A Computational Model for Predicting Music Popularity: A Psychophysiological Study

T.M. Hegodaarachchi

A Computational Model for Predicting Music Popularity: A Psychophysiological Study

T.M. Hegodaarachchi
Index No.: 20000741

Supervisor: Dr. M.I.E Wikramasignhe
Co-Supervisor: Mr. N.H.P.I. Maduranga

April 2025

Submitted in partial fulfillment of the requirements of the
B.Sc. in Computer Science Final Year Project (SCS4224)

UCSC

Declaration

I certify that this dissertation does not incorporate, without acknowledgement, any material previously submitted for a degree or diploma in any university and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, be made available for photocopying and for interlibrary loans, and for the title and abstract to be made available to outside organizations.

Candidate: T.M. Hegodaarachchi

....
Signature

This is to certify that this dissertation is based on the work of Mr. T.M. Hegodaarachchi is under my supervision. The thesis has been prepared according to the format stipulated and is of acceptable standard.

Supervisor: Dr. M.I.E. Wrikramasinghe

....
Signature

Co-Supervisor: Mr. N.H.P.I. Maduranga

....
Signature

Acknowledgements

I would like to thank my research supervisor, Dr. Manjusri Wickramasinghe, and co-supervisor, Mr. Isuru Nanayakkara, for their invaluable guidance, support, and mentoring throughout the duration of this research project. Their contribution, expert knowledge, and insightful feedback were crucial in determining the direction and scope of this study.

I am indebted to the University of Colombo School of Computing (UCSC) for providing me with the opportunity to undertake this research project in the area of EEG and for funding the Emotiv Flex EEG Scanner and their software subscription, which were essential for this study.

Finally, I would like to express my sincere appreciation to the participants and subjects who generously volunteered their time and efforts for this study. Their willingness to participate in this research project has made a significant contribution to the advancement of knowledge in this field, and their contribution is highly valued and appreciated.

Abstract

Hit Song Science (HSS) is pivotal in hit song identification. Music producers and composers often rely on new musical pieces to succeed with no prior knowledge. Hit song science refers to the ability of identifying potential hits relying on various information such as no of streams, no of album sales, likes and listener responses. Among various techniques, Electroencephalogram (EEG) has gained significant attention in the research community as a potential source for gathering lister responses.

This research focuses on development of a computation model for predicting hit songs by analyzing psychophysiological responses. This research took a methodological approach in measuring psychophysiological responses to a set of songs provided by a music chart. The identified musical pieces were divided into hits and flops. Their responses were utilized in developing new deep learning models using Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) techniques for identifying hits and flops using listener responses.

We evaluated the models on two tasks: hit song classification and musical chart ranking prediction. The CNN model trained in Week 13 achieved the highest classification accuracy at 65.43%. For ranking prediction, the best performance was observed with the Week 10 model, which achieved a Mean Squared Error (MSE) of 179.03. These results highlight the potential of deep learning techniques in leveraging psychophysiological signals to improve the accuracy of hit song prediction.

The dataset acquired during the experiment is made available to the public, and researchers are encouraged to use it to test their own hit music identification techniques.

Contents

List of Figures

Acronyms

MIR	Music Information Retrieval
HSS	Hit Song Science
EEG	Electroencephalogram
fMRI	functional Magnetic Resonance Imaging
MER	Music Emotion Recognition
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
ICA	Independent Component Analysis
DWT	Discrete Wavelet Transform
CWT	Continuous Wavelet Transform
PCA	Principal Component Analysis
CNN	Convolutional Neural Networks
LSTM	Long Short Term Memory

Chapter 1

Introduction

This chapter aims to introduce the study by providing with a comprehensive background and contextual information on the study domain. The research problem along with the objectives and questions will be introduced and discussed. The significance of the research will also be highlighted accordingly.

1.1 Background

Music is a universal language that utilizes sequences of tones, duration, and qualities to create various patterns. Music evokes emotions and conveys meaning to the listeners (Murugappan, Ramachandran, and Sazali 2010; Juslin and Västfjäll 2008). With the advancement of technology, the music industry has undergone several significant transformations. Music industry once had its albums sold in CDs and the reports were generated according to its reach of the sales. Yet, with the advancement in technology streaming platforms have gathered most of the attention in the music industry. As shown in *Global Music Report 2024*[1], the overall streaming music share from the global revenue is 67%. With the continuous changes in the industry, understanding what makes a song popular and how to predict its success has become a topic of interest to the community.

Music Information Retrieval (MIR) is the discipline of extracting and organizing meaningful information from music (Burgoyne, Fujinaga, and Downie 2015). MIR can be iden-

[1] https://globalmusicreport.ifpi.org/

tified as a field of study with various practical applications. There are aspects of MIR that still need to be explored. Academic musicology, sociology, signal processing, informatics, computational intelligence, machine learning, or a combination of these disciplines may be the background of MIR (Agres et al. 2021). Moreover, with the advancement of technology, machine learning techniques have been a subject of interest to improve MIR domain (Corrêa and Rodrigues 2016). Therefore, another growing area in MIR is Hit Song Science (HSS), which enhances data analysis techniques to forecast the popularity of a song before it is published on the market (Ni et al. 2013).

Researchers have utilized various metrics to measure music popularity (J. Lee and J. S. Lee 2018). These metrics include charts and streaming data. Hit Song Science (HSS) problems do require to carry out different steps in order to proceed to conclusions. Among these phases, one of the main tasks is to define "Hit Song" in the industry (Seufitelli et al. 2023). In recent research, various matrics and methodologies have been utilized in order to identify popular music. These various methodologies have a significant weight in the conclusions (Berns and Moore 2012; Leeuwis et al. 2021; Soares Araujo, Pinheiro de Cristo, and Giusti 2019). By utilizing those identified popular music, researchers tend to analyze various correlation among different indicators in order to dive deep in to the elusive essence of what makes a song popular.

Hit Song Science (HSS) studies then utilize features identified in songs to compute its popularity using various analytical methods such as statistical analysis and machine learning algorithms (Soares Araujo, Pinheiro de Cristo, and Giusti 2019; Berns and Moore 2012; Rajagopalan and Kaneshiro 2023; Leeuwis et al. 2021). These identified features can be divided into two major branches,

1. Methods based on retrieving Music Features
2. Methods based on retrieving Listeners' Responses

Methods based on retrieving music features are carried out by getting acoustic features of music such as pitch, tempo and etc. These studies then analyze these features to identify the potential Hit Music using analytical methods (Soares Araujo, Pinheiro de Cristo, and Giusti 2019; J. Lee and J. S. Lee 2018).

However, the popularity of a song is largely dependent on the individuals who listen to it, since popularity is achieved when a song appeals to a wide audience (Berns and

Moore 2012). Various sensory techniques are available for measuring a listener's response to stimuli. Cai, X. Li, and Jinsong Li 2023 identified that Facial Expressions and Speech Analysis are some widely used behavioral methods that gather responses. Moreover, psychological methods such as evaluating measurements of Skin Conductance, Heart Rate, and Brain Activity can also be identified as good measurements in identifying music-elected changes in humans. Researchers employ a wide range of techniques to gather listeners' responses to a given musical stimuli (Cai, X. Li, and Jinsong Li 2023).

In recent years, music popularity has been an emerging topic among the research community. Yet only a few researches have been conducted related to the hit song science using psychophysiological measurements. However, the challenge of predicting the popularity of music based on physiological measurements is of broad research interest. The study proposed in this document aims to overcome the existing challenge of predicting the popularity of the song based on the psychophysiological method Electroencephalogram (EEG).

1.2 Problem Definition

Hit song identification using EEG signals involves multiple steps, including signal processing, feature extraction, and machine learning techniques. Hit Song Science (HSS) research follows a structured process, beginning with selecting hit and flop songs. Identifying relevant songs involves using various popularity metrics, such as music charts and streaming views for the specific song.

After identifying the relevant songs the next step involves presenting the stimuli, the songs, to the participants in order to capture their psychophysiological changes using Electroencephalogram (EEG). Once EEG data is collected, the preprocessing step starts which involves cleaning and filtering the signals to remove noise and artifacts (Dowding et al. 2015).

1.3 Research Aim, Questions and Objectives

1.3.1 Research Aim

Develop a computational model that predicts Hit Songs by analyzing listeners' psychophysiological responses.

1.3.2 Research Question

RQ1 How can Hit Songs be identified based on data from streaming platforms?

RQ2 What psychophysiological differences emerge when individuals listen to identified Hit Songs?

RQ3 Which machine learning and deep learning algorithms are most effective for identifying hit songs using psychophysiological responses?

1.3.3 Research Objective

RO1 Formulate a methodology for the identification of Hit Songs from music streaming platforms. (RQ1 Objective)

RO2 Identify Hit Songs and Flop Songs within a defined context using the established measurement criteria. (RQ1 Objective)

RO3 Identify the differences in psychophysiological responses elicited by Hit Songs and Flop Songs. (RQ2 Objective)

RO4 Extract relevant features from the observed differences in psychophysiological responses to the given songs. (RQ2 Objective)

RO5 Obtain the effectiveness of various machine learning and deep learning algorithms in predicting popular music using psychophysiological responses. (RQ3 Objective)

1.3.4 Justification of the Research

Hit music identification is a crucial component of Hit Song Science (HSS). Music producers and composers often create musical pieces without prior knowledge of their potential for commercial success. Despite the importance of accurately identifying songs with hit potential, this task remains challenging due to the complexity of the metrics used to evaluate musical success.

This research aims to develop a computational model for predicting hit songs by analyzing listeners' psychophysiological responses measured through electroencephalography (EEG). The study addresses several key questions: How can hit songs be identified using data from streaming platforms, What are the distinguishing psychophysiological response patterns elicited by hit music, And finally, how can these insights be integrated into a computational model for accurately identifying hit songs based on psychophysiological signals.

The results of this research will have a significant impact on several fields such as Hit Song Science (HSS), music producing and music structure analysis, by improving accuracy of identifying hit music and providing a good understand of the relationship between psychophysiological responses and music listening.

Chapter 2

Literature Review

Music Information Retrieval (MIR) is the discipline of extracting and organizing meaningful information from music (Burgoyne, Fujinaga, and Downie 2015). MIR can be identified as a field of study with various practical applications. There are aspects of MIR still need to be explored. Another growing area in MIR is Hit Song Science (HSS), which enhances data analysis techniques to forecast the popularity of a song before it is published on the market. HSS is a growing field in the research community that focuses on forecasting the popularity of a song before it is published (Seufitelli et al. 2023). The prediction of music popularity in HSS studies involves the analysis of data and the use of other modern computational mechanisms to intersect the domains of science and music. Hit songs in the music industry can be identified as a success in the industry. In HSS studies, hit can be classified using various measurements. The main task of an HSS study is to predict whether a given song is a potential hit. Hence, it can be said that these studies certify the distinction between hits and nonhits.

2.1 Music and Music Popularity

Music is a universal language that utilizes sequences of tones, durations, and qualities to create various patterns. Music often evokes emotions and conveys meaning in various cultures. In Antony, Priya.V, and Gayathri 2018, music is identified as an art form that organizes sound pieces in time series. This new art form has been shown to significantly induce pleasure and emotions in listeners who consume the art piece. The widespread popularity of this art form has paved the way for the emergence of a new industry, now valued in the billions in revenue.

With the advancement of technology, the music industry has undergone several significant transformations. As shown in *GLOBAL MUSIC REPORT* 2024, the overall streaming music share from the global revenue is 67%. The report further illustrates that the global music industry is valued at $28 billion, showing how it establishes significance in the current population. With the continuous changes happening in the industry, understanding what makes a song popular and how to predict its success has become a topic of interest to the community. Therefore, HSS in MIR studies have significance the dominance in hit song prediction.

In HSS researches, defining a success metric is a fundamental requirement as these researches involve predicting hits. Hence, defining metrics and measuring the success of a song can also be a broader topic in HSS research (Seufitelli et al. 2023). Therefore, one of the fundamental challenges is defining "Popular Music". In contrast to objective measurements such as sales data, popularity is a complex and subjective topic (Cillessen and Marks 2011). What appeals to audiences is influenced by cultural trends, individual tastes, and even transient feelings. It is difficult to capture this elusive essence in a dataset that is appropriate for machine learning analysis (Pachet 2012). Moreover, understanding music popularity remains a topic of great interest for music-related industries and researchers within the MIR community.

One of the measurements used to measure the popularity of music is record charts. A record chart, which is usually referred to as a music chart, is a system of ranking music based on its level of popularity over a specific period. These record charts are created with the consideration of several factors like the number of CDs, cassettes, and Long Plays sold; radio airplay; requests made to radio DJs; song selections made by radio listeners; and, more recently, the number of downloads and streams. The generated charts are mostly geographical-based, while some are dedicated to a certain musical genre. The most frequent period that a chart is being generated is a week. This allows the ability to generate summary charts for years and decades making those a significant tool for evaluation of performance of songs. Some of the popular music charts are the UK Singles

Chart, Billboard Hot 100, IMI International Top 20 Singles, Canadian Hot 100, etc[1].

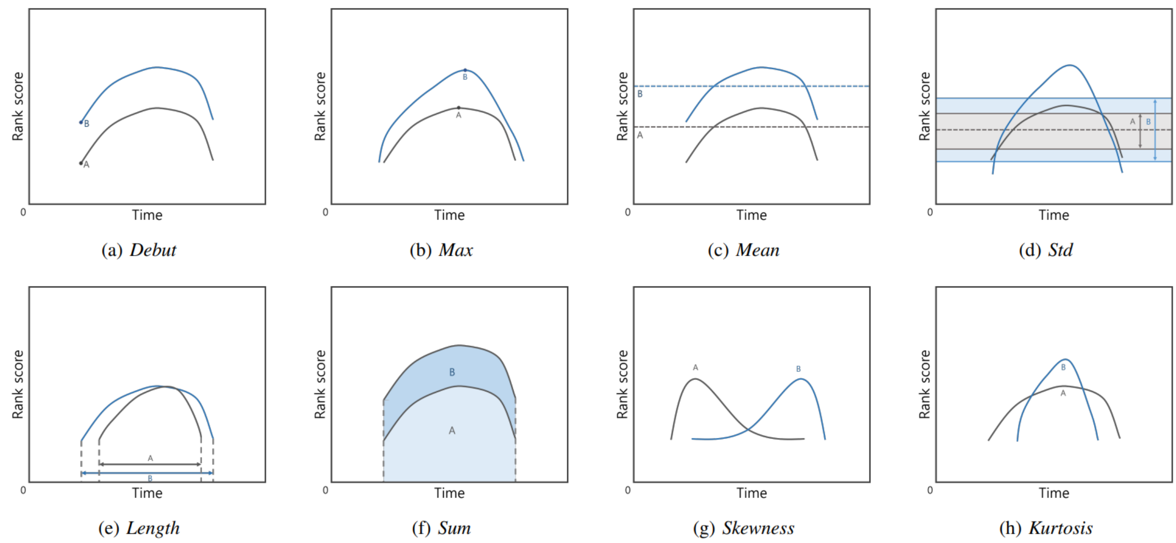
The entertainment magazine Billboard has been prominent in acknowledging the most popular music hits with the release of Billboard Hot 100 since 1958. Initially, the chart was created with the analysis of three main measurements. Initially, these measurements were, the number of times the single was played on the jukebox, the sales number per single, and the number of times the single was played on the radio (Napier and Shamir 2018). The 100 of the most appreciated songs are showcased by the Billboard Hot 100 music chart. The implementation of this record chart with the relevant measurements allows the community to capture both the initial popularity and assess the longevity of a single. With the advancements and the widespread admiration for the Billboard charts, the Billboard 200 for albums and Billboard Hot 100 for singles are the two most significant Billboard charts in the modern day. Other charts may be specialized to a particular genre, such as R&B, country, or rock, or they may encompass all genres. In recent times Billboard rankings take album sales, downloads, radio plays, social media interactions with songs, and online streaming figures to express the rank of a music[2].

Using music charts has been the most common evaluation of the population for most of the studies carried out in Hit Song Science (HSS) domain. In J. Lee and J. S. Lee 2018, authors have used music charts to generate popularity metrics. The idea of these matrices is to examine the both instantaneous and dynamic aspects of popularity. A rank score is introduced by the authors to propose eight popularity metrics that capture various features of popularity derived from the songs on the Billboard Hot 100. The introduced metrics are debut, max, mean, std, length, sum, skewness, and kurtosis. Figure 2.2 shows an illustration of these metrics. These metrics take different aspects of a song to measure popularity. In the debut metric, the use of the rank score when a song first appears on the chart is evaluated. The maximum position that a song has reached is evaluated in Max metrics to analyze the songs. Mean, std, and length are used to describe the average rank, the variation of the rank, and the stability of the popular songs. The sum, skewness, and kurtosis metrics describe the overall popularity of the song, a dynamic pattern in which a song gains or loses popularity, and patterns of growth and declining popularity accordingly.

Popularity measurements in HSS studies do carry a variety of methods in the lit-

[1]https://en.wikipedia.org/wiki/List_of_record_charts

[2]https://www.billboard.com/billboard-charts-legend/



(a) *Debut* (b) *Max* (c) *Mean* (d) *Std*
(e) *Length* (f) *Sum* (g) *Skewness* (h) *Kurtosis*

Figure 2.1: Billboard Chart Based Popularity Metrics J. Lee and J. S. Lee 2018

erature. These methods include music charts, albums sold, number of streams, and popularity matrices. However, a universally accepted method for measuring popularity in HSS studies has not yet been identified. Therefore there exists a gap in the literature for identifying popular music in HSS studies.

2.2 Music as a Stimuli

Most HSS studies leverage musical features from widely available datasets, APIs, and other task-based evaluations to identify hit songs using computational models or statistical methods (Zangerle et al. 2019; Middlebrook and Sheik 2019; Yee and Raheem 2022).

In Zangerle et al. 2019 the the authors define a “Hit” as any song appearing in the Billboard Hot 100 and utilize the Million Song Dataset for audio analysis. Audio features were extracted via the Essentia toolkit and these features were categorized in to two feature labels namely, low-level features and high-level features. These features were used to train a neural network for the prediction of hits. The proposed neural network architecture separates low-level features into a “deep” component and high-level features into a “wide” component. These were concatenated and passed through dense layers for regression, predicting a song’s peak Billboard rank. Results demonstrated that combining low- and high-level features with release year achieved the best performance:

Root Mean Squared Error (RMSE) of 55.45, Mean Absolute Error (MAE) of 43.84, and 75.04% precision in classifying hits vs. nonhits. However, this research carries the limitation of Western music bias, class imbalance in real-world data, and exclusion of external factors. Moreover, to calculate popularity, human intervention is much needed to capture subjective responses as in a population.

Middlebrook and Sheik 2019 predicts Billboard Hot 100 hits using Spotify audio features and metadata. Spotify API gives the data on songs which includes features like danceability, valence, energy, and track metadata like duration, explicitness. The authors compiled a dataset of 1.8 million Spotify tracks and 16,000 Billboard hits, merging them via track-artist matching. By exploiting these features, authors used four classification models: Logistic Regression, Neural Network, Random Forest, and Support Vector Machine. Training used scikit-learn with hyperparameter tuning via grid search. The Random Forest model achieved the highest accuracy: 88.7%. SVM delivered exceptional precision of 99.5% but lower recall 70.6%, making it suitable for minimizing false positives which is critical for music labels investing in potential hits. The inclusion of artist past-performance improved predictions, indicating established artists' higher hit likelihood. However, this study does have limitations in exclusion of social factors and a focus on Billboard charts, which may not reflect global or genre-specific success.

Another such popularity-predicting feature is exploring the acoustic features of a given music. The main features used in recent works are acousticness, danceability, energy, instrumentalness, key, liveness, duration, mode, speechiness, tempo, time signature, and valence. Such features are now available on Spotify API since March 2014[3]. By analyzing these acoustic features, research on predicting popularity has been conducted. In Soares Araujo, Pinheiro de Cristo, and Giusti 2019, authors have investigated acoustic features from Spotify API for predicting the popularity of a given song. They collected data from Spotify's Top 50 Global chart in order to conduct the research. By utilizing those acoustic features they tried predicting whether a song would be popular two months later.

These studies indicate that most HSS approaches prioritize musical features over human behavioral or emotional responses. This exclusion of human-centric data may represent a significant limitation in existing HSS frameworks. Consequently, integrating human responses into HSS research requires identifying reliable methods to measure

[3] `https://developer.spotify.com/documentation/web-api/reference/get-audio-features`

reactions to musical stimuli. Additionally, the limited exploration of human response collection in HSS studies underscores the need to advance MIR domains toward innovative mechanisms for capturing such responses.

In the field of music research, Music Emotion Recognition (MER) is recognized as a significant domain. These studies integrate Music Information Retrieval (MIR), machine learning, and computational methods to identify emotional responses elicited by music (Huq, Bello, and and 2010). Moreover, human responses are influenced by a wide range of stimuli and interferences, which manifest in diverse ways. These responses can be measured and analyzed to identify bodily changes, which may vary based on cultural background, social context, and living conditions, among many other factors. Within the emotion recognition domain, various methods are employed to capture emotional responses. In Cai, X. Li, and Jinsong Li 2023, several sensory techniques were reviewed for their application in emotion recognition. The study categorized these sensors into distinct groups, including visual sensors, audio sensors, radar sensors, and other physiological sensors. However, these methodologies can broadly be classified into two main categories: methods based on behavioral responses and methods based on physiological responses. Additionally, as HSS novel domain that has been emerging in recent years, it employs fewer physiological response evaluation techniques. Therefore, to gain a comprehensive understanding of how music functions as a stimulus to elicit responses, this study will examine both MER and HSS.

Behavioral-based methods involve facial expression, speech, gestures, and body movement to recognize emotions. The widely used methodology is facial expression analysis for emotion recognition in this method.

When it comes to facial expression certain parts need to be considered. The accuracy of these facial emotion recognition models decreases as the light intensity of the picture decreases (Hasinoff and Kutulakos 2011). Furthermore, humans can influence and express facial expressions to their liking which may result in capturing incorrect emotion (Zhao, Adib, and Katabi 2016). To illustrate an example we can relate to a moment where a particular person is engaging in a social activity where we usually express a pleasant mood even though we are not in a pleasant mood (J. Zhang et al. 2020). Furthermore, some systems experience difficulty in identifying facial expressions while an individual speaks continually, which results in inaccurate data measurement. Even when we consider actors who are well experienced in faking expression, the expression they have at the moment changes with what they actually feel. So when evaluating and processing these images, the continuous support of the user will be needed. Furthermore, segmenting and categorizing expressions is particularly challenging sometimes. None of the techniques produce consistent outcomes across all categories; for instance, anger is frequently mistaken for disgust. Factors like culture, different age groups, and several other social aspects also affect the accuracy of the system as they elicit variation in similar expressions (Saeed, Mahmood, and Y. D. Khan 2018).

Using speech recognition for expression recognition has its downsides as well. Identifying emotions from analyzing speech is a complex task. People tend to have different speaking styles from each other. This will result in acoustic variability from person to person. Hence this results in making it difficult for speech feature extraction and labeling (Burmania and Busso 2017). Furthermore, different emotions can be present in the same phrase while speaking. Some of these variations may be specific to the speaker's living environment or the culture in his local area, which further complicates speech emotion detection. In Fahad et al. 2021, according to the author, when speakers differ throughout the training and testing stages of the speech recognition model, the model's ability to detect emotions in speech can decrease. This scenario often occurs in natural settings. Therefore, when a speaker who has not been a part of the training process encounters the system, the system tends to lack the ability to recognize the emotion. Moreover, the author has discussed the downgrades of speech emotion recognition systems by dividing the systems into four categories.

In conclusion, behavioral methods can be unreliable in accurately categorizing emotions and capturing human responses to music. Existing literature indicates that these measurements can be intentionally altered by individuals and may vary across different cultural backgrounds. Consequently, relying solely on these methods can lead to misleading conclusions. Therefore, it is essential to employ a measurement approach that captures bodily changes in a manner that prevents participants from consciously modifying their responses.

Physiological methods involve different methods to capture body signals to recognize stimuli responses. Here these methods can be categorized into two main categories namely invasive and non-invasive methods. Invasive methods are carried out by insertion of an instrument through the layer of the skin or into a body orifice, whereas non-invasive methods do not require penetration of the body's surface or openings. When dealing with invasive methods the need for an expert or an operator is essential. Therefore, dealing with a reliable non-invasive method will improve the safety of experiments carried out.

In physiological methods, researchers peek into the realm of emotion recognition by monitoring various bodily changes. This involves various several measuring techniques. This mostly involves measuring different types of human body parameters. Another mostly involved measuring method is electrical pulses in the nervous system. After capturing these parameters, the signals then can be analyzed to understand the human changes. Among the available techniques, some of the most popular techniques are skin resistance measurements, electroencephalography (EEG), blood pressure, eye activity, heart rate, and motion analysis (Dzedzickis, Kaklauskas, and Bucinskas 2020).

In analysis of the functionality of the heart, one of the best and most powerful diagnostic tools used in the medical field is considered to be Electrocardiography (ECG). Although it is a useful tool for analyzing feedback in stimuli response the main drawback of ECG is that, when used for a long period, it produces huge amounts of data (Khatib et al. 2007). The involvement of ECG analysis in real-world settings has encouraged its frequent use in combination with other measuring methods for emotion recognition. Analysis of relevant research demonstrates the efficacy of the ECG technique in accurately recognizing emotions within controlled laboratory settings and predefined stable environments. However, its inherent limitations rule out its application for real-time, contactless emotion recognition (Dzedzickis, Kaklauskas, and Bucinskas 2020).

Heart rate variability can be measured using EEG as well. Evaluating heart rate variability is a technique that can be used to evaluate responses to different stimuli. An alternative way of measuring heat rate variability is using photoplethysmography (PPG). Moreover, PPG is a method used to detect alterations in microvascular blood volume in tissues (Dzedzickis, Kaklauskas, and Bucinskas 2020). Despite their convenience and non-invasiveness, PPG sensors hold significant limitations for accurately measuring physiological responses to stimuli. Firstly, their susceptibility to motion artifacts, even slight movements, can mask subtle changes in blood volume caused by stimuli (Allen 2007). Additionally, external factors like ambient light, temperature, and contact pressure can introduce noise, making it challenging to isolate the true response to the stimulus.

Skin resistance measurements involve in analysis of continuous measurements gathered from human skin which induce the electrical parameters. Most often skin conductions are used as the main parameters in this technique. Emotional changes in humans often trigger observable sweat reactions in the skin, which will then lead to changes in the electrical resistance of the skin (Ayata, Yaslan, and Kamasak 2017). One of the disadvantages of skin conductance is that it is mainly related to the level of arousal in the human body. Therefore measuring valance with this technique leads to difficulties which most of the time are solved by implementing additional emotional recognition methods with this technique (Dzedzickis, Kaklauskas, and Bucinskas 2020).

Measuring brain activity changes is a method that has been gathering attention from the research community. Electroencephalography (EEG) is a noninvasive method used for recording electrical changes that occur in the human brain (Louis et al. 2016). EEG directly taps into the electrical activity of the brain, offering insights into the neural underpinnings of emotion. This provides richer information for understanding emotional response mechanisms. Different EEG features can be analyzed to explore various aspects of emotion, including valence, arousal, and specific emotions like fear or happiness. This versatility allows for more comprehensive emotional profiling compared to methods limited to single dimensions (Picard 2000).

2.3 EEG as Stimuli Response Gathering Machenism in MIR Research

The electroencephalogram (EEG) is a non-invasive technique used to record the brain's electrical activity, offering valuable insights into neural processes without the need for surgical procedures. Over the years, EEG technology has evolved significantly, shaping our understanding of brain function and its connection to human emotions. In this section we will explore the history and development of EEG, the fundamental principles behind its ability to capture brain activity, and how it can be used to study human responses for music.

Hans Berger, a German psychiatrist, recorded the first human EEGs in 1924 (Jung and Berger 1979). He captured these signals during a neurosurgical procedure on his

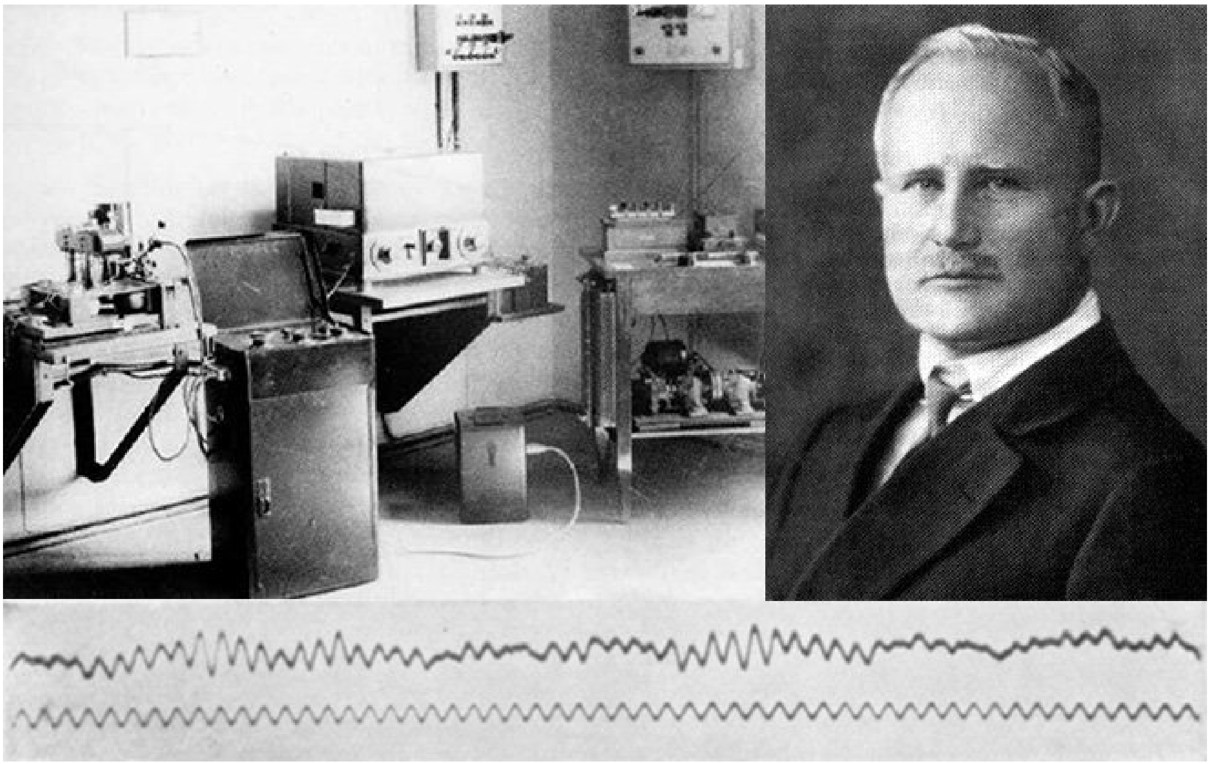


Figure 2.2: Picture of the first EEG Device Setup, EEG Inventor: Hans Berger and, early EEG Recording

som. He utilized a highly sensitive double-coiled galvanometer in recording these signals. Using a string galvanometer he observed that brain waves slowing down among other observations he made later. These groundbreaking discoveries laid the foundation for modern EEG research and its applications in neurology.

Today, EEG devices have evolved into capturing high temporal data using multiple electrodes that places on the scalp. These electrodes recodes electrical differences in scalp that identify different part of the brain. Moreover, the devices can be further divided based on the electrode type namely: Wet electrode based devices, Saline based devices and Dry electrode base devices (Soufineyestani, Dowling, and A. Khan 2020).

2.3.1 EEG Bands and Brain Activity

EEG signals are the result of electrical impulses that are generated when neurons communicate with each other. EEG electrodes helps in measuring these electrical impulses by recording difference in two points. These resulted EEG signals can be divided into different categories based on their frequency. These frequency ranges can be identified as delta

(0.5 - 4 Hz), theta (4 - 8 Hz), alpha (8 - 13 Hz), beta (13 - 30 Hz), and gamma ($\geq$ 30Hz). These ranges are associated with different cognitive functions and brain states. Alpha and beta brain waves are commonly associated with conscious states, whereas theta and delta waves are predominantly linked to unconscious states, such as deep relaxation and sleep. Additionally, gamma rhythms have been identified as playing a crucial role in the perception and processing of sensory stimuli. Table 2.1 presents overview of the frequency ranges and key characteristics of these brain wave patterns (Yasin et al. 2021).

EEG Waves	**Frequency Range (Hz)**	**Brain States**	**Mostly Found In**
Delta waves	0.1–3 Hz	Unconscious/Sleeping	Newborns and deep sleep phases
Theta waves	4–8 Hz	Imagination	Drowsiness and sleep
Alpha waves	8–13 Hz	Relaxed/Conscious	Normal and relaxed subjects
Beta waves	13–30 Hz	Conscious/Focused/Problem solving	Attentive or nervous subjects
Gamma waves	30–40 Hz	Conscious perception/Peak performance	Attentive subjects

Table 2.1: Electroencephalogram (EEG) bands and their characteristics.

2.3.2 Participant Recruitment

Subject recruitment plays a pivotal role in EEG-based emotion recognition research, as it directly impacts the generalizability and reliability of the findings. Studies have commonly sourced participants from diverse populations, including university students, hospital patients, and community members. To minimize potential sampling bias, recruitment strategies typically emphasize random and balanced participant selection. Additionally, ensuring an adequate sample size is essential for achieving statistical significance and representing the broader population effectively.

Various recruitment methods have been employed across the literature. Yuvaraj et al. 2014 gathered data from 20 parkinson's disease patients and 20 healthy individuals for the detection of emotions in Parkinson's disease using higher order spectral features. Moreover, Soraia M. Alarcão and Manuel J. Fonseca 2019 recruited 40 healthy male and female participants via advertisements on social media platforms and bulletin boards.

These examples illustrate the common reliance on convenience sampling methods while highlighting the importance of transparent and systematic recruitment protocols in EEG studies.

Informed consent is a fundamental aspect of ethical research, ensuring that participants are fully aware of the study's purpose, potential risks, and their rights, including the right to withdraw at any time. Prior studies have consistently emphasized obtaining written informed consent, allowing participants sufficient time to review and understand the consent form before signing. These forms typically include assurances of data confidentiality and clear explanations of participant rights. For example, Baur et al. 2019 and Koelstra et al. 2012 documented the use of written informed consent to uphold ethical standards and transparency throughout the research process.

2.3.3 EEG Devices and Data Acquisition

EEG devices vary widely in terms of the number of electrodes, software capabilities, and price ranges. Therefore number of electrodes is not the only variable that is considered when choosing a device. Several manufacturers offer EEG systems tailored to different needs and applications, including Emotiv, OpenBCI, and Neuroscan.

There are two types of electrodes that can be found in EEG devices, Wet and Dry. Wet electrodes provide more higher quality information in recording EEG signals as they involve in inserting a conductive gel between scalp and electrode. However, the setup time for these EEG devices is longer compared to dry electrode devices because of the gel insertion time and the conformation of the participant to carry out the experiment without discomfort. Dry electrodes, on the other hand, do not require conductive gel, which makes them more comfortable for the subjects. However, they provide lower signal quality due to the higher impedance and lower contact area with the scalp (Soufineyestani, Dowling, and A. Khan 2020).

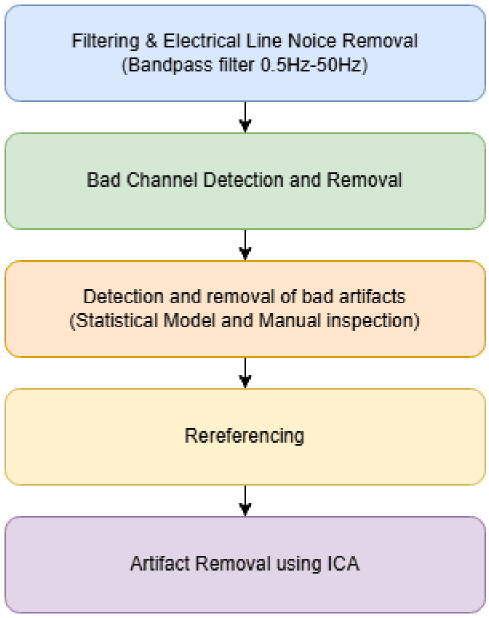


Figure 2.3: EEG Preprocessing Pipeline Workflow

2.3.4 Preprocessing

EEG signals are afflicted with a number of types of noise, such as electrical (line) noise, Muscle movement, eye blinking, and other actions can significantly influence how precise things are and the accuracy of emotion recognition outcomes. Hence, a proper pre-processing pipeline is needed to remove these sources of noise and problems and to increase the signal-to-noise ratio of the EEG data. The pre-processing process typically includes some significant steps, including filtering, electrical (line) noise removal, bad channel rejection, bad artifact detection and removal, and re-referencing. Figure 2.3 illustrates the general pre-processing Pipeline for EEG processing.

Filtering

Filtering is an essential pre-processing process in EEG-based research. The primary function of filtering is to eliminate noise from the recorded EEG signals and enhance the critical frequency bands associated with the brain. EEG signals are frequently infested with a wide range of noise, both biological and environmental noise, and motion artifacts that can impact the validity and reliability of the emotion recognition system (Jeunet et al. 2018).

In EEG-based studies, two types of filters are usually utilized: high-pass and low-pass filters. High-pass filters eliminate the low-frequency components of the EEG signals, and low-pass filters eliminate the high-frequency components.

A bandpass filter is typically employed to obtain the desired frequency range[4]. The cut-off frequencies of the filters rely on the frequency range of Interest and the character of the EEG signal.

Several EEG signal processing tools, such as EEGLAB (Delorme and Scott Makeig 2004) and Python-based libraries like MNE (Gramfort et al. 2013), offer a variety of filtering options for EEG data. Users can configure parameters such as filter order, type, and cut-off frequencies. The choice of filter typically depends on the specific characteristics of the EEG signal and the type of noise to be removed, including the frequency range of interest and the desired level of attenuation in the stop band.

Electrical noise removal

Many techniques have been suggested and tried to minimize the Electrical line noise. The most common are filtering and regression methods.

Filtering is the most common technique used for noise reduction. Some of the filters that are widely used by individuals include notch filters, bandpass filters, and adaptive filters. Notch filters are very effective at removing narrow-band interference, especially the common 50 Hz or 60 Hz electrical noise from power lines.

In addition to filtering techniques, regression-based methods offer an effective approach for noise reduction in EEG signals. Common techniques include linear regression, least mean square (LMS) adaptive filtering, and principal component analysis (PCA) (Widrow and Stearns 1985; Pearson 1901). These methods operate under the assumption that electrical noise can be distinguished from neural signals, enabling its identification and removal. LMS adaptive filtering has shown promising results in denoising EEG data, as demonstrated by studies such as Jian Li, W. Zhang, and Wang (2020) and I. Khan, Khalid, and Javaid (2021). This approach involves estimating the power spectrum of the noise and subtracting it from the contaminated EEG signal. PCA has also proven effec-

[4] https://en.wikipedia.org/wiki/Band-pass_filter

tive by decomposing EEG signals to isolate and remove noise components, as reported in Jian Li, W. Zhang, and Wang (2020) and H. Zhang, Liu, and Chen (2021).

Bad channels Detection and Removal

EEG signals are often contaminated by artifacts that can obscure genuine brain activity, particularly in emotion recognition models. One common issue is "bad" channels, which can result from electrode movement, poor contact with the scalp, or external noise. These faulty channels can be identified visually by technicians, though this process can be time-consuming and subjective. Automatic methods, such as kurtosis-based detection (O'Reilly, Nielsen, and Hansen 2007), correlation analysis (Nolan, Whelan, and Reilly 2010), and variance-based detection (Viola, Orozco, Pinaya, et al. 2009), offer more efficient alternatives.

Once problematic channels are identified, they are usually corrected or removed before further analysis. Common correction techniques include interpolation, where data from surrounding electrodes is used to reconstruct the faulty channel (Perrin, Pernier, and Bertrand 2011; Kayser and Tenke 2006). Other methods like robust averaging (Pernet, Wilcox, and Rousselet 2011) and threshold-based rejection (Mognon, Grapperon, and Pernier 2011) help remove bad channels by excluding data with poor signal quality or low signal-to-noise ratio.

Detecting and correcting bad channels is a crucial step in EEG preprocessing for accurate emotion recognition. Platforms such as EEGLAB (Delorme and Scott Makeig 2004), and MNE-Python (Gramfort et al. 2013) offer robust tools for these tasks, making it easier for researchers to clean EEG data.

Re-referencing

Re-referencing is an essential pre-processing step in EEG data analysis that standardizes the signals by converting them to a common reference point, helping to eliminate bias introduced by the initial reference electrode. This process is crucial as the choice of reference can significantly affect the recorded signals. Common reference points, such as earlobes, the nose, or the average of all electrodes, may yield different results depending on the specific characteristics of the data. Re-referencing methods like the Average Reference (AR) and the Reference Electrode Standardization Technique (REST) are widely used to address these issues and improve the consistency and reliability of EEG analysis (Sara M. Alarcão and Maria J. Fonseca 2017; Zhu et al. 2017).

Artifcat Detection and Removal

EEG signal preprocessing is crucial for emotion recognition applications, as unwanted noise can degrade the signal quality and the reliability of results. Common artifacts in EEG include eye movements, muscle contractions, and electrode movements, all of which can significantly affect the signal. Eye blinks typically alter the frontal and temporal regions, while muscle contractions, particularly in the facial area or neck, introduce interference in the adjacent electrodes. Additionally, electrode movement results in unstable and unclear recordings. Identifying and removing these artifacts is vital for accurate analysis. Techniques like visual inspection are useful but time-consuming, and automated methods like Independent Component Analysis (ICA) and wavelet analysis offer more efficient solutions. ICA separates EEG signals into independent components, allowing for the identification and removal of artifact-related components (S. Makeig et al. 2004), while wavelet analysis decomposes signals into frequency bands, effectively eliminating high-frequency noise or brief artifacts (Cohen 2014).

Several software tools, including EEGLAB, FieldTrip, and BrainVision Analyzer, provide end-to-end solutions for artifact removal. EEGLAB, in particular, offers the extended Infomax ICA algorithm, which decomposes EEG signals into independent components and removes mutual dependencies. This method is especially effective in isolating and eliminating artifacts like eye blinks and muscle movements due to their distinct frequency patterns. Overall, artifact removal methods such as ICA and wavelet decomposition, along with software like EEGLAB, enhance the quality of EEG signals for subsequent analysis, making them essential for emotion recognition and other EEG-based studies.

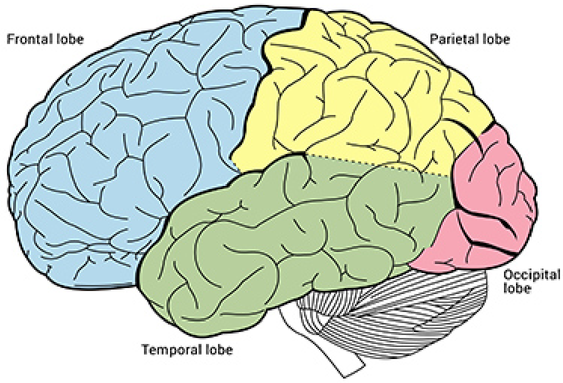


Figure 2.4: Regions in brain's cerebral cortex

2.4 Music Elected Psychophysiological Changes

The cerebral cortex, forming the brain's outermost layer, is a network of densely packed neurons known collectively as the gray matter. Positioned just beneath the protective meninges, this intricate structure is divided into four distinct lobes: frontal, temporal, parietal, and occipital, each responsible for a wide range of essential cognitive and sensory functions (Javed, Reddy, and Lui 2025).

The frontal lobe, the largest of the brain's lobes, is situated in front of the brain. It is anatomically and functionally divided into three principal regions: the primary motor cortex, the supplementary and premotor areas, and the prefrontal cortex. Each of these regions plays a distinct role in motor control, planning, and higher cognitive functions. The frontal lobes are critical for more difficult decisions and interactions that are essential for human behavior (Pirau and Lui 2025). Furthermore, current literature indicates that, frontal lobe is identified for being useful in cognitive tasks during music listening, including decision-making and attention allocation. Additionally, frontal lobe tend to have higher engagement in emotional responses to music (AmplifyYou 2021; University of Central Florida 2021).

The temporal lobe is essential for processing auditory information, memory formation, language comprehension, and aspects of emotion and object recognition. Additionally, it analyzes the basic elements of music: pitch, rhythm, melody, and harmony, allowing us to perceive and appreciate musical structure. The right temporal lobe specializes in recognizing nonverbal sounds, such as musical tones, timbre, and environmental sounds, which are fundamental for distinguishing instruments and musical genres (Samson 1999).

The parietal lobe plays a key role in helping us make sense of the world around us. It integrates sensory information from different sources like touch, temperature, pain, and body position to create a cohesive understanding of our environment. Additionally to sensory processing, it also supports spatial awareness, attention, navigation, and even contributes to certain aspects of language (Bisley 2022). In musical responses, parietal lobe helps in identifying spacial elements. These include locating of musical elements, or movement of these elements. Furthermore, musical engagement frequently involves movement. Therefore, the parietal lobe's role in coordinating movement is essential for synchronizing actions to rhythm and beat (Bellmann and Asano 2024).

The occipital lobe serves as the brain's primary center for visual processing, receiving and interpreting signals from the eyes. Its main functions include processing visual information, visual recognition, spatial mapping (which supports depth perception), distance assessment, and color perception. As such, it does not play a central role in musical perception, which is primarily processed in other areas of the brain (Cleveland Clinic 2023).

2.5 Psychophysiological Responses Based Popularity Prediction

Analyzing responses from listeners can give insights into the listener's perspective of the song. These responses can be self-assessments or recorded physiological changes. In contrast to acoustic features, which give insight into music specific features, listener-based feature extraction can be used for analysis of the population's insights. As music tends to elicit physiological responses in humans (Bernardi, Porta, and Sleight 2006), analyzing those responses can give insights into how a popular song is different from the rest.

In Berns and Moore 2012, the study aims to find the correlation between the neural response and the commercial success of music based on functional Magnetic Resonance Imaging (fMRI). To capture a quality fMRI scan, the movement of the subject should be controlled and minimized. Therefore from the 32 participants, five were excluded as they tended to show high movement and artifacts. Participants were exposed to 15-second songs in two stages where first one involved scanning the fMRI and rating the song in familiarity and liking scores and the second stage had the likability rating but the popularity of the song was displayed before the rating. The study contains an analysis of the nucleus accumbens activity. The nucleus accumbens is a part of the human brain which is mostly responsible for pleasure. The study found that the correlation between the nucleus accumbens activity and sales was integrated over the entire 15-second listening period, indicating the importance of sustained neural engagement in predicting popularity.

In Rajagopalan and Kaneshiro 2023, an analysis of publicly available EEG data was conducted to analyze the correlation between the choruses that are spread across the music. This study mainly focuses on music structure analysis (MSA), which is a part of the MIR. The study aims of using EEG to understand the neural response to music structure, particularly in popular music choruses. The findings reveal a significant correlation between neural response choruses in the music. Furthermore, findings suggest that brain activity synchronizes even across non-identical chorus instances.

In Leeuwis et al. 2021 researchers have carried out an experiment on thirty-one people using EEG to predict music streams of a song. The device used in the experiment had nine channels to capture EEG responses to the played stimuli. Twenty-four-second fragments of 24 songs were selected. The selected songs were from a pop album and an R&B album. The albums were selected just a few days after the release. The linear regression model was constructed using the computed EEG neural synchrony as the primary predictor variable. This was used to evaluate the general popularity based on the number of streams. Both group-level and individual-level analyses of the result are presented in the paper. The group-level analysis indicates that neural synchrony measured within the sample was significantly correlated with public appreciation of Spotify. However, it was not correlated with the subjective likability scores provided by the sample itself. Individual-level analyses indicated that the matrix of pleasantness was associated with subjective likability ratings. Moreover, engagement emerged as a significant predictor when combined with factors such as artist and single release.

The current literature on hit song prediction problem highlights a gap in research, particularly in the analysis of psychophysiological signals to tackle this problem. When looking at different psychophysiological responses, Electroencephalogram (EEG) can be identified as a method for gathering brain activity for a given stimulus. EEG utilizes electrodes to capture the electrical impulses generated in the brain when neurons communicate with each other. EEG can be identified as a non-invasive, low-cost method for gathering psychophysiological responses (Dzedzickis, Kaklauskas, and Bucinskas 2020). It is evident from an initial analysis of the existing literature that the application of Electroencephalogram (EEG) for investigating changes in popular music is a domain that has received less attention from the community. Despite the different approaches used for predicting the popularity of a song, there exists a gap in identifying neural changes in popular songs.

Chapter 3

Methodology

As previously discussed in the introduction chapter, the aim of the research is to develop a computational model that predicts hit songs by analyzing listeners' psychophysiological responses. To achieve this aim, we have addressed a set of research questions. The first step involves identifying hit songs based on data from the streaming platform. Then we need to identify psychophysiological differences in hit songs. Finally, we have proposed a model for identifying hit songs using psychophysiological responses of the listener. This chapter offers a detailed explanation of each stage of the research process, outlining the specific design decisions and implementation strategies employed. It also discusses the limitations encountered in the approach and the measures taken to minimize their potential impact.

3.1 Music Selection

Initially, Billboard Hot 100 chart and Popnable Top 40 Sri Lanka chart were selected for music selection. From each chart 20 musical stimuli were selected. These selections were done based on the rank assigned to each song and was conducted on musicals charts that appeared a week earlier to the experiment. The highest ranked 10 songs and the lowest rank 10 songs were selected based on ranking system of each chart. Each charts gives the ranking for each song alongside the metric "Number of weeks in the chart". Thereafter, these metrics were used to select the songs that we in the chart for a long time. Moreover, each chart gives the peak ranking of each song, which was also considered when selecting the song. Tables **??** and 3.2 show the pool of songs selected from the weekly charts. The highlighted songs indicate those that were chosen for the experiment.

After selecting the relevant music from the charts, each stimulus was assigned a label to support both classification based on the label and regression for ranking prediction. As identified in the literature, the most common labels for music pieces are "Hit" and "Flop." The label "Hit" denotes songs that are successful, while "Flop" refers to songs that are not. The lowest-ranked music stimuli were labeled as "Flop," whereas the highest-ranked were labeled as "Hit."

Additionally, we utilized Spotify's "New Music Friday" chart to select 10 songs. These selections aimed to identify new music that could potentially emerge in the Billboard Hot 100 and Popnable Top 40 charts. Songs were selected randomly from the chart based on language.

3.2 Participants

A total of 31 participants were included in the experiment. Participants were from the university. Participants' ages ranged from 21 - 25 an average of 23.4. Both female and male participants were involved in the experiment, and a total of 19 male participants and 12 female participants were involved. All the participants are from Sri Lanka. Each participant was briefly explained about the experiment and the EEG device. Moreover, the response gathering tool was also introduced before the experiment to avoid confusion while annotating the stimulus. The consent from each participant was gathered before the experiment was carried out. With the consent, the experiment was carried out, and each experiment consisted of 6 music stimuli per user, resulting in 186 EEG recordings for music pieces.

Song	Artist	Last Week	This Week	Peak	Weeks on Chart
A Bar Song (Tipsy)	**Shaboozey**	**1**	**1**	**1**	**17**
I Had Some Help	**Post Malone Featuring Morgan Wallen**	**2**	**2**	**1**	**13**
Not Like Us	**Kendrick Lamar**	**3**	**3**	**1**	**14**
Espresso	**Sabrina Carpenter**	**4**	**4**	**3**	**17**
Million Dollar Baby	**Tommy Richman**	**5**	**5**	**2**	**15**
Good Luck, Babe!	Chappell Roan	8	6	6	18
Birds Of A Feather	Billie Eilish	10	7	7	12
Please Please Please	Sabrina Carpenter	9	8	1	9
Lose Control	Teddy Swims	6	9	1	52
Too Sweet	Hozier	7	10	1	20
My Kink Is Karma	Chappell Roan	New	91	91	1
Euphoria	Kendrick Lamar	79	92	3	15
Tough	**Quavo & Lana Del Rey**	**78**	**93**	**33**	**5**
Liar	Jelly Roll	New	94	94	1
Sweet Dreams	**Koe Wetzel**	**83**	**95**	**35**	**12**
Alibi	Sevdaliza, Pabllo Vittar & Yseult	98	96	95	3
Parking Lot	**Mustard & Travis Scott**	**81**	**97**	**57**	**4**
Wine Into Whiskey	**Tucker Wetmore**	**95**	**98**	**68**	**19**
Love You, Miss You, Mean It	Luke Bryan	New	99	99	1
We Ride	**Bryan Martin**	**92**	**100**	**56**	**18**

Table 3.1: Billboard Hot 100 Chart Music Selection

Song	Artist	Last Week	This Week	Peak	Weeks on Chart
Seedevi	**Piyath Rajapakse**	**1**	**1**	**1**	**6**
Mandire Hade	**Dulan Arx**	**2**	**2**	**1**	**15**
Labunothin Waram	**Shanuka Ekanayake**	**3**	**3**	**2**	**7**
Ape Kathandare	**Dhyan Hewage**	**10**	**4**	**4**	**4**
Akkalage Wenna	**Manjula Sewwandi**	**4**	**5**	**1**	**19**
Ran Thodu	Chathumi Dihara	6	6	5	8
Surangana	Dj Jnk	New	7	7	1
Dagakaari	Jtsp Boy	9	8	8	2
Nuwara Kumari	Nipun Rajapaksha	5	9	4	3
Kasi Saban Pena	Sarith & Surith	8	10	1	32
Sinhala Wedakam	Maduwa	44	40	8	72
Piuma's Sweet Drink	**2Forty2**	**41**	**39**	**23**	**11**
Rosa Batiththi	Mangala Denex	46	38	2	78
Visabjay	Shan Putha	37	37	5	49
Poddak Saiko	Gayya	36	36	2	21
Bandimu Suda	Piyath Rajapakse	31	35	1	28
Labandi Komaliya	Santhush Weeraman	33	34	5	27
Maga Haree	**Mihiran**	**34**	**33**	**6**	**34**
Ill Mahe Kurullo	Nisala Kavinda	29	32	3	34
Kaari Naa Sanda	Methun Sk	35	31	1	81
Katharina	**Dinuka Jayasinghe**	**28**	**30**	**25**	**9**
Mala Kada Kada	Dinesh Gamage	30	29	1	58
Ummah	**Chanuka Mora**	**27**	**28**	**4**	**45**
Mata Inna Hithuna	Amandi Sulochana	25	27	2	53
Ran Muduwaka	**Supun Perera**	**24**	**26**	**15**	**27**

Table 3.2: Popnable Chart Music Selection

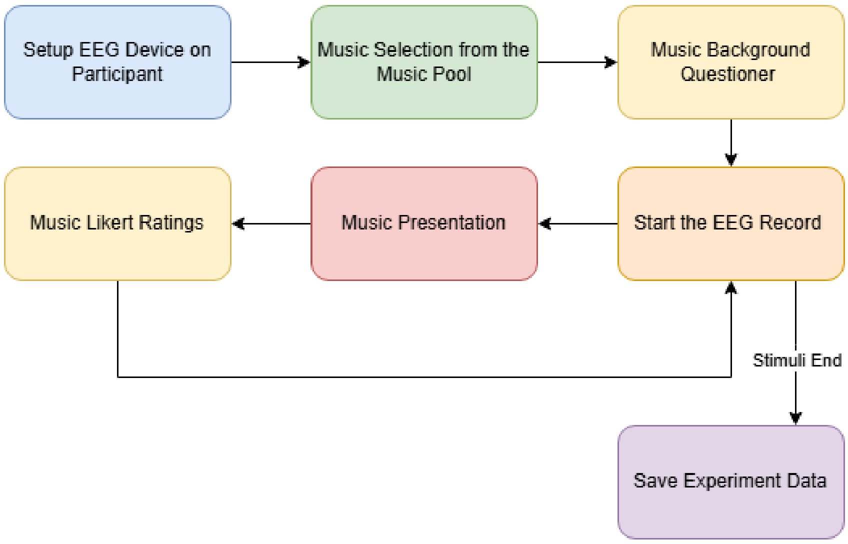


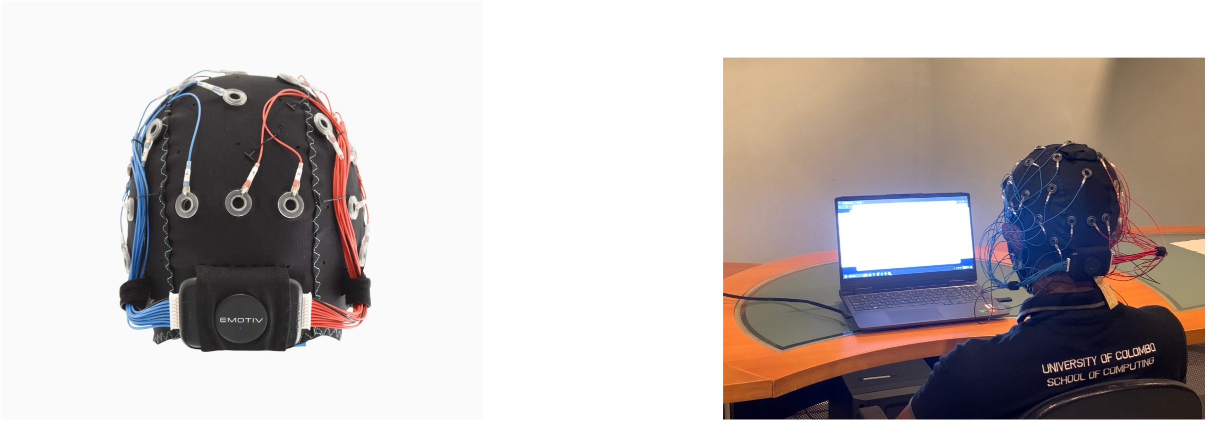
Figure 3.1: Experiment Workflow

3.3 Experiment

Total of 31 participants participated in the experiment and the high level workflow of the experiment can be seen on Figure 3.1. The experiment started with the placement of the EEG headset on the participant. Then the musical stimuli were selected by the algorithm followed by the music background questioner. Afterwards the EEG recording starts and the stimuli is presented followed by a stimuli rating. The last two steps were repeated for each stimuli. When all the stimuli were over the data will be saved in a database for futher processing. The following sections will dive deep into each step explaining design choices.

3.3.1 EEG Device Placement and Recorder

After the consent, each participant was introduced to the EEG device. The EEG device used to capture data is named "Emotiv Epoc Flex". This device is a Gel electrode device with 32 electrodes. The EEG recorder consisted of high-quality Ag/AgCl sensors, and conductive gel was applied between the sensor and the scalp to improve impedance. The device recorded the signal at 128 Hz. Each electrode sensor is placed with the standard 10-20 system. Figure 3.3 illustrates the sensor placement in the "Emotiv Epoch Flex" device on the scalp according to the 10-20 placement system. These placements help in identifying electrodes that are in different regions of the brain, which will be helpful in



(a) Emotiv Epoc Flex Gel device with 32 Electrodes

(b) A participant in the experiment with the EEG device on

Figure 3.2: Emotiv Epoc Flex device and a participant wearing the EEG device

feature selection.

After the placement, the “Emotiv PRO” software was used in capturing the EEG data. The software provided APIs for setting markers on the recording. We utilized this API in order to set markers in EEG recording when the stimuli begin. These markers carries a ID which helps in identifying which song is represented by the marker. The Figure 3.4 illustrates a marker which is placed on a recording when it was captured and the ID can be seen at the bottom of the marker in the recording. The experiment was carried out in a laboratory environment where sound and illumination was controlled in order to avoid any distractions on the participant. The music stimuli was presented using a pair of studio speakers.

3.3.2 Music Stimuli Selection

A music stimulus pool is available within the application, containing all the selected music pieces. An algorithm is used to select a stimulus for each user. It searches for music stimuli with the fewest existing recordings and selects one accordingly. Before the start of the experiment, the algorithm identifies the six music pieces with the least number of selections and prepares them for use in the application. This approach ensures that for every five participants, each piece of music in the pool will receive one EEG recording along with corresponding questionnaire annotations.

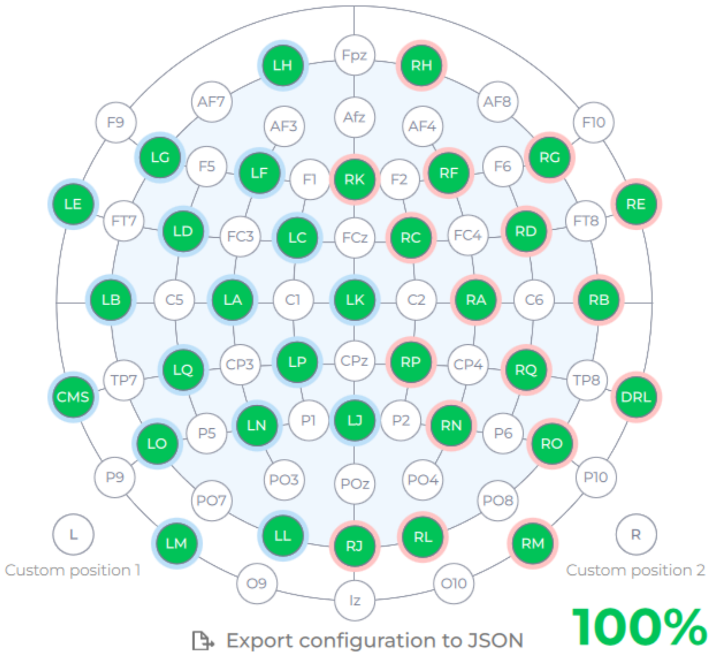


Figure 3.3: International 10-20 system for 32 electrodes placement (Green Colored Electrodes)

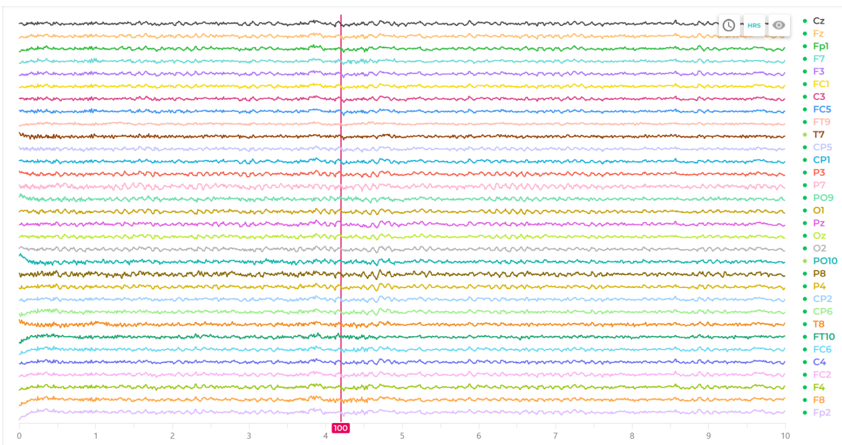


Figure 3.4: Marker in the EEG Recording

3.3.3 Experiment Start

The experiment began with a brief explanation of the EEG device and its functioning to reduce any confusion among participants. The device was then placed on the participants, and the electrodes were calibrated to ensure better contact with the scalp for improved EEG signal accuracy. Afterward, participants were instructed to provide their age and gender before proceeding with the experiment.

3.3.4 Music Background Questioner

Music has been associated with cognitive, emotional, and social benefits; however, measuring these effects at the individual level can be challenging. To address this, a questionnaire was designed to capture the musical background of each participant. It consists of five questions, each annotated using radio buttons. The questionnaire covers various aspects to capture different indices for each participant.

- **Index for Music Listening**
  1. On average, how many hours do you listen to music in a day?
- **Index for Music Training**
  1. What is the highest level of formal music training you have received?
  2. Did you receive any other type of music training?
- **Index for Instrument Playing**
  1. Have you played / do you play a music instrument?

The questionnaire was designed to capture the engagement of the participants across multiple dimensions of music use. The captured annotations can be further analyzed to understand the correlation between subjective knowledge of the domain and the aspect of music popularity (Chin and Rickard 2012).

3.3.5 Music Likert Ratings

For each music stimulus, a Likert rating scale was placed on the application to annotate each stimulus preference by the participant. The ratings were 5 scale ratings ranging from "Very Low" to "Very high". Two questions were placed with the Likert scale which is represented below. Another question was placed to identify the novelty of the participant's response as it captures whether the participant has already heard the song or not.

- Have you heard this music before?
- How much do you like the music?
- How familiar are you with this music/music pattern?

3.3.6 Annotation Application

The first user interface that is presented involve in recording the user age and gender. Afterwards the annotation proceeds to the music background questioner. The music questionnaire consists of above mentioned questions what aims to gather the musical background knowledge of the participant. With the questionnaire complete, one of the selected music is presented to the user. Before playing the music the user is instructed to close the there eyes in order to avoid any distractions and artifacts. Before playing the music the user is given a 5 seconds of silence to relax. Then the music is played where the application is connected with the Emotiv PRO application via its API. As soon as the music starts playing the marker will be placed on the recording. This allows to capture the start of the recording in order for further processing. After each music a stimuli wise rating was given to identify participant's response towards the relevant music piece. This process is carried out for the selected 6 music stimuli. After the application greets the participant marking the end of experiment. Participant could end the experiment if they wish to at any given point.

Annotation application was design in React. Voting and music selection algorithms were called with API calls and the data was passed for relevent calculations. To implement the annotation tool, we modified the stimuli selection annotation tool and added the ability to make an API call to the marking server when starting the stimuli. The marking server was implemented using Flask, a microweb framework developed in Python. The server encodes the stimuli ID/Time and stores it as an analog pulse in an analog EEG channel of the EEG recorder using Python serial communication and Emotiv built-in APIs of the Emotiv Software. This will help in marking the EEG recording with the stimuli ID which will help in recognizing the music piece related to the recording. Several user interfaces can be seen in Figure 3.5, Figure 3.6, Figure 3.8, and Figure 3.7. An example of a marker being placed on the recording can be seen in the Figure 3.4

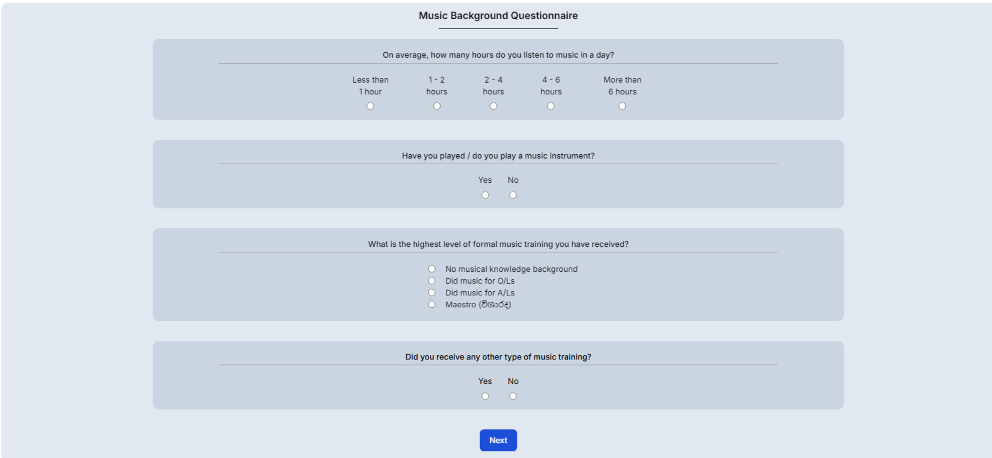


Figure 3.5: Music Background Questioner User Inferface

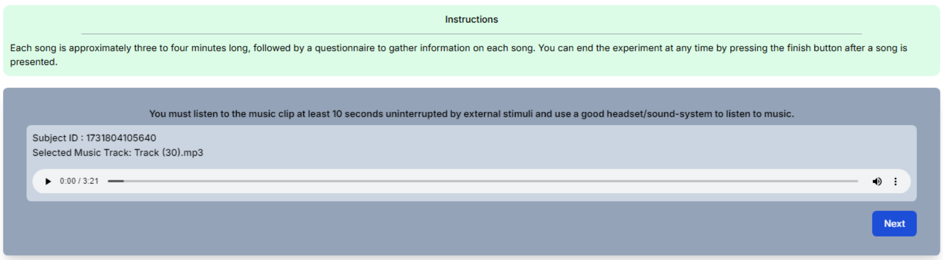


Figure 3.6: Music Stimuli User Interface

3.4 Data Preprocessing

The preprocessing step plays a crucial role in enhancing the quality of EEG data by removing various types of noise that can interfere with the accuracy of the data. In this study, we performed a series of preprocessing steps to improve the quality of the

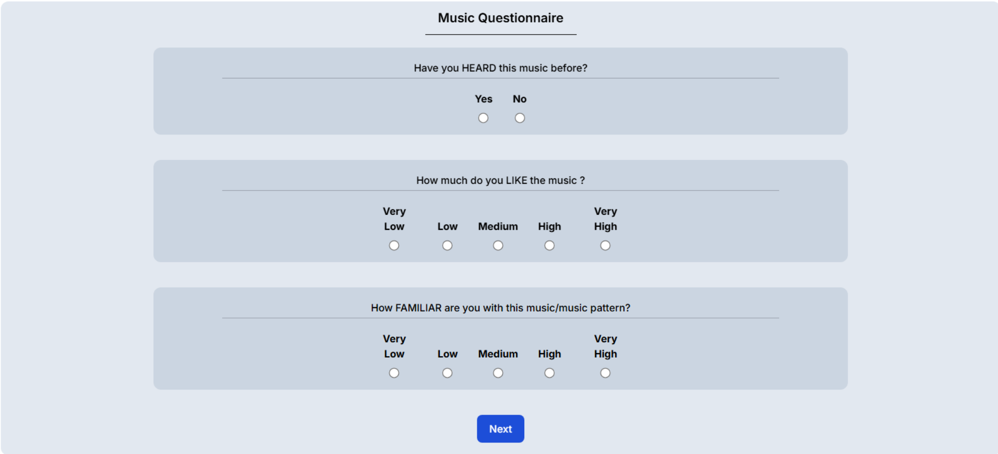


Figure 3.7: Music Stimuli Wise Questionnaire User Interface

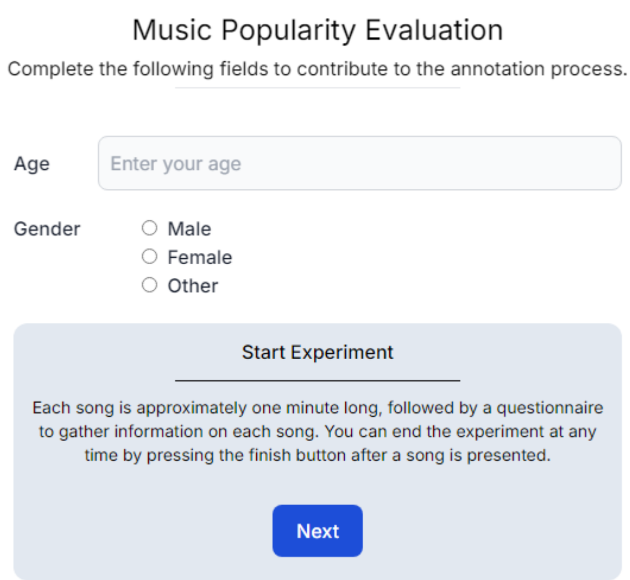


Figure 3.8: Starting User Interface

EEG data. Figure 3.9 depicts an overview of the preprocessing procedure. We used the preprocessing tool, EEGLAB for preprocessing. EEGLAB is an open-source MATLAB toolbox designed for processing, analyzing, and visualizing electrophysiological data, such as EEG. It offers a comprehensive suite of functions for preprocessing, including artifact removal, filtering, epoch extraction, and Independent Component Analysis (ICA), enabling researchers to enhance signal quality and isolate neural activity from noise. Widely adopted in neuroscience research, EEGLAB supports customizable workflows, integration with other toolboxes, and a user-friendly interface, making it a versatile tool for advancing EEG data analysis in both clinical and experimental settings (Delorme and Scott Makeig 2004).

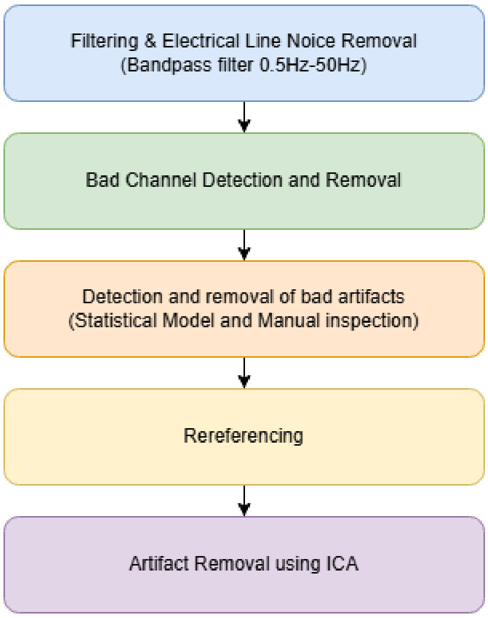


Figure 3.9: Preprocessing Steps

The first step in preprocessing involves removing line noise. We remove line noise with a band-pass filter using a lower limit of 0.5 Hz and a higher limit of 50 Hz. This ensures that electrical noise is removed since in Sri Lanka, electrical lines are 60 Hz. This step is critical in eliminating high-frequency noise that is often present in the data, thereby enhancing the signal-to-noise ratio and obtaining more accurate results.

The next step involved identifying and removing bad channels from the EEG data. Bad channels are those that have poor signal quality or are contaminated by noise, and they can be detected through various techniques such as visual inspection, statistical analysis, or machine learning algorithms. In this study, we used a combination of visual inspection and statistical analysis to identify and remove bad channels, thus improving the overall quality of the data by eliminating any channels that do not contribute to the analysis.

Artifact removal was performed using a two-stage process combining automated and manual techniques. Initially, a statistical model implemented in the EEGLAB toolbox was applied to detect and remove artifacts, including eye movements, muscle activity, and electrical interference. This model-based approach utilized statistical parameters to attenuate non-neural components from the EEG recordings. Following this automated preprocessing, the data were manually inspected to identify and exclude any artifacts not captured by the statistical method. This combined approach ensured the preservation of neural signals while minimizing contamination, thereby improving the overall quality and reliability of the EEG data.

- Remove channel is it is flat for more than 50 seconds
- Max acceptable high frequency noise standard deviation 10.
- Min acceptable correlation with nearby channels 0.5
- Max acceptable 0.5 second window standard deviation 25

In the next step, we carried out a re-referencing process to adjust the reference signal used for analysis. Re-referencing is a crucial step in EEG preprocessing because EEG signals are measured as voltage differences between electrodes, and the choice of reference significantly impacts data interpretation. A poorly chosen reference can distort signals across all channels, while proper re-referencing helps neutralize bias and improve signal quality. We used average referencing to re-reference all the channels in every recording. Average referencing improves EEG data by distributing the reference signal across all electrodes, reducing bias from a single noisy or asymmetric reference. This approach provides a more balanced representation of neural activity, as it assumes the average of all electrodes approximates a neutral baseline. To average reference the data we used the function given by the EEGLAB [1].

In the final step of our data preprocessing, we implemented an Independent Component Analysis (ICA) to eliminate any remaining artifacts from the EEG data. ICA is a

[1] `https://eeglab.org/tutorials/05_Preprocess/rereferencing.html`

source separation technique that decomposes EEG signals into statistically independent components, each representing distinct neural or non-neural sources (e.g., eye blinks, muscle activity, or brain). Labeling these components involves analyzing their spatial topography, time-course patterns, and spectral properties to classify their origins. Tools such as EEGLAB's ICLabel plugin automate this process by assigning probabilistic labels (e.g., "brain," "eye," "muscle," "heart," or "noise") to each component, quantifying the likelihood of its source. This allows researchers to objectively identify and exclude components dominated by artifacts while retaining those reflecting neural activity [2].

Rejecting eye and muscle artifacts through ICA significantly enhances signal quality by removing high-amplitude, non-cortical interference. For instance, eye-blink artifacts exhibit characteristic frontal scalp distribution temporal waveforms, while muscle artifacts often manifest as high-frequency, broadband noise localized to temporal or occipital regions. By isolating and discarding these components, the integrity of the underlying neural signals is preserved. This step is critical for improving the signal-to-noise ratio, ensuring that interpretations of brain activity are not surprised by various physiological or environmental noise.

Following preprocessing, the EEG data were divided into epochs aligned to each music stimulus. Each epoch included a 5-second baseline period before the stimulus began, capturing pre-stimulus brain activity, and extended until the end of the music to fully encompass the neural response. This ensured that both before the music and during the music were preserved. By structuring the data this way, unrelated background noise or brain activity outside the stimulus period was minimized, allowing clearer analysis of how the brain reacted to the music. The epochs were then baseline-corrected using the pre-stimulus interval to standardize the data, ensuring reliable comparisons across trials and preparing the dataset for further analyses.

3.5 Feature Extraction

EEG signals arise from electrical activity produced by neurons communicating in the brain. Electrodes measure these signals by detecting voltage differences between two points. These signals are categorized into frequency bands: delta (0.5–4 Hz), theta (4–8

[2] `https://eeglab.org/tutorials/06_RejectArtifacts/RunICA.html`

Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma ($\geq$ 30Hz). Each band is linked to specific brain states—alpha and beta with wakeful activities like focus, and theta/delta with sleep or deep relaxation (see Table 2.1) (Yasin et al. 2021). Identifying which features to analyze is essential for understanding brain activity and improving classification accuracy in EEG studies.

Discrete Wavelet Transform

Time domain feature exploration, frequency domain feature exploration and time-frequency domain feature exploration can be done in order to carry out further processing in feature classification. Discrete Wavelet Transform (DWT) can be identified as one of the time-frequency domain feature that extract both time and frequency domains information simultaneously.

DWT is a powerful tool for analyzing EEG signals, which are non-stationary and contain time-varying frequency components. DWT decomposes EEG data hierarchically into approximation and detail coefficients using low-pass and high-pass filters with downsampling. Doing this process iteratively helps decomposition of the signal into multiple levels which contains different frequency information at each level. At each level the signal is decomposed into detailed and approximation coefficients which has higher half of the frequency information in detailed coefficients and lower frequency information in approximation coefficients. This frequency information follows the Nyquist sampling theorem which implies that the highest frequency component that can be accurately represented in a digital signal must be less than half of the sampling rate. This results in fewer data points in each level. As half of the frequency information is removed at each level, the sample size also decrease in half at each level.

$$y_{\text{low}}[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot g[2n - k] \tag{3.1}$$

$$y_{\text{high}}[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[2n - k] \tag{3.2}$$

Discrete Wavelet Transform filter bank equations, where $x[k]$ is the input signal, $g[\cdot]$

is the low-pass filter (generating approximation coefficients), $h[\cdot]$ is the high-pass filter (generating detail coefficients), and n is the downsampled time index. The factor of 2 in the filter indices implements the dyadic downsampling characteristic of DWT. $y_{\text{low}}[n]$ are the Approximation coefficients representing smoothed, low-frequency components. $y_{\text{high}}[n]$ are the Detail coefficients capturing high-frequency components.

In this research for feature extraction we utilized 4 level DWT data decomposition using Python and PyWavelets using Coiflets wavelet object. PyWavelets is open source wavelet transform software for Python. It combines a simple high level interface with low level C and Cython performance.

Continuous Wavelet Transform

Time-frequency domain feature exploration provides a comprehensive perspective for analyzing non-stationary signals like EEG. Continuous Wavelet Transform (CWT) serves as a fundamental time-frequency analysis technique that maintains continuous resolution across both domains, unlike its discrete counterpart. This transform proves particularly effective for capturing short neural events that traditional Fourier methods might struggle.

CWT operates by continuously scaling and translating a mother wavelet across the signal, enabling detailed examination of time-varying components. Unlike discrete decomposition, CWT preserves the complete time-frequency structure without downsampling, making it ideal for identifying short-duration changes in EEG recordings.

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\psi^* \left(\frac{t-b}{a}\right) dt \tag{3.3}$$

Continuous Wavelet Transform equation, where $x(t)$ represents the input signal, $\psi(t)$ denotes the mother wavelet function, a controls the scale (inversely related to frequency), and b determines the temporal position. The normalization factor $1/\sqrt{a}$ ensures consistent energy distribution across different scales. $C(a,b)$ coefficients represent the correlation between the signal and scaled/shifted wavelet versions.

For this research's implementation, we employed Python with PyWavelets using Morlet wavelets for CWT computation. PyWavelets is open source wavelet transform software for Python. It combines a simple high level interface with low level C and Cython performance.

Principal Component Analysis

Principal component analysis (PCA) is a technique used to simplify large datasets by shrinking them down to a smaller size while keeping the most important information. It works by combining many variables into fewer, making the data easier to work with without losing key insights.

$$\mathbf{Y} = \mathbf{X}\mathbf{W} \tag{3.4}$$

Principal Component Analysis (PCA) provides an effective method for reducing the dimensionality of EEG data while preserving its most meaningful patterns. The reduced dimensions are orthogonal to each other and are called principal components. In this research we utilized this technique to reduce the original 32-channel electrode signals into a new set of uncorrelated principal components. These components are linear combinations of the original electrodes, ordered by their contribution to the total variance in the data. X is the original data matrix (with dimensions n samples $\times$ 32 electrodes), W contains the eigenvectors that define the principal components and Y represents the transformed data in the new component space. This is particularly valuable when analyzing how the brain responds to different musical stimuli, as it helps us to understand which neural patterns are most strongly associated with specific auditory experiences. The components themselves can be interpreted by examining their weighting across electrodes, often revealing distinct spatial patterns that correspond to different functional networks or artifact sources in the EEG signals.

After reducing the data using PCA, we calculated the average activity for each principal component to make the analysis easier. We also kept track of the lowest and highest activity values to better understand how each component helps determine a song's ranking. This gives us a clearer picture of which brain activity patterns matter most when people judge different pieces of music. Moreover, the eigenvectors of each principal component were extracted and visualized as scalp topographic plots to identify distinct regions of brain activation during exposure to hit musical stimuli.

Pearson Correlation Coefficient

Pearson Correlation Analysis is a statistical technique used to measure linear relationships between variables. It works by quantifying how strongly two sets of data move together, ranging from -1 (perfect negative correlation) to +1 (perfect positive correlation), with 0 indicating no linear relationship.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.5}$$

Pearson correlation provides a straightforward method for examining relationships between EEG signals and music ranking. The correlation coefficients measure how neural activity patterns vary with musical rankings in the charts. In this research, we applied this technique to analyze associations between principal components gathered from PCA and continuous song rankings. Further we analyzed the raw activations on electrodes with rankings using this technique. These correlations will help in revealing which electrodes show activity patterns correlates with musical ratings on which week. Here x represents the variable that we used in each time, principal component activations or raw activations and y represents the ranking of the individual music. r_{xy} represents the correlation value.

After computing the correlations, we examined both the strength and direction of significant relationships to understand how different electrodes contribute to musical rankings. This approach provided insights into which neural signals can be helpful in predict subjective musical rankings and at what time points these associations emerge.

3.6 Feature Analysis

With the feature extraction feature analysis is done in order to identify patterns in Electroencephalogram (EEG) with relation to rank and label of the songs. Since classification is more accurate when the pattern is simplified through representation by important features, feature extraction and selection play an important role in classifying systems such as neural networks (Übeyli 2009). These feature analysis is done in several stages where each stage compare different extracted feature with song label and rank.

Raw Activation Analysis

As the first stage of the analysis, we performed correlation analysis on raw electrode activations where microvolt-volt value was considered for correlation calculation.

First, we imported the preprocessed EEG recordings into Python using MNE, which is a preprocessing tool available in Python for MEG and EEG data analysis (Gramfort et al. 2013). As identified in the literature, we can divide the brain into different sections, which are known as lobes. Moreover, with 10-20 electrode placement, we can identify which electrode is in each lobe. Therefore, we divided the electrodes accordingly to section electrodes for each lobe.

Each data sequence represents a 10-second segment of the EEG recording, centered around the 30-second mark of the stimulus presentation, with a tolerance of ± 5 seconds. Therefore, each recording has a 10-second window around the 30-second mark of the EEG recording. For each electrode, signals were aggregated by cerebral lobe, and the mean raw activation was computed over the 10-second window following the onset of the auditory stimulus. This yielded EEG recording-wise averaged raw activation values. Subsequently, activations were averaged across trials of the same stimulus to derive a stimulus-specific activation profile. The resulting data matrix encapsulated lobe-wise mean activations for each stimulus. Finally, Pearson correlation coefficients, along with their associated ranks, were computed for each lobe to assess the strength and direction of correlations.

PCA on EEG Activation Analysis

In PCA analysis, we first import the preprocessed EEG recordings into Python using MNE, which is a preprocessing tool available in Python for MEG and EEG data analysis (Gramfort et al. 2013). After importing, a 10-second window around the 30-second mark of the EEG recording was extracted from each recording. The PCA was calculated using scikit-learn to reduce the dimensionality of electrodes into 5 principal components (Pedregosa et al. 2011).

After performing Principal Component Analysis (PCA), the mean, minimum, and maximum values of each principal component were computed. Subsequently, the Pearson correlation coefficients between these summary statistics (mean, min, and max) and the corresponding stimulus rankings were calculated. These correlation values were then analyzed to assess the relationship between the derived component features and the ranking data.

Furthermore, the eigenvectors obtained from the principal components were analyzed to identify the contribution of individual electrodes to each component. The corresponding eigenvector loadings were then used to compute correlations with stimulus rankings, providing additional insights into the relationship between spatial electrode patterns and perceptual evaluation.

Moreover, for identifying which lobe correlates with the stimuli rankings, electrodes were grouped according to their corresponding lobe using 10 -20 electrode placement system. After categorizing them PCA was carried out in order to reduce the dimensionality across electrodes. The dimensionality was reduce to two components. The reduced 2 principal components' mean, min and max were used for analysis. These reduced principal components were further analyzed to gather insights.

CWT Intensities Analysis

For time-frequency analysis of the EEG data, the Continuous Wavelet Transform (CWT) was employed to extract features. The CWT was applied to each preprocessed EEG recording using Morlet wavelet, enabling a detailed decomposition of the signal across time and frequency domains. For doing wavelet transformation we in-cooperated PyWavelets, a Python package for wavelet analysis (G. R. Lee et al. 2019).

The frequency spectrum was divided into three canonical EEG bands: alpha (8–13 Hz), beta (13–30 Hz), and theta (4–8 Hz). These ranges were defined by identifying the corresponding frequency index after computing the wavelet transform. For identifying indexes, we used an available function in the PyWavelets Python Library. For each recording, the sum of CWT intensities within each frequency band was computed for every electrode, yielding frequency-specific power estimates.

To analyze the relationship between spectral power and perceptual rankings, the average CWT intensity within each band was calculated over a 10-second window centered at the 30-second mark of the recording. These intensity values were then used to compute Pearson correlation coefficients with the stimulus rankings for each EEG channel, enabling the identification of spatial patterns of frequency-specific engagement.

Additionally, the correlations were computed separately for each week, allowing for a longitudinal analysis of how spectral features evolved over time and how they aligned with the participants' perceptual evaluations. These correlations were visualized across all channels and weeks to observe consistent or varying trends in frequency-band activity associated with subjective rankings.

3.7 Feature Classification

CNN Classification and Regression Models

The proposed model is a Convolutional Neural Network (CNN) tailored for EEG signal classification. It is composed of two sequential one-dimensional convolutional layers, each followed by a max pooling layer that progressively reduces the spatial dimensions of the feature maps while preserving the most significant features. Following these convolutional and pooling operations, the resulting feature maps are flattened into a one-dimensional vector and passed through a fully connected (dense) layer. This layer contains two output neurons for binary classification tasks, while a single output neuron is used when the model is adapted for regression tasks. Rectified Linear Unit (ReLU) activation functions are applied after each convolutional layer to introduce non-linearity and enhance learning of complex EEG patterns.

CNN-LSTM Classification and Regression Model

The proposed model is a hybrid CNN-LSTM architecture specifically developed for EEG signal classification. It begins with a one-dimensional convolutional layer that applies to the 32-channel input, effectively capturing salient temporal patterns in the EEG data. This is followed by a max pooling layer that reduces the dimensionality of the feature maps, helping to highlight the most significant activations while reducing computational complexity. The processed features are then fed into a single-layer Long Short-Term Memory (LSTM) network, which models the temporal dependencies across sequential time steps. The output from the LSTM is flattened and passed through a fully connected layer with two output neurons, enabling binary classification of the EEG signals. For regression tasks, the final layer was adjusted to a single output neuron. Rectified Linear Unit (ReLU) activation is employed after the convolutional layer to introduce non-linearity and enhance the model's ability to learn complex patterns.

Model Evalutaion

In this section, we will look into the algorithms used for classification. We utilized Convolutional Neural Networks (CNN) with Discrete Wavelet Transform (DWT) and Continuous Wavelet Transform (CWT) for classification. Models were trained for both classification tasks and regression tasks to gain insights into what could perform better. Moreover, for evaluation, we use accuracy, precision, and F1 score for classification, while utilizing Mean Absolute Error (MAE) and Mean Squared Error (MSE) for regression models.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.6}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.7}$$

$$\text{F1 Score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{3.8}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3.9}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{3.10}$$

The equations used for calculating each evaluation metric are as follows. Here, **TP** indicates True Positives and **TN** indicates True Negatives. **FP** denotes False Positives, while **FN** represents False Negatives. For regression tasks, y_i represents the true value and $\hat{y}_i$ denotes the predicted value, with n being the total number of samples.

Accuracy serves as a fundamental metric for assessing the overall performance of a classification model. It is the proportion of correctly predicted instances, including positive (Hits) and negative (Flops) outcomes, relative to the total number of instances. Although accuracy offers a general indication of a model's effectiveness, it can be misleading in the presence of class imbalance. For instance, in music classification tasks where Hit tracks vastly outnumber Flops, a model may achieve high accuracy by predominantly predicting the majority class. Therefore, additional evaluation metrics are essential for a more nuanced understanding of model performance.

Precision measures the proportion of correctly identified positive instances (true positives) among all instances predicted as positive. In the context of music classification, it reflects the model's ability to accurately predict Hit tracks while minimizing the misclassifications of Flops as Hits (false positives). High precision is particularly critical in applications such as music recommendation systems, where the erroneous suggestion of a non-Hit track may adversely impact user experience and trust.

The F1 score provides a score integrating both the model's ability to correctly identify all actual positives (recall) and its tendency to avoid false positives (precision). This metric is particularly informative in imbalanced datasets, as it penalizes models that exhibit disproportionately high precision or recall. For example, in a dataset dominated by Flops, a model that overlooks a significant number of true Hits (false negatives) may still achieve high precision, but its F1 score would be lower. A high F1 score thus indicates that the model maintains a balanced and reliable performance in identifying Hit tracks across both dimensions.

Mean Absolute Error (MAE) quantifies the average magnitude of the errors in a set of predictions, without considering their direction. It is computed as the mean of the absolute differences between predicted and actual values. MAE offers a clear and intuitive interpretation of prediction accuracy. A lower MAE signifies superior model performance.

The mean squared error (MSE) computes the average of the squared differences between predicted and actual values. By squaring the errors prior to averaging, MSE disproportionately penalizes larger deviations, thereby emphasizing the importance of minimizing significant prediction errors. This characteristic makes MSE particularly valuable in scenarios where large mispredictions.

To make sure our models work well with limited data available, we used K-Fold Cross-Validation. This method splits the data into K parts. The model is trained K times, each time using K-1 folds for training and the remaining fold for testing. This helps check if the model performs consistently across different splits, reducing the risk of overfitting. Finally, we average the results to get a reliable performance measure. This makes our evaluation more fair and accurate.

Convolutional Neural Networks (CNN) model is used to classify hit music using extracted DWT and CWT data. Here CNN convolve in each filter of the CWT and DWT data in order to find patterns in extracted scales. These patterns are identified during the learning process of the model. In prediction model tried to match identified patterns with new scales in order to classify the musical piece as a “Hit” or “Flop”.

Chapter 4

Results & Discussion

In this chapter we will dive deep into each stage of decisions taken to address the questions in the introduction. First stage involve in identifying hit music and flop music based on streaming platform data. In the next stage we analyzed the EEG recordings to classify hit and flops using psychophysiological responses. As the final stage we classified the musical pieces using the identified differences to find best model to predict hit songs. In this chapter, we will analyze the results of each stage, the methodology used to derive these results, and their impact on the subsequent research path. Additionally, we will investigate how the choices made at each stage influenced the overall direction of the study.

4.1 Stage 01: Analyzing Popular Music using Music Charts and Streaming Data

In this stage we aim to identify "Hit" songs through music charts. We examined both available music charts and the current literature in order to identify hit and flop musical pieces.

Initially, we analyzed the current literature in Hit Song Science (HSS) to identify how the literature labels musical pieces. With the analysis of the current literature we identified two main charts to used in our research. Therefore we used "Billboard Hot

100" and "Popnable Top 40 - Sri Lanka" for music stimuli selection.

To ensure a complete selection of music for our analysis, we incorporated both globally and locally recognized musical pieces. For global-level music, we refer to the Billboard Hot 100 chart, which is widely considered reliable source of global music trends. This chart evaluates multiple metrics, including album sales and the number of streams, which are critical in determining the commercial success and popularity of songs on a global scale[1]. To identify local-level music trends, we used the Popnable Top 40 chart. This platform primarily tracks YouTube view counts, supplemented by engagement metrics such as likes and streaming data, using bots to ensure data accuracy[2]. As Sri Lankan users use YouTube as one of the main song streaming platform, as supported by existing literature (Gunawardana and Thamarasee 2024), this chart served as an appropriate source for identifying popular local tracks. The selected songs for the analysis are available in Tables 3.1 and 3.2 in bold text.

After identifying the relevant music charts, we applied a threshold to classify musical pieces as either "Hit" or "Flop". As identified in the current literature most Hit Song Science (HSS) studies use a threshold value to identify "Hit" music (Soares Araujo, Pinheiro de Cristo, and Giusti 2019; Middlebrook and Sheik 2019; Zangerle et al. 2019; Yee and Raheem 2022). The threshold was set at the midpoint of each chart. Songs that appeared above this midpoint were labeled as "Hits", while those ranked below were considered "Flops". This classification was carried out weekly, and data was collected and analyzed over a period of 20 weeks to label the songs for each week.

4.2 Stage 02: Psychophysiological differences Hit Music

Following the identification of hit music and collection of EEG recordings, we conducted multiple analysis in order to identify neural correlations for song rankings. First, we computed per week Pearson correlations between EEG activations and song rankings for each musical stimulus to identify variations in neural activations. To address the high-

[1] https://www.billboard.com/pro/how-billboard-formulated-new-global-charts/

[2] https://popnable.com/terms

dimensional nature of the EEG data, we performed Principal Component Analysis (PCA) to reduce dimensionality while preserving variance, while calculating correlations between principal components and song rankings.

4.2.1 Raw Activation Analysis

Pearson correlation coefficients, computed for the raw activation data, were visualized graphically to examine the relationship between neural activations and the ranking of musical stimuli. This visualization facilitated the identification of temporal patterns in neural responses, highlighting the weeks during which neural activity exhibited the strongest correlation with rankings of the musical pieces on the relevant week. Figure 4.1 displays the correlation values for rank on each lobe plotted against week after the EEG was recorded.

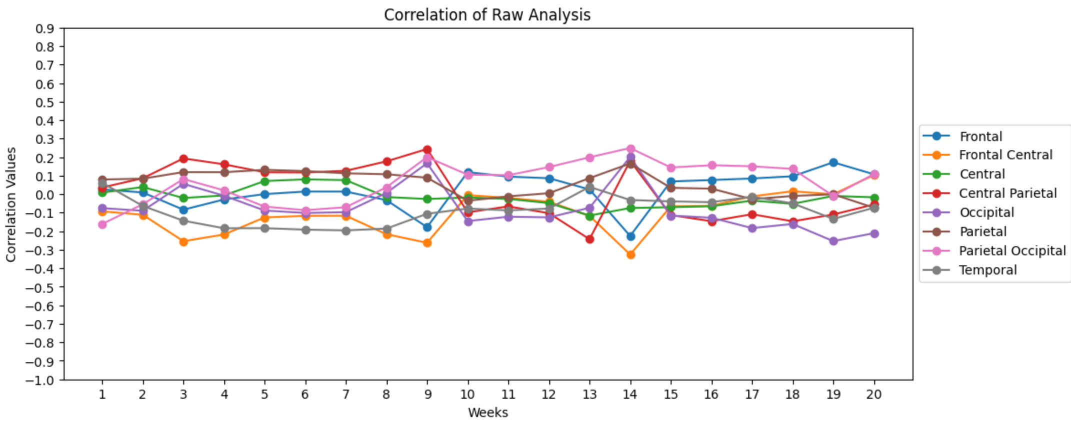
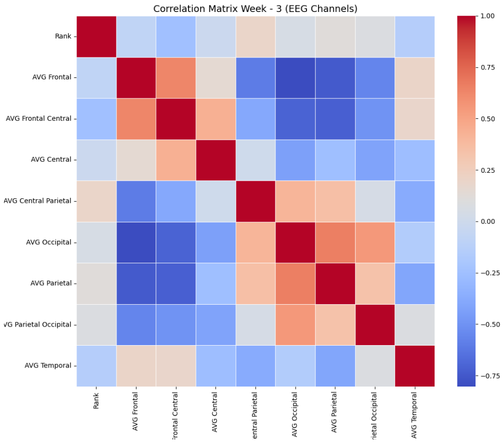
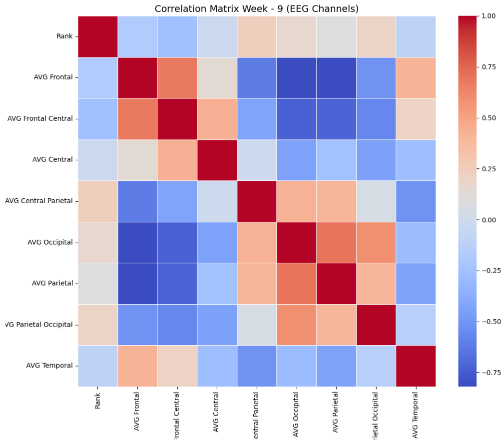


Figure 4.1: Pearson correlation coefficients (per lobe) for the relationship between raw EEG activation and stimulus rankings, plotted against weeks following data collection.

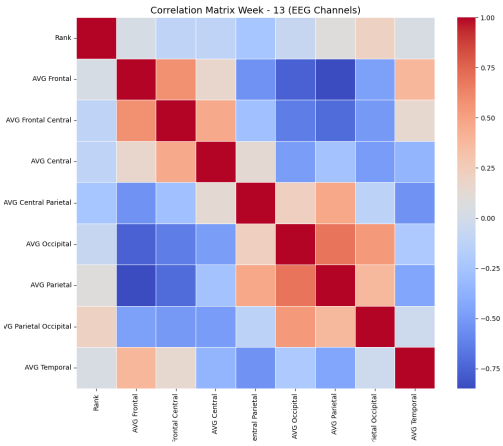
The plotted correlations revealed distinct patterns in neural activation across weeks. Notably, certain weeks exhibited stronger correlations between activations and musical rankings compared to others. Specifically, Week 3 Week 9, Week 13 and Week 14 demonstrated a divergent activation pattern that differed significantly from the correlations observed in other weeks. Correlation Matrices for each week that showed divergent activation patterns are illustrated in figures 4.2a, 4.2b, 4.2c, and 4.2d.



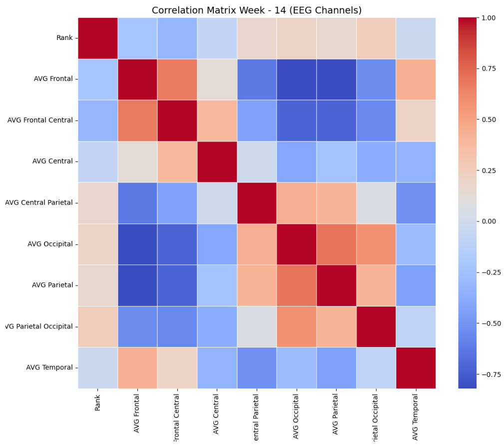
(a) Week 03 Correlation Matrix



(b) Week 09 Correlation Matrix



(c) Week 13 Correlation Matrix



(d) Week 14 Correlation Matrix

Figure 4.2: Correlation matrices of raw EEG activation and stimulus rankings across different weeks.

4.2.2 PCA on EEG Activation Analysis

Principal Component Analysis (PCA) performed on the EEG data was further analyzed to extract deeper insights into the underlying structure of the data. Specifically, correlation values along the principal components were examined to assess their relationship with stimulus rankings.

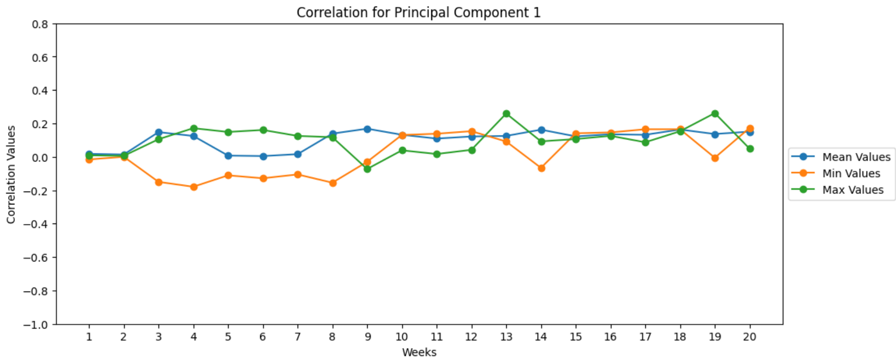
Figures 4.3 and 4.4 illustrate these findings, showing the correlation between the mean, minimum, and maximum values of each principal component and the associated stimulus rankings. These visualizations highlight how different components contribute to the perceptual evaluation over time.

The plotted correlation matrices reveal notable patterns across sessions. Specifically, a marked increase in correlation strength is observed during Week 13 and Week 14, indicating potential alignment between EEG features and stimulus rankings. Furthermore, Principal Components 1 and 4 exhibit distinguishable differences in Week 03. While multiple components show potential correlation in Week 13 and Week 14, some patterns in Week 03 also suggest early associations.

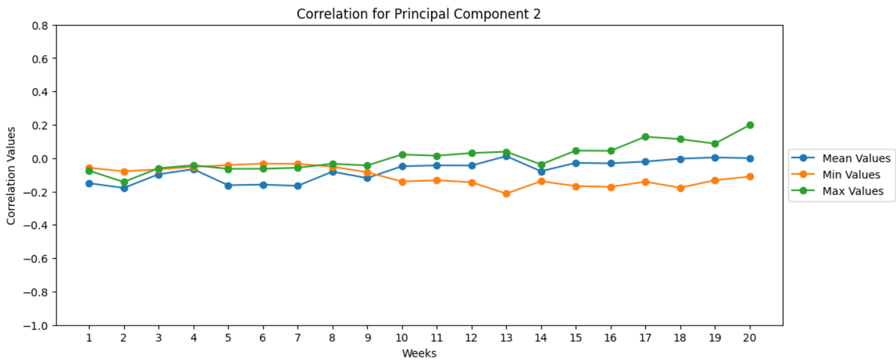
As expected from the PCA output, Principal Component 1 accounted for the largest proportion of variance, reinforcing its significance in capturing the most prominent features of the EEG signal. Given this, the corresponding eigenvector for Principal Component 01 was analyzed to evaluate the contribution of individual electrodes. The correlation between each eigenvector loading and stimulus rankings was computed, offering insights into which electrodes may be most involved in encoding perceptual relevance. Figure 4.5 illustrates the correlation between Principal Component 01 electrode contributions and stimulus rankings, highlighting regions with the strongest influence.

Figure 4.5 illustrates the temporal variation in correlation values between the eigenvector values of Principal Component 1 and stimulus rankings across all EEG electrodes over a 20-week period. Each colored line represents a different electrode, and the y-axis indicates the Pearson correlation coefficient.

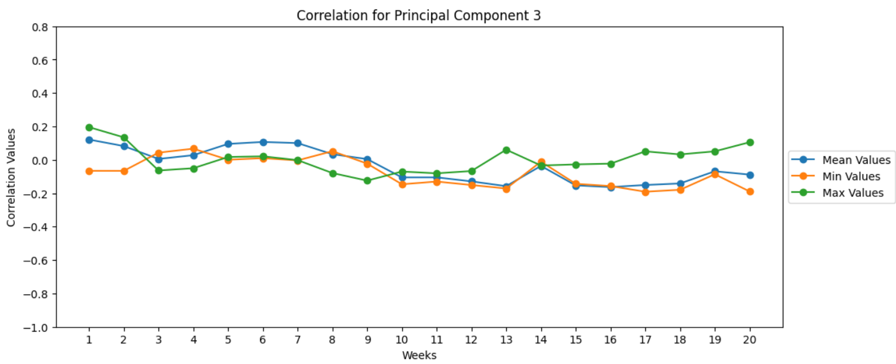
Notably, the correlation values show a pronounced increase during Weeks 13 and 14, with several electrodes exhibiting moderate correlation coefficients (both positive and negative). This suggests that during these weeks, the patterns captured by Principal



(a) Principal Component 1 Correlation

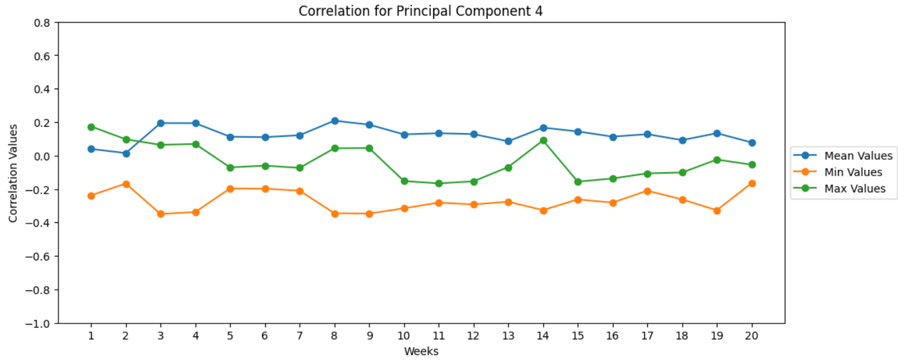


(b) Principal Component 2 Correlation

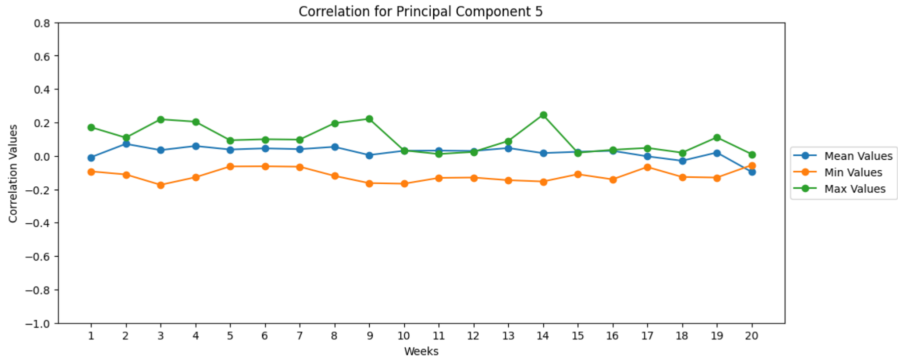


(c) Principal Component 3 Correlation

Figure 4.3: Correlation matrices across different weeks showing the relationship between principal component values and stimulus rankings.



(a) Principal Component 4 Correlation



(b) Principal Component 5 Correlation

Figure 4.4: Correlation matrices across different weeks showing the relationship between principal component values and stimulus rankings.

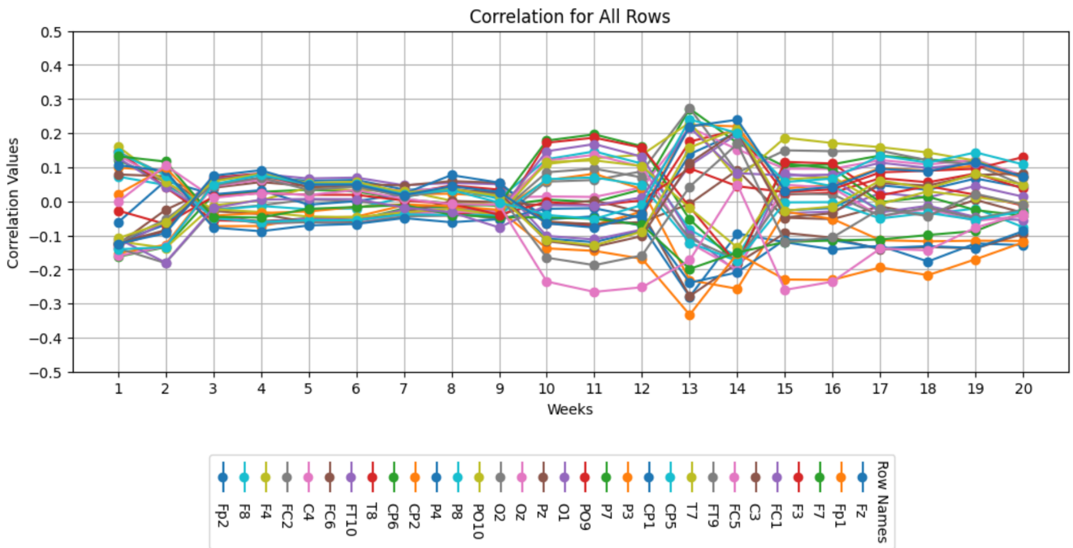


Figure 4.5: Electrode-wise correlation between Principal Component 1 eigenvector values and stimulus rankings.

Component 1 were more strongly aligned with stimulus rankings. Additionally, the period from Week 10 to Week 15 shows generally elevated correlation levels across many electrodes, compared to earlier and later sessions. Outside of this interval, particularly from Week 3 to Week 9, correlations are generally lower and more dispersed, indicating a weaker or more variable relationship between EEG activation patterns and stimulus rankings. The trend stabilizes again after Week 15, but the correlation strength continues to decrease.

To further explore the spatial dynamics of EEG activity in relation to stimulus rankings, lobe-wise correlation analysis was performed. Electrodes were grouped according to the 10–20 international electrode placement system, allowing analysis based on cortical lobes. The mean, minimum, and maximum values of the first two principal components were computed for each lobe and correlated with the stimulus rankings, similar to the previous analyses.

Among all lobes, the parietal lobe exhibited the highest correlation with the stimulus rankings, suggesting a stronger involvement of this region. Figure 4.6 presents the week-wise correlation trends for the parietal lobe, using the mean, minimum, and maximum values of Principal Components 1. Notably, a peak in mean correlation is observed around Week 13, further reinforcing findings from earlier analyses. Overall, the mean activation values from the parietal lobe showed the most consistent and strongest correlation with stimulus rankings when compared to other lobes.

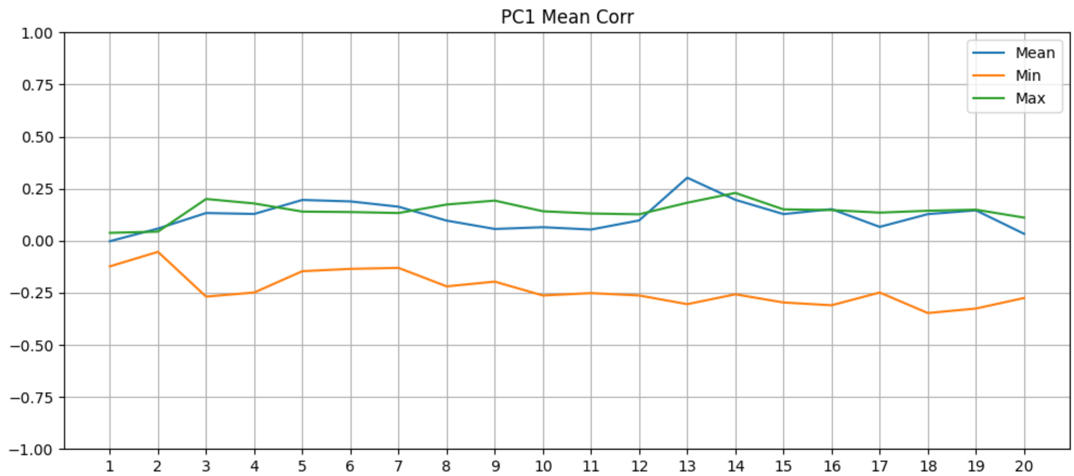


Figure 4.6: Parietal lobe principal component 01 correlation with stimuli rankings

To further explore the spatial contributions of electrodes to the principal components, topographic visualizations of the PCA eigenvectors were generated. These scalp maps, presented in Figures 4.7 and 4.8, illustrate the relative significance of each electrode in forming the principal components. Notably, clear differences in activation patterns are observed in the eigenvectors corresponding to Principal Components 03 to 05 when comparing hit and flop musical pieces identified during Week 13. These differences are evident in both the spatial distribution of the topographic lines and the variations in color intensity across the scalp across all principal components. Such observations suggest that the brain's spatial response patterns vary distinctly between stimuli that are perceived as hits versus flops, providing further insight into the neural encoding of music preference.

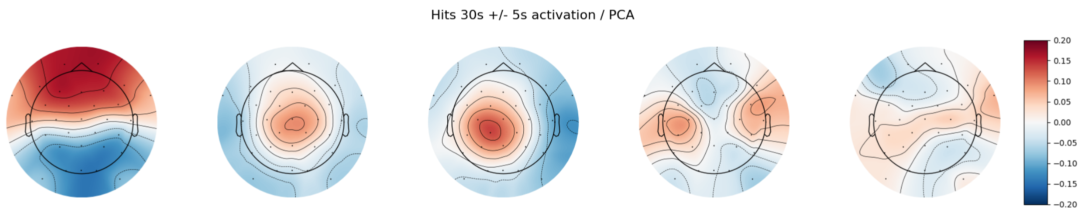


Figure 4.7: Topographic Visualization of PCA Eigenvectors for Hit Musical Pieces

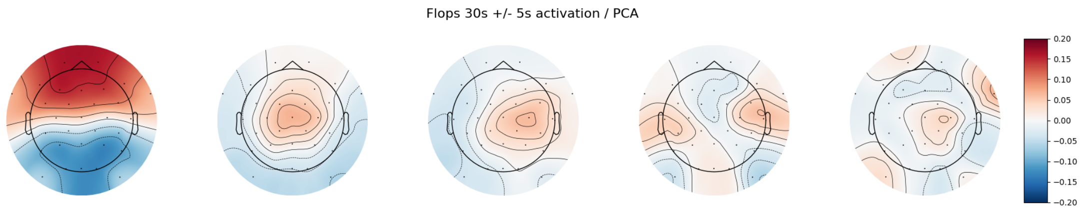


Figure 4.8: Topographic Visualization of PCA Eigenvectors for Flop Musical Pieces

These findings further support the observation that Weeks 10 to 15 demonstrate the strongest alignment between EEG features and musical rankings, as they exhibit consistently higher correlation values compared to other time periods.

4.2.3 CWT Intensities Analysis

The Continuous Wavelet Transform (CWT) analysis was conducted to explore the relationship between EEG frequency bands and participants' perceptual rankings of music stimuli. Specifically, the goal was to identify which frequency components exhibit the strongest correlation with subjective rankings across time. For each frequency band, the average intensity per electrode was extracted using CWT and subsequently correlated with stimulus rankings over a 20-week period.

Figures 4.9, 4.10, and 4.11 illustrate the correlation patterns between the power of the

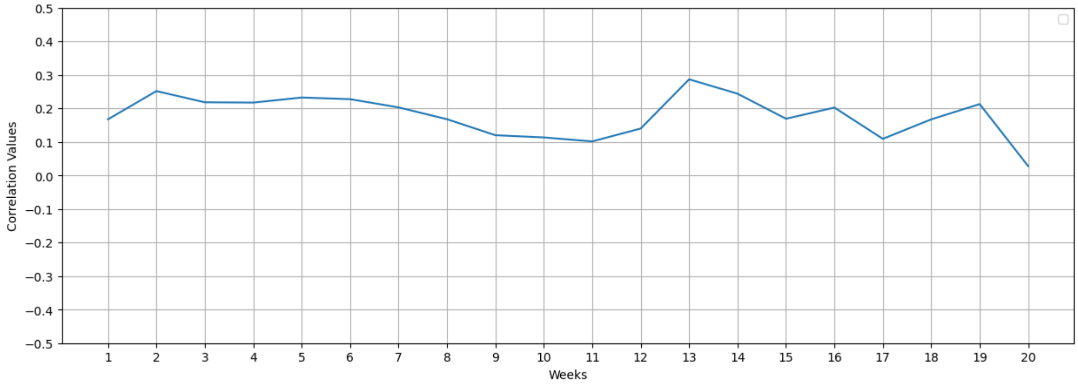


Figure 4.9: Pearson Correlations Between Alpha Band EEG Power and Music Stimulus Rankings Across 20 Weeks

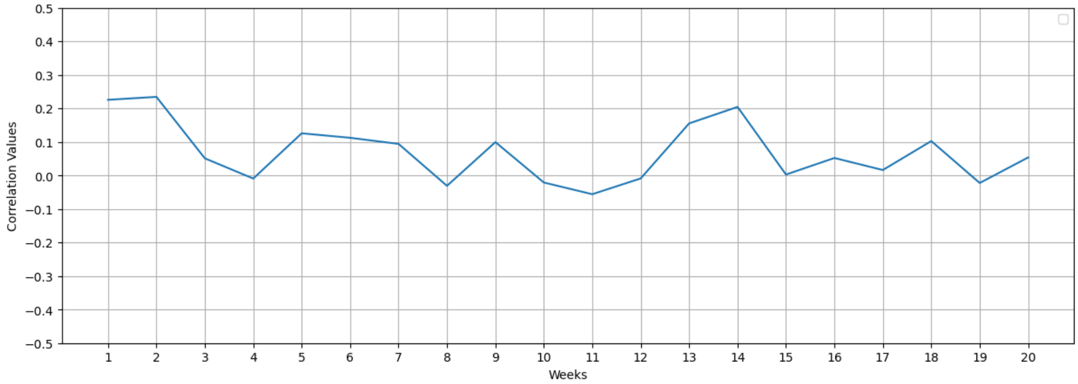


Figure 4.10: Pearson Correlations Between Beta Band EEG Power and Music Stimulus Rankings Across 20 Weeks

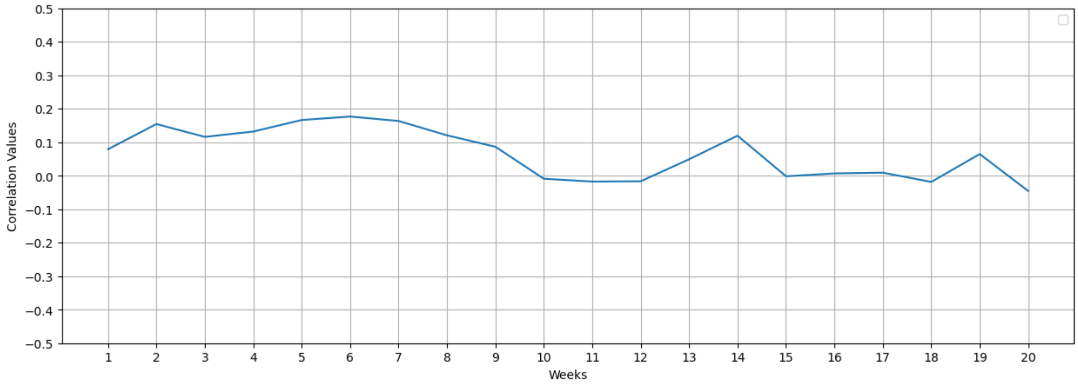


Figure 4.11: Pearson Correlations Between Theta Band EEG Power and Music Stimulus Rankings Across 20 Weeks

Alpha, Beta, and Theta frequency bands, respectively, and the stimulus rankings across weeks. A prominent and consistent observation across all three frequency bands is the marked increase in correlation values during Week 13 and Week 14. This spike suggests a significant neural response to the stimuli during that specific time. Among the frequency bands, the Alpha band demonstrates the widest range and strongest correlations, with Week 13 showing the peak values in comparison to all other weeks. The Alpha band correlations exhibit considerable variation across electrodes.

This aligns with previous findings from PCA and preliminary statistical analyses, which highlighted Weeks 10 to 15 as a period that has an observable correlation. Although there can be seen mild spikes on other weeks, current CWT-based results also further support this observation, notably in Alpha and Beta activations.

The identified differences, particularly those observed between Weeks 10 and 15, highlight the potential for distinguishing hit songs based on neural responses. All the analysis indicates Weeks 13 and 14 or one of them as higher correlation time period. These findings provide a strong foundation for developing predictive models aimed at classifying hit songs within this critical time window Week 10 - Week 15.

4.3 Stage 03: Models to Predict Hit Music

Following the identification of psychophysiological differences associated with hit music prediction, model development and training were undertaken to further explore the feasibility of classification. As outlined in the Methodology section, regression models were evaluated using the Mean Squared Error (MSE) metric, while classification models were assessed based on their overall accuracy. These evaluation metrics were selected to provide a comprehensive understanding of each model's predictive performance, highlighting both their strengths and limitations.

For model training, features extracted using both Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) techniques were utilized. Accordingly, each predictive model was implemented in two variants, one using CWT-based features and the other using DWT-based features, to examine the impact of the chosen feature extraction method on classification and regression performance.

4.3.1 CNN Classification Models

CWT CNN Classification Models

Figure 4.12 illustrates the classification accuracies of the CNN model utilizing CWT input over a six-week period, from Week 10 to Week 15. The test accuracies for each respective week were as follows: 62.97% (Week 10), 63.01% (Week 11), 60.36% (Week 12), 65.58% (Week 13), 58.66% (Week 14), and 57.83% (Week 15). The model achieved its highest performance in Week 13, with an accuracy of 65.58%. Weeks 10 and 11 also demonstrated relatively strong performance, with accuracies of 62.97% and 63.01%, respectively. However, a decline in performance was observed in the final weeks, with the accuracy falling to 58.66% in Week 14 and further decreasing to 57.83% in Week 15.

Figure 4.13 presents the precision scores of the CNN model using continuous wavelet transform (CWT) inputs across Weeks 10 to 15. The precision values for each respective week were 0.548 (Week 10), 0.580 (Week 11), 0.554 (Week 12), 0.574 (Week 13), 0.538 (Week 14), and 0.539 (Week 15). The highest precision was recorded in Week 11 at 0.580, indicating that the model was most effective during this week in minimizing false positives when predicting Hit tracks. Performance remained relatively stable in 13, with a score of 0.57. However, a slight decline in precision was observed in Weeks 14 and 15, suggesting a reduced ability to avoid misclassifying Flops as Hits during the latter stages of evaluation.

Figure 4.14 depicts the F1 scores of the CNN model utilizing CWT representations over the same six-week period. The F1 scores were 0.554 (Week 10), 0.581 (Week 11), 0.566 (Week 12), 0.586 (Week 13), 0.545 (Week 14), and 0.551 (Week 15). The model achieved its highest F1 score in Week 13, reaching 0.586, indicating an optimal balance between precision and recall during this period. Weeks 11 and 12 also showed competitive performance, with F1 scores above 0.56.

DWT CNN Classification Models

Figure 4.15 displays the test accuracies of the CNN model employing discrete wavelet transform (DWT) input across Weeks 10 to 15. The corresponding accuracies were 62.03%

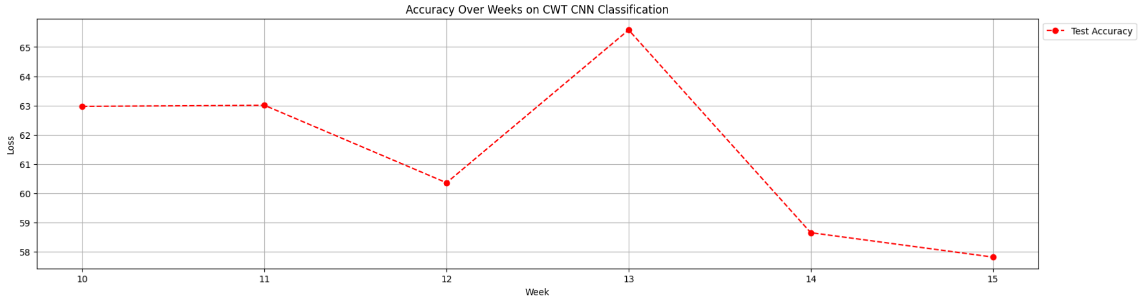


Figure 4.12: Accuracy of CWT CNN Classification model from Week 10 to Week 15

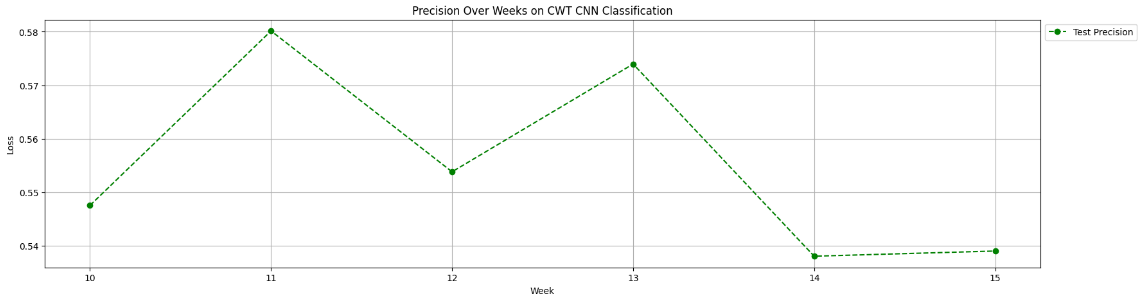


Figure 4.13: Precision of CWT CNN Classification model from Week 10 to Week 15

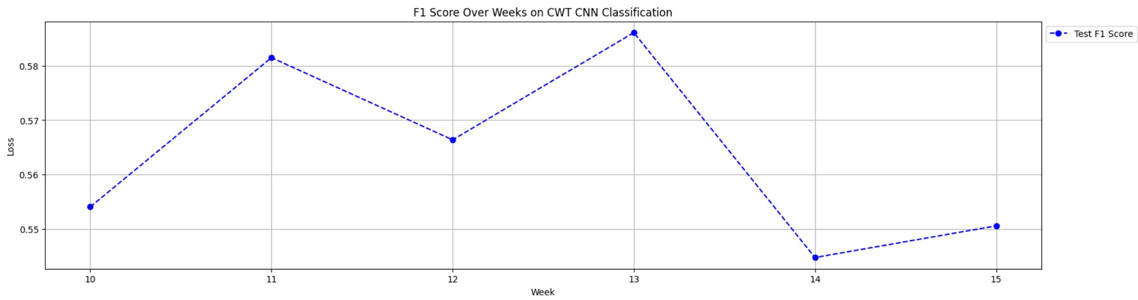


Figure 4.14: F1 Score of CWT CNN Classification model from Week 10 to Week 15

(Week 10), 54.28% (Week 11), 59.46% (Week 12), 65.43% (Week 13), 56.09% (Week 14), and 59.42% (Week 15). Among these, the highest accuracy was observed in Week 13, reaching 65.43%, suggesting that the model was most effective during this period in correctly classifying both Hit and Flop tracks. In contrast, performance in Week 11 declined significantly to 54.28%, indicating challenges in generalization during that week.

As shown in Figure 4.16, the model's precision values across the six-week span were 0.522 (Week 10), 0.442 (Week 11), 0.496 (Week 12), 0.603 (Week 13), 0.594 (Week 14), and 0.642 (Week 15). The peak precision was recorded in Week 15 at 0.642, indicating a strong capability to avoid false positives late in the evaluation period. However, the precision in Week 13, with a value of 0.603, coincided with the model's highest accuracy. The relatively low precision in Week 11 (0.442) suggests a high rate of false positive predictions during that period.

Figure 4.17 illustrates the F1 scores for the DWT-based CNN model from Week 10 to Week 15. The scores were 0.556 (Week 10), 0.475 (Week 11), 0.520 (Week 12), 0.620 (Week 13), 0.489 (Week 14), and 0.510 (Week 15). The model achieved its highest F1 score in Week 13 at 0.620, underscoring its optimal balance between precision and recall during this period. This reflects the model's strong performance not only in avoiding false positives but also in capturing a majority of true positive cases. In contrast, the low F1 score of 0.475 in Week 11 further confirms the decline in predictive balance and overall effectiveness in that week.

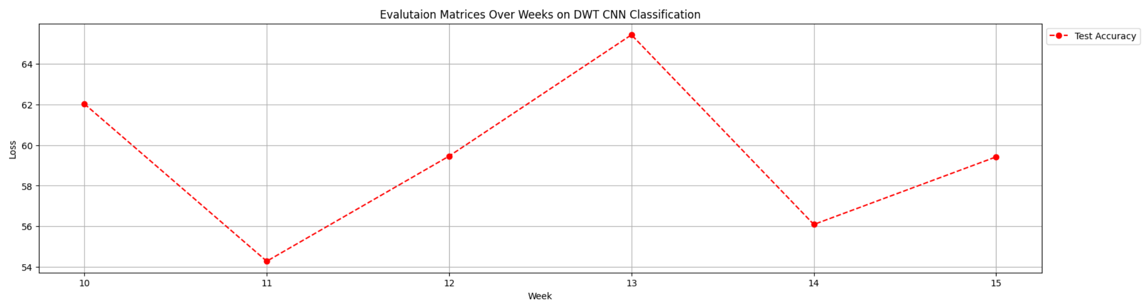


Figure 4.15: Accuracy of DWT CNN Classification model from Week 10 to Week 15

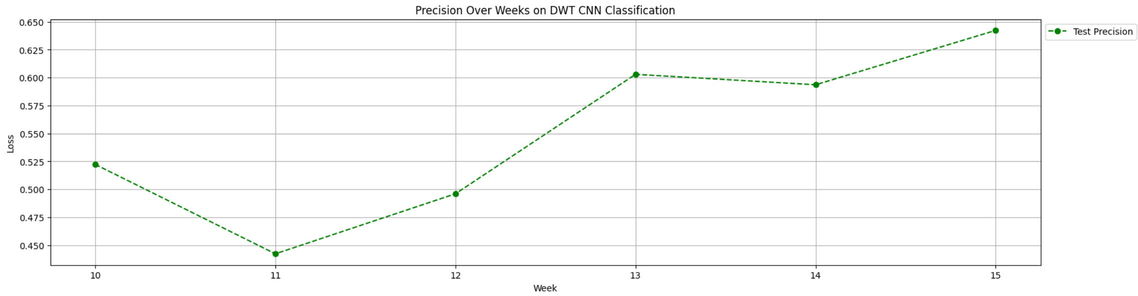


Figure 4.16: Precision of DWT CNN Classification model from Week 10 to Week 15

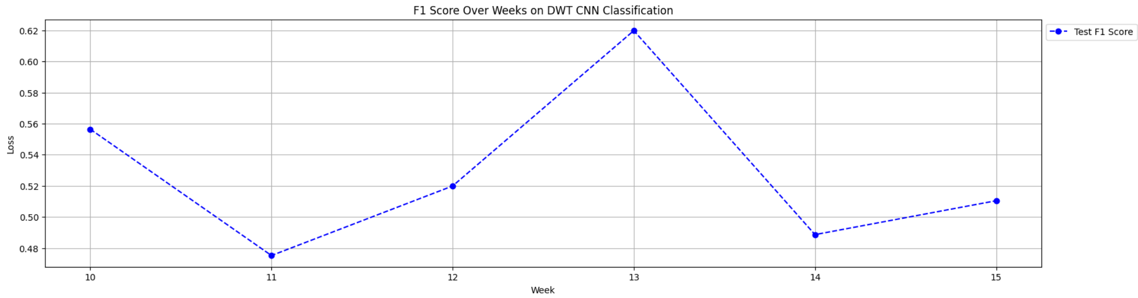


Figure 4.17: F1 Score of DWT CNN Classification model from Week 10 to Week 15

4.3.2 CNN-LSTM Classification Models

CWT CNN-LSTM Classification Models

Figure 4.18 illustrates the test accuracies of the CNN-LSTM model utilizing continuous wavelet transform (CWT) input across Weeks 10 to 15. The highest accuracy was recorded in Week 13, reaching 61.16%. The model also performed relatively well in Weeks 10 and 12, both yielding an accuracy of 58.73%. In contrast, Week 11 showed the lowest accuracy at 54.31%, followed closely by Week 15 with 54.28%.

Figure 4.19 presents the precision values of the CNN-LSTM model using CWT input over the same period. The precision scores for Weeks 10 to 15 were 0.545, 0.529, 0.631, 0.587, 0.575, and 0.493, respectively. The highest precision was achieved in Week 12 (0.631), indicating that during this period, the model most effectively minimized false positives in Hit classification. Precision remained above 0.57 in Weeks 13 and 14, reflecting relatively strong predictive reliability. However, the precision dropped in Week 15 to 0.493, suggesting a decrease in the model's ability to correctly identify Hit tracks without misclassifying Flops.

Figure 4.20 depicts the F1 scores of the CNN-LSTM model over the evaluation period. The scores across Weeks 10 through 15 were 0.537, 0.505, 0.567, 0.546, 0.555, and 0.502, respectively. The highest F1 score was recorded in Week 12 at 0.567, indicating a well-balanced performance between precision and recall. Week 13 also exhibited a relatively high F1 score, underscoring the model's capacity to maintain equilibrium between minimizing false positives and capturing true positives.

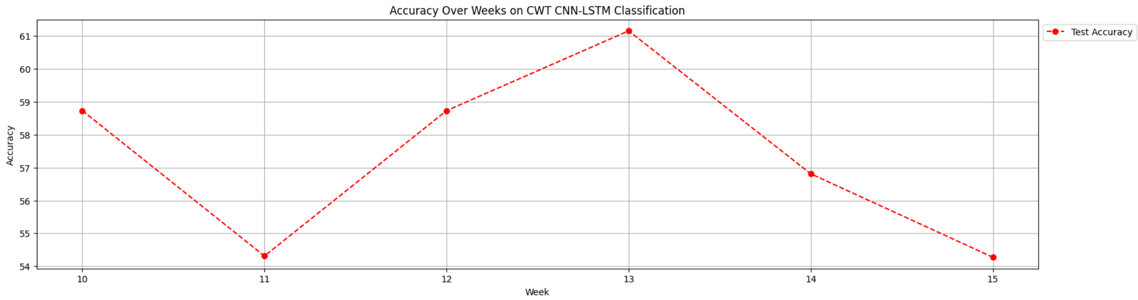


Figure 4.18: Accuracy of CWT CNN-LSTM Classification model from Week 10 to Week 15

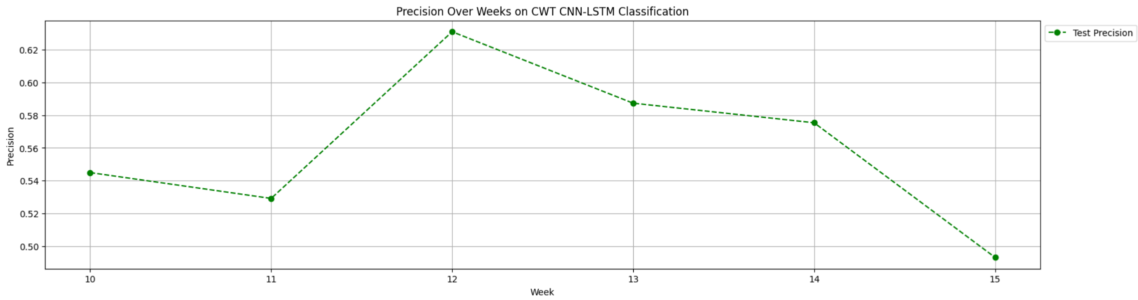


Figure 4.19: Precision of CWT CNN-LSTM Classification model from Week 10 to Week 15

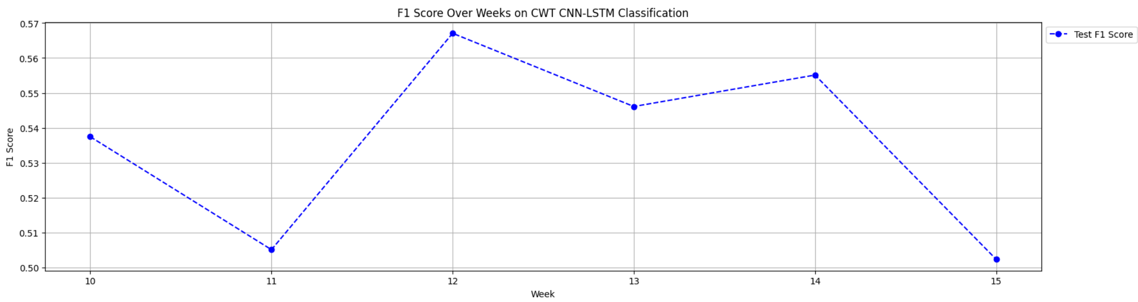


Figure 4.20: F1 Score of CWT CNN-LSTM Classification model from Week 10 to Week 15

DWT CNN-LSTM Classification Models

Figure 4.21 presents the classification accuracies of the CNN-LSTM model trained on discrete wavelet transform (DWT) input from Week 10 to Week 15. The model achieved its highest accuracy in Week 13 at 59.31%, indicating relatively strong performance in correctly identifying both Hit and Flop tracks during that period. Other weeks showed moderate to low performance, with accuracies of 54.46% (Week 12), 53.55% (Week 10), and 52.57% (Week 11). The lowest accuracies were observed in Week 14 and Week 15.

As shown in Figure 4.22, the precision scores for the CNN-LSTM model using DWT input were 0.443 (Week 10), 0.452 (Week 11), 0.508 (Week 12), 0.596 (Week 13), 0.445 (Week 14), and 0.446 (Week 15). The highest precision was achieved in Week 13 (0.596), indicating that the model was most successful in limiting false positive predictions during that week. These results suggest that, except for Week 13, the model struggled to consistently differentiate Hit tracks from Flops without misclassification.

Figure 4.23 illustrates the F1 scores across the six-week evaluation period. The scores were 0.472 (Week 10), 0.443 (Week 11), 0.502 (Week 12), 0.557 (Week 13), 0.447 (Week 14), and 0.432 (Week 15). The highest F1 score was recorded in Week 13 at 0.557, indicating a relatively well-balanced trade-off between precision and recall during this period. The model's performance in other weeks was comparatively weaker, with F1 scores falling below 0.51. The lowest F1 score occurred in Week 15 (0.432).

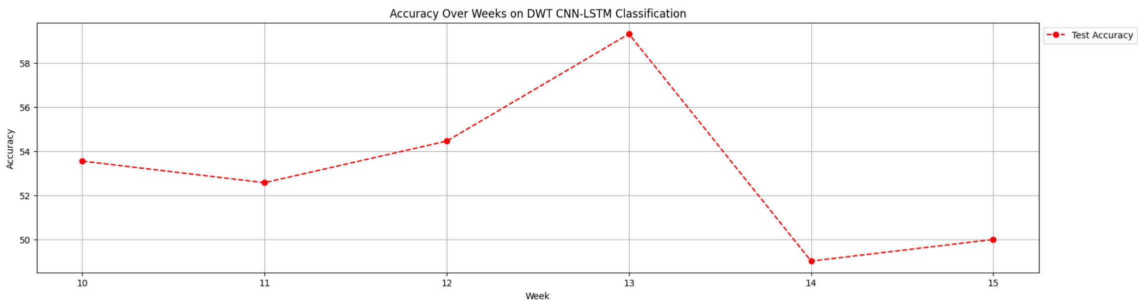


Figure 4.21: Accuracy of DWT CNN-LSTM Classification model from Week 10 to Week 15

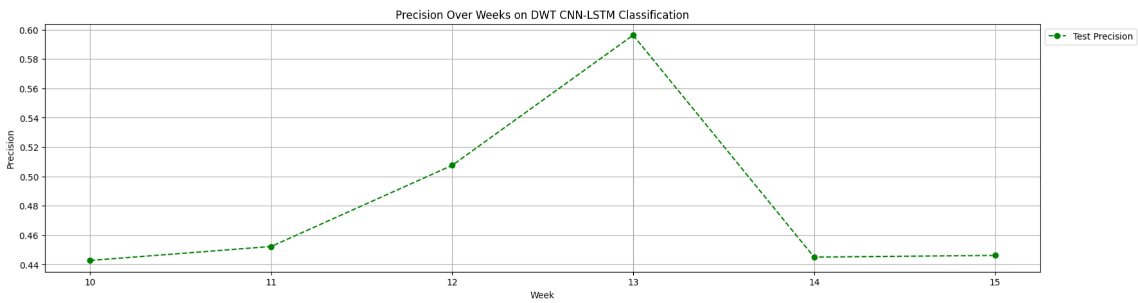


Figure 4.22: Precision of DWT CNN-LSTM Classification model from Week 10 to Week 15

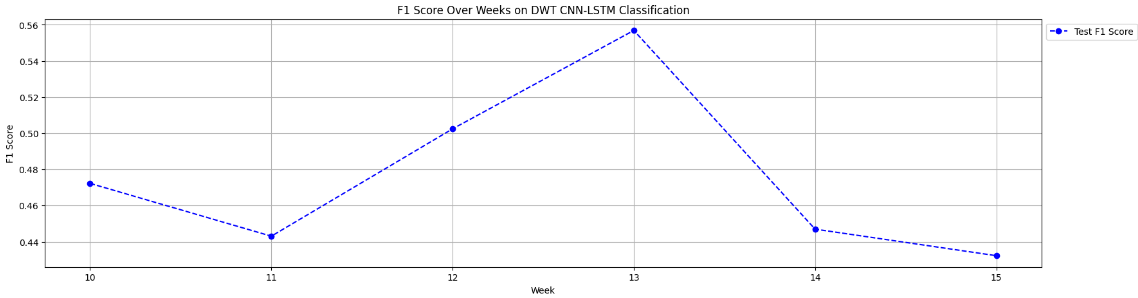


Figure 4.23: F1 Score of DWT CNN-LSTM Classification model from Week 10 to Week 15

4.3.3 CNN Regression Models

CWT CNN Regression models

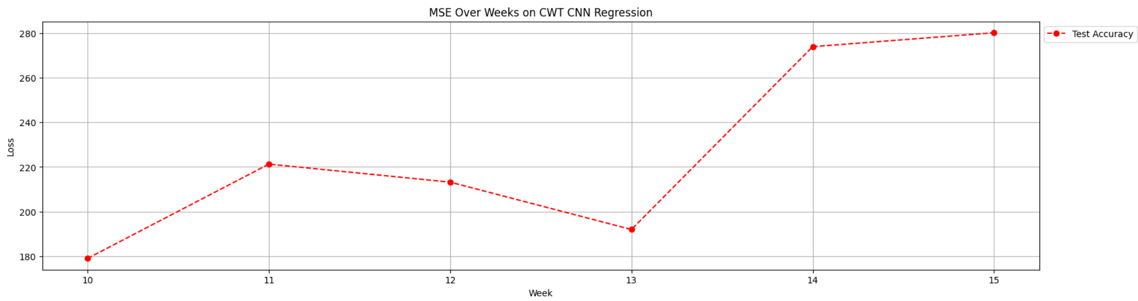


Figure 4.24: MSE of CWT CNN Regression model from Week 10 to Week 15

Figure 4.25 illustrates the test mean squared error (MSE) values of the CNN regression model utilizing discrete wavelet transform (DWT) input features across Weeks 10 to 15. The model attained its best performance in Week 10, yielding the lowest MSE value of 179.03, indicating superior accuracy in predicting target values during this period. Week 10 also demonstrated commendable performance, with an MSE of 191.97—closely approximating the optimal value and representing the second-lowest error across all evaluated weeks. In contrast, Week 15 recorded the highest MSE value of 280.14, suggesting considerable deviation between predicted and actual values and highlighting the model's weakest performance.

DWT CNN Regression Model

Figure 4.24 presents the test mean squared error (MSE) results for the CNN regression model employing continuous wavelet transform (CWT) input features from Week 10 to

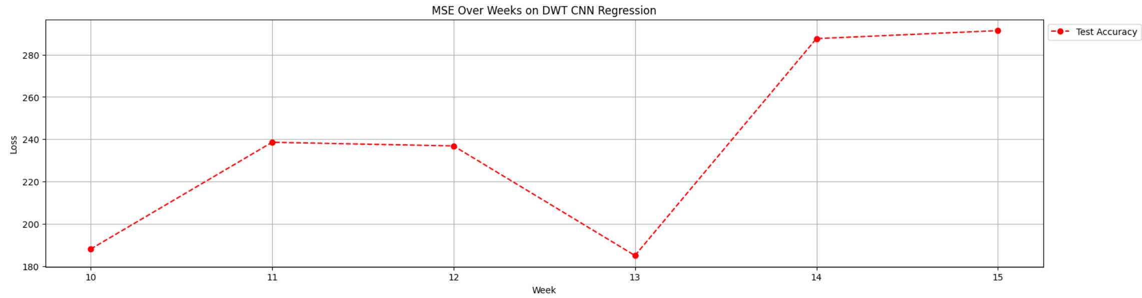


Figure 4.25: MSE of DWT CNN Regression model from Week 10 to Week 15

Week 15. The model achieved its best predictive performance in Week 13, producing the lowest MSE value of 185.01. The next most accurate result was observed in Week 10, with an MSE of 188.07, reflecting a similarly high level of precision. On the other hand, Week 15 marked the model's poorest performance, as evidenced by the highest MSE value of 291.34, indicating a substantial prediction error during that period.

4.3.4 CNN-LSTM Regression Models

CWT CNN-LSTM Regression Model

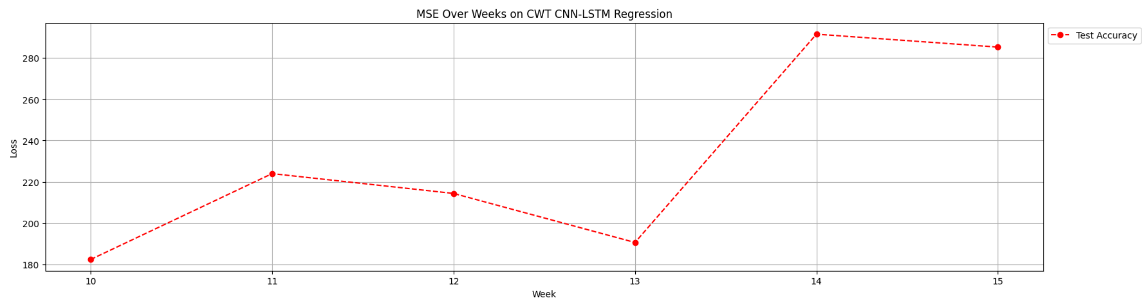


Figure 4.26: MSE of CWT CNN-LSTM Regression model from Week 10 to Week 15

Figure 4.26 presents the test mean squared error (MSE) values for the CNN-LSTM regression model utilizing continuous wavelet transform (CWT) input features from Week 10 to Week 15. The model exhibited optimal performance in Week 10, achieving the lowest MSE value of 182.33. Week 13 also demonstrated strong predictive accuracy, with an MSE of 190.57, representing the second-lowest error and indicating stable model generalization. In contrast, Week 14 yielded the highest MSE value of 291.59, suggesting substantial prediction discrepancies and marking the model's weakest performance across the evaluated period.

DWT CNN-LSTM Regression Model

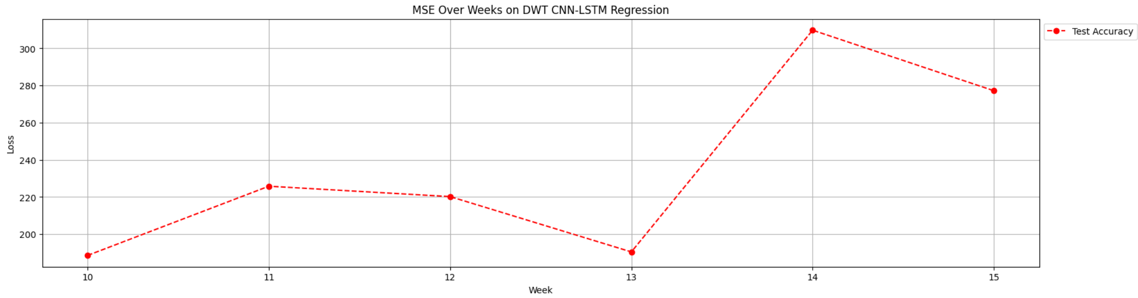


Figure 4.27: MSE of DWT CNN-LSTM Regression model from Week 10 to Week 15

Figure 4.27 displays the test mean squared error (MSE) values for the CNN-LSTM regression model trained on discrete wavelet transform (DWT) input features. The model achieved its best performance in Week 10, with the lowest MSE value of 188.58, indicating effective estimation of the target variable. Similarly, Week 13 produced the second-best result, with an MSE of 190.48, reflecting consistent and robust predictive capabilities during that interval. Conversely, the model's poorest performance occurred in Week 14, where the MSE peaked at 309.76, indicating considerable prediction error and limited model reliability in that initial testing phase.

4.3.5 Discussion

The primary aim of the research was to develop a computational model that predicts Hit Songs by analyzing listeners' psy-chophysiological responses. Therefore following the selection and feature extraction, we used existing deep learning algorithms to predict hit songs.

In summary, the trained models consistently demonstrated improved predictive performance during Week 13 across all classification tasks, with most models achieving either their highest or second-highest classification accuracy in that week. This trend suggests that the data characteristics in Week 13 may have been particularly conducive to model generalization.

Among the classification models, the CWT-based CNN classifier emerged as the best-performing model in terms of accuracy, achieving the highest score of 65.58% in Week 13. However, the DWT-based CNN classifier, while slightly lower in accuracy at 65.43%, demonstrated higher precision and F1 score during the same week, with values of 0.60 and 0.62, respectively. This highlights its effectiveness in correctly identifying Hit songs while maintaining a better balance between false positives and false negatives.

In the regression task, where the goal was to predict song rankings based on streaming performance, the CWT-based CNN model trained on Week 10 data yielded the most accurate results with the lowest mean squared error (MSE) of 179.03, indicating strong predictive accuracy with minimal deviation from actual values. Additionally, the DWT-based CNN regression model demonstrated excellent performance in Week 13, achieving the substantially best MSE of 185.01, further reinforcing Week 13's significance across model types.

These results suggest that, while multiple architectures contributed valuable predictive capabilities, the DWT CNN model in Week 13 stands out as the most robust for classification tasks, and the CWT CNN model in Week 10 is the most reliable for regression. The consistent success observed in Week 13 across models may warrant further investigation into temporal characteristics specific to that week.

Overall, the proposed mechanism, and computational model can be effective for EEG-based hit music prediction. However, future research should focus on improving the accuracy of the computational models, exploring the generalization of the findings to different populations and contexts, and investigating the effectiveness of the proposed tools in detecting more complex hit music classification.

Chapter 5

Conclusion

The objective of this research was to overcome the challenges in identifying Hit Songs by analyzing listeners' psychophysiological responses. The research focus on three main research questions:

RQ1 How to measure the Hit Songs based on streaming platform data?

RQ2 What psychophysiological differences can be identified when listening to identified Hit Songs?

RQ3 What machine learning and deep learning algorithms are best for identifying Hit Songs using psychophysiological responses?

To address RQ1, we examined relevant literature to identify the most appropriate musical charts for the selection process. By incorporating both globally recognized and locally relevant charts, we ensured a balanced representation of hit music at both the international and local levels. This approach allowed for a comprehensive musical selection that reflects diverse listening contexts. The choice of charts was guided by preliminary investigations and insights drawn from current literature.

To address RQ02, we conducted a comprehensive analysis of EEG recordings to explore psychophysiological differences associated with the EEG recordings for identified hit songs. The investigation involved multiple analytical approaches, including raw activation analysis, dimensionality reduction via Principal Component Analysis (PCA), and time-frequency analysis through Continuous Wavelet Transform (CWT).

The raw activation analysis revealed clear correlations with music rankings, with Weeks 3, 9, 13, and 14. These patterns were further validated through correlation matrices, which highlighted changes in neural responses aligned with changes in musical stimulus rankings. PCA enabled the identification of components associated with musical rankings. Moreover, Principal Component 1 captures the largest variance, and indicates a strong correlation with stimulus rankings, particularly during Weeks 13 and 14. Lobe-wise analysis revealed that the parietal lobe exhibited the strongest correlations, indicating its key role. Additionally, scalp topography visualizations illustrated spatial differences in PCA eigenvectors between hit and flop songs, especially for Principal Components 3 to 5, further highlighting neural distinctions. Moreover, eigenvector on principal component 01 indicates a higher correlation on musical stimuli rankings.

Complementing these findings, CWT-based frequency domain analysis revealed that alpha, beta, and theta frequency bands showed higher correlation with song rankings during Week 14. This pattern was consistent with earlier findings from the PCA and raw activation analysis, reinforcing the significance of the Week 10 to Week 15 time window as a period of heightened neural sensitivity to hit songs.

In conclusion of RQ2, the analyses underscore the existence of measurable psychophysiological differences in brain activity when individuals are exposed to hit songs. The consistent indication of Weeks 13 and 14 as time points of strong neural correlation with music rankings can be identified as a time period where higher psychophysiological differences are captured.

In addressing RQ3, We used deep learning models to predict hit musics and their rankings for the weeks after each recording was gathered. The trained models were evaluated on gathered EEG data. 48 models were trained, and Week 13 on the DWT CNN model performed best in identifying Hit songs as a classification problem with an accuracy of 65.43% with 0.60 precision and 0.62 F1-score. Moreover, the Week 10 model on CWT CNN performed best in predicting ranks as a regression problem where it had an MSE of 179.03.

This study presents an effective and efficient approach to EEG-based hit song prediction, offering valuable insights for researchers and practitioners in the field. As part of this work, a novel dataset comprising 179 EEG recordings was developed, which can serve as a valuable resource for future research. The current methodology and algorithms provide a strong foundation but also leave room for further refinement. With continued development, these techniques can be enhanced to achieve higher accuracy and more generalizable outcomes across diverse social and cultural backgrounds, paving the way for more inclusive and robust applications in music cognition and preference prediction.

Bibliography

Agres, Kat R. et al. (2021). “Music, Computing, and Health: A Roadmap for the Current and Future Roles of Music Technology for Health Care and Well-Being”. In: *Music & Science* 4, p. 2059204321997709. DOI: 10.1177/2059204321997709. eprint: https://doi.org/10.1177/2059204321997709. URL: https://doi.org/10.1177/2059204321997709.

Alarcão, Sara M. and Maria J. Fonseca (2017). “Re-referencing in EEG: A review of methodologies and their impact on data analysis”. In: *Journal of Neuroscience Methods* 281, pp. 1–9.

Alarcão, Soraia M. and Manuel J. Fonseca (2019). “Emotions Recognition Using EEG Signals: A Survey”. In: *IEEE Transactions on Affective Computing* 10.3, pp. 374–393. DOI: 10.1109/TAFFC.2017.2714671.

Allen, John (Feb. 2007). “Photoplethysmography and its application in clinical physiological measurement”. In: *Physiological Measurement* 28.3, R1. DOI: 10.1088/0967-3334/28/3/R01. URL: https://dx.doi.org/10.1088/0967-3334/28/3/R01.

AmplifyYou (Sept. 2021). *How Does Our Brain Process Sound?* Accessed: 2025-04-17. URL: https://amplifyyou.amplify.link/2021/09/how-does-our-brain-process-sound/.

Antony, Monica, Vishnu Priya.V, and R. Gayathri (2018). “Effect of music on academic performance of college students”. In: *Drug Invention Today*, pp. 2093–2096. URL: https://api.semanticscholar.org/CorpusID:186351545.

Ayata, Deger, Yusuf Yaslan, and Mustafa Kamasak (2017). “EMOTION RECOGNITION VIA GALVANIC SKIN RESPONSE: COMPARISON OF MACHINE LEARNING ALGORITHMS AND FEATURE EXTRACTION METHODS”. In: 17 (1), pp. 3129–3136.

Baur, Dominik et al. (2019). "A survey on EEG-based emotion recognition: recent advances, current limitations and future trends". In: *Journal of Neural Engineering* 16.3, p. 031002. DOI: 10.1088/1741-2552/ab260c.

Bellmann, Oliver Tab and Rie Asano (June 2024). "Neural Correlates of Musical Timbre: An ALE Meta-Analysis of Neuroimaging Data". In: *Frontiers in Neuroscience* 18. Section: Auditory Cognitive Neuroscience, Original Research Article. DOI: 10.3389/fnins.2024.1373232. URL: https://doi.org/10.3389/fnins.2024.1373232.

Bernardi, Luciano, C. Porta, and P. Sleight (Apr. 2006). "Cardiovascular, cerebrovascular, and respiratory changes induced by different types of music in musicians and non-musicians: The importance of silence". In: *Heart* 92 (4), pp. 445–452. ISSN: 13556037. DOI: 10.1136/hrt.2005.064600.

Berns, Gregory S. and Sara E. Moore (Jan. 2012). "A neural predictor of cultural popularity". In: *Journal of Consumer Psychology* 22 (1), pp. 154–160. ISSN: 10577408. DOI: 10.1016/j.jcps.2011.05.001.

Bisley, James W. (2022). "Parietal Lobe". In: *Encyclopedia of Animal Cognition and Behavior*. Ed. by Jennifer Vonk and Todd K. Shackelford. Cham: Springer. DOI: 10.1007/978-3-319-55065-7_1252. URL: https://doi.org/10.1007/978-3-319-55065-7_1252.

Burgoyne, John Ashley, Ichiro Fujinaga, and J. Stephen Downie (2015). "Music Information Retrieval". In: *A New Companion to Digital Humanities*. John Wiley Sons, Ltd. Chap. 15, pp. 213–228. ISBN: 9781118680605. DOI: https://doi.org/10.1002/9781118680605.ch15. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118680605.ch15. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118680605.ch15.

Burmania, Alec and Carlos Busso (2017). "A stepwise analysis of aggregated crowdsourced labels describing multimodal emotional behaviors". In: *A stepwise analysis of aggregated crowdsourced labels describing multimodal emotional behaviors*. Vol. 2017-August. International Speech Communication Association, pp. 152–156. DOI: 10.21437/Interspeech.2017-1278.

Cai, Yujian, Xingguang Li, and Jinsong Li (Mar. 2023). *Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review*. DOI: 10.3390/s23052455.

Chin, Tan Chyuan and Nikki S. Rickard (Apr. 2012). "The music USE (MUSE) questionnaire: An instrument to measure engagement in music". In: *Music Perception* 29 (4), pp. 429–446. ISSN: 07307829. DOI: 10.1525/mp.2012.29.4.429.

Cillessen, Antonius and P.E.L. Marks (Jan. 2011). "Conceptualizing and measuring popularity". In: *Popularity in the peer system*, pp. 25–56.

Cleveland Clinic (2023). *Occipital Lobe.* https://my.clevelandclinic.org/health/body/24498-occipital-lobe. Accessed: 2025-04-17. URL: https://my.clevelandclinic.org/health/body/24498-occipital-lobe.

Cohen, A. (2014). *Wavelet Methods for Time Series Analysis.* Cambridge University Press.

Corrêa, Débora C. and Francisco Ap. Rodrigues (2016). "A survey on symbolic data-based music genre classification". In: *Expert Systems with Applications* 60, pp. 190–210. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2016.04.008. URL: https://www.sciencedirect.com/science/article/pii/S095741741630166X.

Delorme, Arnaud and Scott Makeig (2004). *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.* URL: http://www.sccn.ucsd.edu/eeglab/.

Dowding, Irene et al. (Aug. 2015). "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP". In: vol. 2015. DOI: 10.1109/EMBC.2015.7319296.

Dzedzickis, Andrius, Artūras Kaklauskas, and Vytautas Bucinskas (Feb. 2020). *Human emotion recognition: Review of sensors and methods.* DOI: 10.3390/s20030592.

Fahad, Md Shah et al. (Mar. 2021). *A survey of speech emotion recognition in natural environment.* DOI: 10.1016/j.dsp.2020.102951.

GLOBAL MUSIC REPORT (2024).

Gramfort, Alexandre et al. (2013). "MEG and EEG data analysis with MNE-Python". In: *Frontiers in Neuroscience* (7 DEC). ISSN: 1662453X. DOI: 10.3389/fnins.2013.00267.

Gunawardana, C.L. and K.D. Thamarasee (2024). "Popularity Prediction of Sinhala YouTube Videos". In: *2024 9th International Conference on Information Technology Research (ICITR)*, pp. 1–5. DOI: 10.1109/ICITR64794.2024.10857802.

Hasinoff, Samuel W and Kiriakos N Kutulakos (2011). "Light-efficient photography". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.11, pp. 2203–2214.

Huq, Arefin, Juan Pablo Bello, and Robert Rowe and (2010). "Automated Music Emotion Recognition: A Systematic Evaluation". In: *Journal of New Music Research* 39.3, pp. 227–244. DOI: 10.1080/09298215.2010.513733. eprint: https://doi.org/10.1080/09298215.2010.513733. URL: https://doi.org/10.1080/09298215.2010.513733.

Javed, K., V. Reddy, and F. Lui (2025). "Neuroanatomy, Cerebral Cortex". In: *StatPearls [Internet]*. Updated 2023 Jul 25. Treasure Island (FL): StatPearls Publishing. URL: https://www.ncbi.nlm.nih.gov/books/NBK537247/.

Jeunet, Camille et al. (Mar. 2018). "Using Recent BCI Literature to Deepen our Understanding of Clinical Neurofeedback: A Short Review". In: *Neuroscience* 378. DOI: 10.1016/j.neuroscience.2018.03.013.

Jung, Richard and Wiltrud Berger (Dec. 1979). "Hans Bergers Entdeckung des Elektrenkephalogramms und seine ersten Befunde 1924–1931". In: *Archiv für Psychiatrie und Nervenkrankheiten* 227.4, pp. 279–300. ISSN: 1433-8491. DOI: 10.1007/BF00344814. URL: https://doi.org/10.1007/BF00344814.

Juslin, Patrik N. and Daniel Västfjäll (2008). "Emotional responses to music: The need to consider underlying mechanisms". In: *Behavioral and Brain Sciences* 31 (5). ISSN: 0140525X. DOI: 10.1017/S0140525X08005293.

Kayser, J. and C. Tenke (2006). "Spherical spline interpolation of EEG data for artifact correction". In: *Clinical Neurophysiology* 117, pp. 2329–2342.

Khan, Imran, Saeed Khalid, and Nadeem Javaid (2021). "Hybrid EEG noise removal technique using adaptive LMS and ICA". In: *Computers in Biology and Medicine* 131, p. 104263.

Khatib, Iyad Al et al. (2007). "Hardware/software architecture for real-time ECG monitoring and analysis leveraging MPSoC technology". In: *Hardware/software architecture for real-time ECG monitoring and analysis leveraging MPSoC technology.* Vol. 4050 LNCS, pp. 239–258. ISBN: 3540715274. DOI: 10.1007/978-3-540-71528-3_16.

Koelstra, Sander et al. (2012). "A review of affective computing: from unimodal analysis to multimodal fusion". In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2.3, 14:1–14:30. DOI: 10.1145/2362394.2362405.

Lee, Gregory R. et al. (2019). "PyWavelets: A Python package for wavelet analysis". In: *Journal of Open Source Software* 4.36, p. 1237. DOI: 10.21105/joss.01237. URL: https://doi.org/10.21105/joss.01237.

Lee, Junghyuk and Jong Seok Lee (Nov. 2018). "Music popularity: Metrics, characteristics, and audio-based prediction". In: *IEEE Transactions on Multimedia* 20 (11), pp. 3173–3182. ISSN: 15209210. DOI: 10.1109/TMM.2018.2820903.

Leeuwis, Nikki et al. (July 2021). "A Sound Prediction: EEG-Based Neural Synchrony Predicts Online Music Streams". In: *Frontiers in Psychology* 12. ISSN: 16641078. DOI: 10.3389/fpsyg.2021.672980.

Li, Jian, Wei Zhang, and Lei Wang (2020). "A hybrid method for EEG signal denoising using adaptive filtering and wavelet thresholding". In: *Biomedical Signal Processing and Control* 57, p. 101819.

Louis, Erik K. St. et al. (2016). *Electroencephalography (EEG) : an introductory text and atlas of normal and abnormal findings in adults, children, and infants.* Placeholder Publisher, p. 95. ISBN: 9780997975604.

Makeig, S. et al. (2004). "Independent Component Analysis of EEG and MEG". In: *NeuroImage* 12.6, pp. 1615–1625.

Middlebrook, Kai and Kian Sheik (Aug. 2019). "Song Hit Prediction: Predicting Billboard Hits Using Spotify Data". In: URL: http://arxiv.org/abs/1908.08609.

Mognon, A., J. Grapperon, and J. Pernier (2011). "Threshold-based rejection of bad EEG channels". In: *Journal of Neuroscience Methods* 199, pp. 130–140.

Murugappan, Murugappan, Nagarajan Ramachandran, and Yaacob Sazali (2010). "Classification of human emotion from EEG using discrete wavelet transform". In: *Journal of Biomedical Science and Engineering* 3.4, pp. 390–396. DOI: 10.4236/jbise.2010.34054. URL: http://www.SciRP.org/journal/jbise/.

Napier, Kathleen and Lior Shamir (Dec. 2018). "Quantitative sentiment analysis of lyrics in popular music". In: *Journal of Popular Music Studies* 30 (4), pp. 161–176. ISSN: 15331598. DOI: 10.1525/jpms.2018.300411.

Ni, Yizhao et al. (2013). "Hit Song Science Once Again a Science?" In: Intelligent Systems Lab, University of Bristol, UK; Signal Theory and Communications Department, Universidad Carlos III de Madrid, Spain.

Nolan, H., R. Whelan, and R. Reilly (2010). "Automatic detection of bad EEG channels using correlation". In: *Computers in Biology and Medicine* 40.4, pp. 270–276.

O'Reilly, C., F. Nielsen, and L. Hansen (2007). "Detection of bad EEG channels using kurtosis". In: *Journal of Neuroscience Methods* 159, pp. 40–47.

Pachet, François (2012). *Hit Song Science.*

Pearson, Karl (1901). "LIII. On lines and planes of closest fit to systems of points in space". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11, pp. 559–572.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pernet, C., R. Wilcox, and G. Rousselet (2011). "Robust averaging for EEG signal denoising". In: *NeuroImage* 60, pp. 2098–2109.

Perrin, F., J. Pernier, and O. Bertrand (2011). "Spherical spline interpolation of EEG data". In: *NeuroImage* 22, pp. 535–543.

Picard, Rosalind W. (2000). *Affective computing*. Placeholder Publisher, p. 292. ISBN: 9780262161701.

Pirau, L. and F. Lui (2025). "Frontal Lobe Syndrome". In: *StatPearls [Internet]*. Updated 2023 Jul 17. Treasure Island (FL): StatPearls Publishing. URL: https://www.ncbi.nlm.nih.gov/pubmed/30422576.

Rajagopalan, Neha and Blair Kaneshiro (2023). *CORRELATION OF EEG RESPONSES REFLECTS STRUCTURAL SIMILARITY OF CHORUSES IN POPULAR MUSIC*. URL: https://github.com/dmochow/rca.

Saeed, Somia, M. Khalid Mahmood, and Yaser Daanial Khan (May 2018). *An exposition of facial expression recognition techniques*. DOI: 10.1007/s00521-016-2522-2.

Samson, Sylvie (1999). "Musical Function and Temporal Lobe Structures: A Review of Brain Lesion Studies". In: *Journal of New Music Research* 28.3, pp. 217–228. DOI: 10.1076/jnmr.28.3.217.3107. URL: https://doi.org/10.1076/jnmr.28.3.217.3107.

Seufitelli, Danilo B. et al. (2023). "Hit song science: a comprehensive survey and research directions". In: *Journal of New Music Research*. ISSN: 17445027. DOI: 10.1080/09298215.2023.2282999.

Soares Araujo, Carlos Vicente, Marco Antônio Pinheiro de Cristo, and Rafael Giusti (2019). "Predicting Music Popularity Using Music Charts". In: *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pp. 859–864. DOI: 10.1109/ICMLA.2019.00149.

Soufineyestani, Mahsa, Dale Dowling, and Arshia Khan (Nov. 2020). *Electroencephalography (EEG) technology applications and available devices*. DOI: 10.3390/app10217453.

Übeyli, Elif Derya (Aug. 2009). “Statistics over features: EEG signals analysis”. In: *Computers in Biology and Medicine* 39 (8), pp. 733–741. ISSN: 00104825. DOI: 10.1016/j.compbiomed.2009.06.001.

University of Central Florida (2021). *Your Brain on Music*. Accessed: 2025-04-17. URL: https://www.ucf.edu/pegasus/your-brain-on-music/.

Viola, F., P. Orozco, W. Pinaya, et al. (2009). “Variance-based EEG channel detection and removal”. In: *NeuroImage* 45.3, pp. 782–789.

Widrow, Bernard and Samuel D Stearns (1985). *Adaptive Signal Processing*. Prentice-Hall.

Yasin, Sana et al. (2021). *EEG based Major Depressive disorder and Bipolar disorder detection using Neural Networks:A review*. URL: https://www.who.int/news-room/fact-sheets/detail/.

Yee, Yap Kah and Mafas Raheem (Sept. 2022). “Predicting Music Popularity Using Spotify and YouTube Features”. In: *Indian Journal Of Science And Technology* 15 (36), pp. 1786–1799. ISSN: 09746846. DOI: 10.17485/IJST/v15i36.2332. URL: https://indjst.org/articles/predicting-music-popularity-using-spotify-and-youtube-features.

Yuvaraj, Rajamanickam et al. (2014). “Detection of emotions in Parkinson’s disease using higher order spectral features from brain’s electrical activity”. In: *Biomedical Signal Processing and Control* 14, pp. 108–116.

Zangerle, Eva et al. (2019). “HIT SONG PREDICTION: LEVERAGING LOW-AND HIGH-LEVEL AUDIO FEATURES”. In: DOI: 10.5281/zenodo.3258042. URL: https://doi.org/10.5281/zenodo.3258042.

Zhang, Hui, Ming Liu, and Yu Chen (2021). “PCA-based artifact removal from EEG signals for improved BCI performance”. In: *Journal of Neuroscience Methods* 347, p. 108961.

Zhang, Jianhua et al. (2020). “Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review”. In: *Information Fusion* 59, pp. 103–126. ISSN: 1566-2535. DOI: https://doi.org/10.1016/j.inffus.2020.01.011. URL: https://www.sciencedirect.com/science/article/pii/S1566253519302532.

Zhao, Mingmin, Fadel Adib, and Dina Katabi (2016). “Emotion recognition using wireless signals”. In: *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. MobiCom ’16. New York City, New York: Association for

Computing Machinery, pp. 95–108. ISBN: 9781450342261. DOI: 10.1145/2973750.2973762. URL: https://doi.org/10.1145/2973750.2973762.

Zhu, Y. et al. (2017). “Impact of reference electrode on EEG signal quality and data analysis”. In: *NeuroImage* 148, pp. 47–55.