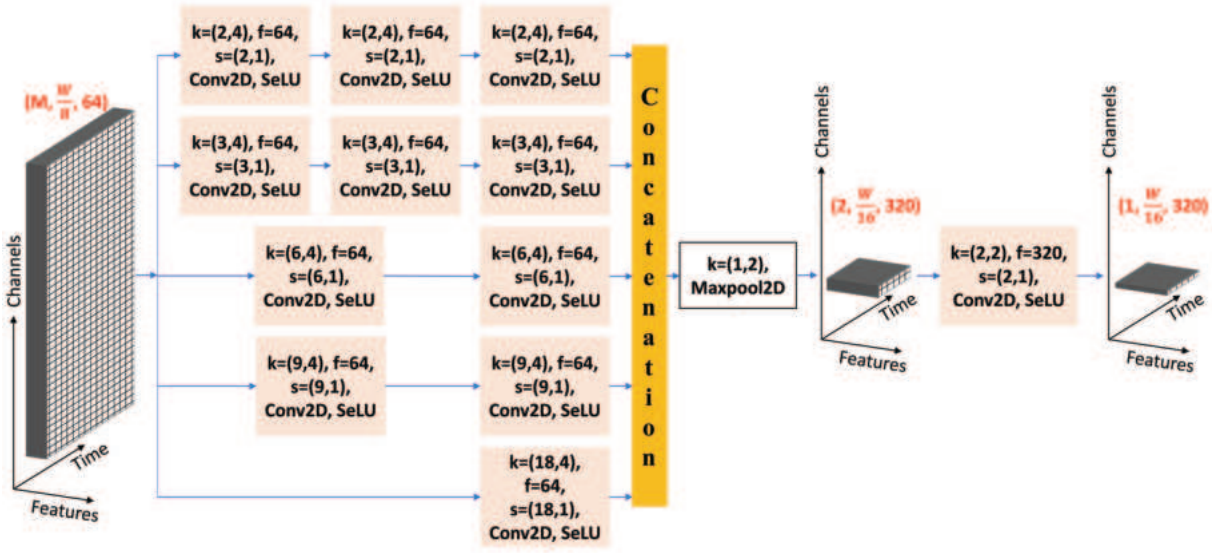


Figure 2.7: Architecture of the proposed Inter-Channel Relation (ICR) block

signal and 12.8 Hz to the above signal respectively. As a result, the α and the β band signals are approximated to the frequency band from 8 Hz to 12.8 Hz and from 12.8 Hz to 21.3 Hz respectively. After extracting the approximated α and the β band signals using FBI kernels, the signals are downsampled to reduce the number of samples in the signal and similar lengths of FBI kernels are used to extract different temporal features from these frequency bands. As downsampling increases the sampling frequency and the frequency bands using the same rate, the length of the FBI kernel is constant for each downsampling stage. In the proposed method, in total three stages of downsampling operation are performed with a downsampling rate of 2 for each stage. As a result, the length of the α and the β band signals are reduced by 8 times than the initial. Then the temporal features from both frequency bands are concatenated in the feature dimension to explore the interrelationship between the frequency bands together.

Inter-Channel Relation (ICR) Block

In this work, the ICR Block consists of five parallel ICR branches. Moreover, each ICR branch has multi-scale convolutional kernels, namely hemisphere kernels whose sizes are related to the location of the EEG channels and have multiple stages of extracting the informative features from different combinations of the channels. As a result, the number of stages for all five ICR branches is not necessarily equal to having the spatial features of both left and right hemispheres individually in the spatial dimension for each ICR branch. For the purpose of extracting the global relationship between the left and the right hemispheric channels, the individual spatial features from all five ICR branches are concatenated in the feature dimension and a global kernel is used to extract the global spatial feature. The demonstration of the ICR block is depicted in Figure 2.7.

Finally, a global average pooling is performed in the temporal dimension to extract an efficient

CNN-based feature vector that goes through a dense classifier to classify the valence and arousal types of emotions.

2.1.6 Results and Discussion

Experimental Setup

In this subsection, to validate the effectiveness of the proposed scheme, the performance analysis of an extensive experiment on the DEAP database is presented. From the 63 seconds of recorded EEG data (X_R), data from the first three seconds (X_B) is used as the baseline. The remaining one-minute-long EEG data (X_E) is considered as the baseline included raw EEG signal for the experimentation purpose. In this work, a 2 second non-overlapping window frame is applied to the preprocessed signal to increase the total number of EEG trials. The proposed method is designed to classify emotions for valence and arousal dimensions and valence and arousal labels are categorized into binary classes in the proposed work. Depending on participants' ratings, for binary class, rating < 5 corresponds to low class and otherwise high class. The training and validation of the proposed method are performed in the Google Colaboratory Platform.

Model Training, Validation and Testing

The proposed method is trained and tested on 32 subjects individually, as in this study, a subject-dependent experiment is conducted. The data are randomly split into 90% training and 10% testing. In order to validate the model, a 10-fold cross-validation scheme is employed. Furthermore, the network runs for 100 epochs per fold in the training stage. The learning rate is set to 0.0001 with $Adam$ optimizer and batch size is set to 64. The loss function is categorical cross-entropy and the metrics are accuracies.

Performance Evaluation

The performance analysis of the proposed method is assessed with following performance metrics

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{2.4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{2.5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2.6}$$

$$\text{F-1 score} = 2.\frac{\text{Precision.Recall}}{\text{Precision} + \text{Recall}}, \tag{2.7}$$

where TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative. The results for binary classification of both valence and arousal dimensions for each subject are recorded in Table 2.2. The average accuracy and F-1 scores are consistently very high for each subject (with an average greater than 94%). Furthermore, the resultant standard deviation is lower, indicating consistent performance among different subjects. The result demonstrates the efficacy of the proposed method.

Effect of Channel-Wise Attention

In order to validate the efficacy of the channel-wise attention, the performance of the network is also observed without applying the attention mechanism. The classification performances (accuracy and f1-score) are displayed in Figure 2.8 respectively. It is observed that the proposed CNN-based architecture with channel-wise attention exhibits higher accuracy and f1-score compared to that without attention mechanism.

2.2 Bidirectional LSTM-Based (BiLSTM-Based) Deep Neural Network

In this part, a bidirecional LSTM (BiLSTM) based deep neural network is proposed to investigate the effects of sequence learning of EEG data on emotion classification. Along with that, a channel-wise attention mechanism is employed to learn the varying contribution of the EEG channels on emotion recognition. The major steps included in this section are grouping of EEG channels, pre-processing and windowing, BiLSTM-based deep feature extraction and multi-head attention mechanism. The structure of the proposed BiLSTM-based method with attention mechanisms is displayed in Figure 2.9. The multi-head attention layer combines knowledge from different EEG trials and sets more weight to the significant trial. Similar to the previous section, the baseline is removed from the raw EEG signal after organizing the EEG channels and then windowing operation is performed to increase the number of EEG trials. Next, each EEG trial is fed into a bidirectional LSTM-based deep neural network. Subsequently, the extracted feature vector is applied to the multi-head attention layer to obtain discriminative temporal information by exploring each EEG trial's relative importance. Finally, a dense classifier is applied to classify different types of emotions. The performance of the proposed scheme with and without attention mechanism (both channel-wise and multi-head attention) is also analyzed in this section.

As the grouping of EEG channels, the pre-processing and windowing and the channel-wise attention mechanism are described in the subsection 2.1.2, 2.1.3 and 2.1.4 respectively, the

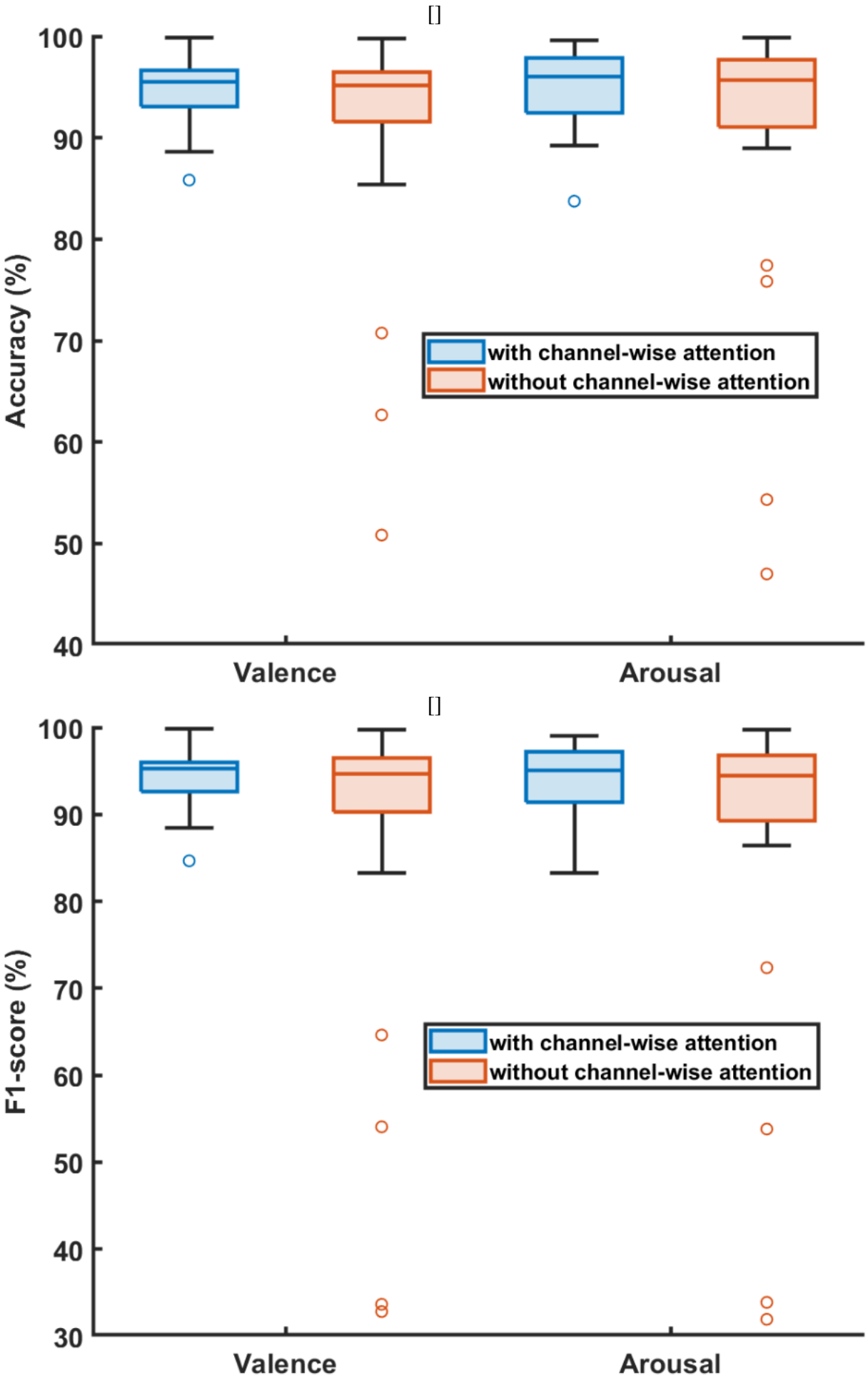


Figure 2.8: Box plot for performance comparison (accuracy and f1-score) of the proposed scheme with and without channel-wise attention mechanism

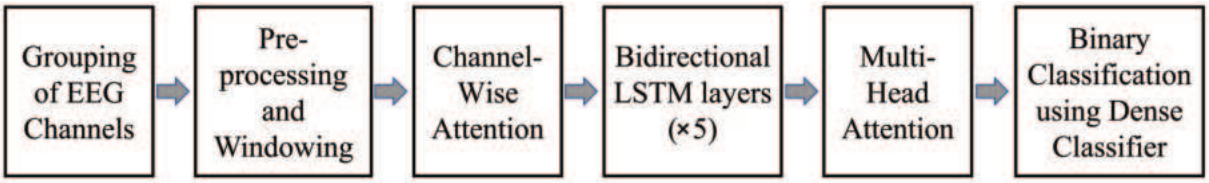


Figure 2.9: Workflow of the proposed BiLSTM-based methodology

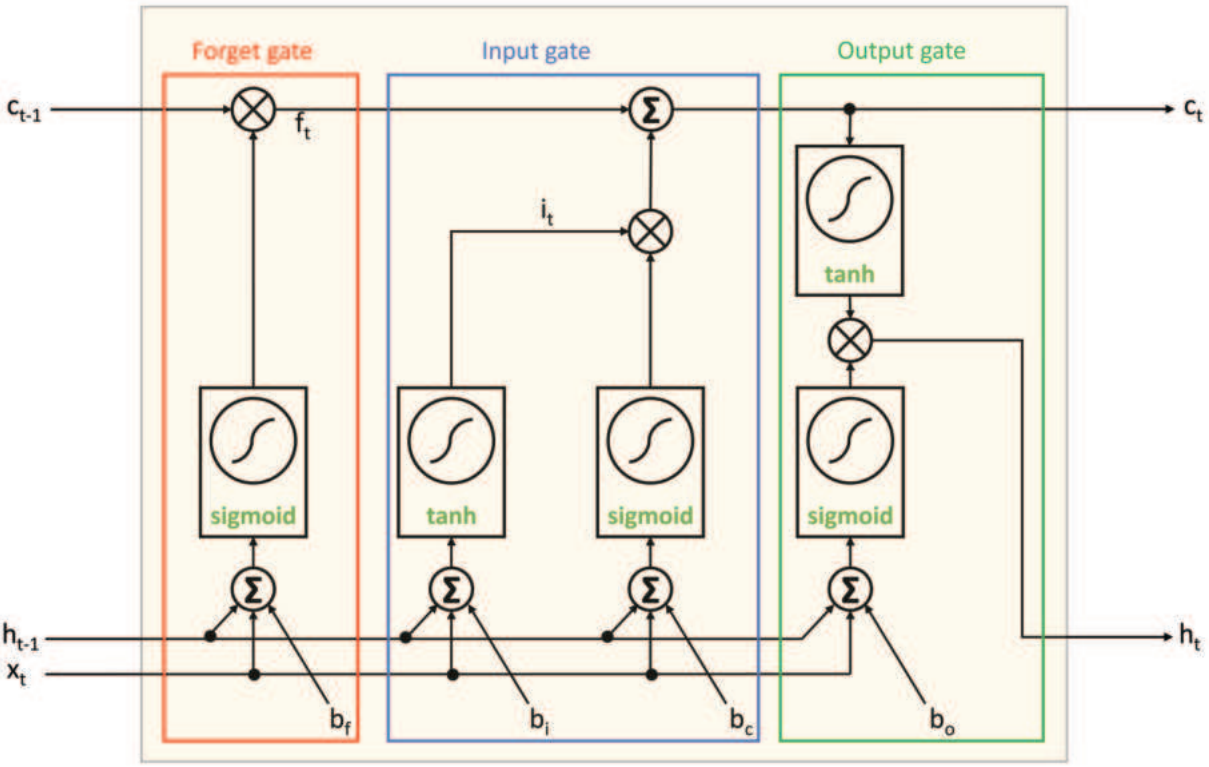


Figure 2.10: Architecture of LSTM unit

BiLSTM-based deep feature extraction and the multi-head attention mechanism are explored in this section. In what follows, the steps are explained in detail.

2.2.1 BiLSTM-Based Deep Feature Extraction

The BiLSTM-based deep neural architecture can learn the temporal information of the EEG signal due to the recurrent structure of the network. Hence, it can learn features from EEG data based on temporal dependence.

An LSTM layer is composed of recurrently connected memory blocks where each block contains one or more memory cells. Furthermore, each cell consists of three multiplicative gate units: the input, output, and forget gates, where they perform functions analogous to read, write, and reset operations. The multiplicative input gate units prevent the adverse effects created by the uncorrelated inputs. The input and output gate control a memory cell's inflow and outflow data stream to other LSTM blocks by using the sigmoid and tanh activation functions.

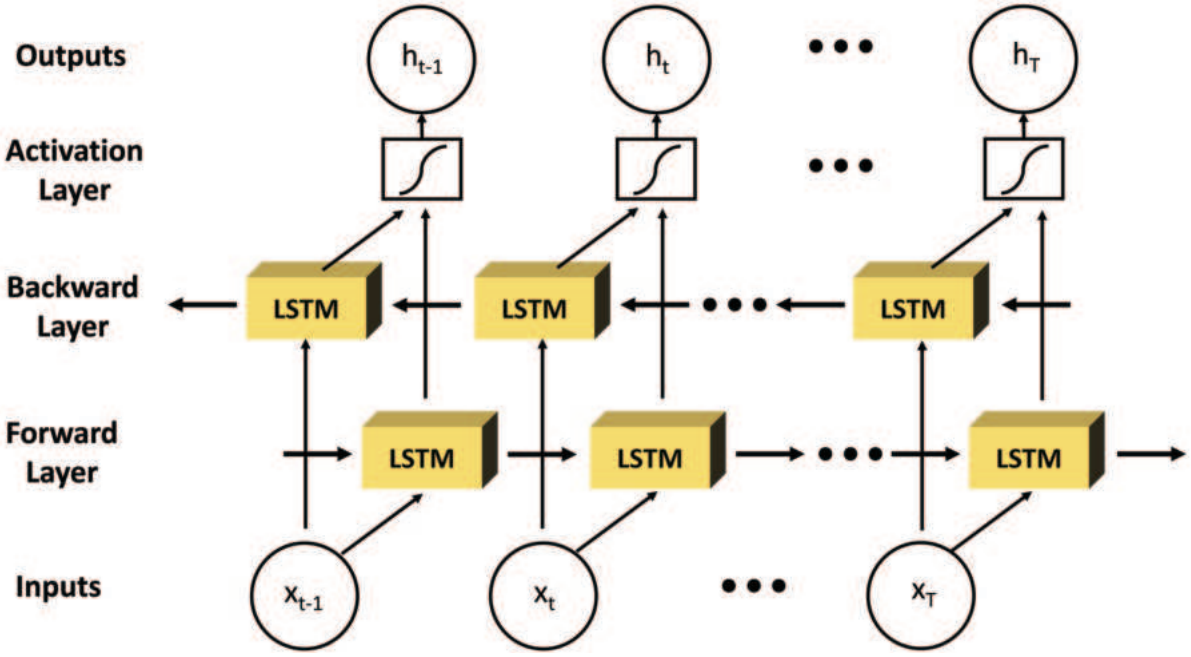


Figure 2.11: Basic structure of BiLSTM layer

In general, the cell input is multiplied by the activation of the input gate, the cell output by that of the output gate, and the previous cell values by the forget gate. As a result, the context information over long periods can be stored and retrieved by the network. However, the activation of the memory cell will not be overwritten by new inputs as long as the input gate remains closed and so it can be made available to the network for future sequences by opening the output gate. The internal structure of an LSTM memory block is illustrated in Figure 2.10.

An LSTM cell receives three inputs: input x_t at the current time t, memory C_{t-1} of previous time $t-1$, and h_{t-1} denotes the hidden state of the previous time $t-1$ and transmits two outputs: memory C_t at the current time t and hidden state h_t represented as the t-th temporal feature extracted from LSTM. The forget gate in the memory block structure is controlled by a simple one-layer neural network with an activation function which can be calculated as

$$f_t = \sigma(W[x_t, h_{t-1}, C_{t-1}] + b_f), \tag{2.8}$$

where W denotes separate weight vectors for each input, b_f is the bias vector and σ is the logistic sigmoid function which is the output of the forget gate and applied to the previous memory block by element-wise multiplication. As a result, the previous memory block will be effective on the current LSTM. If the activation output vector is close to zero, then the previous memory will be forgotten. Moreover, the new memory is created using a simple neural network with the tanh activation function and the effect of the previous memory block in the input gate, which can be calculated as

$$i_t = \sigma(W[x_t, h_{t-1}, C_{t-1}] + b_i), \tag{2.9}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot tanh(W[x_t, h_{t-1}, C_{t-1}] + b_c). \tag{2.10}$$

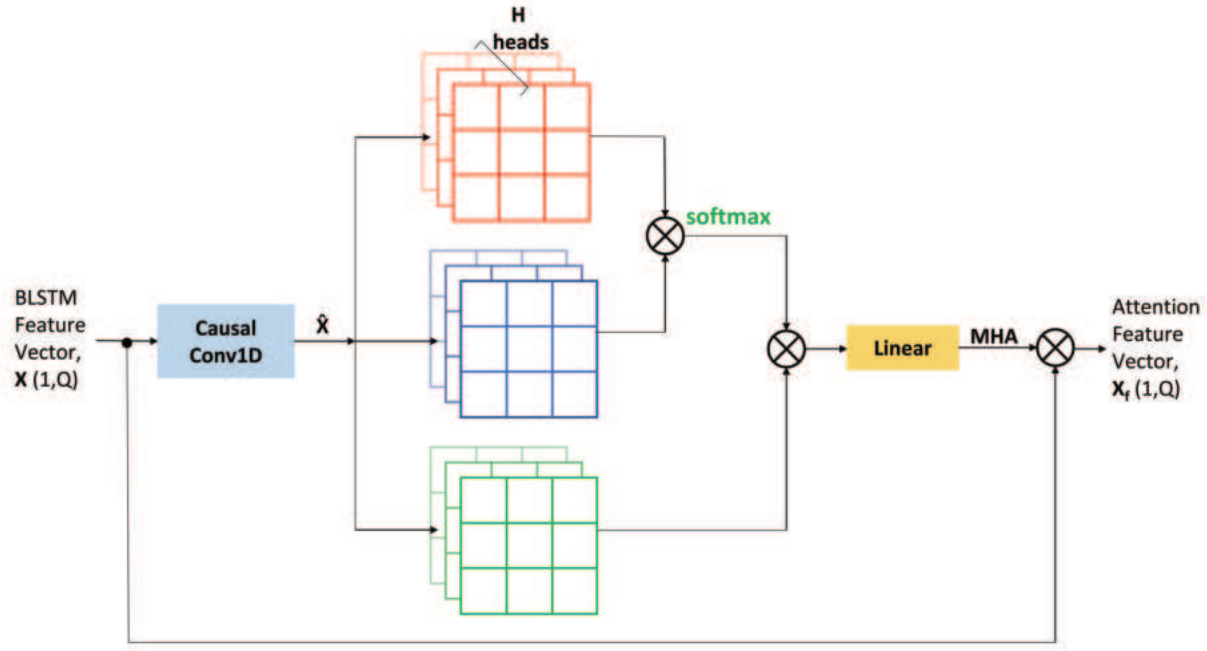


Figure 2.12: Configuration of the proposed multi-head attention (MHA) block

Finally, the output gate generates the current output or the current hidden state of the LSTM cell, which can be calculated as

$$h_t = tanh(c_t) \cdot \sigma(W[x_t, h_{t-1}, C_t] + b_o). \tag{2.11}$$

However, a problem with LSTM block is that they have access to past but not to future context, which can be overcome by using bidirectional LSTM (BiLSTM) block. In a BiLSTM layer, two separate recurrent hidden layers scan the input sequences in opposite directions and are connected to the same output layer, which is illustrated in Figure 2.11. As a result, it can extract context information in both directions where forward and backward contexts are learned independently.

In this study, the number of BiLSTM units in each layer is the same as the number of EEG channels and the number of BiLSTM layers is five. As a result, the i-th output of the BiLSTM network is the hidden state of the fifth recurrent layer and the output in each time step can be considered the temporal information extracted from each EEG trial.

2.2.2 Multi-Head Attention (MHA)

In this study, to extract more discriminative temporal information, the multi-head attention (MHA) mechanism is adopted by exploring the intrinsic importance to assign weights to each EEG signal trial. The structure of multi-head attention is illustrated in Figure 2.12. Furthermore, it can better describe the specific meaning by computing the similarity within each recurrent en-

coded slice.

MHA is inspired by the transformer model, which shows excellent success in many machine learning applications. MHA improves the self-attention mechanism in two main aspects. First, it expands the model's capability to focus on different positions. The encoding of each head knows about the encodings of the other heads, which improves the model's ability to learn temporal dependencies. Second, splitting the input features into different partitions increases the formation of the subspaces which generate attention weights for each subspace that represent each partition's importance, and concatenating these representations may extract better overall features to enhance the classification accuracy. And so, the output feature vector from the BiLSTM block $\mathbf{X} = [x_1, x_2, \ldots, x_Q] \in \mathbf{R}^{1\times Q}$ where Q is the length of the BiLSTM feature vector, serves as the input of MHA layer as shown in Figure 2.12. In general, MHA takes three copies of $\mathbf{X}$ as inputs and all three of them are transformed into $\hat{\mathbf{X}}$ by using causal convolution and then the attention is calculated as

$$ATT(\hat{\mathbf{X}}, \hat{\mathbf{X}}, \hat{\mathbf{X}}) = softmax(\frac{\hat{\mathbf{X}} \cdot \hat{\mathbf{X}}^T}{\sqrt{d}}) \cdot \hat{\mathbf{X}}. \tag{2.12}$$

Each of the three $\hat{\mathbf{X}}$ are further expanded for the attention over H heads where each $\hat{\mathbf{X}}$ is split into H subspaces, which forms $\hat{\mathbf{X}} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_H], \mathbf{X}_h \in \mathbf{R}^{1\times \frac{Q}{H}}, (h = 1, 2, \ldots, H)$. The attention $\mathbf{A}_h$ in each subspace h can be calculated as

$$\mathbf{A}_h = ATT(\hat{\mathbf{X}}_h, \hat{\mathbf{X}}_h, \hat{\mathbf{X}}_h), \tag{2.13}$$

where $\mathbf{A}_h \in \mathbf{R}^{1\times \frac{Q}{H}}$ and all the H representations are concatenated together to produce the final attention vector as

$$MHA(\hat{\mathbf{X}}, \hat{\mathbf{X}}, \hat{\mathbf{X}}) = Concat(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_H) \in \mathbf{R}^{1\times Q}. \tag{2.14}$$

In order to extract discriminative temporal attention of the BiLSTM feature vector, the attention vector is multiplied by the BiLSTM feature vector to set more weight to the significant one and it can represent more extensive feature vector denoted by $\mathbf{X}_f \in \mathbf{R}^{1\times Q}$. Finally, a dense classifier is used to classify the types of emotions.

2.2.3 Results and Discussion

The experimental framework and other setup for model training, validation and testing remain similar to the subsection 2.1.6.

Performance Evaluation

The recorded results (binary classification of both valence and arousal dimensions for each subject) with and without the attention mechanism are recorded in Table 2.3 and 2.4 respectively. The average accuracy and F-1 scores are slightly higher for the case of BiLSTM-based classification without attention. Furthermore, the resultant standard deviation is considerably higher for the case of BiLSTM-based classification with attention mechanism.

2.3 Performance Comparison

2.3.1 Performance Comparison of Different Proposed Methods

The classification performance of different proposed methods is compared with each other in Table 2.5 in the case of the DEAP dataset. The average accuracies and the f1-scores are above 84% for all the four proposed schemes. However, in the case of the CNN-based method without the channel-wise attention or shortly CA, the resultant standard deviation is much higher, indicating inconsistent performance among different subjects. Nevertheless, the performance improves drastically after adding the CA, ensuring the channel-wise attention provides generalized feature information. Though the BiLSTM-based method, both with and without attention mechanism, shows approximately similar performance with consistency among different subjects, the average accuracies are lower than the CNN-based architecture. Finally, the proposed CNN-based classification method with the CA outperforms all the other proposed architectures in this chapter.

2.3.2 Performance Comparison with Other Approaches

The classification performance of the proposed CNN-based method with channel-wise attention mechanism is compared with that reported very recently by some other methods. The results reported by some other methods are presented in Table 2.6 where the experimental analysis are carried out on DEAP dataset. The major variations in the experimental setup, such as the number of channels used and number of classes considered in the emotion experiment are also shown. Similar to the proposed methods in this chapter, subject-dependent performance is investigated in the methods proposed by others which are used to compare. In [32], CapsNet and other classifiers (1DCNN, 2DCNN, KNN, RDF and SVM) are used with multiband feature matrices. They consider all subjects together and 3 second window. In [13], mixed subject analysis is considered with 4s window and 5 pairs of EEG channels. In [23], 6s window with 50% overlap is used and subject-dependent performance is reported.

2.4 Conclusion

In this chapter, different types of deep learning architectures are proposed to classify EEG-based emotion recognition. But only the proposed CNN-based method with channel-wise attention (CACNN) outperforms the other proposed architectures in terms of accuracy and f1-score with consistency among all subjects. Moreover, the proposed CNN is productive in capturing detailed spatial and temporal information on the EEG data and the proposed CA can extract the attentive information among the channels. Finally, extensive experimental results in the case of the DEAP dataset have demonstrated that the proposed CACNN achieved an average accuracy of 94.74% and 95.17% on the valence and arousal classification tasks respectively. Furthermore, the proposed CACNN improved EEG-based emotion recognition accuracy compared with some existing methods.

Table 2.2: Performance of the proposed CNN-based method with channel-wise attention mechanism in binary class

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	95.92	95.91	96.25	96.15
02	97.00	96.56	92.33	91.53
03	95.75	95.71	98.83	98.06
04	89.83	89.17	91.75	91.19
05	91.83	91.81	98.83	97.26
06	95.00	93.41	93.50	93.36
07	96.33	95.78	95.75	95.79
08	96.08	96.04	97.33	97.23
09	94.67	94.65	96.08	95.80
10	99.83	99.83	98.33	98.29
11	85.83	84.66	83.75	83.25
12	91.83	91.81	98.83	97.26
13	90.17	90.01	95.83	87.82
14	91.58	91.57	92.50	90.25
15	97.67	97.66	92.25	92.06
16	98.67	98.58	98.08	98.08
17	88.58	88.42	91.83	89.64
18	94.42	94.30	96.08	95.92
19	95.58	95.54	97.83	97.19
20	98.25	98.22	99.58	99.06
21	95.42	95.35	97.58	94.42
22	97.83	97.81	98.92	98.68
23	95.83	95.19	95.75	95.28
24	95.92	95.91	96.50	94.48
25	94.25	94.03	96.58	94.93
26	94.75	94.59	93.83	93.60
27	97.00	95.07	95.17	94.00
28	95.83	95.53	90.42	90.36
29	98.00	97.93	97.00	95.91
30	95.17	94.38	97.83	97.82
31	91.58	91.42	89.25	89.12
32	95.33	95.32	91.08	89.56
Average	**94.74±3.08**	**94.44±3.19**	**95.17±3.46**	**94.17±3.70**

Table 2.3: Performance of the proposed BiLSTM-based method with channel-wise and multi-head attention mechanism in binary class

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	84.75	84.65	86.17	85.82
02	76.75	73.71	78.83	77.63
03	93.33	93.24	96.17	93.43
04	82.42	81.37	86.08	85.29
05	86.58	86.22	85.08	84.92
06	83.83	79.74	79.83	79.05
07	80.92	78.05	81.42	77.86
08	85.25	85.11	85.50	85.08
09	77.75	77.58	81.75	80.13
10	92.67	92.67	90.50	90.20
11	79.08	78.00	74.50	73.22
12	84.25	84.16	91.67	80.90
13	74.17	73.82	86.83	71.01
14	80.83	80.73	86.42	83.26
15	86.83	86.78	87.08	86.91
16	89.00	88.38	90.92	90.91
17	75.92	75.73	74.83	69.27
18	91.42	91.02	91.25	90.85
19	85.67	85.47	86.67	83.22
20	89.67	89.54	92.58	84.80
21	88.25	88.03	93.08	84.71
22	89.42	89.36	94.25	93.15
23	89.42	87.67	85.42	83.77
24	85.00	84.98	82.08	86.25
25	83.00	82.77	93.08	90.16
26	83.83	83.49	84.83	84.07
27	90.42	84.79	88.08	85.07
28	86.25	85.39	81.33	81.28
29	85.00	84.56	93.17	91.03
30	93.00	92.32	90.83	90.76
31	83.08	79.95	83.00	82.89
32	85.67	85.65	83.08	80.63
Average	**85.11±4.89**	**84.21±5.16**	**86.45±5.41**	**83.99±5.90**

Table 2.4: Performance of the proposed BiLSTM-based method without any attention mechanism in binary class

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	85.17	84.96	89.00	88.84
02	79.67	77.46	82.50	81.53
03	93.67	93.59	97.80	94.93
04	87.08	87.17	87.42	86.45
05	88.08	87.75	87.92	87.82
06	85.50	82.32	81.92	81.46
07	85.58	83.28	84.25	81.76
08	87.00	86.86	86.50	86.15
09	80.83	80.59	85.00	83.77
10	93.17	93.11	91.50	91.32
11	80.50	79.32	75.58	74.44
12	83.17	83.00	92.67	84.14
13	78.67	78.38	90.17	76.97
14	77.58	77.37	86.25	83.26
15	89.33	89.31	87.00	86.69
16	93.75	93.39	90.17	90.12
17	76.67	75.91	77.83	71.50
18	92.00	91.75	90.83	90.44
19	87.17	86.85	88.33	85.72
20	89.67	89.50	93.17	86.44
21	88.67	88.38	94.42	88.87
22	91.50	91.40	95.42	94.59
23	87.00	85.30	88.33	87.25
24	84.92	84.87	94.58	90.86
25	87.75	87.63	92.67	89.49
26	85.08	84.94	86.83	86.21
27	92.75	88.75	88.75	84.87
28	86.50	85.79	79.75	79.36
29	85.92	85.40	92.58	90.46
30	92.42	91.60	90.92	90.86
31	80.67	80.42	83.00	82.88
32	86.08	86.05	86.75	84.47
Average	**86.36±4.71**	**85.70±4.83**	**88.10±5.01**	**85.75±5.23**

Table 2.5: Performance of the proposed BiLSTM-based and CNN-based method in binary class

2*Proposed method	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
BiLSTM	86.36±4.71	85.70±4.83	88.10±5.01	85.75±5.23
CA+BiLSTM+MHA	85.11±4.89	84.21±5.16	86.45±5.41	83.99±5.90
CNN	91.53±10.47	88.29±16.93	91.32±11.82	88.30±16.60
CA+CNN (CACNN)	**94.74±3.08**	**94.44±3.19**	**95.17±3.46**	**94.17±3.70**

Table 2.6: Comparative performance analysis

2*Study	2*No. of channels	2*Class	Accuracy	
			Valence	Arousal
Chao et al. [32]	32-EEG	Binary	68.28%	66.73%
Mohammadi et al. [13]	10-EEG	Binary	86.75%	84.05%
Yin et al. [23]	32-EEG	Binary	90.45%	90.60%
Proposed CACNN method	**32-EEG**	**Binary**	**94.74%**	**95.17%**

Chapter 3

CNN-Based Emotion Recognition Using Different Band Frequencies of EEG Signal

In this chapter, different band frequencies of the EEG signal and spectral power are explored to classify emotion. As neural firing provides a pathway to elicit emotions, a thorough analysis on EEG frequency bands and spectral power can uncover salient information to categorize different classes of emotion. Firstly, the effect of different sub-bands is investigated on the classification performance of emotion recognition. Following the spectral decomposition of the EEG signal into different sub-bands, a 3D frame is formatted considering the spatial and temporal information of all frequency bands of all channels. As the constructed frame contains significant details of the brain signal, the proposed scheme shows substantial performance in classification process. Along with the frequency band signals, the inspection of spectral power is also important in the study of emotion analysis. Spectral power reflects the frequency content of a signal and it is closely associated with the strength and intensity information of each frequency band. In this regard, the feature vector comprising the spectral power of EEG sub-bands, effectively maps distinctive attributes from the raw data and provides a satisfactory performance in emotion recognition. In this study, detailed and extensive experimentations are carried out on a publicly available DEAP dataset.

3.1 Effects of different band frequencies

Since emotional states are caused by neural oscillations, different frequency bands are expected to have significant impacts on elicitation of emotion. In this regard, a comprehensive analysis is performed on different frequency bands of the EEG signal.

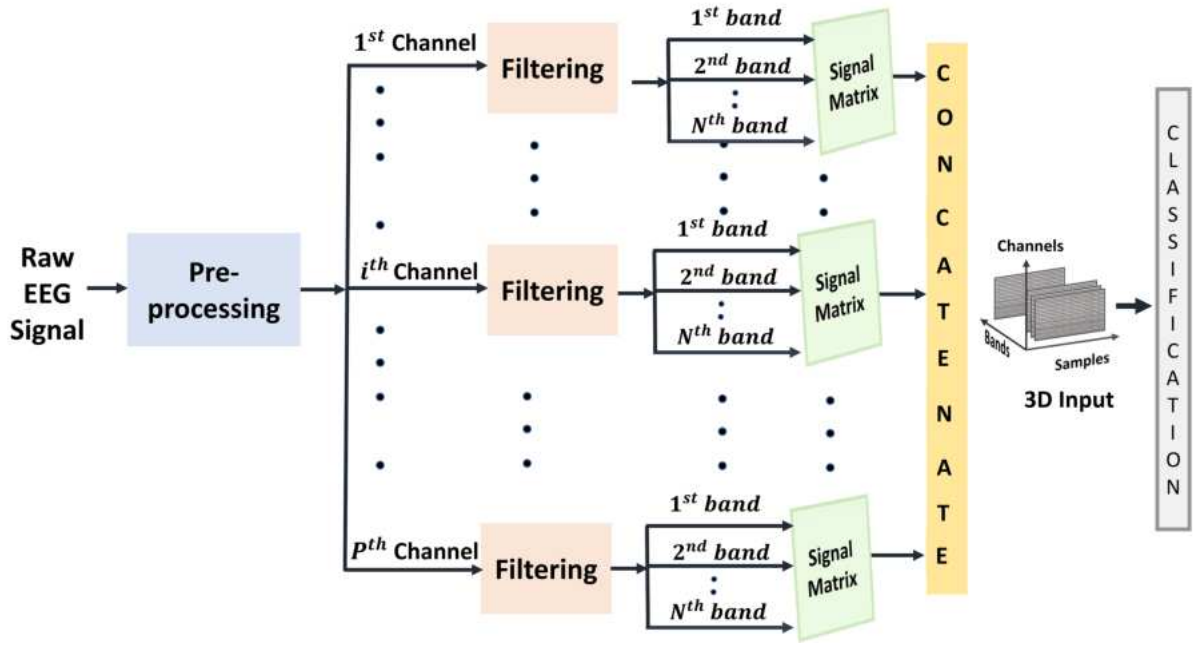


Figure 3.1: Major steps involved in the proposed method

3.1.1 Proposed Method

The major stages involved in the proposed approach are illustrated in Fig 3.1 which include data pre-processing, windowing and formation of 3D frame and 2D CNN-based classification. Firstly, the baseline of the raw EEG data is removed in the pre-processing stage. The processed EEG signal is then divided into different frequency bands (δ, θ, α, β, γ) and a matrix of filtered signal is formed with the extracted rhythms. Next, each frequency band is windowed with a fixed frame length and a signal matrix is formed utilizing the information of the frequency rhythms. Subsequently, a 3D frame is formatted by concatenating all signal matrices obtained from the available channels. Finally, the constructed frame is applied as an input to a 2D deep neural network. As the 3D frame captures the temporal and spatial information of different frequency bands, the proposed method shows a substantial performance in binary classification and categorizes emotion in valence and arousal domain efficiently. In the classification stage, both valence and arousal dimensions are considered binary classes (high and low). The detailed analysis of each stage of the proposed method is presented below:

Pre-processing

According to [33], baseline removal of the raw EEG signal can improve the performance of emotion recognition task. In this context, in the pre-processing stage of the proposed work, baseline data ($\mathbf{X}_{bs} \in \mathbf{R}^{1\times L_{bs}}$) of the acquired raw EEG signal ($\mathbf{X}_r \in \mathbf{R}^{1\times L_r}$) is first removed. If the raw EEG signal (X_r) is recorded as F Hz sampling frequency and X_i denotes the i^{th} second baseline signal ($X_i \in \mathbf{R}^{1\times F}$) the mean value of the baseline data can be calculated as follows:

$$\overline{\mathbf{X}}_{bs} = \frac{F}{L}\sum_{i=1}^{\frac{L}{F}} \mathbf{X}_{bs}. \tag{3.1}$$

The trial signal ($\mathbf{X}_t \in R^{1\times L}$) is obtained as:

$$\mathbf{X}_t = \mathbf{X}_r - \overline{\mathbf{X}}_{bs} \tag{3.2}$$

The obtained trial signal is then band pass-filtered to extract the frequency content of the EEG signal. As the spectral analysis of the EEG signal provides more spatial and temporal information, the decomposition is applied on the baseline excluded signal. The main sub-bands of the EEG signal are delta ($0.5 - 3.5$ Hz), theta ($3.5 - 7.5$), alpha ($7.5 - 13$ Hz), beta ($13 - 30$ Hz) and gamma ($30 - 50$ Hz). Following the extraction of N useful frequency band signals ($N \in [2, 5]$) for a particular channel i, a matrix of filtered signal X_f^i is formed where $X_f^i = [\mathbf{X}_{b_1}^{iT}, \mathbf{X}_{b_2}^{iT}, \ldots, \mathbf{X}_{b_N}^{iT}] \in \mathbf{R}^{L\times N}$, $i \in [1, 2, \ldots, P]$ where P denotes the number of available channels and $\mathbf{X}_{b_n}^i \in \mathbf{R}^{1\times L}$ refers to the n^{th} frequency band EEG signal of i^{th} channel. The formatted matrix of the filtered signal contains significant details of EEG sub-bands and hence it is expected to capture the characteristic information for the purpose of emotion classification.

Windowing and Formation of 3D Frame

In view of incorporating the time and frequency band information of all available channels, a three dimensional representative frame is required. In this regard, the filtered signal matrices obtained from all available channels needs to be considered. In order to maintain the stationarity constraint, each frequency band signal of a particular channel is windowed with a frame length of W. The windowing operation divides the full signal in M number of segments and 2D frame is formed combining the information of N different frequency bands. Following the operation, a 2D frame is formed and for the k^{th} segment of the i^{th} channel, $X_{fr_k}^i = [\mathbf{X}_{w_{b_1}}^{iT}, \mathbf{X}_{w_{b_2}}^{iT} \cdot\cdot, \mathbf{X}_{w_{b_N}}^{iT}] \in R^{W\times N}$ where $\mathbf{X}_{w_{b_n}}^t$ is a window segment for the n^{th} band of i^{th} channel and $\mathbf{X}_{w_{b_n}}^i \in \mathbf{R}^{1\times W}$. Finally, the 2D frames acquired from P available channels are concatenated and a 3D frame (X_{fr}) is formatted where $X_{fr} \in \mathbf{R}^{P\times W\times N}$. As the 3D frame ($X_{fr}$) maps the spatial and temporal contents of all channels in a three-dimensional space, it encapsulates salient information of EEG signal for the purpose of emotion recognition. The workflow of the formation of 3D frame is illustrated in Figure 3.2.

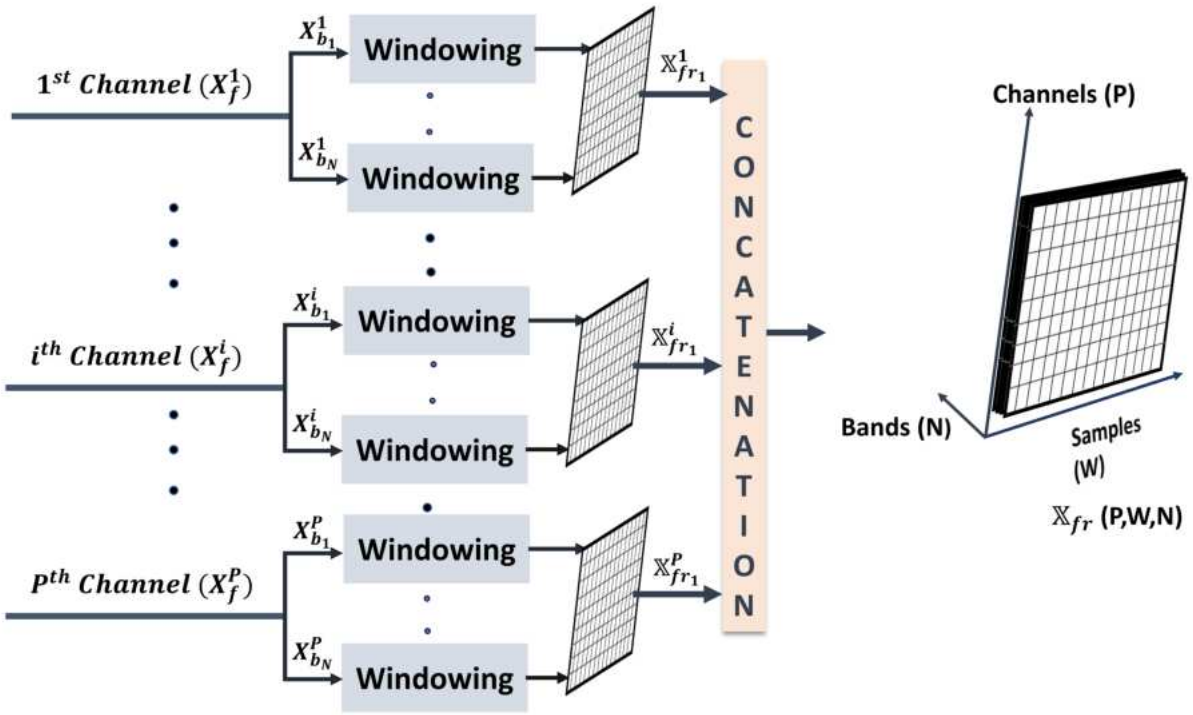


Figure 3.2: Formation of proposed 3D frame utilizing different frequency band signals

Architecture of the Proposed Neural Network

For the purpose of emotion classification, the proposed 3D frame ($X_{fr} \in \mathbf{R}^{P\times W\times N}$) is applied to a 2D CNN model. The CNN used here contains one input layer, one output layer and 16 hidden layers. The output shape of each layer and total number of parameters are shown in Table 3.1. The 2D CNN is trained to take the input shape of (P,W,N). In the layers of 2D convolutions, stride is kept 2 with similar padding. For the dropout layers, the dropout rate is set to 0.2 and 0.4. In the dense layer, $tanh$ and $Relu$ activation functions are employed. For the final dense layer, $Softmax$ activation is used. For the 2-class classification, the hidden neuron is set to 2.

3.1.2 Results and Discussion

Experimental Setup

In this section, performance analysis of the proposed scheme is presented. For the validation of the proposed approach, extensive experimentation is carried out on DEAP dataset. From the 63 seconds of recorded EEG signal from each channel (X_r), data from the first three seconds (X_b) are discarded as the baseline, and the remaining one minute long EEG data (X_t) is processed for the experimentation purpose. In this work, all 5 sub-bands of the EEG signal ($\delta, \theta, \alpha, \beta, \gamma$) are considered and a 2 second non-overlapping window frame is applied on the filtered signal.

Table 3.1: Details of the proposed 2D CNN model

Layer Type	Layer Parameters	Output Shape
Conv2D	f=16, k=5, s=2	(16, 128, 16)
BatchNormalization	-	(16, 128, 16)
MaxPooling2D	*pool size* = 2	(8, 64, 16)
Dropout	*rate* = 0.2	(8, 64, 16)
Conv2D	f=32, k=7, s=2	(4, 32, 32)
BatchNormalization	-	(4, 32, 32)
MaxPooling2D	*pool size* = 2	(2, 16, 32)
Dropout	*rate* =0.2	(2, 16, 32)
Conv2D	f=64, k=9, s=2	(1, 8, 64)
BatchNormalization	-	(1, 8, 64)
Dropout	*rate*=0.2	(1, 8, 64)
Flatten	-	(512,)
Dense	*units* = 256	(256,)
Dropout	*rate*=0.4	(256,)
Dense	*units* = 32	(32,)
Dropout	*rate*=0.4	(32,)
Dense	*units* = 2	(2,)
Total no. of parameters		331, 554

Following the process of the formation of 3D frame, the dimension of the data is $(32, 256, 5)$. The proposed scheme is designed to classify emotions for valence and arousal dimensions. Like most studies, in the proposed work, valence and arousal labels are categorized into binary class. Depending on participants' ratings, for binary class, rating < 5 corresponds to low class and otherwise high class. The training and validation of the proposed method are performed in the Google Colaboratory Platform.

Model Training, Validation and Testing

In the proposed study, the network is trained and tested on 32 subjects individually. Firstly, the data are randomly split into 80% training and 20% testing. For the validation purposes, a 5-fold cross-validation scheme is employed. In the training stage of each fold, the network is run for 200 epochs. The learning rate is set to 0.001 with $Adam$ optimizer and batch size is set to 128.

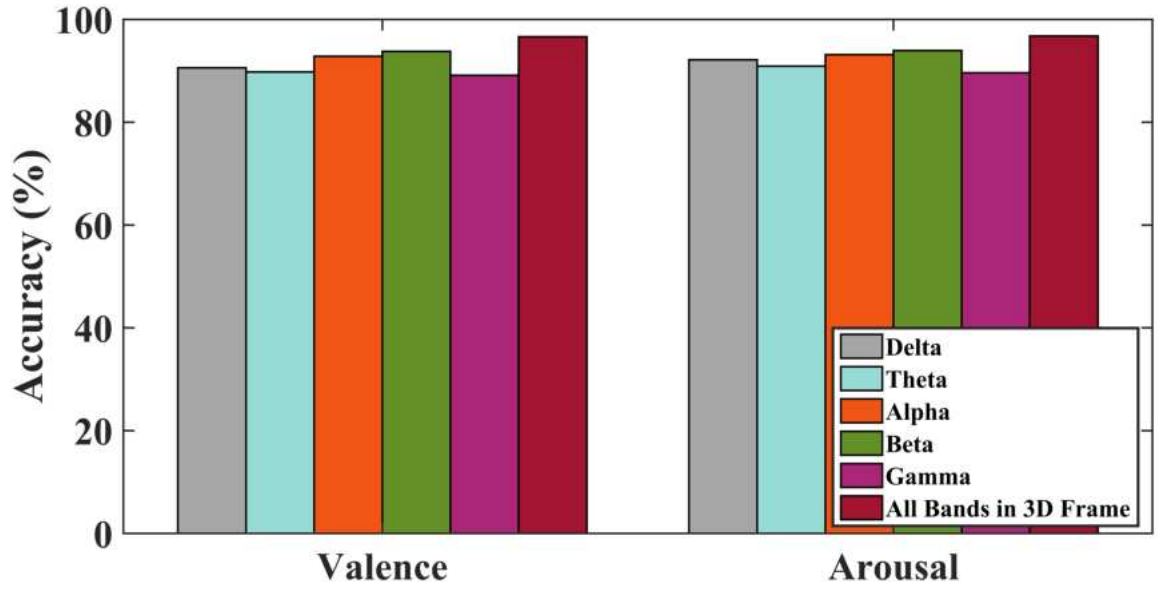


Figure 3.3: Classification performance (accuracy) for different frequency bands for valence and arousal cases

The loss function is categorical cross-entropy and metrics are accuracies.

Performance Evaluation

The performance of the proposed scheme is evaluated with following performance metrics

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{3.3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.4}$$

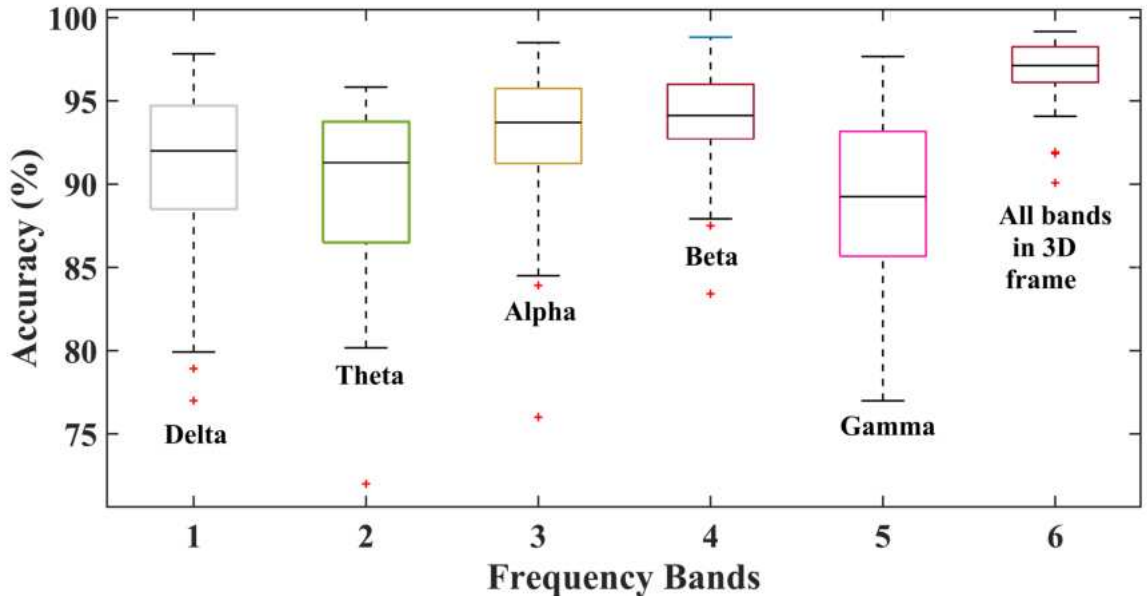
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.5}$$

$$\text{F-1 score} = 2.\frac{\text{Precision.Recall}}{\text{Precision} + \text{Recall}} \tag{3.6}$$

where TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative. The results for 2-class classification are recorded in Table 3.2.

The result for binary class classification is recorded in Table 3.2. The average accuracy and F-1 scores are found consistently very high for each subject (with an average greater than 96%). Along with that, the resultant standard deviation is considerably lower which indicates consistent performance among different subjects. The result demonstrates the efficacy of the proposed 3D frame with the band extraction scheme.

[Valence case]



[Arousal case]

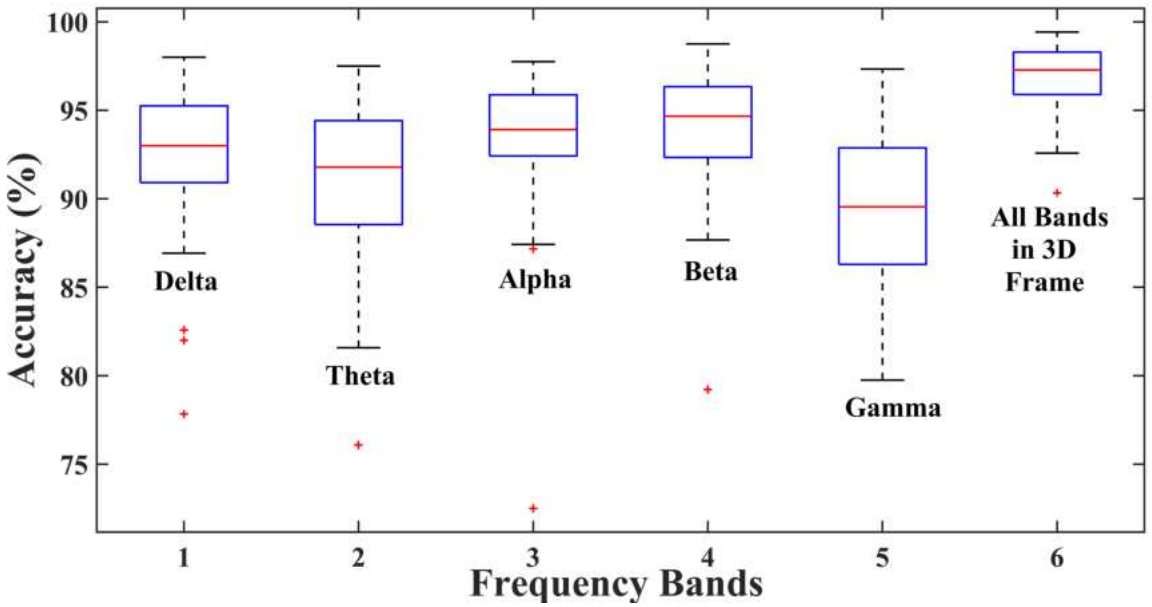


Figure 3.4: Box plot for accuracy comparison

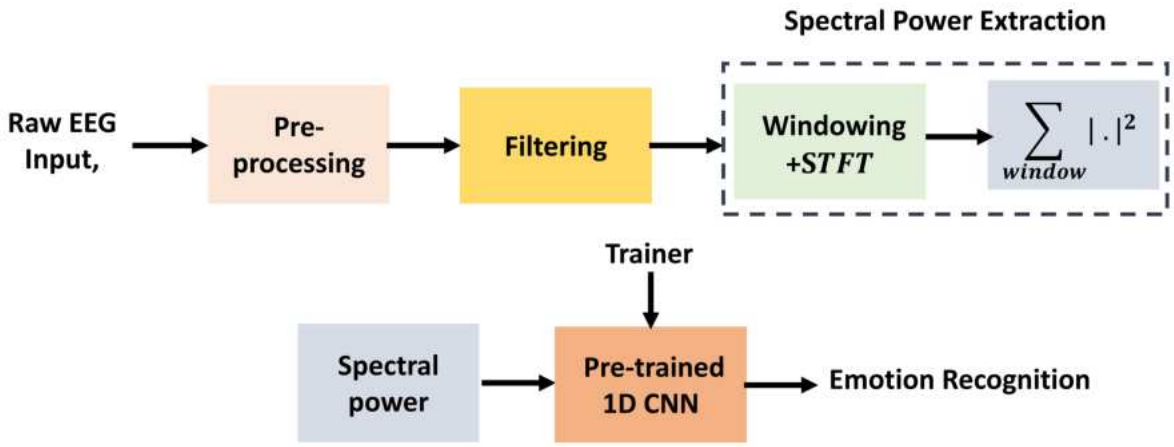


Figure 3.5: Workflow of the proposed method

Effect of Different Band Frequencies

One important aspect of the proposed method is to consider the effect of all frequency bands in a 3D frame instead of considering a spectral band. In this subsection, the effect of selecting other bands separately or taking all the bands together is explored. The proposed scheme is verified for different band frequencies of the EEG signal and the classification performance (accuracy) is shown in Fig. 3.3. It is observed that the the proposed 3D frame exhibits the highest accuracy in comparison to other cases. α and β bands also perform well in the proposed study. According to different accepted studies, α, β bands are dominant at awaken states [30]. The results obtained by the proposed scheme support the above findings. The performance of individual frequency band and the proposed scheme is illustrated in a box plot in Figure 3.4.

3.2 Effect of Spectral Power

In this section, detailed analysis and experimentation on emotion recognition are carried out utilizing the spectral power of different frequency bands of the EEG signal. Spectral power encapsulates the information of EEG signal in frequency domain and hence efficient in the context of emotion recognition. At first stage, after removing the baseline of the data, the EEG signal is decomposed into five different frequency bands. For each frequency band, the signal is divided into multiple overlapping sub frames. Subsequently, spectral power is extracted for each sub-frame of the EEG signal and a feature vector is formed combining the spectral power of all sub-bands of all available channels. Finally, the feature vector is applied as an input to a deep neural network. As the spectral power reflects the strength and the intensity information associated with different sub-bands, the proposed feature vector can provide a satisfactory classification performance The proposed workflow is illustrated in Figure 3.5.

Pre-processing and Windowing

In the pre-processing stage of the proposed study, baseline is excluded from the raw EEG signal (X_r) where $X_r = [X_b, X_t]$ using the steps described in 2.1.3. Let, $X_r \in \mathbf{R}^{P \times L_r}$, $X_b \in \mathbf{R}^{P \times L_b}$ and $X_t \in \mathbf{R}^{P \times L_t}$ where L_r, L_b and L_t denotes the length of the raw EEG signal, baseline and baseline excluded signal respectively. In order to analyze the spectral content of the EEG signal, band-pass filtering is performed on the acquired signal (X_t) and N different frequency bands are extracted. With the extracted frequency bands, a signal matrix X_p^i is formed for a particular channel i, where $X_p^i = [\mathbf{X}_{b_1}^{iT}, \mathbf{X}_{b_2}^{iT} \ldots, \mathbf{X}_{b_N}^{iT}]^T$, $i \in [1, 2, \ldots P]$ and $X_p^i \in \mathbf{R}^{N \times L_t}$. Subsequently, each frequency band of the signal matrix is divided into M overlapping segments of W length and a matrix obtained after the windowing operation is $X_w^i = [\mathbf{X}_{w_1}{}^{iT}, \mathbf{X}_{w2}{}^{iT}, \ldots, \mathbf{X}_{wN}{}^{iT}]^T$, where $X_w^i (i = 1, 2, \ldots, P) \in \mathbf{R}^{N \times W}$ represents the trial signal matrix at the i-th EEG channel. All the trials (X_w^i) obtained from all channels are assigned the same labels as X_t. For a ΔW amount of shift, the total number of trial matrices obtained from a given X_t will be $1 + (L - W)/\Delta W$.

Feature Extraction and Formation of Feature Vector

Since feature encapsulates the characteristics of a signal, a set of relevant and representative information is obtained with the feature extraction process. In this regard, analysis of spectral power can play an vital role in the context of emotion recognition as it carries the energy and intensity associated with each sub-band of EEG signal.
In view of incorporating the frequency domain information, short-time fourier transform (STFT) is performed on each window segment of the trial signal matrix (X_w^i) and spectral power is extracted from the corresponding window segment. The extraction of spectral power from a signal can be expressed by equation (3.7) and (3.8)

$$\mathbf{X_{feat}} = STFT\{x(t)\} = X(\tau, \omega) = \int_{-\infty}^{\infty} x(t) \cdot w(t - \tau) e^{-j\omega t} \tag{3.7}$$

$$F = \sum_{w} |\mathbf{X_{feat}}|^{\mathbf{2}} \tag{3.8}$$

Following the feature extraction process, a set of feature vectors $[F_1, F_2, \ldots F_P]$ are formed where $F_i \in \mathbf{R}^{N \times 1}$ represents the feature vector from the i^{th} channel. The final feature vector X_{FEAT} is formatted after combining the information of all features from all channels and $X_{FEAT} \in \mathbf{R}^{F \times 1}$ where $F = P \times N$. The formation of the feature vector is demonstrated in Figure 3.6.

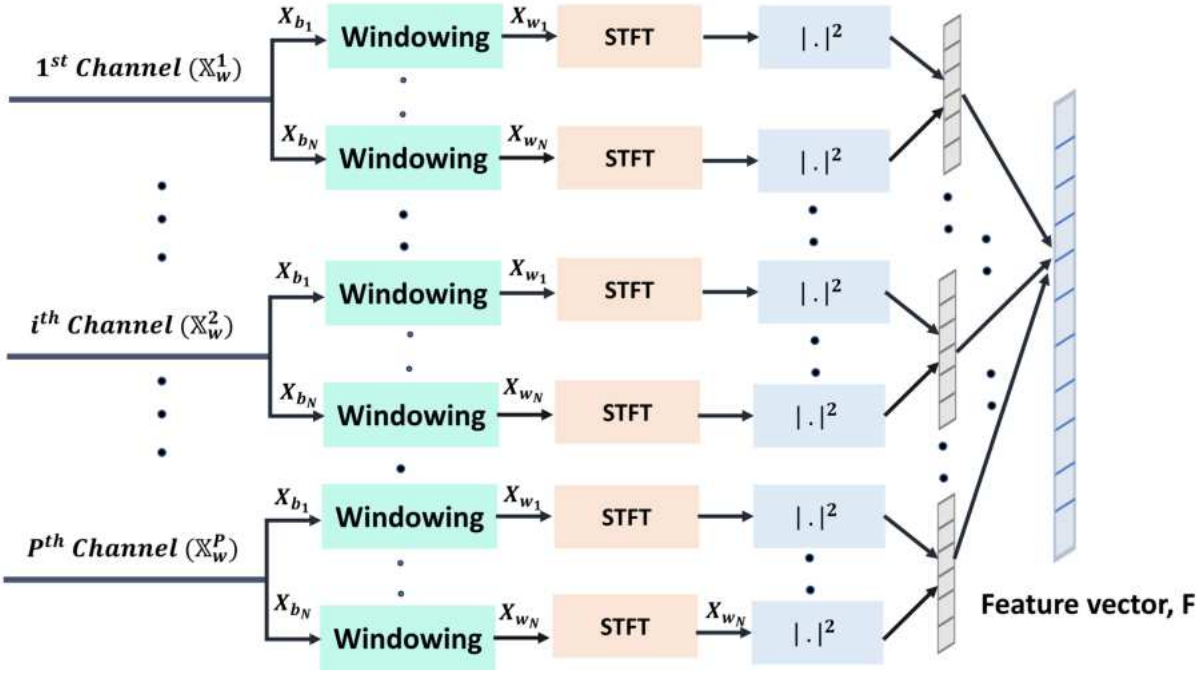


Figure 3.6: Formation of the proposed feature vector using spectral power of different frequency bands

Classification

For the purpose of emotion classification, the proposed feature vector (X_{FEAT}) is applied to a 1D CNN. The CNN used in the study consists of one input layer, one output layer and 12 hidden layers. The output shape of each layer and total number of parameters of the proposed network is displayed in Table 3.3. In all 1D convolutions, stride is kept 1. The dropout rate of two dropout layers in the classification head is 0.4. is obtained which is fed to a 1D deep neural network. In the dense classifier, *Softmax* activation function is used with two hidden units for binary classification.

3.2.1 Results and Discussion

Experimental Setup

In this section, performance analysis of the proposed method for emotion recognition is presented. From the recorded 63 seconds data from each channel, the first 3 second was discarded as baseline. The remaining 60 second EEG data is divided into 2 second long frames with 1.875 second overlap. In the study, all five bands of the EEG data (δ, θ, α, β and γ) are considered. Following the feature extraction process, the dimension of the feature vector is (160, 1) and the dimension of the final data is (18600, 160, 1). The proposed method is designed to classify emotions of binary class for both valence and arousal dimension. The threshold of the binary

classification process is kept the same as the previous work. The training and validation are performed in the Google Colaboratory Platform.

Model Training, Validation and Testing

The method conducted in the proposed study aims to be subject-dependent. So, the experiment is trained and tested on 32 subjects individually. Firstly, the data is randomly split into 80% training and 20% testing. A 10-fold cross-validation scheme is employed in the work for validation purpose. During the training stage, the network is run for 200 epochs. The loss function is categorical cross-entropy and metrics were accuracy.

Performance Evaluation

The performance of the proposed scheme is evaluated with two performance metrics: Accuracy and F-1 score. The results for binary classification is displayed in Table 3.4. The average accuracy and F1-scores for both valence and arousal dimensions are found consistently higher for each subject. In accordance with that, the lower standard deviation denotes the consistent performance among different subjects. The result demonstrates the efficacy of the proposed emotion classification scheme with spectral power.

3.3 Comparison between Different Frequency Band Signals and Spectral Power on Classification Performance

The performance comparison (accuracy) between the two schemes is presented in Figure 3.7. The studies for both methods are considered to be subject-dependent and the accuracy of the 3D frame based analysis is more consistent compared to the spectral power analysis of the signal. In spite of the higher median value in case of spectral power-based analysis, the frequency analysis offers superior performance as the classification result is more consistent and the standard deviation is lower compared to the spectral power-based method. For both valence and arousal cases, the spectral power demonstrates a larger deviation compared to the 3D frame based approach. Since the 3D frame contain significant temporal details of all sub-bands, the method offers a consistent and superior performance in comparison with the spectral power.

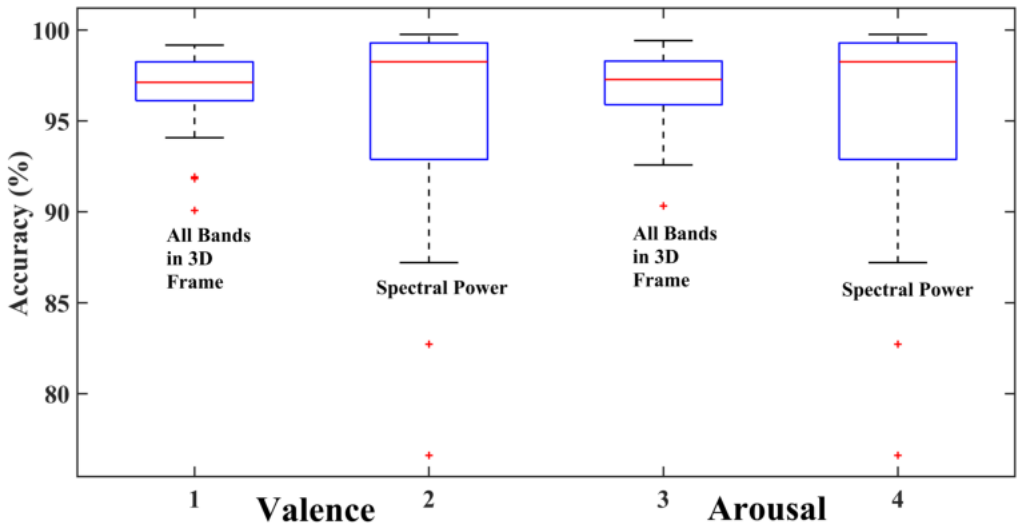


Figure 3.7: Accuracy comparison between all frequency bands considered in 3D frame and spectral power based approach

3.4 Conclusion

In this work, efficient emotion recognition schemes are proposed utilizing different frequency bands of the EEG signal. As different sub-bands contain distinctive spatial and temporal information, band analysis exhibits a better classification performance. The proposed 3D frame discussed in the study is constructed utilizing the details of five sub-bands $(\delta, \theta, \alpha, \beta, \gamma)$. As a consequence, it captures spatial, temporal as well as frequency contents and hence is very effective in categorizing different emotional states with a 2D deep neural network. As the effects of all sub-bands are considered instead of a single frequency band, the performance of emotion recognition gets better which demonstrates the significant contribution of different sub-bands in classifying emotions. The proposed study is also compared with the approach of spectral power analysis. As spectral power reflects the information of intensity and power associated with each sub-band, the method demonstrates a significant improvement in classification performance compared to traditional method of classifying emotion with raw EEG signal. The proposed feature vector containing the intensity information of all sub-bands provides a satisfactory performance in emotion classification. From extensive experimentation in a subject-dependent study, the proposed frequency band analysis approach shows superior performance in binary classification for both valence and arousal dimension.

Table 3.2: Performance of the proposed method in binary class

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	98.83	98.83	98.0	97.94
02	96.33	95.89	97.25	97.11
03	96.33	95.88	96.25	96.06
04	91.92	91.59	92.58	92.18
05	98.41	98.41	98.92	98.90
06	97.42	96.49	94.92	94.52
07	97.42	96.99	94.00	93.72
08	96.25	96.21	97.33	97.26
09	97.17	97.16	96.75	96.45
10	98.33	98.33	97.5	97.49
11	90.08	89.33	90.33	89.87
12	96.00	95.99	98.67	97.48
13	91.83	91.60	96.92	93.05
14	94.08	94.08	95.83	95.27
15	98.00	97.98	97.50	97.50
16	97.08	96.95	97.92	97.91
17	97.33	97.30	97.25	97.00
18	98.25	98.04	97.28	97.08
19	98.92	98.89	99.00	98.86
20	98.00	97.96	96.75	94.00
21	96.83	96.83	99.49	99.11
22	99.17	99.16	98.33	98.11
23	97.25	96.79	97.84	97.46
24	98.25	98.25	99.25	98.76
25	95.25	95.25	95.58	93.79
26	94.25	93.77	93.08	92.91
27	99.08	98.56	98.17	97.78
28	95.00	94.73	94.58	94.58
29	98.42	98.36	99.42	99.30
30	96.92	96.42	99.00	98.99
31	96.75	96.51	96.08	96.08
32	96.23	96.24	99.08	98.40
Average	**96.61±2.19**	**96.40±2.29**	**96.75±2.15**	**96.40±2.36**

Table 3.3: Details of the proposed 1D CNN model for spectral power-based classification

Layer Type	**Layer Parameters**	**Output Shape**
Conv1D	f=128, k=3, s=1	(160, 128)
BatchNormalization	-	(160, 128)
MaxPooling1D	*pool size* = 2	(80, 128)
Conv1D	f=128, k=3, s=1	(80, 128)
BatchNormalization	-	(80, 128)
MaxPooling1D	*pool size* = 2	(40, 128)
Conv1D	f=64, k=3, s=1	(40, 64)
MaxPooling1D	*pool size* = 2	(20, 64)
Flatten	-	(1280,)
Dense	*units* = 32	(32,)
Dropout	*rate*=0.4	(32,)
Dense	*units* = 32	(32,)
Dropout	*rate*=0.4	(32,)
Dense	*units* = 2	(2,)
Total no. of parameters		117,010

Table 3.4: Performance of the spectral power-based method in binary classification

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	99.52	99.52	99.10	99.06
02	76.61	74.05	73.45	67.23
03	93.25	93.10	96.61	94.6
04	76.71	74.53	78.44	76.89
05	85.25	84.10	85.44	85.35
06	95.25	93.65	94.85	94.74
07	99.15	98.98	99.67	99.64
08	94.77	94.68	95.04	94.84
09	81.95	81.75	78.47	74.97
10	95.37	95.37	95.73	95.70
11	70.27	65.62	69.40	65.16
12	93.56	93.54	97.32	95.37
13	90.03	89.74	96.31	92.41
14	86.15	86.13	86.01	83.69
15	95.40	95.40	96.17	96.16
16	98.70	98.62	97.96	97.96
17	58.38	46.40	68.81	64.71
18	95.11	94.68	92.99	92.39
19	88.66	88.18	92.02	90.65
20	93.21	93.02	96.05	94.17
21	96.07	96.05	99.04	98.48
22	95.62	95.62	95.73	95.40
23	99.47	99.40	99.70	99.64
24	92.53	92.51	97.4	95.55
25	93.14	93.13	97.02	94.83
26	92.29	91.13	89.99	89.74
27	99.64	99.51	99.59	99.51
28	87.21	86.11	66.15	66.01
29	99.54	99.53	99.04	98.89
30	99.47	99.40	98.51	98.51
31	99.29	99.21	98.75	98.75
32	95.78	95.78	97.35	96.88
Average	**95.61±5.55**	**95.41±5.94**	**95.17±7.37**	**94.63±8.38**

Chapter 4

CNN-Based Emotion Recognition Using Wavelet Transform of EEG Signal

In this chapter, a thorough analysis on emotion recognition is performed using multi-lead EEG signal in the wavelet domain. The time-frequency information provides a better representative characteristics of a signal and the wavelet coefficients or the feature extracted it captures more salient information compared to the raw EEG signal. Extensive experimentations are carried out on the publicly available DEAP dataset.

In the first section of the chapter, the effects of discrete wavelet transform (DWT) are explored. Firstly, after taking the baseline excluded raw EEG signal, the acquired data are divided into multi-level DWT coefficients. With the extracted co-efficients 3D frames are constructed and applied as the input to a 2D deep neural network. The 3D frame contains significant details and hence offers a satisfactory classification performance in emotion recognition.

The subsequent section deals with the effect of continuous wavelet transform (CWT) on emotion classification. At first stage of the proposed method, a particular frequency band (α) is extracted from the baseline excluded raw EEG data. The filtered signal is then divided into multiple sub-frames and for each sub-frame, strength-to-entropy component ratio (SECR) is calculated with the extracted CWT coefficients. The 2D feature matrix maps significant information of the EEG signal in wavelet domain and is proved to be very effective in categorizing emotions.

The combined effects of discrete wavelet transform (DWT) and continuous wavelet transform (CWT) are investigated in the final section of the chapter. In this part, multi-level discrete wavelet coefficients are extracted from the pre-processed raw EEG signal. For a particular decomposition level, the transformed signal is divided into multiple overlapping subframes and strength-to-entropy component ratio (SECR) is calculated with the extracted CWT coefficients. Finally, 2D frames are constructed in the SECR feature domain combining the information of the available channels. As both DWT and CWT preserve the time and frequency domain in-

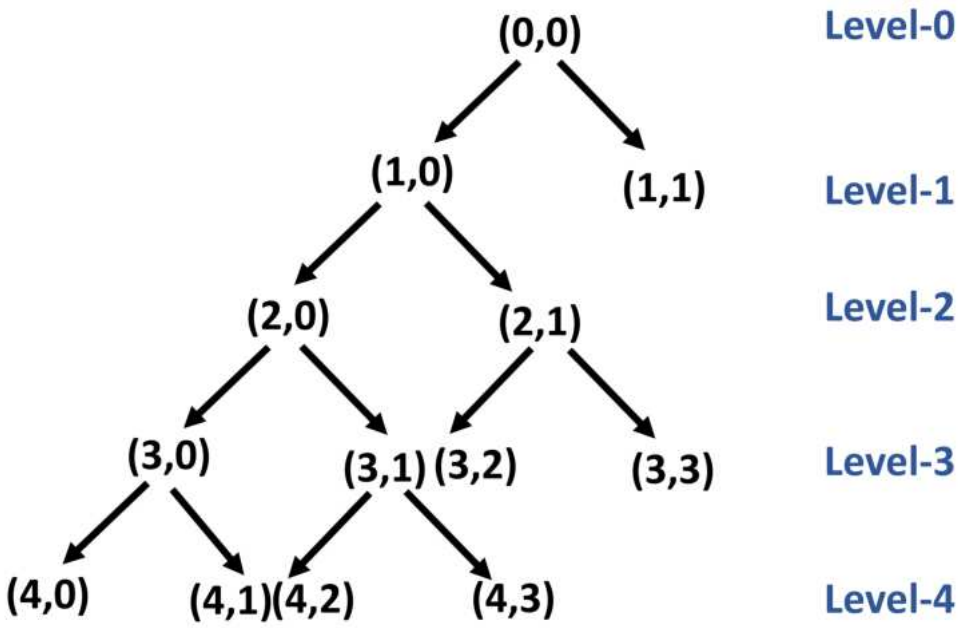


Figure 4.1: Tree Deccomposition of a signal

formation, the feature is expected to provide better information about the neural activities and hence capture different emotional states effectively.
In order to reduce the computational complexity associated with the CWT process, a novel method of channel and scale selection is introduced.

4.1 Discrete Wavelet Transform

In view of obtaining the spectral characteristics of a signal, it is a common approach to divide the EEG signal into multiple frequency bands namely- delta ($0.5 - 3.5$ Hz), theta ($3.5 - 7.5$ Hz), alpha ($7.5 - 13$), beta ($13 - 30$ Hz) and gamma (30-50 Hz) bands and perform analysis in each bands. As the filtering process does not preserve the time-frequency domain information, discrete wavelet transform (DWT) is getting much attention in the context of emotion recognition. DWT decomposes a signal into several subsets of approximation and detail coefficients that may be regarded as different frequency bands. As neural firing in the form of EEG signal can reflect different emotional states, it is necessary to analyze the properties by maintaining its time-frequency resolution in order to obtain the better representations. Any types of signal can be divided using the process of tree decomposition which is illustrated in Figure 4.1. In this section, extensive experimentation is carried out on the EEG signal using discrete wavelet transform.

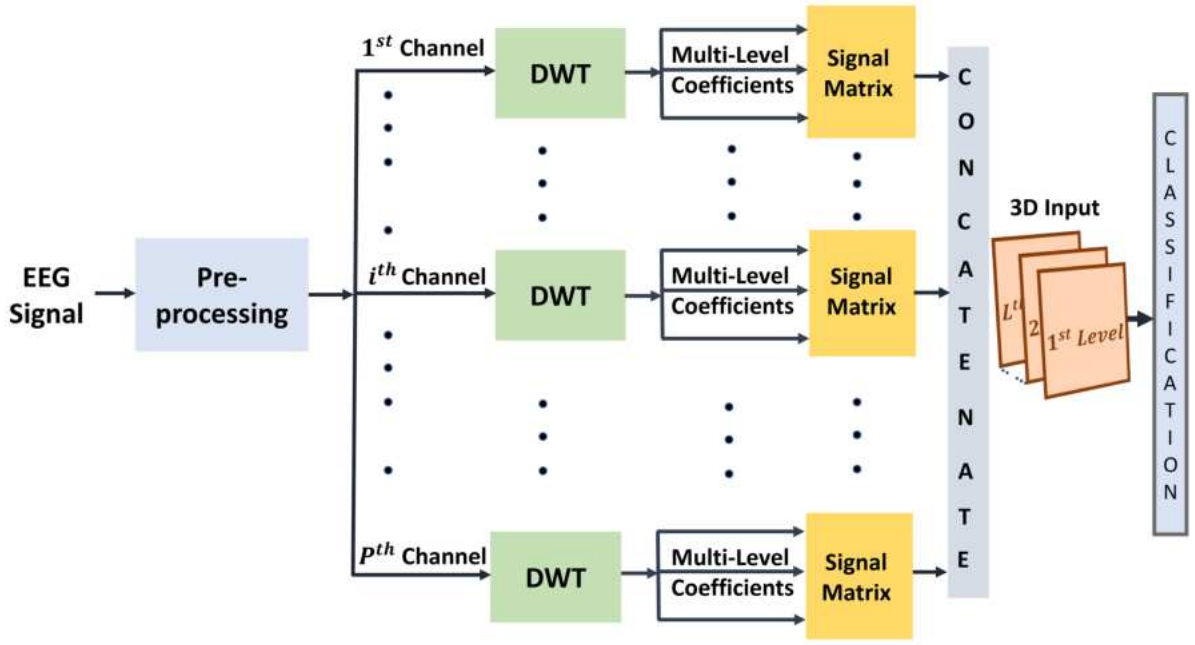


Figure 4.2: Major steps involved in the proposed method

4.1.1 Proposed Method

The major stages involved in the proposed method are depicted in Figure 4.2. The main steps are data pre-processing, DWT based signal decomposition scheme, 3D frame formation and classification with a 2D CNN. Firstly, after removing the baseline of the raw EEG data, the signal for each channel is divided into multi-level DWT coefficients. For a particular level decomposition, the signal is divided into multiple non-overlapping subframes and a 2D signal matrix is formed combining the information of all decomposition levels. After concatenating the signal matrix from all available channels, a 3D input is formed which is subsequently applied to a 2D CNN. In the classification stage, both valence and arousal dimensions are considered 2 classes (positive and negative).

Pre-processing

Pre-processing stage of the proposed study includes removing the baseline signal from the raw EEG data. The EEG data acquired from each channel $(\mathbf{X}_r)$ contains the baseline $(\mathbf{X}_b)$ and trial signal $(\mathbf{X}_t)$. Since the baseline data contains irrelevant information, removing it from the raw data can improve the emotion recognition task. The trial data is obtained as follows:

$$\mathbf{X}_t = \mathbf{X}_r - \mathbf{X}_b \tag{4.1}$$

In (4.1), $\mathbf{X}_r$ denotes the raw EEG data where $\mathbf{X}_r \in R^{P \times L_r}$, $\mathbf{X}_t$ is the trial signal where $\mathbf{X}_t \in R^{P \times L_t}$ and $\mathbf{X}_b$ is the baseline data where $\mathbf{X}_b \in R^{P \times L_b}$.

Wavelet Decomposition Scheme

In wavelet transform, a bank of filters enable each subband to be analyzed at a resolution matched to its scale. The scheme provides a range of advantages in case of non-stationary signal by offering simultaneous localization in time and frequency domain. DWT co-efficients of i^{th} channel signal ($x(t)$) can be expressed as

$$\gamma_{jk} = \int x(t) \cdot \psi_{j,k}(t)\, dt \tag{4.2}$$

In equation (4.2), γ_{jk} can be viewed as convolution between $x(t)$ and dilated, reflected and normalized version of mother wavelets.

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \cdot \psi(\frac{t - k \cdot 2^j}{2^j}) \tag{4.3}$$

In equation (4.3), $\psi(t)$ denotes the mother wavelet and $\psi_{j,k}(t)$ refers to the child wavelets obtained from different scaling and shifting parameters of the DWT's basis function. With the appropriate choice of scaling and shifting parametrs of the basis function, co-efficients of different levels can be obtained.

In discrete wavelet transform, the decomposed time-series co-efficients describes the time-evolution of signal in corresponding frequency band. The DWT of a discrete signal $x[n]$ can be defined as:

$$DWT(m,k) = \frac{1}{\sqrt{a}} \sum_n x[n] \cdot g(\frac{k - nb_0a_0^m}{a_0^m}) \tag{4.4}$$

where $g(\cdot)$ denotes the mother wavelet.

DWT offers the advantage of geometric scaling i.e, $\frac{1}{a}, \frac{1}{a^2}, \cdots \frac{1}{a^n}$ and translation by $0, n, 2n, \cdots$ which provides the logarithmic frequency coverage in contrast to the uniform frequency range like STFT. The multi-resolution analysis (MRA) through multi-stage filter implementation enables to explore the signal at various scales. In this regard, the wavelets can unfold the fine details of a signal by analyzing at different resolution.

Following the decomposition of the EEG data into L-level wavelet co-efficients for a particular channel i, a matrix of wavelet co-efficients X_f^i is formed where $X_f^i = [\mathbf{X}_{l_1}^{iT}, \mathbf{X}_{l_2}^{iT}, \cdots \mathbf{X}_{l_L}^{iT}] \in R^{L_t \times L}$,$i \in [1, 2, \cdot\cdot, P]$ and $\mathbf{X}_{l_L} \in R^{1 \times L_t}$.

The wavelet matrix (X_f^i) contains the significant approximate and detail coefficients of L level wavelet decomposition and hence it is expected to capture more distinctive information in time-frequency domain for emotion classification.

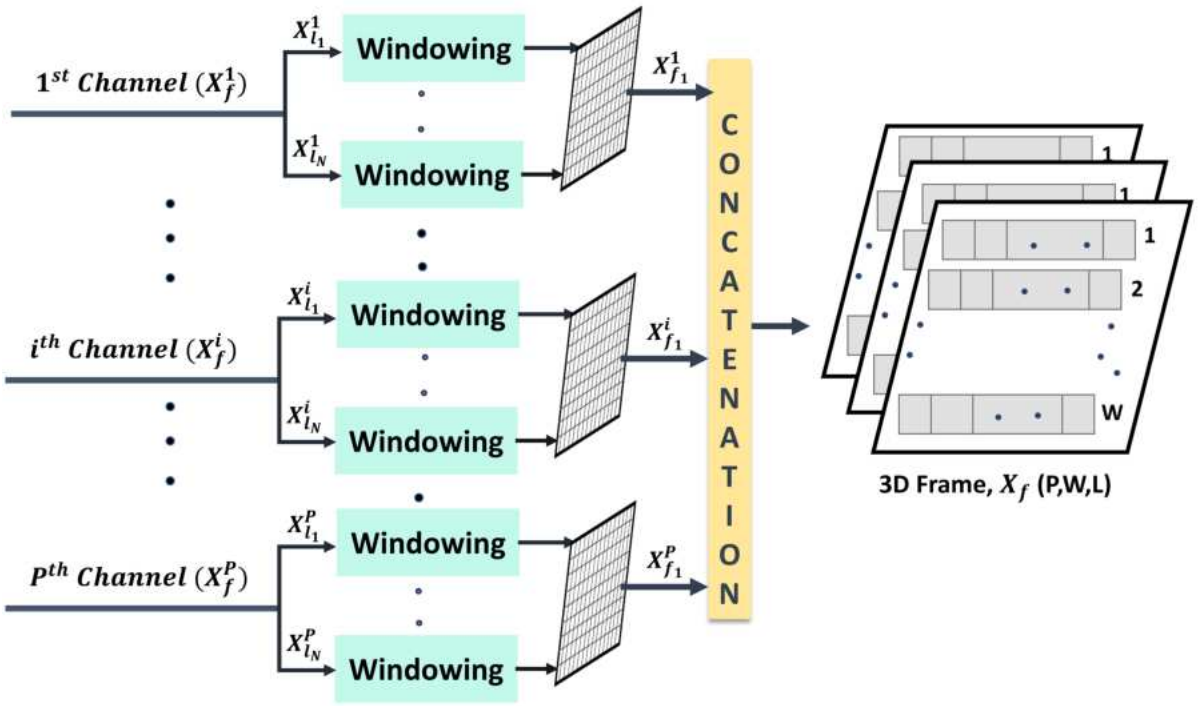


Figure 4.3: Formation of 3D frame with multi-level DWT co-efficients

Windowing and Formation of 3D Frame

In view of incorporating the information of all available channels in wavelet domain, a three dimensional representative frame is required. The wavelet matrix (X_f^i), obtained after the decomposition process, contains L significant levels of wavelet coefficients of channel i and each level of the decomposed signal is windowed with a frame of length W for the purpose of increasing total number EEG trials. Following the operation, a 2D frame $X_f^i = [\mathbf{X}_{w_{l_1}}^{iT}, \mathbf{X}_{w_{l_2}}^{iT} \cdot\cdot, \cdot\mathbf{X}_{w_{l_L}}^{iT}] \in R^{W\times L}$ is formed for each channel i of an EEG trial, where $\mathbf{X}_{w_{l_L}}^i$ is the L^{th} level decomposed signal and $\mathbf{X}_{w_{l_L}}^i \in R^{1\times W}$.
Finally, the 2D wavelet matrices acquired from P available channels are concatenated and a 3D frame (X_f) is formatted where $X_f \in R^{P\times W\times L}$. As the 3D frame (X_f) maps the time-frequency contents of all channels in the form of wavelet coefficients in a three-dimensional space, it encapsulates discriminative information of the EEG signal for the purpose of emotion recognition. The workflow of the formation of 3D frame is illustrated in Figure 4.3.

Architecture of the Proposed Neural Network

In order to classify emotion, the constructed 3D frame, X_f $(X_f \in R^{P\times W\times L})$ is applied to a 2D deep neural network. The network used in the study comprises one input layer, one output layer and 16 hidden layers. The CNN described in Table 4.1 is trained to take the input shape of (W, L). In all 2D convolutions, the stride is kept 2 and the activation function is $Relu$. In

the base model of the CNN, the dropout rate is 0.2. On the other hand, the dropout rate is set to 0.4 in the dense classifier. The activation functions employed in the classification heads are $Relu$ and $tanh$. For the final dense layer, the hidden units are set to 2 for binary classification and $Softmax$ activation function is employed.

Table 4.1: Details of the proposed 2D CNN model for 5-level DWT decomposition scheme

Layer Type	**Layer Parameters**	**Output Shape**
Conv2D	f=16, k=5, s=2	(16, 128, 16)
BatchNormalization	-	(16, 128, 16)
MaxPooling2D	*pool size* = 2	(8, 64, 16)
Dropout	*rate* = 0.2	(8, 64, 16)
Conv2D	f=32, k=7, s=2	(4, 32, 32)
BatchNormalization	-	(4, 32, 32)
MaxPooling2D	*pool size* = 2	(2, 16, 32)
Dropout	*rate* =0.2	(2, 16, 32)
Conv2D	f=64, k=9, s=2	(1, 8, 64)
BatchNormalization	-	(1, 8, 64)
Dropout	*rate*=0.2	(1, 8, 64)
Flatten	-	(512,)
Dense	*units* = 256	(256,)
Dropout	*rate*=0.4	(256,)
Dense	*units* = 32	(32,)
Dropout	*rate*=0.4	(32,)
Dense	*units* = 2	(2,)
Total no. of parameters		333, 154

4.1.2 Results and Analysis

Experimental Setup

In order to validate the effectiveness of the proposed scheme, extensive and detailed experimentation is carried out on a publicly available emotion dataset (DEAP). From the recorded 63 seconds data of the raw signal (X_r), 3 seconds data are discarded as baseline (X_b). The remaining 60 seconds data is regarded as the trial signal (X_t). The trial signal is divided into 2

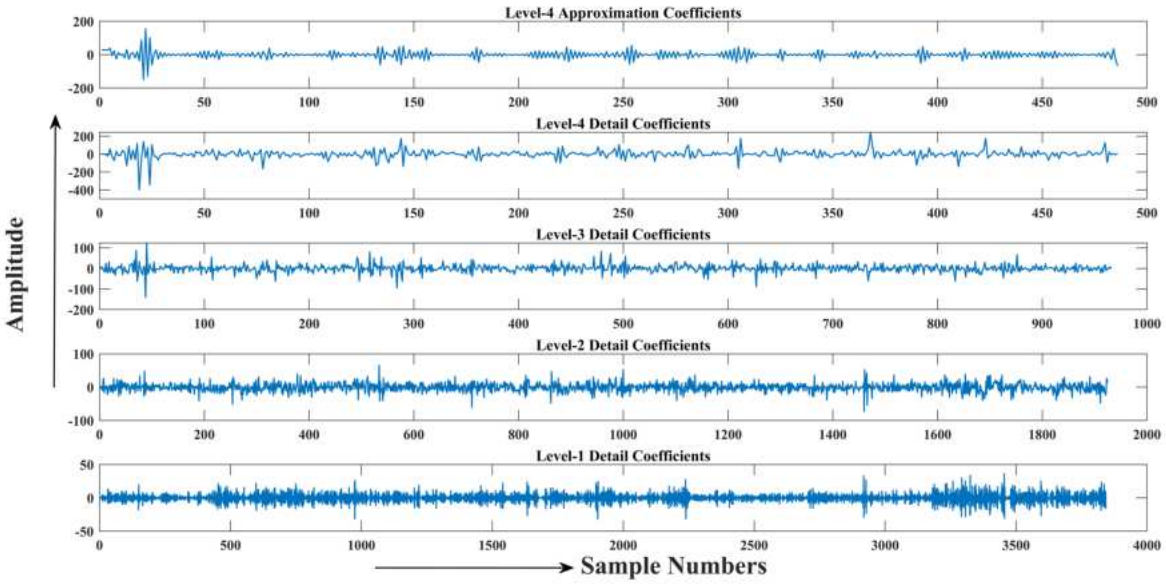


Figure 4.4: Wavelet decomposition of EEG signal

seconds window frame with no overlap. Following the process of 3D frame formation, the final dimension is (1200, 32, 256, 5). The proposed scheme is designed to categorize the emotion into binary class.

Depending on the participants' rating of DEAP dataset, high class (valence/arousal) is defined for rating ≥ 5 and low is defined otherwise. The training and validation is performed in the Google Colaboratory platform.

Model Training, Validation and Testing

The proposed work aims to be subject dependent and the network is trained and tested on 32 subjects individually. Firstly, the data is randomly split into 80% training data and 20% testing data. For validation purpose, a 10-fold cross-validation scheme is adopted. The network is run for 200 epochs. The learning rate is 0.001 and batch size is set to 128. The loss function is categorical cross-entropy and metrics are accuracy.

Performance Evaluation

The proposed approach utilizing DWT co-efficients is evaluated with two performance metrics: Accuracy and F-1 score. The results for binary classification is recorded in Table 4.2. For both valence and arousal dimension, the average accuracy and F-1 scores are found consistently very high for each subject. In addition, the standard deviation is found lower which denotes consistent performance among different subjects. The result demonstrates the efficacy of the proposed scheme with a 2D deep neural network.

4.2 Continuous Wavelet Transform

In this section, an efficient feature representation of multi-channel EEG data in the continuous wavelet transform (CWT) domain is proposed and the extracted features are employed in a deep learning model for automatic emotion recognition. Instead of using the raw EEG signal, its CWT coefficients that preserve relevant time-frequency domain information in each scale are considered in the feature extraction process. For a given channel of EEG data, each CWT coefficient from different scales is mapped into a corresponding strength-to-entropy component ratio plane to obtain a 2D feature representation. Finally, by concatenating these representations from different channels, proposed 2D feature matrix is generated, namely CEF2D, which is fed into a deep convolutional neural network architecture. In order to reduce the computational complexity, effective channel and CWT scale selection schemes are proposed based on the energy-to-entropy ratio in the CWT domain. Extensive experimentation is carried out on a publicly available EEG emotion (DEAP) dataset and very satisfactory classification performance is obtained for valence and arousal types of emotions both in 3-class and 2-class scenarios.

4.2.1 Proposed Method

The major steps involved in the proposed method are illustrated in Fig. 4.5, which are data pre-processing, proposed CWT based feature extraction scheme and 2D CNN based classification. Firstly, after taking the baseline excluded raw EEG signal, a band-limited signal containing the alpha (α) band frequency is extracted and the filtered signal is then windowed to increase the total number of EEG trials. Next, CWT is performed on each EEG trial and the proposed entropy-based feature is calculated from the extracted CWT coefficients for each scale. Finally, a 2D feature matrix (**C**WT domain **e**ntropy-based **f**eature, CEF2D) is formed and employed as the input to a CNN model to categorize different classes of emotions. In order to reduce the computational complexity, an analysis is performed considering the energy and entropy in the CWT domain and an efficient channel and scale selection scheme is designed. In the classification stage, both valence and arousal dimensions are classified into 3 classes (positive, neutral and negative) as well as binary classes (high and low). In what follows, steps involved in the proposed method are presented in detail.

Pre-processing and Windowing

Pre-processing stage of the proposed method includes exclusion of baseline signal and application of a bandpass filter to select a specific band. Let $X_r = [X_b, X_t] \in \mathbf{R}^{P\times H}$ be the recorded EEG signals with F Hz sampling frequency, where X_b indicates the baseline signal, P is the total number of EEG channels and H is the length of the EEG data. In addition, $X_t \in \mathbf{R}^{P\times L}$

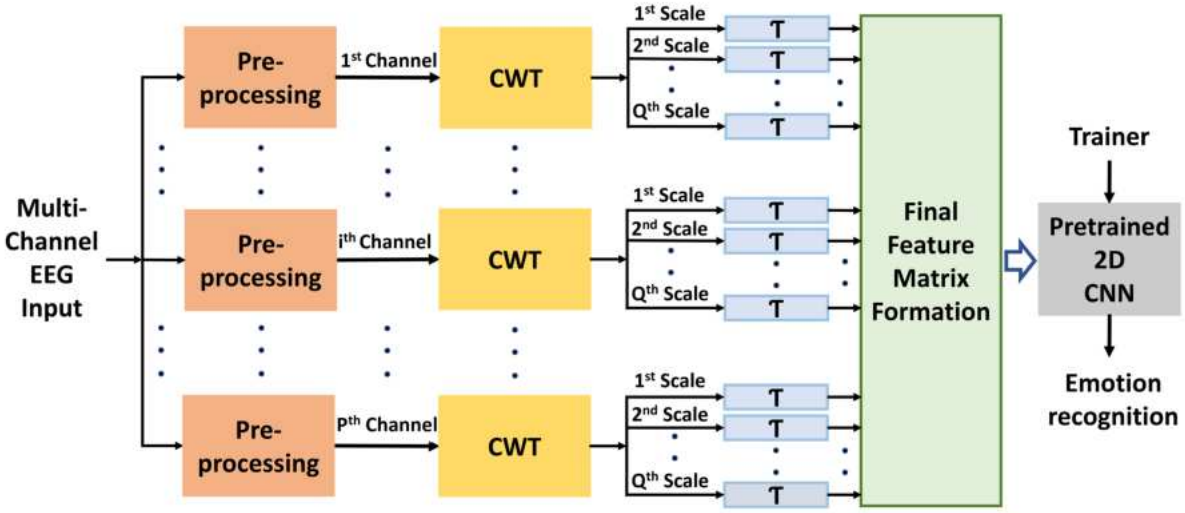


Figure 4.5: Major steps involved in the proposed method

denotes baseline excluded raw EEG signal to be used, where L denotes its length. In an EEG signal, main frequency bands are delta ($0.5 - 3.5$ Hz), theta ($3.5 - 7.5$ Hz), alpha ($7.5 - 13$ Hz), beta ($13 - 30$ Hz) and gamma ($30 - 50$ Hz). According to some studies, δ and θ waves are mostly associated with deep sleep, relaxation and they are less relevant to the cognitive tasks of the human brain [30]. On the other hand, α, β and γ waves are dominant in the cases of information processing, conscious thoughts, learning and emotional tasks. As β and γ waves contain high-frequency bands, they are involved in the processing of multi-dimensional complex tasks [31]. Hence, the α wave is expected to capture different emotional states better than β and γ waves. As a result, in the proposed method, only the α band signal is extracted from the raw EEG signal employing a 4^{th} order bandpass Butterworth filter. In order to increase the total number of trials to perform the task of emotion recognition, the given baseline excluded EEG signal (X_t) is divided into several small frames using a proper window and overlap. The trial signal $X_w = [\mathbf{X}_w^{1\,T}, \mathbf{X}_w^{2\,T}, \ldots, \mathbf{X}_w^{P\,T}]^T$ is obtained from X_t, where $\mathbf{X}_w^i (i = 1, 2, \ldots, P) \in \mathbf{R}^{1\times W}$ represents the trial signal at the i-th EEG channel, W denotes the window length and X_w is assigned the same emotion label as X_t. For a ΔW amount of shift, the total number of trials obtained from a given X_t will be $1 + (L - W)/\Delta W$.

Proposed Feature Extraction Scheme

In view of incorporating both time and frequency domain information, on each channel of EEG data, CWT-based feature extraction is performed. CWT of the i-th channel EEG signal $x(t)$ can be expressed as [34]

$$WT_\psi\{x\}(a_j, b_l) = \langle x, \psi_{a_j,b_l}\rangle = \int_R x(t) \cdot \psi_{a_j,b_l}(t)\, dt, \tag{4.5}$$

where the CWT's basis functions $\psi_{a_j,b_l}(t)$ are scaled at a scale $(a_j > 0)\,\{j = 1, 2, 3, \ldots, Q\}$, shifted by a translational value $(b_l \in \mathbf{R})\,\{l = 1, 2, 3, \ldots, W\}$ and Q refers to the total number of scales. The basis functions can be expressed as

$$\psi_{a_j,b_l}(t) = \frac{1}{\sqrt{a_j}} \cdot \psi_0 \left(\frac{t - b_l}{a_j} \right). \tag{4.6}$$

Here $\psi_0(t)$ denotes the mother wavelet. The CWT of a discrete time finite duration signal with uniform sampling can be considered as the convolution between that signal and the scaled and normalized wavelet [35]. In the proposed method, Morlet mother wavelet is used, which provides a good balance between time and frequency localization and defined as

$$\psi_0(t) = \pi^{-1/4} \cdot e^{imt} \cdot e^{-t^2/2}. \tag{4.7}$$

In the case of EEG signals, the signal strength varies in different portions of the brain depending on the triggered neurons associated with the neural signalling. The numbers and the pattern of neuron firing provide vital information to classify different types of emotions [36]. Since our objective is to analyze the time and frequency information of the recorded EEG signal, the strength of the wavelet coefficients can effectively serve the purpose, which can be defined as

$$E_{a_j,b_l} = |WT_\psi\{x\}(a_j, b_l)|^2, \tag{4.8}$$

and the energy of a particular scale of CWT coefficients can be calculated as

$$E_{a_j} = \sum_{l=1}^{W} E_{a_j,b_l}. \tag{4.9}$$

It is well known that different scales of CWT encapsulate the information of different frequencies [34]. For EEG signal, different frequency bands have their own significance and hence it is expected that the energy of a particular scale of CWT coefficients (E_{a_j}) contains necessary information of a particular frequency band. Another essential factor to be considered is the variation of information within a CWT scale of EEG signal, which can be evaluated with the help of entropy. Considering the inherent fluctuating nature of EEG signal due to the variations of emotions, the entropy measure corresponding to the strength of the CWT coefficients of a particular scale can be obtained as

$$H_{a_j} = \sum_{l=1}^{W} H_{a_j,b_l} = \sum_{l=1}^{W} P_{a_j,b_l} \cdot \log \frac{1}{P_{a_j,b_l}}. \tag{4.10}$$

Here (P_{a_j,b_l}) of each CWT coefficient $(b_l; l = 1, 2, 3, \ldots, W)$ corresponding to a scale $(a_j; j = 1, 2, 3, \ldots, Q)$ refers to the relative contribution of each coefficient and can be computed as $E_{a_j,b_l}/(E_{a_j})$. In (4.10), the entropy is expressed as a sum of entropy components, and these

entropy components for a particular scale (H_{a_j,b_l}) capture the randomness associated with each time point. The strength of the EEG signal provides an information about the number of neuron firing, whereas the entropy of the EEG signal contains the information of the random pattern of neurons triggered [37]. Hence, the combination of relative strength and randomness of the strength of CWT coefficients can provide a representative feature characteristics.

In the proposed study, we consider the strength-to-entropy component (SEC) ratio of CWT coefficients as an informative feature for EEG signals to classify different types of emotions. In this method, CWT coefficients for the pre-processed EEG signal of i^{th} channel $(\mathbf{X}_w^i)$ for a given trial (X_w) is extracted and the coefficient matrix X_{g_i} is formed where $X_{g_i} = [\mathbf{X}_{g_i}^{1\ T}, \mathbf{X}_{g_i}^{2\ T}, \ldots, \mathbf{X}_{g_i}^{Q T}]^T \in \mathbf{R}^{Q\times W}$. Here, $\mathbf{X}_{g_i}^j$: CWT coefficients of j^{th} scale $(j = 1, 2, 3, \ldots, Q)$ and i^{th} channel $(i = 1, 2, 3, \ldots, P)$. Following the process, extracted X_{g_i} is mapped into a 2D feature representation $X_{f_i} \in \mathbf{R}^{Q\times W}$ by transforming each CWT coefficient $x_{g_i}^j$ into the SEC ratio defined as, $T\{x_{g_i}^j\} = \frac{E_{a_j,b_l}}{H_{a_j,b_l}}$. After concatenating each X_{f_i} from all the channels, the proposed CEF2D feature matrix $X_f \in \mathbf{R}^{F\times W}$ is obtained, where $F = P \times Q$. The SEC ratio of each CWT coefficient (a_i) at each scale (b_j) for i^{th} channel can be calculated as

$$SECR_{a_i,b_j} = T\{x_{g_i}^j\} = \frac{E_{a_i,b_j}}{H_{a_i,b_j}}. \tag{4.11}$$

The formation of 2D feature matrix is illustrated in Fig. 4.6.

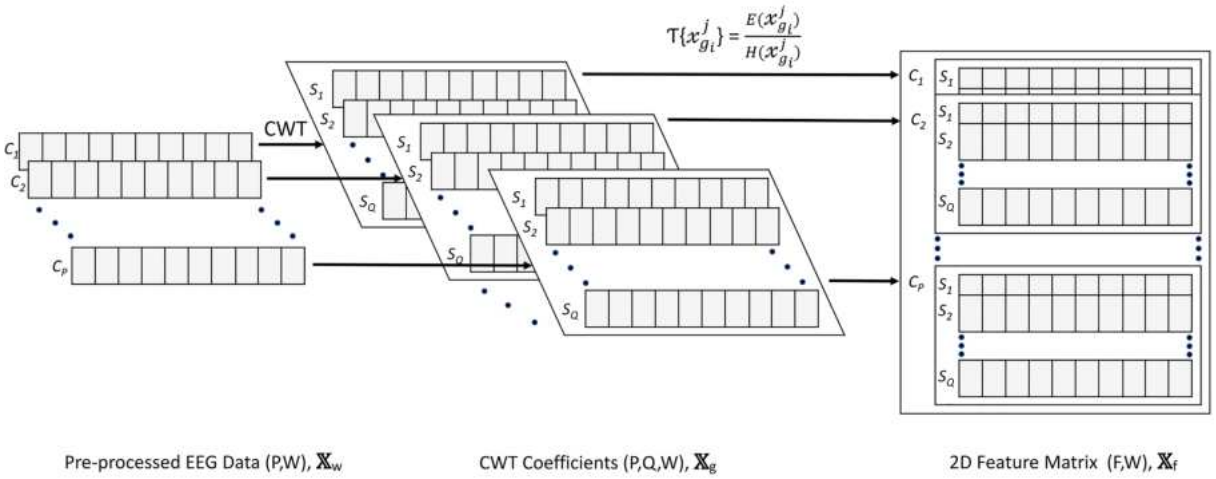


Figure 4.6: Formation of proposed 2D feature matrix

Architecture of the Proposed Neural Network

For the purpose of emotion classification, the proposed 2D feature matrix $(X_f \in \mathbf{R}^{F\times W})$ is applied to a 2D CNN model. The CNN used here contains one input layer, one output layer and 17 hidden layers. The output shape of each layer and total number of parameters are shown in

Table 4.3. The 2D CNN is trained to take the input shape of (F,W,1). In the 2D convolutions, stride is kept 2 with similar padding. For the dropout layers, the dropout rate is set to 0.2. In the dense classifier, *tanh* and *Relu* activation functions are employed. For the final dense layer, *softmax* activation is used. For 3-class and 2-class classification, the hidden neurons are set to 3 and 2 respectively. The algorithm of the proposed method is summarized in Algorithm 1.

Algorithm 1 The pseudo-code of the proposed method

Input: Raw EEG data of a subject, ground truth label y;
$data\ (X_t) \in \mathbf{R}^{P\times L}$; $//baseline\ excluded\ EEG\ signal$
$n \leftarrow Total\ number\ of\ audio\&visual\ stimulus\ for\ a\ subject$;
$\mathbf{C} \leftarrow [c_1, c_2, c_3, \ldots, c_P]$; $//channels$
$window\ size \leftarrow W$;
$CEF2D\ (X_f) \in \mathbf{R}^{F\times W}$;
$//s_a : s_b = (range\ of\ scales)$
$j \leftarrow 1\ to\ n\ i \leftarrow 1\ to\ P\ \mathbf{X}_{t_j}^{c_i} \leftarrow Butterworth\ Filter(\mathbf{X}_{t_j}^{c_i}$,$\alpha$ band);
start $\leftarrow 0$;
$start+W \leq L\ \mathbf{X}_{w_k}^{c_i} \leftarrow \mathbf{X}_{t_j}^{c_i}(start : start+W)$; $//k\ used\ to\ indicate\ k-th\ number\ of\ trial$
$X_{g_k}^{c_i} \leftarrow CWT(\mathbf{X}_{w_k}^{c_i}, s_a : s_b, \text{“}Morlet\text{”})$;
$X_{f_k}^{c_i} \leftarrow T\{x_{g_k}^{c_i}\}$;
$start \leftarrow start + skip$;
$X_{f_k} \leftarrow Append\{X_{f_k}^{c_i}\}$;
$Apply\ 2D\ CNN\ on\ CEF2D\ feature\ matrix\ (X_{f_k})$

4.2.2 Results and Discussion

Experimental Setup

In this section, performance analysis of the proposed scheme is presented. From the 63 seconds of recorded EEG data from each channel (X_r), data from the first three seconds (X_b) are discarded as the baseline, and the remaining one minute long EEG data (X_t) is divided into 2 second long frames with 1.75 second overlap. The proposed method is designed to classify emotions for valence and arousal dimensions. Like most studies, in the proposed work, valence and arousal labels are first categorized into binary and three classes. Depending on participants' ratings, low and high valence (or arousal) labels are assigned when the rating<4 and rating ≥ 6, respectively, and otherwise, it is medium valence/arousal. For binary class, rating < 5 corresponds to low class and otherwise high class. The training and validation of the proposed method are performed in the Google Colaboratory Platform.

Model Training, Validation and Testing

In the proposed study, the network is trained and tested on 32 subjects individually. Firstly, the data are randomly split into 80% training and 20% testing. For the validation purposes, a 10-fold cross-validation scheme is employed. In the training stage of each fold, the network is run for 300 epochs. The learning rate is set to 0.001 with $Adam$ optimizer and batch size is set to 128. The loss function is categorical cross-entropy and metrics are accuracies.

Performance Evaluation

The performance of the proposed scheme is evaluated with following performance metrics

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \tag{4.12}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.13}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.14}$$

$$\text{F-1 score} = 2.\frac{\text{Precision.Recall}}{\text{Precision} + \text{Recall}} \tag{4.15}$$

where TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative. The results for 3-class classification are recorded in Table 4.4. The average accuracy and F-1 scores for both valence and arousal dimensions are found consistently very high for each subject (with an average greater than 98%). In the case of binary classification, the results are shown in Table 4.5. The average accuracy and F-1 scores for valence and arousal in binary cases are also found consistently very high for each subject. Moreover, the overall performance is found better in binary case than that is obtained for 3-class case. For both binary and 3-class performance, the standard deviation is considerably lower which indicates consistent performance among different subjects. The result demonstrates the efficacy of the proposed CWT-based 2D feature matrix used in the 2D-CNN network.

Effect of Different Band Frequencies

One important aspect of the proposed feature extraction scheme is to consider only one spectral band (α) out of the conventional five bands. In this subsection, the effect of selecting other bands or taking all the bands together is explored. The proposed scheme is verified for different band frequencies of the EEG signal and the classification performance (accuracy) is shown in Fig. 4.7. It is observed that the α band exhibits the highest accuracy in comparison to other cases. γ and β bands also perform well in the proposed study, but the performances of δ and

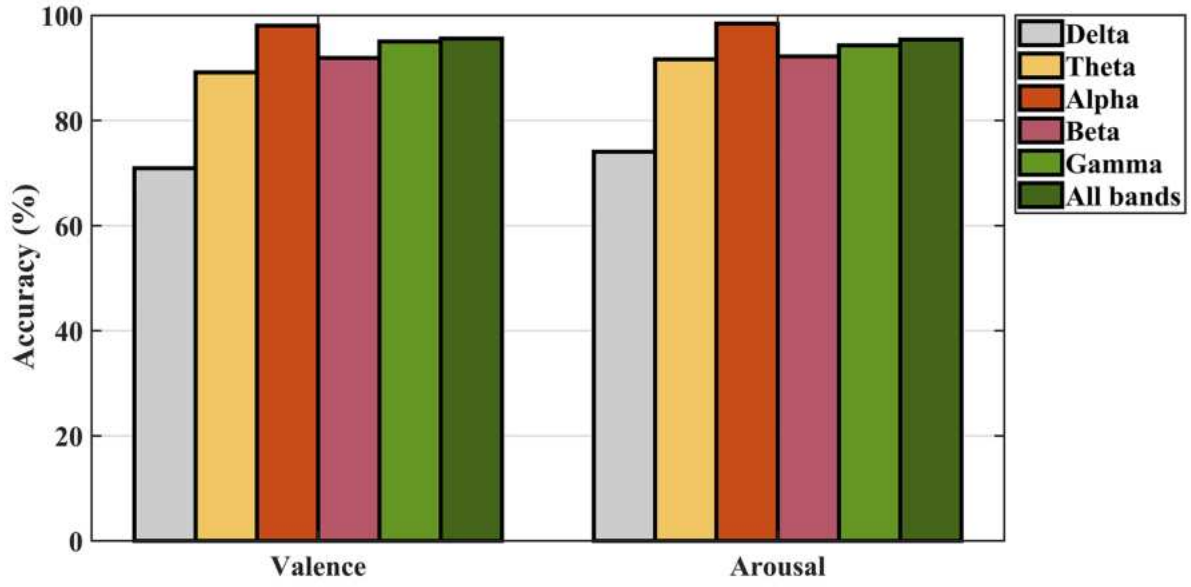


Figure 4.7: Classification performance (accuracy) for different frequency bands for valence and arousal cases

θ bands are found considerably lower. According to different accepted studies, α, β, and γ bands are dominant at awaken states and δ band is closely related to sleep stages and θ band is associated with intuition and relaxation as explained before [30, 38]. The results obtained by the proposed scheme support the above findings. The performance of the proposed scheme is also observed on the raw EEG signal containing all the frequency bands. As expected, the performance is found very satisfactory when no frequency selection is performed (i.e. all bands are present).

4.3 Discrete Wavelet Transform (DWT)+ Continuous Wavelet Transform (CWT)

In this section, a thorough analysis is carried out taking the advantages of both discrete wavelet transform (DWT) and continuous wavelet transform (CWT). The proposed 3D frame effectively captures time-frequency domain behaviour that provides an efficient feature in the context of emotion recognition.

4.3.1 Proposed Method

In this section, an efficient feature extraction scheme of emotion recognition is proposed combining discrete wavelet transform (DWT) and continuous wavelet transform (CWT). As both

DWT and CWT preserve time-frequency content of the raw EEG signal, this proposed approach is effective in classifying different emotional states. The raw EEG signal for each channel is first decomposed into multi-level discrete wavelet coefficients using DWT and third-level detail coefficients are selected for subsequent operation. Later the decomposed signal is divided into multiple overlapping frames. For each frame, an efficient feature representation namely strength-to-entropy component ratio (SECR) is obtained. CWT co-efficients of the decomposed signal effectively captures the information in time-frequency domain which contributes to the emotion classification process. Finally, by concatenating the feature from all selected channels, a feature matrix is generated which is fed to a deep neural network. In this work, extensive experimentation is carried out on DEAP dataset.

Pre-processing

In the pre-processing stage of the proposed method, the baseline $(\mathbf{X}_b)$ of the raw EEG signal is excluded. The acquired EEG signal from each channel $(\mathbf{X}_r)$ contains the baseline and trial data where $(\mathbf{X}_r = [\mathbf{X}_t, \mathbf{X}_r] \in R^{P\times L})$ and L denotes the length of the raw data. Since baseline data does not contain the actual stimuli involved in different emotional states, removing it from the signal can improve the emotion recognition performance. The trial data $(\mathbf{X}_t)$ can be obtained as follows:

$$\mathbf{X}_t = \mathbf{X}_r - \mathbf{X}_b \tag{4.16}$$

where $\mathbf{X}_t \in R^{P\times L_t}$ and $\mathbf{X}_b \in R^{P\times L_b}$.

Wavelet Decomposition and Windowing

Wavelet transform offers simultaneous localization in time and frequency domain which enables to uncover the fine details of the signal. DWT coefficients of signal $(x(t))$ can be expressed as:

$$\gamma_{jk} = \int x(t) \cdot \psi_{j,k}(t)\, dt \tag{4.17}$$

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \cdot \psi(\frac{t - k \cdot 2^j}{2^j}) \tag{4.18}$$

In (4.18), γ_{jk} denotes the wavelet coefficients that can be viewed as convolution between $x(t)$ and dilated, reflected and normalized version of mother wavelets and $\psi(t)$ in (4.18) denotes the mother wavelet. The child wavelets at different translation and scales $(\psi_{j,k}(t))$ can be obtained with the appropriate choice of scaling and shifting parameters.

Discrete wavelet transform offers the advantage of multi-resolution analysis (MRA) through the multi-stage implementation of the filter banks.

Following the decomposition of EEG signal into L-level DWT coefficients, a set of wavelet vec-

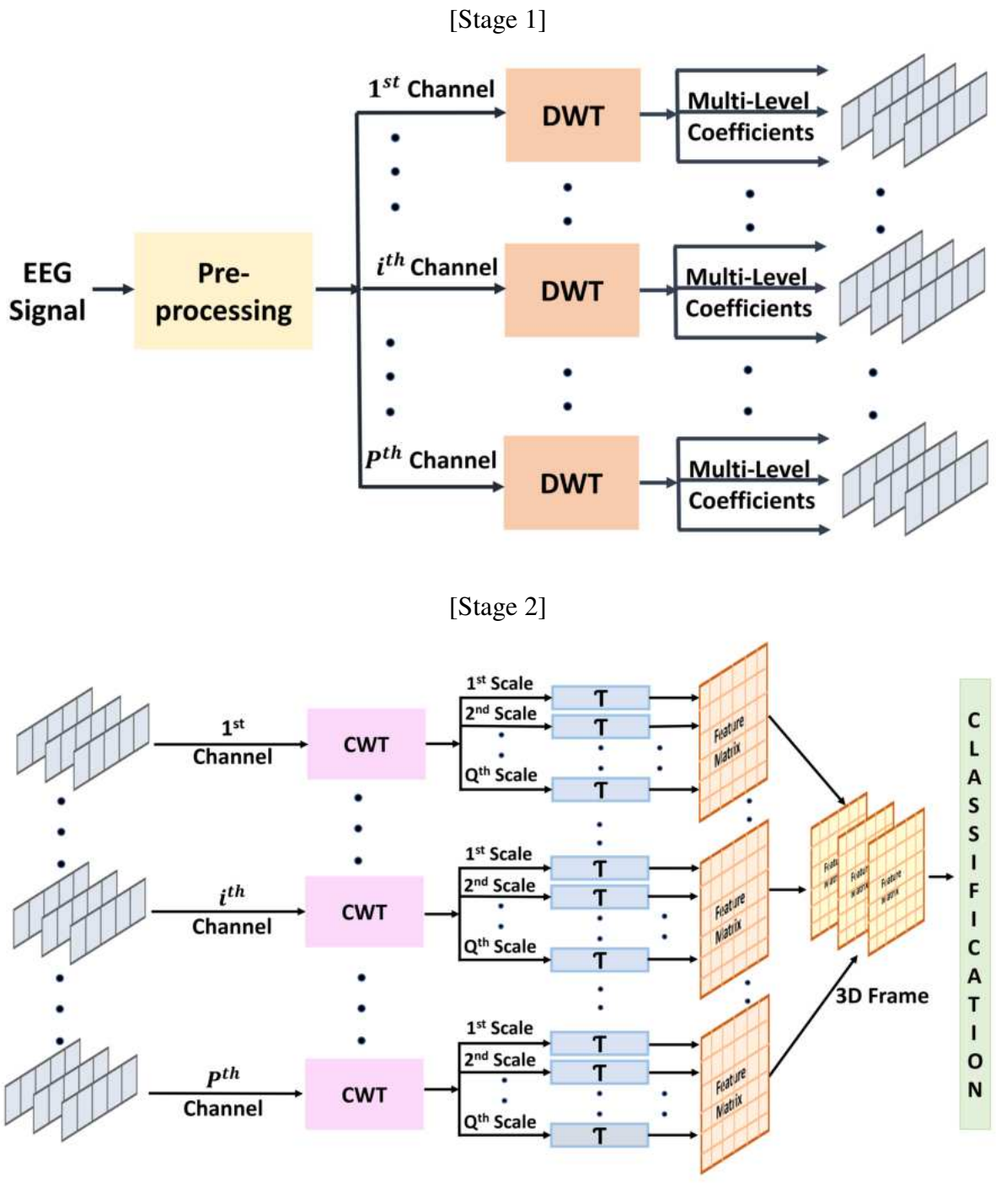


Figure 4.8: Workflow in the proposed method

tors $\mathbf{X}_f^i$ is formed for a particular channel i where $\mathbf{X}_f^i = \{\mathbf{X}_{l_1}^T, \mathbf{X}_{l_2}^T, \ldots, \mathbf{X}_{l_L}^T\}$, $i \in [1, 2, \ldots, P]$ and $\mathbf{X}_{l_L} \in R^{1\times L_t}$. The wavelet vectors preserve the time-frequency resolution of the EEG data which provide inherent characteristics of the signal and is effective in capturing different emotional states.

Each level wavelet vector is then windowed with a frame length of W and a set of window segments $(\mathbf{X}_{w_k}^i)$ are formed for the k^{th} segment of the i^{th} channel where $\mathbf{X}_{w_k}^i = \{\mathbf{X}_{w_{l_1}}^{iT}, \mathbf{X}_{w_{l_2}}^{iT}, \ldots, \mathbf{X}_{w_{l_L}}^{iT}\}$ where $\mathbf{X}_{w_{l_L}}$ is a window segment for the L^{th} level and $\mathbf{X}_{w_{l_L}} \in R^{1\times W}$.

Proposed Feature Extraction Scheme

In this part, continuous wavelet transform (CWT) based feature extraction scheme is proposed. In the case of EEG signal, signal strength varies in different portion of the brain that depends on the neural firing. In this regard, the strength of the wavelet co-efficients can reflect the effective time-frequency domain information that can be expressed as

$$E_{a_j,b_l} = |WT_\psi\{x\}(a_j, b_l)|^2, \tag{4.19}$$

and the energy associated with a particular frequency band can be calculated as

$$E_{a_j} = \sum_{l=1}^{W} E_{a_j,b_l}. \tag{4.20}$$

Since, different scales of CWT encapsulate the information of different frequencies, energy of a particular scale of CWT coefficients (E_{a_j}) contain necessary information associated with that scale. Considering the inherent random variation of EEG signal due to different types of emotions, the entropy corresponding to the strength component can be calculated as

$$H_{a_j} = \sum_{l=1}^{W} H_{a_j,b_l} = \sum_{l=1}^{W} P_{a_j,b_l} \cdot \log \frac{1}{P_{a_j,b_l}}. \tag{4.21}$$

In (4.21), the entropy component (H_{a_j}) of a particular scale captures the randomness associated with each time point and (P_{a_j,b_l}) of each CWT coefficient $(b_l; l = 1, 2, \cdots W)$ corresponding to a scale $(a_j; j = 1, 2, 3, \cdots, Q)$ and entropy of a particular scale H_{a_j} can be expressed as the sum of entropy components.

In this regard, the combination of strength-to-entropy component ratio (SECR) can provide a better representative feature characteristics of EEG signal.

In the proposed method, the strength-to-entropy component ratio (SECR) of the extracted CWT coefficients are considered as they provide an informative feature in the context of emotion recognition. The CWT coefficients are extracted from the set of decomposed wavelet coefficients $(\mathbf{X}_{f_k}^i)$ of the i^{th} channel and the coefficient matrix X_g^i is formed where $X_g^i = [\mathbf{X}_{g_i}^{1\ T}, \mathbf{X}_{g_i}^{2\ T}, \cdot\cdot$

$\cdot\cdot \mathbf{X}_{g_i}^{QT}]^T \in R^{Q\times W}$. Here, $\mathbf{X}_{g_i}^{j}$: CWT coefficients of j^{th} scale $(j = 1, 2, 3, \ldots, Q)$ and i^{th} channel $(i = 1, 2, 3, \ldots, P)$. Following the process, extracted X_{g_i} is mapped into a 2D feature representation $X_{m_i} \in \mathbf{R}^{Q\times W}$ by transforming each CWT coefficient $x_{g_i}^{j}$ into the SEC ratio defined as, $T\{x_{g_i}^{j}\} = \frac{E_{a_j,b_l}}{H_{a_j,b_l}}$. After concatenating each X_{m_i} from all the channels, the proposed CEF2D feature matrix $X_c \in \mathbf{R}^{F\times W}$ is obtained for a particular level DWT co-efficients $(\mathbf{X}_{l_d})(d = 1, 2 \cdot\cdot\cdot L)$;, where $F = P \times Q$.
After combining the DWT coefficients of all levels $(1, 2 \cdot\cdot\cdot L)$, a 3D frame X_z is obtained, where $X_z \in R^{F\times W\times L}$

4.3.2 Architecture of the Proposed Neural Network

The proposed 3D frame (X_z) is applied to a 2D CNN model for the purpose of emotion classification. The architecture follows the same as that of Table 4.3. The 2D deep neural network is trained to take the input shape of (F, W, L). The proposed network is trained to perform the binary classification task.

4.3.3 Results and Analysis

Experimental Setup

In this stage of experimentaion, first 3 second data is discarded as baseline $(\mathbf{X}_b)$ from the recorded 63 seconds of raw EEG data $(\mathbf{X}_r)$. The remaining 60s data is divided into multiple sub-frames with a window length of 2 second and 1.75 second overlap. In this study, third level detail coefficients are utilized for the experimentation purpose. Following the feature extraction process from all window segments, the final dimension of the data is $(9320, 200, 256, 1)$. The proposed approach is designed to classify the valence and arousal dimensions into binary classes.
The high labels (high valence/arousal) are assigned when the rating is ≥ 5 and low class is defined otherwise.
All experimentations are performed in Google Colaboratory Platform.

Model Training, Validation and Testing

The proposed method is designed for the subject-dependent study. At first, the data is divided into 80% training and 20% testing data. In the training stage is run for 200 epochs. During validation, 10-fold cross-validation scheme is employed. The learning rate is set to 0.001 with $Adam$ optimizer. The loss function is categorical cross-entropy and metrics are accuracy.

4.3.4 Performance Evaluation

The performance of the proposed scheme is evaluated with accuracy and F1-score. The results for binary classification are recorded in Table 4.6. The average accuracy and F1-score for both valence and arousal dimensions are found high for each subject. The lower standard deviation indicates the consistent performance and low variability among different subjects.

4.3.5 Conclusion

In conventional time and frequency domain analysis, the significant time-frequency localization property of the EEG signal are ignored. The extracted temporal and spectral features do not truly reflect the inherent characteristics of the EEG signal which is crucial to map the neural activities for the purpose of emotion classification. The discrete wavelet transform discussed in section 4.1 maps the coefficients from time domain to wavelet domain which captures the salient characteristics of the EEG data in the proposed 3D frame. As the frame effectively preserves the spatial and time-frequency characteristics the proposed scheme offers a consistent performance.

In section 4.2, an entropy-based efficient 2D feature representation of EEG signal in CWT domain is proposed for emotion recognition using a CNN model. In the proposed method, CWT coefficients of the multi-channel EEG data are mapped into a strength-to-entropy-component (SEC) ratio and CEF2D feature matrix is obtained, which effectively optimizes emotion recognition performance. As CWT coefficients contain information in the time-frequency domain, the mapped 2D feature is efficacious in capturing detailed information of the EEG data. Here the choice of α band compared to other frequency bands exhibits better emotion recognition performance. Instead of using all EEG channels and CWT scales, an efficient approach of channel and scale selection is introduced, for which the energy-to-entropy ratio in the CWT domain is considered. It is observed that forming a 2D feature matrix with a reduced number of scales and channels can capture discriminative features for different emotional states and provide better classification performances. An optimal selection of 10 channels and 20 scales has decreased the dimension of the feature matrix by 10.24 times and hence reduced the computational burden. Utilizing fewer channels also lessens the discomfort of wearable EEG devices. From extensive experimentation in a subject-dependent study, a very consistent performance is obtained both in 3-class and binary class tasks of valence and arousal dimensions for all subjects, and the proposed method outperforms other state-of-the-art approaches.

Table 4.2: Performance of the proposed DWT-based decomposition method in binary class

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	96.83	96.83	99.24	99.25
02	95.67	95.11	96.17	95.98
03	98.83	98.82	98.75	98.96
04	90.25	89.93	93.50	93.28
05	97.83	97.76	97.66	97.66
06	97.17	96.10	91.75	91.47
07	97.5	97.10	96.75	96.52
08	95.25	95.19	97.58	97.51
09	95.5	95.46	96.5	96.15
10	98.58	98.58	98.5	98.5
11	88.92	88.2	89.5	89.00
12	95.83	95.82	98.17	98.42
13	92.17	91.95	95.75	92.60
14	92.58	92.57	95.25	94.6
15	98.08	98.08	98.42	98.41
16	97.83	97.74	98.33	98.33
17	97.17	97.13	97.08	96.89
18	98.75	98.61	97.75	97.64
19	99.00	98.99	97.50	97.12
20	98.33	98.31	98.08	96.75
21	96.25	96.24	99.42	98.99
22	98.42	98.42	98.00	97.70
23	97.00	96.49	98.75	98.53
24	99.25	99.25	98.42	97.32
25	95.17	95.17	94.83	92.97
26	94.25	93.78	92.17	91.97
27	98.5	97.69	98.33	98.01
28	93.42	92.96	94	93.99
29	98.42	98.36	97.92	97.45
30	97.99	97.75	97.92	97.92
31	97.00	96.81	95.67	95.67
32	96.25	96.25	96.83	96.28
Average	**96.37±2.57**	**96.17±2.66**	**96.67±2.38**	**96.31±2.54**

Table 4.3: Details of the proposed 2D CNN model

Layer Type	**Layer Parameters**	**Output Shape**
Conv2D	f=16, k=5, s=2	(100, 128, 16)
BatchNormalization	-	(100, 128, 16)
MaxPooling2D	*pool size* = 2	(50, 64, 16)
Dropout	*rate* = 0.2	(50, 64, 16)
Conv2D	f=32, k=7, s=2	(25, 32, 32)
BatchNormalization	-	(25, 32, 32)
MaxPooling2D	*pool size* = 2	(12, 16, 32)
Dropout	*rate* =0.2	(12, 16, 32)
Conv2D	f=64, k=9, s=2	(6, 8, 64)
BatchNormalization	-	(6, 8, 64)
MaxPooling2D	*pool size*= 2	(3, 4, 64)
Dropout	*rate*=0.2	(3, 4, 64)
Flatten	-	(768,)
Dense	*units* = 256	(256,)
Dropout	*rate*=0.4	(256,)
Dense	*units* = 32	(32,)
Dropout	*rate*=0.4	(32,)
Dense	*units* = 3	(3,)
Total no. of parameters		397, 123

Table 4.4: Performance of the proposed method in 3-class task

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	99.14	98.97	99.36	99.54
02	98.55	98.36	98.39	98.19
03	99.03	98.97	99.09	98.83
04	95.98	95.75	98.98	98.88
05	99.20	99.16	98.28	98.18
06	97.05	96.71	99.41	99.35
07	99.36	99.24	99.25	99.18
08	97.05	96.96	99.09	98.99
09	99.03	99.02	97.75	97.07
10	98.93	98.94	96.57	96.57
11	98.50	98.46	98.40	98.17
12	98.44	98.42	99.30	99.11
13	98.44	98.44	99.36	99.06
14	99.73	99.65	99.41	99.41
15	98.55	98.51	99.30	99.24
16	98.71	98.63	99.09	99.00
17	97.59	97.60	97.91	97.91
18	99.46	99.18	98.98	98.71
19	98.87	98.86	99.41	99.42
20	99.14	99.04	99.46	99.52
21	98.93	98.90	99.14	98.86
22	95.49	95.32	98.55	98.50
23	99.30	99.24	99.25	99.15
24	97.59	97.55	98.07	97.74
25	96.64	96.02	99.14	98.85
26	97.59	97.11	97.00	96.84
27	95.60	95.66	96.51	96.54
28	98.55	98.45	98.82	98.66
29	98.55	98.40	98.28	97.98
30	98.44	98.19	97.69	97.55
31	98.98	98.85	98.93	98.91
32	97.48	97.23	96.84	96.77
Average	**98.25±1.13**	**98.15±1.21**	**98.68±0.87**	**98.46±0.91**

Table 4.5: Performance of the proposed method in binary class

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	99.30	99.30	99.79	99.77
02	98.93	98.92	99.20	99.16
03	99.52	99.51	99.62	99.61
04	98.61	98.54	97.65	97.64
05	99.36	99.33	96.94	96.93
06	99.41	99.22	99.30	99.29
07	99.68	99.61	99.62	99.61
08	97.91	97.88	98.66	98.62
09	99.09	99.09	98.18	98.07
10	99.57	99.57	98.07	98.06
11	98.93	98.89	98.53	98.49
12	98.77	98.76	98.79	98.63
13	99.03	99.02	99.36	99.68
14	99.57	99.57	99.14	99.04
15	98.82	98.82	98.23	98.23
16	99.52	99.48	99.36	99.36
17	93.99	93.92	98.50	98.38
18	99.73	99.71	99.25	99.17
19	98.98	98.96	99.79	99.76
20	99.25	99.23	99.68	99.55
21	99.20	99.20	99.36	98.95
22	96.51	96.51	99.41	99.38
23	99.30	99.21	99.30	99.21
24	97.10	97.09	99.57	99.50
25	98.98	98.97	99.14	98.83
26	97.75	97.46	98.74	98.74
27	99.41	99.19	99.57	99.50
28	99.62	99.60	98.98	98.98
29	99.46	99.45	99.73	99.70
30	99.25	99.15	98.55	98.54
31	99.25	99.15	99.52	99.52
32	98.77	98.78	98.71	98.45
Average	**98.83±1.15**	**98.78±1.15**	**98.95±0.67**	**98.95±0.68**

Table 4.6: Performance of the proposed method in binary class

2*Subject	Valence		Arousal	
	Accuracy(%)	F-1 score(%)	Accuracy(%)	F-1 score(%)
01	98.98	98.99	99.24	99.25
02	99.27	99.23	99.27	99.23
03	99.51	99.50	99.83	99.72
04	97.65	97.54	98.50	98.44
05	99.42	99.39	99.55	99.55
06	97.58	96.73	98.81	98.78
07	98.04	98.85	97.79	97.67
08	98.55	98.53	98.89	98.87
09	99.31	99.31	99.41	99.37
10	99.83	99.83	99.70	99.70
11	96.62	96.52	98.46	98.43
12	98.06	98.06	98.27	96.73
13	94.84	94.76	99.52	99.02
14	98.51	98.51	98.22	97.99
15	99.49	99.49	99.25	99.25
16	99.28	99.23	99.41	99.41
17	94.27	94.23	98.39	98.27
18	98.43	98.30	99.44	99.40
19	99.46	99.45	99.83	99.81
20	99.14	99.12	99.83	99.76
21	98.09	98.09	99.34	99.59
22	99.92	99.92	99.97	99.97
23	98.32	98.10	98.13	97.84
24	99.50	99.49	99.87	99.76
25	98.11	99.11	99.30	99.05
26	96.23	95.75	97.17	97.07
27	98.88	98.47	99.04	98.88
28	97.48	97.32	99.12	99.12
29	99.38	99.37	99.60	99.55
30	99.44	99.38	99.05	99.05
31	98.41	98.22	99.22	99.21
32	98.71	98.71	99.28	99.14
Average	**98.4±1.34**	**98.33±1.41**	**99.08±0.68**	**98.96±0.80**

Chapter 5

Channel Selection and Scale Selection

5.1 Proposed Method

Since data storage and processing from multiple channels and scales are computationally expensive, a scale and channel selection scheme is designed in the CWT domain. In this regard, energy and entropy values of different scales of CWT coefficients for various channels are investigated by using (4.9) and (4.10), respectively. For a particular channel $(i; i = 1, 2, 3, \ldots, P)$, for each scale $(j; j = 1, 2, 3, \ldots, Q)$, the energy-to-entropy ratio EER_j^i is computed, where $EER_j^i = \frac{E_j^i}{H_j^i}$; E_j^i and H_j^i are the energy and the entropy of a particular scale (j) of a particular channel (i), respectively. From the EER values for all channels and scales, an EER matrix X_{m_k} is obtained where $X_{m_k}(k = 1, 2, \ldots, n) \in \mathbf{R}^{P \times Q}$ is for k-th trial and n is the total number of trials. In view of getting the overall behaviour of different subjects with respect to different audio-visual stimuli at different trials, considering the EEG signals in the DEAP dataset, an average of $X_{m_k}(k = 1, 2, \ldots, n)$ is computed and corresponding 3D plot is illustrated in Fig. 5.1. This figure represents an overall variation of EER values with respect to channels and scales in a 3D space considering all trials of the DEAP dataset. It is to be noted that the EER operation is performed in the CWT domain and here in the proposed method, Morlet Wavelet is used, which exhibits Gaussian-shaped properties [39]. From Fig. 5.1, three major observations are

1. The variation of EER values with respect to scales for each channel represents a Gaussian shape.
2. The scale corresponding to the maximum EER value of all channels lies approximately at the same level.
3. The Gaussian-shaped curve of the EER values remains concentrated within the short range of scales instead of expanding all over the scales.

Based on these observations, one may select a range of scales in the neighbourhood of the mean

scale, where the mean scale can be obtained by averaging the peak scale locations obtained from different channels. Depending on the spread of EER variations, if a wide range of scales is chosen around the mean scale, there is a possibility of getting some channels where the EER distribution is not significant. That is why a very close neighbourhood is preferred. In Fig. 5.1, the mean scale number is found to be in-between 19 and 20 and (as illustrated in Fig. 5.1) one possible choice could be 20 scales ranging from $11-30$ out of 64 scales. In order to select the number of channels, the maximum EER value (EER^i_{max} =max$\{EER^i_1, EER^i_2, \ldots, EER^i_j, \ldots, EER^i_Q\}$) among all scales for a particular channel (i) is considered. From the maximum EER value obtained for each channel (EER^i_{max}), a normalized plot of maximum EER values versus channel numbers sorted in the descending order is obtained. In order to reduce the computational com-

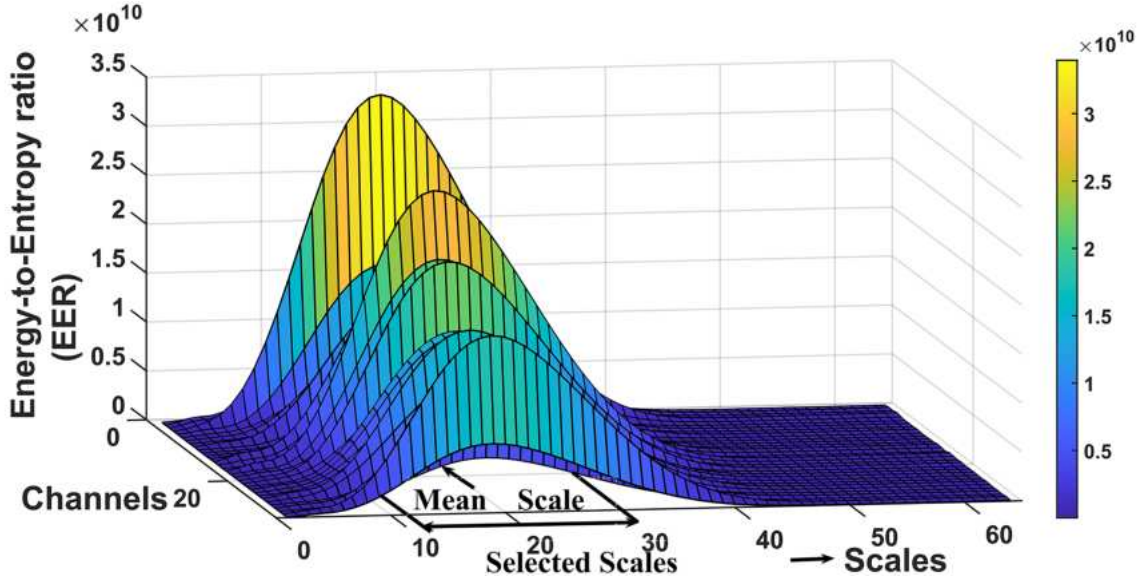


Figure 5.1: EER variations w.r.to channels and scales

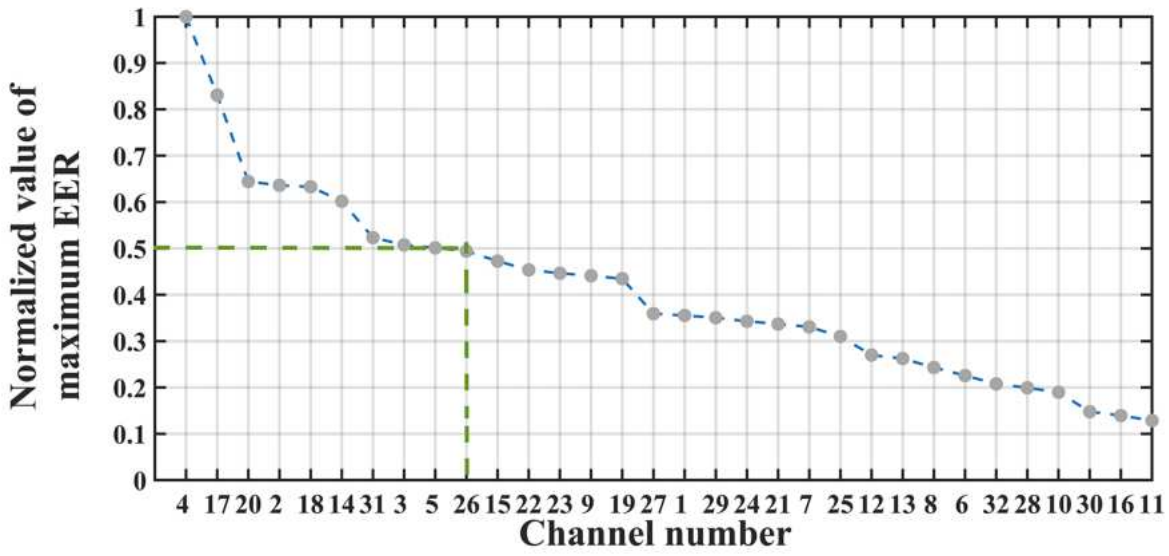


Figure 5.2: Normalized maximum EER variations for different channels

plexity, a threshold is selected from the plot and significant channels are singled out for which the normalized maximum EER has a higher value than the threshold. In the designed study on the DEAP dataset, the normalized plot of maximum EER value for all channels (sorted in the descending order) is displayed in Fig. 5.2. Here if a threshold of 0.5 is selected, a significant reduction in channel number can be obtained. Further reduction in the threshold value will allow selecting more channels with lower EER values. In case of competing EER values near the

Table 5.1: Sorted channels (descending order)

Rank	**Channel**	**Rank**	**Channel**	**Rank**	**Channel**	**Rank**	**Channel**
01	F7(04)	**09**	FC5(05)	**17**	FP1(01)	**25**	T7(08)
02	FP2(17)	**10**	T8(26)	**18**	P4(29)	**26**	FC1(06)
03	F4(20)	**11**	Oz(15)	**19**	Cz(24)	**27**	O2(32)
04	AF3(02)	**12**	FC6(22)	**20**	F8(21)	**28**	CP2(28)
05	AF4(18)	**13**	FC2(23)	**21**	C3(07)	**29**	CP1(10)
06	O1(14)	**14**	CP5(09)	**22**	C4(25)	**30**	P8(30)
07	PO4(31)	**15**	Fz(19)	**23**	P7(12)	**31**	Pz(16)
08	F3(03)	**16**	CP6(27)	**24**	PO3(13)	**32**	P3(11)

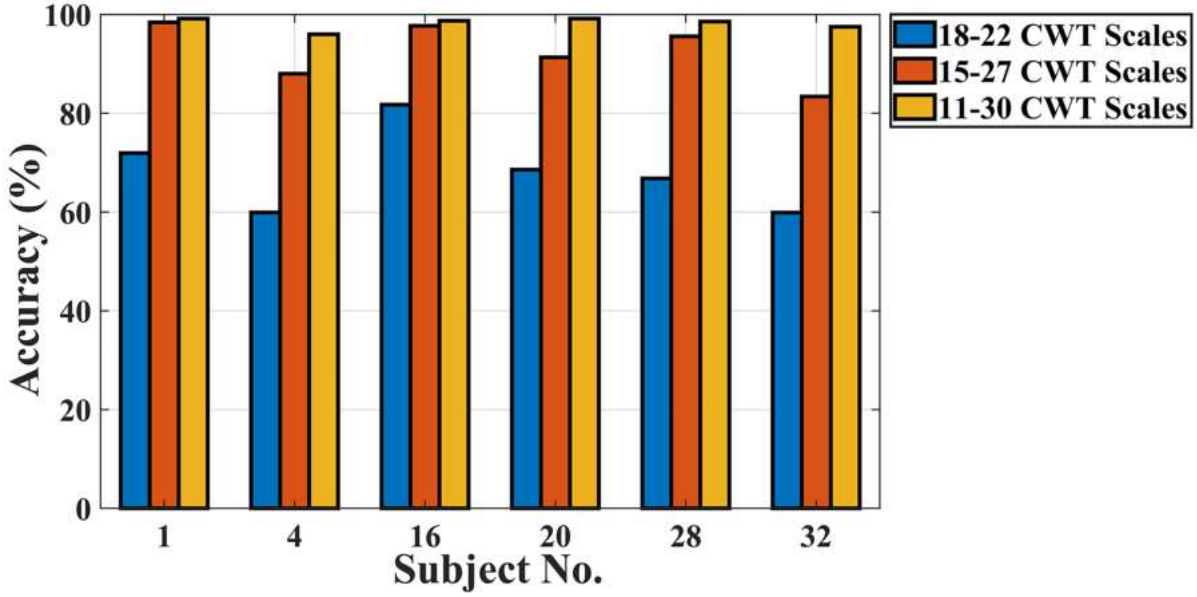
threshold, one may consider the spatial location of the channel on the brain. According to some studies, the frontal cortex plays a significant role in the elicitation of emotion [40, 41]. For this reason, comparatively higher-ranked and channels from the frontal cortex are preferred while choosing the channels with approximately the same EER value at the edge of the threshold. Following the proposed channel selection process, the total number of channels reduces from P to M. In the proposed method, for the DEAP dataset, the first 10 channels are selected based on the channel selection criteria described above. The layout of the sorted channels and the corresponding channel numbers are shown in Table 5.1 according to their rank. From Table 5.1, it is evident that the top five ranked channels are from the frontal region of the brain.
Finally, the proposed channel and scale selection scheme offers a channel reduction of P/M times as well as scale reduction of Q/N times, which combinedly reduces the dimension of the feature matrix by $(P \times Q)/(M \times N)$ times. In the case of the DEAP dataset, as per the selected values described above, channel reduces by 3.2 times and scale reduces by 3.2 times which combinedly reduces the feature dimension by 10.24 times. Such reduction in the feature matrix dimension eases the computation with negligible effects in the classification.

Effect of Scale Selection

Based on an analysis of the EER distribution, a very close neighbourhood around the mean-scale number is recommended for the purpose of CWT scale selection. As discussed in Section 5.1,

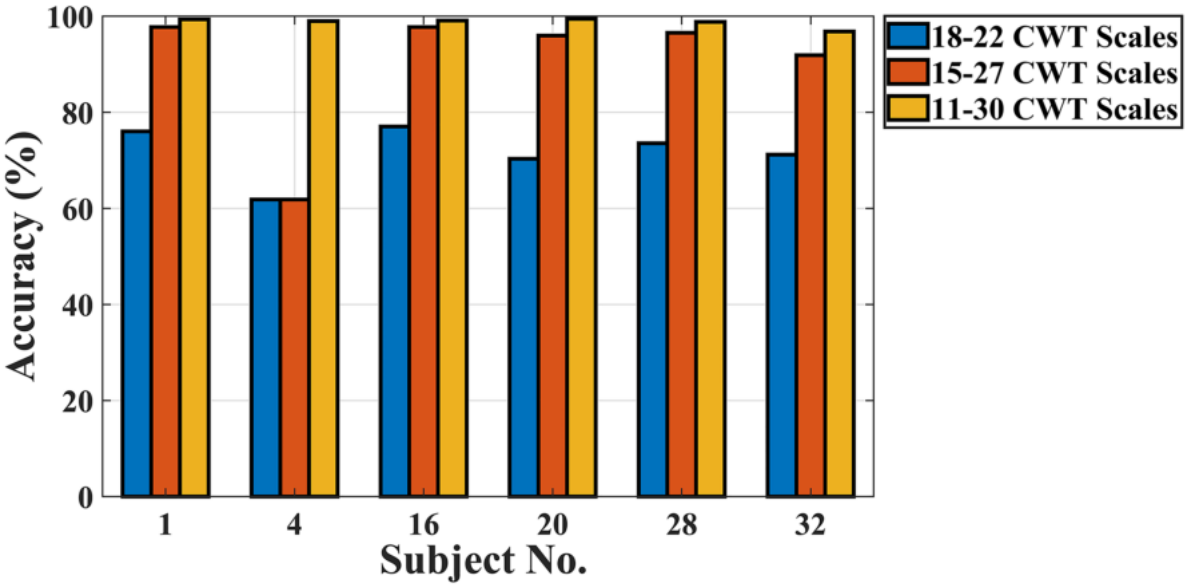
around the mean-scale in Fig. 5.1, twenty scales within $11 - 30$ out of 64 scales are selected and any contribution is rarely found in the 3D EER plot outside this range. We also analyze the choice of selecting more narrow range around the mean-scale, such as thirteen scales (15-27) and five scales (18-22). Classification performance (accuracy) for valence and arousal cases are shown in Fig. 5.3 considering six sample subjects. It is clearly observed that the classification performance deteriorates with the further decrease in number of scales. For example, for the valence dimension, the average accuracy for ranges 18-22, 15-27 and 11-30 are 63.05%, 91.96% and 98.00%, respectively. For arousal dimension, the average accuracy for ranges 18-22, 15-27 and 11-30 are 66.72%, 88.56% and 98.65% respectively. Hence, the choice of 20 scales (ranging from 11-30) out of 64 scales is found very satisfactory in terms of classification performance and computational complexity.

[Valence



case]

[Arousal



case]

Figure 5.3: Classification performance (accuracy) for different ranges of CWT scales for different subjects

Chapter 6

Conclusion

In this thesis, deep learning-based emotion recognition schemes are presented in three different domains, i.e., time, frequency and wavelet domain. Deep neural networks have been used to achieve high performance in many fields and such high performance has also been observed in EEG-based emotion recognition tasks. In this study, it is shown that instead of utilizing an LSTM-based deep neural network, CNN-based architecture extracts the local information better. Furthermore, different time-frequency domains, such as multi-band EEG signals and wavelet packet decomposed EEG signals, have been used since such band-limited signals in different domains can encapsulate neural activities better and thus a difference in terms of neural activity for different classes of emotion can be exploited. Different features, such as power spectral density in the frequency domain and entropy-based feature in the wavelet domain, have been extracted to generate feature variation patterns. Instead of utilizing the EEG signal in time or frequency domain analysis, wavelet analysis has been carried out to extract efficacious local information which provides high classification accuracy. In addition to extracting a more generalized feature, an efficient channel and scale selection scheme is also carried out in the wavelet domain. As a result, computational expenses for data storage and processing from multiple channels and scales are reduced. Detailed analyses and various types of investigation carried out on the publicly available DEAP database verify that the proposed methods are capable of classifying different types of emotion with high accuracy.

6.1 Contribution of this Thesis

The major contributions of the thesis can be summarized as follows:

- One of the main contributions of this work is to show the effective use of different frequency bands of EEG signals to classify different types of emotion. It is shown that the use of frequency information along with the time-domain information can drastically

improve classification accuracy. As a result, the wavelet domain can extract more local information for emotion recognition. Moreover, the different frequency band signals together also show a better classification performance, indicating that the frequency-domain information along with the time-domain information is more efficient for the emotion recognition task.

- Another significant contribution of this work is to introduce an efficient approach channel and scale selection technique to reduce the computational burden. For which the energy-to-entropy ratio in the CWT domain is considered. It is observed that forming a 2D feature matrix with a reduced number of scales and channels can capture discriminative features for different emotional states and provide better classification performances.

- In this work, a frequency band information (FBI) block and an inter-channel relationship (ICR) block are proposed using CNN-based neural architecture for the classification task. It is shown that the use of these blocks is effective for emotion recognition which can extract informative deep features to provide better results.

- In order to extract the information about the spatial location of the EEG electrodes, an effective scheme of grouping EEG channels is proposed in this study. Moreover, a channel-wise attention mechanism is also introduced to distribute the significance of EEG channels. Channel-wise attention can extract more detailed information about channels with the change of the weights of different channels by exploring the information of the feature map. It is to be noted that the use of the channel-wise attention mechanism along with the scheme of grouping EEG channels can drastically improve classification accuracy.

- Detailed simulations have been carried out to investigate the performance of the proposed method for the authentication and identification using EEG signals available from the DEAP database. The performance of the proposed method is compared with state-of-the-art methods using the same database. In this study, significant analyses have been presented to state that the proposed method can outperform the state-of-the-art methods in terms of performance parameters, such as accuracy and F1 score.

6.2 Scopes for Future Work

However, there are still some scopes for future research, as mentioned below:

- The trial-based and inter-subject EEG emotion recognition can be studied as a future work of this thesis.

- Available database other than DEAP database can be utilized for validating efficacy of the proposed method.

- Need to be applied in real-life conditions in order to implement real-time emotion recognition systems.

References

[1] J. Gross and R. Thompson, "Emotion regulation: Conceptual foundations," *Handbook of Emotion Regulation*, pp. 3–27, 01 2007.

[2] H. Huang, Q. Xie, J. Pan, Y. He, Z. Wen, R. Yu, and Y. Li, "An EEG-based brain computer interface for emotion recognition and its application in patients with disorder of consciousness," *IEEE Trans. on Affective Computing*, vol. 12, no. 4, pp. 832–842, 2021.

[3] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine," in *Mayo clinic proceedings*, vol. 87, pp. 268–279, Elsevier, 2012.

[4] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[5] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *Journal of Research in Personality*, vol. 11, no. 3, pp. 273–294, 1977.

[6] P. Sarkar and A. Etemad, "Self-supervised learning for ECG-based emotion recognition," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3217–3221, IEEE, 2020.

[7] K. Wang, N. An, B. N. Li, Y. Zhang, and L. Li, "Speech emotion recognition using fourier parameters," *IEEE Transactions on affective computing*, vol. 6, no. 1, pp. 69–75, 2015.

[8] V. Kehri, R. Ingle, S. Patil, and R. Awale, "Analysis of facial EMG signal for emotion recognition using wavelet packet transform and svm," in *Machine intelligence and signal analysis*, pp. 247–257, Springer, 2019.

[9] Y. Xu, G. Liu, M. Hao, W. Wen, and X. Huang, "Analysis of affective ECG signals toward emotion recognition," *Journal of Electronics (China)*, vol. 27, no. 1, pp. 8–14, 2010.

[10] X. Zhang, J. Liu, J. Shen, S. Li, K. Hou, B. Hu, J. Gao, and T. Zhang, "Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine," *IEEE Trans. on Cybernetics*, vol. 51, pp. 4386–4399, 2020.

[11] P. Pandey and K. R. Seeja, "Subject independent emotion recognition system for people with facial deformity: an EEG based approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2311–2320, 2021.

[12] F. Feradov and T. Ganchev, "Ranking of EEG time-domain features on the negative emotions recognition task," *Annual Journal of Electronics*, vol. 9, pp. 26–29, 2015.

[13] Z. Mohammadi, J. Frounchi, and M. Amiri, "Wavelet-based emotion recognition system using EEG signal," *Neural Computing and Applications*, vol. 28, no. 8, pp. 1985–1990, 2017.

[14] A. Mert and A. Akan, "Emotion recognition from EEG signals by using multivariate empirical mode decomposition," *Pattern Anal. Appl.*, vol. 21, no. 1, pp. 81–89, 2018.

[15] N. Salankar, P. Mishra, and L. Garg, "Emotion recognition from EEG signals using empirical mode decomposition and second-order difference plot," *Biomedical Signal Processing and Control*, vol. 65, p. 102389, 2021.

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

[17] N. Liu, Y. Fang, L. Li, L. Hou, F. Yang, and Y. Guo, "Multiple feature fusion for automatic emotion recognition using EEG signals," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 896–900, 2018.

[18] J. Cheng, M. Chen, C. Li, Y. Liu, R. Song, A. Liu, and X. Chen, "Emotion recognition from multi-channel EEG via deep forest," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 453–464, 2020.

[19] C. Wei, L. lan Chen, Z. zhen Song, X. guang Lou, and D. dong Li, "EEG-based emotion recognition using simple recurrent units network and ensemble learning," *Biomedical Signal Processing and Control*, vol. 58, p. 101756, 2020.

[20] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.

[21] Y. Chen, R. Chang, and J. Guo, "Effects of data augmentation method borderline-smote on emotion recognition of EEG signals based on convolutional neural network," *IEEE Access*, vol. 9, pp. 47491–47502, 2021.

[22] F. Demir, N. Sobahi, S. Siuly, and A. Sengur, "Exploring deep learning features for automatic classification of human emotion using EEG rhythms," *IEEE Sensors Journal*, vol. 21, no. 13, pp. 14923–14930, 2021.

[23] Y. Yin, X. Zheng, B. Hu, Y. Zhang, and X. Cui, "EEG emotion recognition using fusion model of graph convolutional neural networks and LSTM," *Appl. Soft Comput.*, vol. 100, p. 106954, 2021.

[24] M. Algarni, F. Saeed, T. Al-Hadhrami, F. Ghabban, and M. Al-Sarem, "Deep learning-based approach for emotion recognition using electroencephalography (EEG) signals using bi-directional long short-term memory (Bi-LSTM)," *Sensors*, vol. 22, no. 8, 2022.

[25] M. R. Islam, M. M. Islam, M. M. Rahman, C. Mondal, S. K. Singha, M. Ahmad, A. Awal, M. S. Islam, and M. A. Moni, "EEG channel correlation based model for emotion recognition," *Computers in Biology and Medicine*, vol. 136, p. 104757, 2021.

[26] A. S. Rajpoot, M. R. Panicker, *et al.*, "Subject independent emotion recognition using EEG signals employing attention driven neural networks," *Biomedical Signal Processing and Control*, vol. 75, p. 103547, 2022.

[27] Z.-M. Wang, S.-Y. Hu, and H. Song, "Channel selection method for EEG emotion recognition using normalized mutual information," *IEEE Access*, vol. 7, pp. 143303–143311, 2019.

[28] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for emotion analysis; using physiological signals," *IEEE Trans. on Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012.

[29] T. Alotaiby, F. E. Abd El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for eeg signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–21, 2015.

[30] P. A. Abhang, B. W. Gawali, and S. C. Mehrotra, "Technical aspects of brain rhythms and speech parameters," in *Introduction to EEG- and Speech-Based Emotion Recognition*, ch. 3, pp. 51–79, 2016.

[31] J. D. Kropotov, "Beta and gamma rhythms," in *Functional Neuromarkers for Psychiatry*, ch. 2.3, pp. 107–119, San Diego: Academic Press, 2016.

[32] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion recognition from multiband EEG signals using capsnet," *Sensors*, vol. 19, no. 9, 2019.

[33] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network," in *2018 international joint conference on neural networks (IJCNN)*, pp. 1–7, IEEE, 2018.

[34] J. Sadowsky, "The continuous wavelet transform: a tool for signal investigation and understanding," *Johns Hopkins APL Technical Digest*, vol. 15, pp. 306–306, 1994.

[35] S. J. A. Grinsted, J. C. Moore, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear Processes in Geophysics, European Geosciences Union*, vol. 11, p. 561–566.

[36] P. Kennedy, "Changes in emotional state modulate neuronal firing rates of human speech motor cortex: A case study in long-term recording," *Neurocase*, vol. 17, no. 5, pp. 381–393, 2011.

[37] M. D. Nunez, P. L. Nunez, R. Srinivasan, H. Ombao, M. Linquist, W. Thompson, and J. Aston, "Electroencephalography (EEG): neurophysics, experimental methods, and signal processing," *Handbook of neuroimaging data analysis*, pp. 175–197, 2016.

[38] Y. Kurihara, K. Watanabe, T. Nakamura, and H. Tanaka, "Unconstrained estimation method of delta-wave percentage included in EEG of sleeping subjects," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 607–615, 2011.

[39] H.-C. Shyu and Y.-S. Sun, "Construction of a morlet wavelet power spectrum," *Multidimensional Systems and Signal Processing*, vol. 13, no. 1, pp. 101–111, 2002.

[40] L. A. Schmidt and L. J. Trainor, "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions," *Cognition & Emotion*, vol. 15, no. 4, pp. 487–500, 2001.

[41] J. A. Coan, J. J. Allen, and E. Harmon-Jones, "Voluntary facial expression and hemispheric asymmetry over the frontal cortex," *Psychophysiology*, vol. 38, no. 6, pp. 912–925, 2001.

Appendix A

Algorithms

A.1 Sample Algorithm

In Algorithm 2 we show how to calcute $y = x^n$.

Algorithm 2 Calculate $y = x^n$

Require: $n \geq 0 \vee x \neq 0$
Ensure: $y = x^n$
  $y \leftarrow 1$
  if $n < 0$ **then**
    $X \leftarrow 1/x$
    $N \leftarrow -n$
  else
    $X \leftarrow x$
    $N \leftarrow n$
  end if
  while $N \neq 0$ **do**
    if N is even **then**
      $X \leftarrow X \times X$
      $N \leftarrow N/2$
    else $\{N$ is odd$\}$
      $y \leftarrow y \times X$
      $N \leftarrow N - 1$
    end if
  end while

Appendix B

Codes

B.1 Sample Code

We use this code to find out...

```
1 #include <stdio.h>
2 int Fibonacci(int);
3
4 main()
5 {
6   int n, i = 0, c;
7
8   printf("Enter␣the␣value␣of␣n:␣");
9   scanf("%d",&n);
10
11  printf("\nFibonacci␣series\n");
12
13  for (c = 1 ; c <= n ; c++)
14    {
15      printf("%d\n", Fibonacci(i));
16      i++;
17    }
18
19  return 0;
20 }
21
22 int Fibonacci(int n)
23 {
```

```
24   if (n == 0)
25     return 0;
26   else if (n == 1)
27     return 1;
28   else
29     return (Fibonacci(n-1) + Fibonacci(n-2));
30 }
```

B.2 Another Sample Code

```
1  SELECT associations2.object_id, associations2.term_id,
2         associations2.cat_ID, associations2.term_taxonomy_id
3  FROM (SELECT objects_tags.object_id, objects_tags.term_id,
4       wp_cb_tags2cats.cat_ID, categories.term_taxonomy_id
5  FROM (SELECT wp_term_relationships.object_id,
6       wp_term_taxonomy.term_id, wp_term_taxonomy.term_taxonomy_id
7  FROM wp_term_relationships
8  LEFT JOIN wp_term_taxonomy ON
9       wp_term_relationships.term_taxonomy_id =
10      wp_term_taxonomy.term_taxonomy_id
11 ORDER BY object_id ASC, term_id ASC)
12 AS objects_tags
13 LEFT JOIN wp_cb_tags2cats ON objects_tags.term_id =
14      wp_cb_tags2cats.tag_ID
15 LEFT JOIN (SELECT wp_term_relationships.object_id,
16      wp_term_taxonomy.term_id as cat_ID,
17      wp_term_taxonomy.term_taxonomy_id
18 FROM wp_term_relationships
19 LEFT JOIN wp_term_taxonomy ON
20      wp_term_relationships.term_taxonomy_id =
21      wp_term_taxonomy.term_taxonomy_id
22 WHERE wp_term_taxonomy.taxonomy = 'category'
23 GROUP BY object_id, cat_ID, term_taxonomy_id
24 ORDER BY object_id, cat_ID, term_taxonomy_id)
25 AS categories on wp_cb_tags2cats.cat_ID = categories.term_id
26 WHERE objects_tags.term_id = wp_cb_tags2cats.tag_ID
27 GROUP BY object_id, term_id, cat_ID, term_taxonomy_id
28 ORDER BY object_id ASC, term_id ASC, cat_ID ASC)
29 AS associations2
30 LEFT JOIN categories ON associations2.object_id =
```

```
31      categories.object_id
32 WHERE associations2.cat_ID <> categories.cat_ID
33 GROUP BY object_id, term_id, cat_ID, term_taxonomy_id
34 ORDER BY object_id, term_id, cat_ID, term_taxonomy_id
```

Generated using Undegraduate Thesis LaTeX Template, Version 1.4. Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh.

This thesis was generated on Sunday 8th May, 2022 at 5:13am.