2024-04-27

Music Genre Classification and Sentiment Analysis of Bengali Music based on various inherent audio features

Humayra, Atika

IUB, CSE

Music Genre Classification and Sentiment Analysis of Bengali Music based on various inherent audio features

By

Atika Humayra

Student ID: **2030021**

Md Maruf Kamran Sohag

Student ID: **2130278**

Spring, 2024

Supervisor:

Dr. Mohammed Anwer

Professor

Department of Physical Sciences

Co-Supervisor:

Dr. Mahady Hasan

Associate Professor

Department of Computer Science & Engineering

Independent University, Bangladesh

April 27, 2024

Dissertation submitted in partial fulfillment for the degree of Bachelor of Science in Computer Science

Department of Computer Science & Engineering

Independent University, Bangladesh

Attestation

We are conscious of the nature of plagiarism and the university's stringent anti-plagiarism policy. We confirm that this is our original work. Moreover, any software or textual work by others that is utilized in this project is acknowledged appropriately and in accordance with generally accepted academic standards.

Signature

Date

Atika Humayra

Name

Signature

Date

Md Maruf Kamran SOhag

Name

Acknowledgement

We commence by offering our heartfelt gratitude to the Almighty Allah for endowing us with the fortitude, perseverance, and capability to diligently embark on and successfully complete our senior project. Our profound appreciation extends to Dr. Mohammed Anwer, our esteemed Supervisor and Professor, and Dr. Mahady Hasan, our co-supervisor and Associate Professor at Independent University, Bangladesh. Their invaluable guidance, boundless patience, and devoted investment of time have been instrumental in steering us through the intricacies of our project journey and shaping the formulation of this comprehensive report. Under their expert tutelage, our project progressed seamlessly, and their insightful counsel proved indispensable at every juncture.

Moreover, we express our sincere gratitude to the committee members for their invaluable contributions during the defense session, which rendered the experience not only fruitful but also enriching. Their discerning feedback and invaluable recommendations served to augment the depth and quality of our project significantly.

In addition to our academic mentors, we are deeply indebted to our family and friends for their unwavering support and encouragement throughout our tenure in the B.Sc. program, particularly during the more challenging phases of our senior project.

Furthermore, our heartfelt appreciation extends to Independent University Bangladesh for fostering an enriching senior project program that has played an integral role in honing our skills and preparing us for the rigors of the research industry. Their robust support has been instrumental in shaping our academic and professional journey, and for that, we are immensely grateful.

Letter of Transmittal

April 28, 2024
Dr. Mohammed Anwer
Senior Project Supervisor & Professor
Department of Physical Sciences,
Independent University, Bangladesh.

Dr. Mahady Hasan
Senior Project Co-Supervisor & Associate Professor
Department of Computer Science and Engineering,
Independent University, Bangladesh.

Subject: Senior project report on 'Music Genre Classification and Sentiment Analysis of Bengali Music based on various inherent audio features'.

Dear Sir,
We are truly grateful for the chance to submit a senior project report to you on "'Music Genre Classification and Sentiment Analysis of Bengali Music based on various inherent audio features'.. This report is a comprehensive reflection of the knowledge and effort we put into the project under your supervision. During the project, we applied a diverse set of skills and emerging technologies, which have enriched our understanding and proficiency in this field. We believe that this report effectively fulfills its intended purpose. We are extremely grateful for your invaluable guidance, mentorship, knowledge, and support. We hope that this report meets your expectations and is deemed satisfactory.
Thank you for your continuous cooperation and invaluable assistance throughout the semester.

Yours sincerely,
Atika Humayra (ID: 2030021)
Md Mauf Kamran Sohag (ID: 2130278)

Evaluation Committee

Supervision Panel

....	....
Supervisor	Co-supervisor

Internal Examiners

....	....
External Examiner 1	External Examiner 2

External Examiners

....	....
External Examiner 1	External Examiner 2

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The Bengali language is among the most widely used languages globally. With a global speaker population of over 320 million, Bengali is the fifth most spoken language in the world [1]. Being a language with an outstanding literary heritage, Bengali possesses a long and illustrious history. Bengali features songs that cover a wide range of genres and express a wide range of sentiments, making it one of the most musically diverse languages. There is a wide variety of genres within Bengali music. It includes a varied array of religious and secular compositions that have been passed down through generations. Bengali music has also become famous all over the world in the last few years. More and more people are able to hear new music from all over the globe, including Bengali music, due to the proliferation of the Internet and social media. Thus, it's clear that Bengali music's many styles and heartfelt melodies have been enthralling many listeners for generations.

Bengali music encompasses a diverse range of genres and subgenres, including music from Bangladesh and West Bengal, India. The level of detail we want to achieve in our music classification will dictate the total number of categories and subcategories. These songs represent a wide range of sentiments. The relationship between sentiment or emotion and music is deep and diverse. It's a fascinating dynamic that has an impact on both our biological and psychological well-being. Sentiment analysis focuses on identifying specific feelings and emotions. These sentiments can be identified by analysing similar features and themes in songs and compositions, which allows us to classify them. In this article, we have considered some of the major sentiments that are usually presented in Bengali music. It is relevant to mention that, in this paper our principal objective is to analyze Bengali music sentiments in a unique way, which will be advantageous to build an online Bengali Music platform that can suggest the audiences a wide spectrum of musical varieties in terms of their preferences.

1.2 Problem Statement

It is a matter of sorrow to mention that, there are not so much notable work related to Bengali music in the field of artificial intelligence. Currently there are some works that helps to discover new music for other languages, but, there is a notable absence of appropriate tools for the analysis and categorization of Bengali music. In addition, diverse varieties of machine learning approaches and algorithms have been used until the present time, yet, it is necessary to train them properly to get effective result for automated music playlist generator system. Since, it is the arduous task of categorizing Bengali music, however, only a minor number of datasets have been enhanced for Bengali music genre classification music which is not well earned to fulfill the demand of providing playlist recommendation to the listeners of Bengali music. Consequently, the outcome is unsatisfactory.

On the other hand, there are some works that help to analyze sentiments of music for other languages, but, there is a notable absence of appropriate studies for the analysis and categorization of sentiments in Bengali music. Furthermore, in the present works, the categories of sentiments used in the existing works are very limited. Furthermore, many different machine learning algorithms and methodologies have been employed up till now, but in order to acquire good results for automatically recognising music sentiments, it is important to train them correctly. In order to identify the sentiments, previous research relied only on music lyrics. Even though it provides a decent solution, it may not always be sufficient enough to reach a fair level of accuracy since the lyrics do not always have literal meanings. This makes it very challenging to identify the meanings conveyed by musical compositions. Due to the inability of current classifiers to handle complicated sentiments, the result is not satisfactory enough. Each song has some inherent audio features such as mel spectogram, rmse, delta, tempo, mfcc, zero crossing, spectral centroid, spectral bandwidth, chroma frequency. After studying about them, we were able to understand that, songs representing each sentiment categories show similarities in these features. Also, if we use these inherent features to train the machine learning and deep learning classifiers, it would provide a decent perfomance in classifying the music sentiments in much more categories.

1.3 Objectives

In this study, we attempted to improve the performance of Music Genre Classification for Bengali Music by adopting several machine learning and Deep Learning Methods. Our primary goal is to use simple machine learning algorithms and ensemble techniques that are computationally efficient in order to create a simpler and more robust system along with deep learning which can be a resource-intensive technique. We also used dimensionality reduction to remove extraneous data from the datasets. Furthermore, we aim to attempt to create some artificial intelligence based models that are able to classify sentiments of Bengali music using the inherent audio features mentioned above by using several machine learning and Deep Learning methods. Since, this approach is unique and never tried before, there was a lack of appropriate dataset. That is why, in this study, we also used a dataset that we created ourselves to achieve the desired results and functionalities. Our primary objective is to use our dataset to classify Bengali song sentiments using the audio features and categorizing the songs into five distinct sentiment categories. In summary, our plan is to implement two seperate studies, one is to classify Bengali music genre and the other one to analyse sentiments in Bengali music. With a primary objective of classifying Bengali song sentiments and categorizing them into five distinct sentiment categories, our study sets out to unravel the intricate emotions encapsulated within Bengali music. Through this dual-focused investigation, we endeavor to not only contribute to the advancement of music genre classification but also to illuminate the nuanced sentiment landscape of Bengali music, opening doors to innovative applications and insights.

Chapter 2

Literature Review

2.1 Relationship with Undergraduate Studies

In our undergraduate studies, we covered several key courses that greatly contributed to our ability to create a machine learning model. These courses equipped us with essential skills and knowledge necessary for our project.

Firstly, in the Introduction to Programming Language course, we learned the basics of coding, such as declaring variables and using loops and functions. We also delved into more advanced topics like object-oriented programming and creating graphical user interfaces for real-world problems.

Next, the Algorithms course taught us how to write efficient code by understanding time and space complexity. We learned various sorting algorithms and techniques for handling large datasets effectively.

Probability and Statistics for Science and Engineering expanded our understanding of statistical concepts. We mastered calculating measures like mean, median, and mode, and learned about different types of distributions, which are crucial for analyzing data accurately.

Lastly, Numerical Methods introduced us to the Python programming language and its libraries like Pandas, Numpy, and Matplotlib. These tools are essential for data manipulation and visualization, which are central to our research. Overall, these courses provided us with a strong foundation in programming, algorithms, statistics, and numerical methods, all of which are essential for developing our machine learning model.

2.2 Related works

Research on categorizing music has made progress, but when it comes to Bengali music, there's still a lot of ground to cover. One big issue is the lack of good datasets and ways to process all the data efficiently. Similarly, understanding the emotions in Bengali music hasn't been explored much. Most studies so far have focused on looking at the words in song lyrics to figure out what emotions they express. But this approach has its limits because song lyrics can be tricky—they often have more than one meaning, and it's not always easy for computers to understand them.

Relying solely on song lyrics to understand emotions in Bengali music isn't enough. Words in songs can mean different things depending on the situation, culture, and who's listening. This makes it hard for computers to accurately figure out what emotions are being expressed. Also, the current systems for classifying emotions in Bengali music only look at a few basic feelings, which doesn't capture all the different emotions that songs can evoke.

To tackle these challenges and make progress in understanding Bengali music better, we need to try different approaches. One idea is to look at other aspects of music, like the way it sounds and the cultural background it comes from, to get a fuller picture of the emotions it conveys. Also, we need to find better ways to collect and process data related to Bengali music so that computers can handle it more efficiently.

A pioneering study in Bangla music genre classification tackles under-explored area with a deep learning model by focusing on six distinct genres, researchers leverage a dataset of 1742 songs [2]. From this rich resource, they extract audio fingerprints like zero crossing rate and mel-frequency cepstral coefficients, paving the way for genre identification. This innovative approach promises to unlock new possibilities in Bangla music organization and recommendation. The proposed model, built with Keras, outperforms traditional machine learning methods, achieving 74% accuracy. The study emphasizes the model's efficiency and potential for improved accuracy with larger datasets or additional audio features. The paper concludes with the model's promising results for Bangla music genre classification and suggests future enhancements [2].

Another approach to Bangla music genre classification emerges with the development of an integrated framework [3]. This framework leverages the power of ensemble boosting, specifically the CatBoost algorithm, to achieve efficient and accurate genre identification.

CatBoost's inherent strengths, including its remarkable speed and adeptness at mitigating overfitting, make it particularly well-suited for this task. It emphasizes the importance of audio feature extraction from both time and frequency domains, and the reduction of less relevant MFCC features to improve classification performance. The framework shows significant improvement over other classifiers, particularly when excluding the 'Palligeeti' genre due to its overlapping characteristics with other genres2. The proposed model aligns with Industry 4.0 advancements, offering potential benefits for the music industry and enhancing real-time user experience [3].

In a further research, researchers set out to explore the world of Bangla music genres through the lens of machine learning and deep learning [4]. Their investigation aimed to crack the code of genre classification, using these powerful tools to unlock the unique soundscapes of Bangla music. It highlights the diversity of Bangla music and the lack of significant research in this area. The authors propose several models, including SVM, Random Forest, and a deep neural network, to classify music into six categories based on features like zero crossing rate and MFCC. The deep neural network model achieved the highest accuracy of 77.68%. The study emphasizes the potential of these techniques to enhance online music streaming services and suggests that larger datasets could further improve accuracy [4].

Sarma et al., [5] developed a dataset for Bengali music genre classification called BMGCD, comprising 2944 Bengali music clips. Introducing a Gated Recurrent Unit based deep learning model, they achieved an accuracy of 80.6% in predicting music genres from audio signals.

The paper "BanglaMusicStylo" [6], the first stylometric dataset of Bangla music lyrics, aimed at facilitating various stylometric analyses such as authorship attribution, linguistic forensics, and genre classification. The dataset comprises 2824 song lyrics from 211 lyricists, stored in text format and categorized by genre, including classical, folk, and modern music. The authors highlight the significance of lyrics in influencing listeners' preferences and the potential of text analytics in extracting stylometric features. They detail the data collection process, which involved meta-searching techniques and keyword-based searches across multiple platforms. The paper also discusses the dataset's attributes, such as the organization of lyrics by author and genre, and the use of the Siyam Rupali Bangla font. Statistical analysis of the dataset is provided, emphasizing its adequacy for machine learning applications. Finally, the paper outlines potential uses of the dataset, including authorship attribution, linguistic forensics, genre classification, vandalism detection, and emotion classification, underscoring the dataset's contribution to advancing research in Bangla stylometric analysis. The authors conclude by emphasizing the dataset's role in exploring new research avenues in the field of Bangla computing and stylometry [6].

In another research based on "A Dataset of Bangla Words for Sentiment Analysis", the authors introduce "BanglaSenti," [7]a lexicon-based dataset designed for sentiment analysis of Bangla words. As part of the growing field of Bangla Natural Language Processing (BNLP), this dataset addresses the need for reliable resources. "BanglaSenti" comprises 61,582 Bangla words, each associated with sentiment scores and English translations. The dataset's structure mirrors that of the English SentiWordNet, making it compatible with existing codes. The paper outlines the dataset's creation process, including data collection, selection, formatting, and translation. By providing a publicly available resource for sentiment analysis in Bangla, "BanglaSenti" aims to advance research in this domain[7].

Employing Machine Learning (ML) and Neural Network (NN) classifiers, researchers present a novel emotion analysis framework tailored to Bengali songs.[8]. The study addresses the dearth of emotion recognition research in Bengali music and aims to automate sentiment analysis in this linguistically rich domain. To achieve this, the authors curate a custom dataset of Bengali song verses annotated with three emotion labels: love, sadness, and idealism. They apply various ML and NN models, including BERT, for multi-class classification. The best accuracy achieved is 65% for multi-class and 80% for binary classification. Additionally, the paper discusses a user interface integrated with the model for visual representation and outlines future research directions to enhance accuracy and applicability [8].

Examining the feelings conveyed in Bengali songs, the researchers introduce an advanced sentiment analysis system in a paper called "Textual Lyrics Based Emotion Analysis of Bengali Songs".[9]. It introduces a new annotated corpus of 1500 Bengali songs and employs basic language-independent features like Bag-of-Words, tf-idf, and Word2Vec for feature extraction. Various machine learning algorithms, including Naive Bayes, KNN, SVM, and CNN, are tested, with CNN yielding the highest accuracy. The study highlights the challenges of sentiment analysis in Bengali, a low-resourced language, and suggests future work on incorporating more complex emotions and expanding the dataset. The research contributes to the fields of music information retrieval and emotion recognition in Bengali songs[9].

A further prior study utilizes machine learning to conduct a sentiment analysis of Bangla song commentary.[10]. It explores the emotional narratives within Bangla music and its cultural significance, employing various classification algorithms on a dataset of over 2224 comments. The research aims to understand the impact of Bangla songs on mental health and culture. Key algorithms like Random Forest, Decision Tree, and Support Vector Machine were used, with Random Forest achieving the highest accuracy of 63.68%. The study also addresses challenges in data collection and the hesitancy of individuals to share opinions on Bangla songs. The paper concludes with the successful application of the model for predicting sentiments, despite the obstacles faced during data collection [10].

In a subsequent paper titled “Mood Analysis of Bengali Songs Using Deep Neural Networks”, researchers explore emotion classification for Bengali songs [11]. They propose an annotated dataset with positive and negative polarities, achieving over 81% accuracy in classifying song lyrics. The study addresses sentiment analysis challenges in a morphologically rich language like Bengali. Leveraging machine learning and deep neural networks, they shed light on the emotional content of Bengali music. The chapter emphasizes the significance of understanding emotions expressed through song lyrics and contributes to this fascinating domain [11].

Introducing an attention-based Conv1D-BiGRU model, an additional study tackles the challenges of low accuracy and overfitting encountered by prior models in classifying emotions within Bengali music [12]. It introduces a novel approach combining Conv1D, Bi-GRU, and Bahdanau attention mechanism, validated on a Bengali music dataset with an impressive 95% accuracy. The model effectively categorizes music into four emotions: Angry, Happy, Relax, and Sad, using MFCCs for feature extraction. The study compares the proposed model with baseline methods, demonstrating its superior performance in emotion classification. This advancement in music emotion classification using deep learning techniques marks a significant contribution to the field, particularly for Bengali music. Future work includes exploring transformer models and different wav preprocessing methods to further enhance emotion discrimination in music[12].

Another paper ”Music Genre Classification Using MIDI and Audio Features” explores employing MIDI files and derived audio features for genre classification, using normalized compression distance (NCD) to measure distances between MIDI pieces. While individual use of MIDI or audio features yields lower accuracy than previous studies, combining both enhances accuracy. Best accuracies achieved are 0.75 (MIDI), 0.86 (audio), and 0.93 (combination). The study examines the influence of sample rates and sizes on classification and suggests improving results with domain-based features rather than NCD. It concludes that despite current limitations in audio-to-MIDI conversion, advancements could bolster genre classification by merging audio and MIDI features, benefiting music information retrieval (MIR) [13].

Besides, a paper delves into comparing human and automated methods for musical genre classification, focusing on feature extraction and classification techniques [14]. It explores diverse features capturing timbral texture, rhythmic elements, and pitch characteristics using methods like STFT, MFCC, and BH. Additionally, it investigates an auditory model simulating human ear functions for feature extraction. Various classifiers such as GS, GMM, and KNN are employed for pattern recognition. Evaluation highlights the significance of dataset quality and establishes accuracy benchmarks by comparing automated results with human performance and random classification. Overall, the paper aims to deepen genre classification comprehension and enhance automated systems' accuracy by referencing human classification standards [14].

Correspondingly, an investigation investigates Music Genre Classification employing diverse machine learning algorithms to categorize songs into genres such as classical, jazz, metal, and pop[15]. It utilizes the GTZAN Genre Collection, comprising 400 songs for training and testing purposes. Mel Frequency Cepstral Coefficients (MFCC) are employed for waveform representation. Various machine learning techniques including k-NN, k-means, multi-class SVM, and neural networks are employed for classification. The findings reveal neural networks achieving the highest accuracy, while other methods exhibit differing success rates across different genres [15].

On the other hand, a study "Musical Genre Classification: Is it Worth Pursuing and How Can it be Improved?" by Cory McKay and Ichiro Fujinaga examines the challenges and importance of automatic genre classification in Music Information Retrieval (MIR)[16].They oppose abandoning genre classification for broader similarity research, stressing its cultural significance and user familiarity. The authors propose a multidisciplinary approach, integrating musicology and psychology insights to enhance accuracy. They suggest incorporating low-level, high-level, and cultural features, allowing multiple genre labels per recording, and utilizing a diverse array of genre labels. The paper underscores the need for meticulously annotated datasets and further psychological exploration into human genre classification [16].

Furthermore, "A Comparative Study on Content-Based Music Genre Classification" delves into content-based music genre classification, introducing DWCHs (Daubechies Wavelet Coefficient Histograms) as a novel feature extraction technique [17]. DWCHs capture both local and global attributes of music signals, leading to improved classification precision. The study evaluates multiple machine learning algorithms, including SVMs, GMM, LDA, and KNN, to assess their efficacy in genre classification. Experimental findings indicate that combining DWCHs with SVMs notably boosts the accuracy of music genre classification [17].

Moreover, "Temporal Feature Integration for Music Genre Classification" explores methods of incorporating temporal features into music genre classification, particularly focusing on a multivariate autoregressive feature model [18]. It stresses the necessity of capturing temporal dynamics and inter-feature dependencies for precise genre classification. The proposed model includes diagonal autoregressive (DAR) and multivariate autoregressive (MAR) features, compared against mean-variance and other integration techniques. Experimental results showcase the reproducibility and effectiveness of the proposed features across various datasets and classifiers, alongside comparisons with human performance [18].

Henceforth, the research "MUSIC GENRE CLASSIFICATION WITH TAXONOMY" examines automatic music genre classification and underscores the significance of hierarchical classification within taxonomies[19]. It emphasizes the benefits of employing taxonomies in genre classification, citing enhanced usability and greater tolerance for classification errors. Feature extraction methods, including Mel-Frequency Cepstral Coefficients (MFCC) and other timbral features, are discussed for extracting features from music signals. Experimental findings demonstrate that employing taxonomy enhances classification performance, alongside proposing a methodology for automatically generating taxonomies. These results underscore the utility of hierarchical classification frameworks in music genre classification, offering insights into improved usability and classification accuracy through taxonomy integration [19].

In the interim, an article "Multimodal Deep Learning for Music Genre Classification" outlines a method for music genre classification using a multimodal deep learning approach[20]. This approach integrates audio tracks, text reviews, and cover art images to create intermediate representations, which are then combined to improve classification accuracy. The study underscores the value of multimodal data, demonstrating that different modalities offer complementary information that enhances performance. It emphasizes the superiority of deep learning over traditional handcrafted features, particularly when dealing with extensive music collections. Additionally, the paper delves into both single-label and multi-label classification, stressing the significance of representation learning and the effectiveness of a multimodal feature space in achieving a more nuanced categorization of music genres [20].

In the same vein, the "Musical Genre Classification of Audio Signals" explores the automated categorization of audio signals into musical genres by employing three distinct feature sets: timbral texture, rhythmic content, and pitch content [21]. It proposes specific features for each set, including spectral centroid and Mel-frequency cepstral coefficients for timbre, beat histogram for rhythm, and pitch histograms for pitch content. Through the utilization of statistical pattern recognition classifiers, the study achieves a classification accuracy rate of 61% across ten musical genres, a performance on par with human capabilities. The conclusion points towards avenues for future research, advocating for further investigation into alternative feature sets and the development of real-time classification methods within the realm of music information retrieval [21].

In consideration of the publication, a study discusses a novel approach for automated music genre recognition by combining visual and acoustic features [22]. The authors present a system that extracts both types of features from music tracks and fuses them to improve classification accuracy. Visual features are derived from spectrogram images using texture descriptors and bag of features projections, while acoustic features are obtained from audio signals using an ensemble of heterogeneous classifiers. The study demonstrates that the fusion of visual and audio features leads to performance comparable or even superior to state-of-the-art methods, without the need for ad hoc parameter optimization. The paper also highlights the independence of visual and audio features, confirmed by Q-statistics, suggesting that they can be effectively combined in a heterogeneous system. The MATLAB code for the ensemble of classifiers and visual features extraction will be made publicly available for future research comparisons. The paper emphasizes the increased computational cost as a drawback but considers it manageable with the advancement of processing technologies, (rewrite this without grammatical errora)[22].

Chapter 3

Definitions and Problem Characteristics

From our study and literature review, we have found that, the number of notable research in Bengali Music Genre Classification and Bengali Music Sentiment Analysis is very low. Even though there are some works on Bengali music, all the works depend on textual lyrics to classify the music genres and sentiment classification.

The main issue is that, the lyrics may not always be available to classify based on it in production. Also, some lyrics do not have literal meaning, it may have some other meanings depending on the language and synonyms of the words which the machine is not able to understand. That is why, depending only on lyrics is not always a best solution.

On the other hand, to use a lyrics based software, the software must have a speech to text mechanism to convert the song into lyrics while working on production. This heavily depends on the accuracy on the speech to text algorithm and for Bengali language, the accuracy is not very high. However, if the inherent audio features are used to classify the music genres and sentiment analysis, the software can easily detect and extract necessary audio characteristics and use those to classify the music. That is why, our approach will work better on production.

Furthermore, for sentiment classification problems, the present works only focus on two to three sentiments which is not enough to properly analyse song sentiments. That is why, we aim to implement a model that can analyze five categories of sentiments for Bengali music.

We also aim to increase efficiency for both music genre classification problem and song sentiment analysis problem from present works.

In summary, we aim to implement a unique approach for both the music genre classification problem and music sentiment analysis problem for Bengali music. Furthermore, we aim to increase the efficiency from all the present works by studying inherent audio features as well as machine learning and deep learning classifiers.

3.1 Inherent Features of Music

Audio features of music are measurable characteristics or attributes of sound that help describe and analyze musical content. These features provide insights into various aspects of music, including rhythm, pitch, timbre, dynamics, and more.

Such audio features are Zero Crossing, Spectral-Centroid, Spectral-Rolloff, Spectral-Bandwidth, Chroma-Frequency, RMSE, Delta, Melspectogram, Tempo, MFCC etc. The details of these features can be found below:

3.1.1 Zero-Crossing Rate

The Zero-Crossing Frequency is the rate at which the sign of a signal changes or the pace at which the move from negative to positive or vice versa.

ZCR is a technique for detecting human voices in a music signal [23]. The ZCR approach is rapid and easy for determining whether a speech frame is voiced, unvoiced, or silent. The following equation is used to compute it.

$$ZCR = \frac{1}{H-1} \sum_{h=1}^{H-1} 1_{K<0} \left(C_h C_{h-1} \right) \tag{1}$$

Here, C is a signal of length H and 1K¡0 [23]. The librosa.zero crossings() function is used to find the zero crossing for each frame. The means of several classes are shown in Figure 3.1.

We may examine the mean ZCR values for several classes in our dataset. Here, Bangla Band Music has the highest value, whilst Rabindra Sangeet has the lowest.

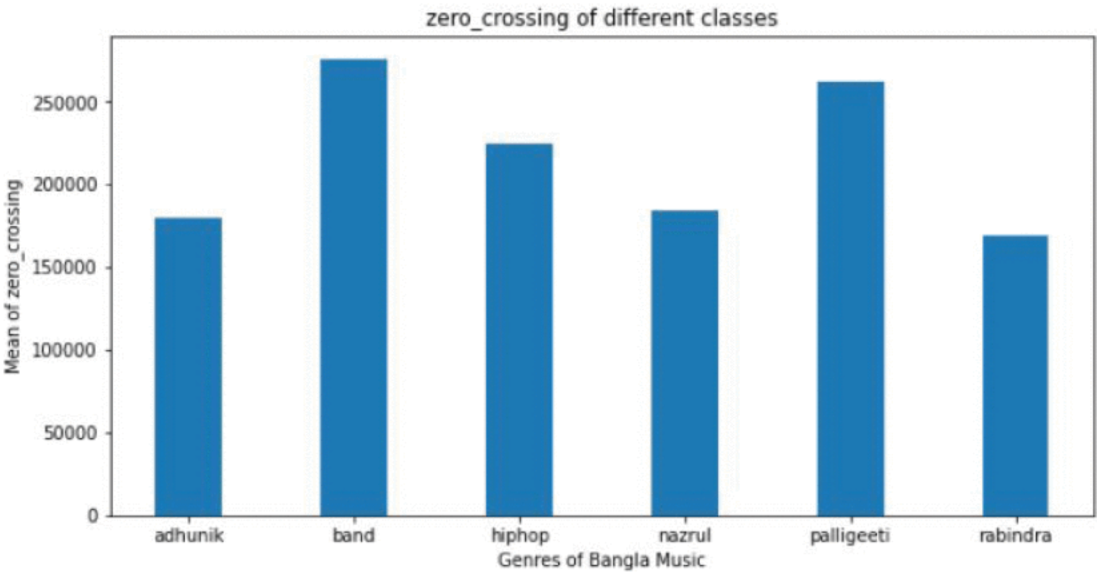

Figure 3.1: Zero-Crossing Rate of Different Classes.

3.1.2 Spectral Centroid

In digital signal processing, a spectrum is characterized by its spectral centroid. It reveals the position of the spectrum's gravitational core.The spectral centroid is a measure used in digital signal processing to characterise a spectrum. It indicates where the center of mass of the spectrum is located. Perceptually, it has a robust connection with the impression of brightness of a sound.

The centroid is determined using the following formula:

$$\text{Spectral Centroid } = \frac{\sum_{m=0}^{M-1} f(m)Y(n)}{\sum_{m=0}^{M-1} Y(m)} \tag{2}$$

where Y(n) is the weighted frequency value, or magnitude, of bin(the range of values) number m and f(m) is the bin's center frequency [24]. librosa.spectral centroid() is used to determine the average spectral centroid value for each value.

3.2 depicts the mean values of Spectral Centroids from various classes in the dataset. The maximum value is assigned to Bangla Band Music, while the minimum value is assigned to Nazrulgeeti.

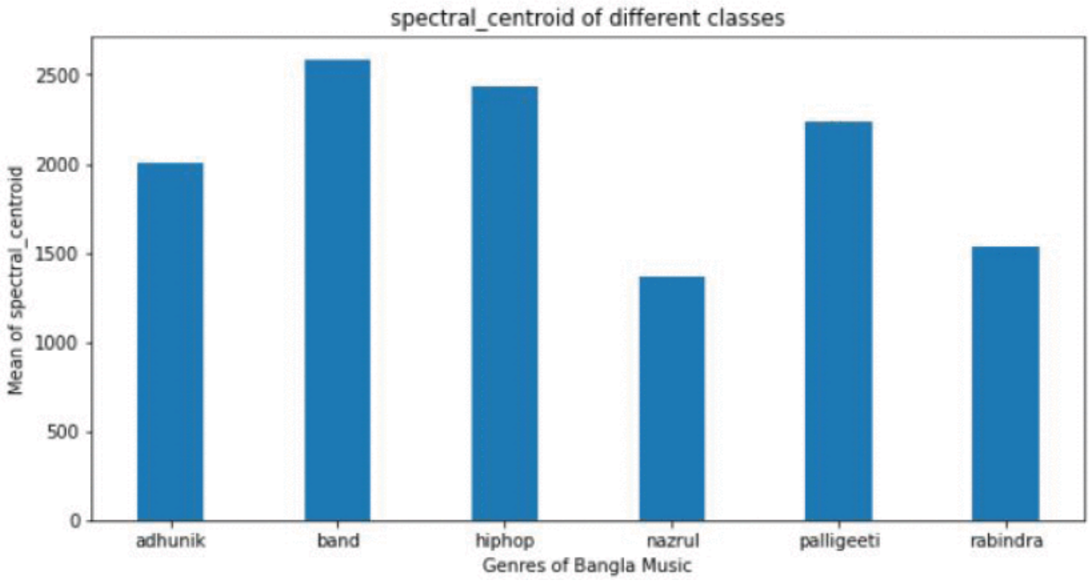

Figure 3.2: Spectral Centroid of Different Classes.

3.1.3 MFCC(MEL-Frequency Cepstral Coefficients)

Mel-frequency cepstral coefficients (MFCCs) are a set of features extracted from audio signals that capture the characteristics of the human auditory system. They are computed by first converting the audio signal into a frequency domain representation using a technique such as the discrete Fourier transform (DFT). Then, the power spectrum of the signal is mapped onto the mel scale, which approximates the human perception of sound frequencies. Finally, the logarithm of the mel-scaled power spectrum is taken, followed by a discrete cosine transform (DCT) to produce a set of coefficients. These coefficients, known as MFCCs, represent the spectral envelope of the audio signal and are commonly used in speech and audio processing tasks such as speech recognition and speaker identification.

A signal's mel frequency cepstral coefficients (MFCCs) are a limited group of characteristics (typically 10-20) that simply define the overall shape of a spectral envelope [25]. It is often used to describe timbre in MIR. The MFCCs are computed using a mel-frequency spectrogram and the discrete cosine transform (DCT). MFCC can normally extract up to 20 features, but 12-13 features are thought to be excellent for feature extraction, thus we pick 13. librosa.feature.mfcc is used to compute the mean and variance for these 13 features.

We picked a sample of mfcc features where from 3.3, we analyzed that Nazrulgeeti has the maximum value while Bangla Band Music has the lowest.

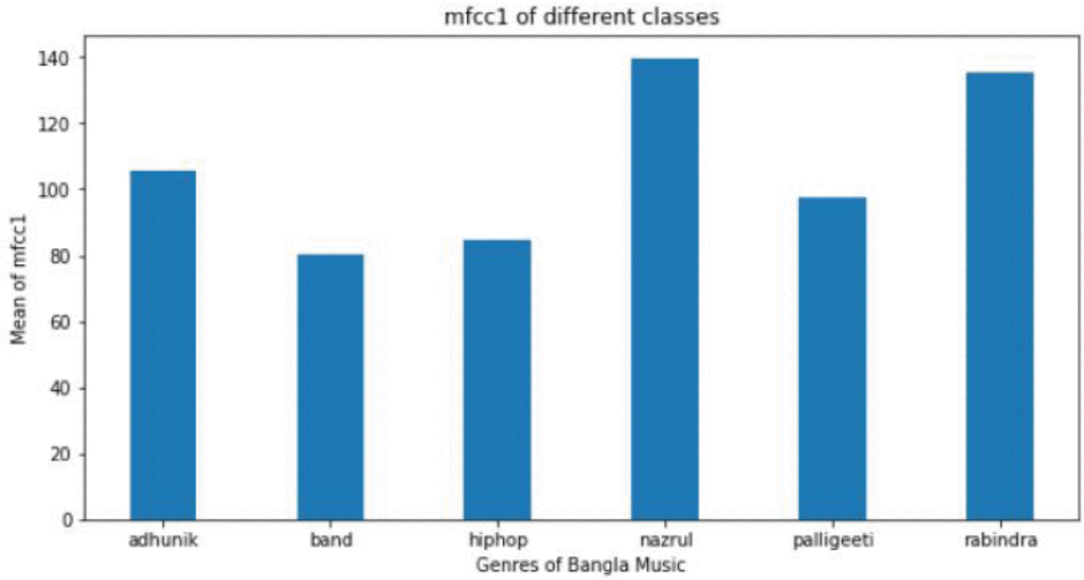

Figure 3.3: MFCC1 of Different Classes.

3.1.4 Spectral Roll-Off

Spectral rolloff is a feature used in audio signal processing to characterize the shape of the spectral distribution. Specifically, it refers to the frequency below which a certain percentage of the total spectral energy is concentrated. Spectral rolloff is calculated by examining the cumulative power spectrum and finding the frequency bin at which the accumulated energy reaches the specified threshold. It provides valuable information about the high-frequency content of the signal and is used in various audio processing applications such as music genre classification and sound event detection.
When the entire spectral energy falls below a certain frequency, this phenomenon is known as spectral rolloff [23]. The spectral rolloff is calculated using the following equation:

$$\text{Spectral roll-off } = \sum N_g[m] = 0.85 * \sum_{m=1}^{M} N_g[m] \tag{3}$$

The bandwidth under which 85% of the magnitude spectrum is focused is referred to as the spectral rolloff, where N g[m] is the Fourier transform magnitude at frame t and frequency bin.

When we look at 3.4 and look at the mean values for Spectral Roll-off, we can see that the values for the classes *Bangla HipHop* and *Bangla Band Music* are very close to each other. *Nazulgeeti* has the minimal value.

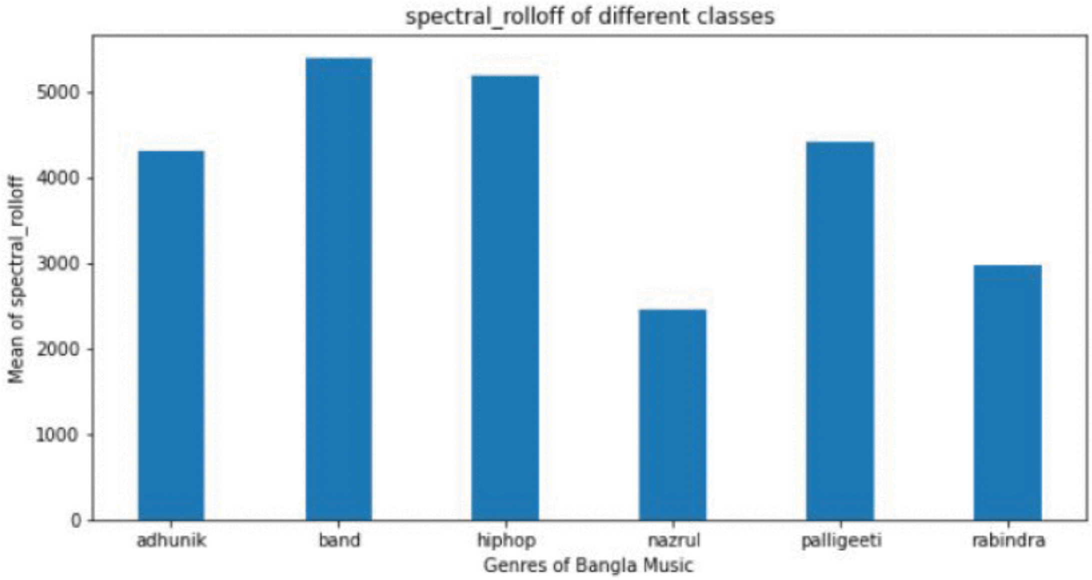

Figure 3.4: Spectral Roll-Off of Different Classes.

3.1.5 Chroma Frequency

Chroma frequency, in the context of audio signal processing, refers to a representation of musical pitch class. It characterizes the distribution of energy across the twelve pitch classes in the equal-tempered scale. Essentially, it captures the tonal content of an audio signal by disregarding the specific octave of each note and focusing solely on its pitch class. Chroma features are calculated by first transforming the audio signal into a spectrogram and then summing the energy in each frequency bin across all octaves that correspond to the same pitch class. This results in a vector representing the strength of each pitch class in the audio signal, which can be used for tasks such as chord recognition, music genre classification, and audio similarity analysis.

The way people hear pitch is periodic, meaning that two pitches that are different by one or more octaves are heard as having the same color, or harmonic role (where, in our scale, an octave is defined as the distance of 12 pitches). The main idea behind chroma features is to combine all spectral information about a given pitch class into a single coefficient [26]. One of the most important things about chroma features is that they capture the harmony and melody of music. This is why we are using this aspect to classify the genre.

This is how we figure out the mean value of a chromagram from a music signal of a given signal. We figure this out with the librosa.feature.chroma stft() function in the frequency domain. 3.5 illustrates the average Chroma Frequencies for each class. Here, Bangla HipHop is the highest and Nazrulgeeti is the lowest.

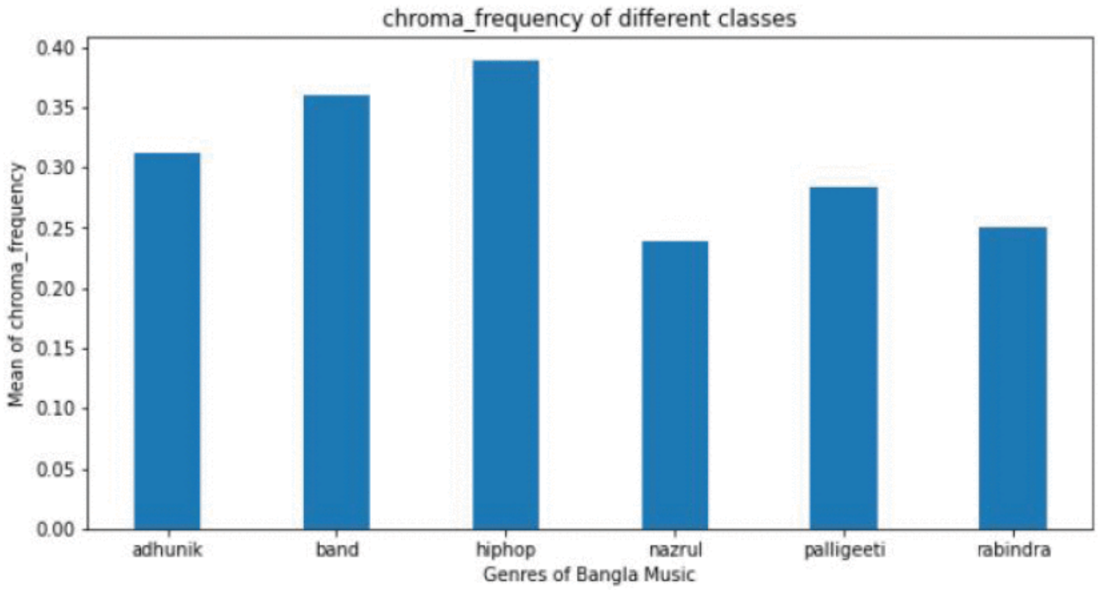

Figure 3.5: Chroma Frequency of Different Classes.

3.1.6 Spectral Bandwidth

Spectral bandwidth is a measure used in signal processing to describe the range of frequencies present in a signal. Specifically, it quantifies the spread or width of the frequency content of a signal around its central frequency. It can be calculated in various ways, such as the difference between the upper and lower frequencies where the signal's power falls to half (-3 dB) of its peak value, or as the standard deviation of the frequency distribution. Spectral bandwidth provides valuable information about the frequency content and complexity of a signal, and it is used in various applications including audio and image processing, communication systems, and spectroscopy.

Spectral Bandwidth identifies the frequency at which the energy of a spectrum is concentrated. A signal's resolution is determined by its bandwidth [27]. Radiated spectral quantities are not less than half their maximum value in this wavelength range.

$$\text{Spectral Bandwidth } = \left(\Sigma_p V(p)\left(z(p)-z_d\right)^q\right)^{\frac{1}{q}} \tag{4}$$

where z(k) is the frequency in bin p, z(p) is the spectral magnitude, and z(d) is the spectral magnitude at z(p). For each file, we compute the mean and standard deviation of spectral Bandwidth using librosa.feature.spectral bandwidth. The order-q spectral bandwidth is calculated using librosa.feature.spectral bandwidth.

Spectral Bandwidth values for several classes are shown in a bar graph in 3.6. Bangla

HipHop and Bangla Band Music have almost identical peak values here, whereas Nazrulgeeti has the lowest mean value.

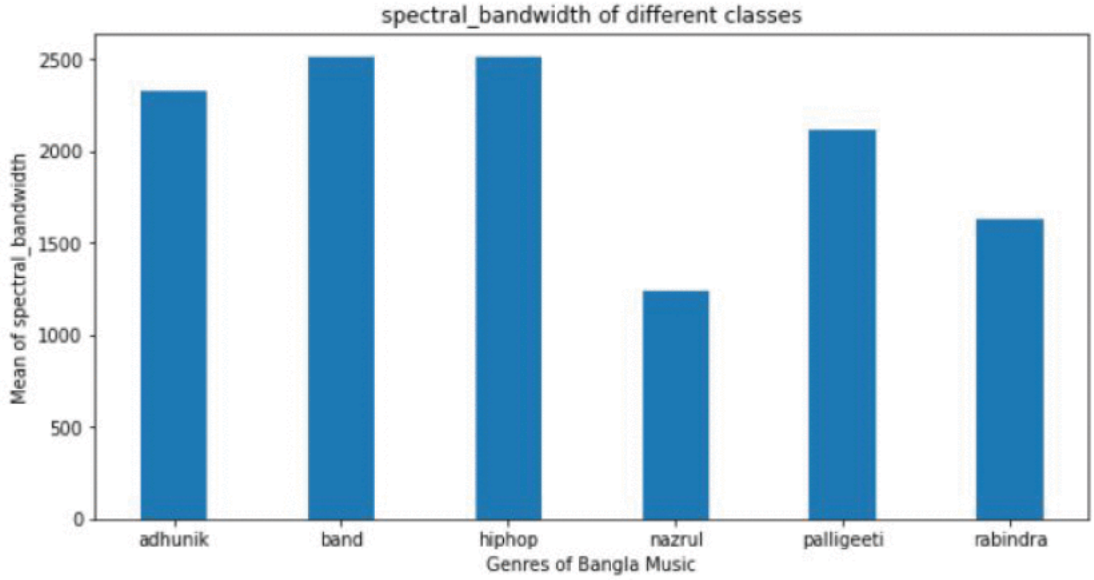

Figure 3.6: Spectral Bandwidth of Different Classes.

3.1.7 Spectral Flux

Spectral flux measures the change in the spectrum between two frames [28]. It is calculated as the squared difference between the two normalized magnitudes of the two frames spectral distributions.

$$Fl(i, i-1) = \sum_{k=1}^{Wf_L} (EN_i(k) - EN_{i-1}(k))^2 \tag{5}$$

where EN i(k) is the kth normalized DFT coeffcient at the ith frame. From the librosa Python package, librosa.onset. onset strength was used to figure out the spectral flux in our research.

3.1.8 Pitch

Most of the time, the pitch of an audio signal is used to describe the vibration of a frequency. Many researcher can say that a high frequency audio wave has a high pitch, and a low frequency audio wave has a low pitch [29]. librosa.piptrack figures out how high or low the sound files are.

3.1.9 Tempo

In music, tempo means how fast or slow a piece of music is played [30]. For example, a tempo of 60 BPM means that there is one beat every second. We used librosa.beat.tempo to find the numerical data for each file based on this feature.

3.7 shows how the averages of Tempos from different classes look. We can see that the values for Tempo are almost the same for each class.

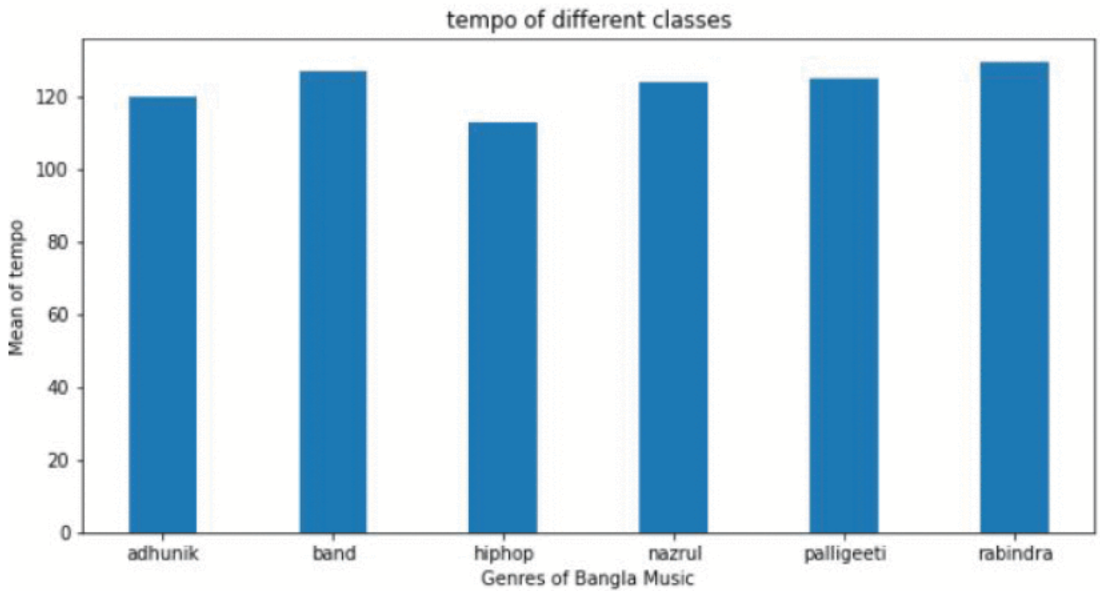

Figure 3.7: Tempo of Different Classes.

3.2 Machine Learning and Deep Learning Classifiers

Since both the objectives such as, genre classification and sentiment analysis are classification problems, we can use various machine learing and deep learning classifiers. Machine Learning and Deep Learning Classifiers are used for classification to automatically categorize data into predefined classes, simplifying operations and extracting valuable insights. On the other hand, Deep Learning algorithms, a subset of Machine Learning, utilize artificial neural networks with multiple hidden layers to process unstructured data like images and text. These classifiers enable automation of tasks that were previously manual, saving time and money while enhancing operational efficiency. Some notable classifiers are detailed below:

3.2.1 Machine Learning Classifiers

Support Vector Machine

Support Vector Machines (SVMs) are powerful tools in machine learning used for classifying data into different categories. They work by finding the best possible line or boundary that separates different groups of data points with the largest possible gap between them. This boundary is known as the hyperplane. SVMs are particularly good at handling situations where the relationship between features and classes is complex. They operate by first transforming the input data into a higher-dimensional space where it's easier to separate the classes. Then, new data points are classified based on which side of the gap they fall on. SVMs can handle non-linear relationships between features and classes through a technique called the kernel trick, which implicitly maps data into higher-dimensional spaces. The effectiveness of SVMs depends on choosing the right parameters, such as the type of kernel used and its associated hyperparameters, as well as a hyperparameter called C, known as regularization parameter that controls the trade-off between maximizing the margin and minimizing classification errors. These parameters are often fine-tuned using methods like grid search or Bayesian optimization. SVMs are unique in their ability to find a balance between maximizing the margin (the space between classes) and minimizing classification errors. This makes them robust classifiers suitable for a wide range of applications [31].

Random Forest

Random Forest is a powerful ensemble learning method widely used in machine learning for classification and regression tasks. It operates by constructing a multitude of decision trees during training and outputs the mode or mean prediction of the individual trees for classification or regression, respectively. Random Forest mitigates overfitting and improves generalization by introducing randomness in the tree-building process. This randomness is achieved by using a random subset of features at each node split and sampling the training data with replacement (bootstrapping). By combining the predictions of multiple trees, Random Forest reduces variance and improves prediction accuracy. According to Breiman's seminal work [32] on Random Forests, the method's strength lies in its ability to handle high-dimensional data with a large number of features while maintaining robustness against noise and outliers. Furthermore, the decision-making process of Random Forest can provide insights into feature importance, aiding in feature selection and interpretation of the underlying data structure. Random Forest's effectiveness has been demonstrated across various domains, including bioinformatics, finance, and remote sensing, making it a popular choice for predictive modeling tasks. Overall, Random Forest stands out as a versatile and reliable machine learning algorithm, offering high performance and interpretability in diverse applications [32].

Decision Tree

Decision trees are fundamental models in machine learning used for both classification and regression tasks. They work by recursively partitioning the input space into regions, with each partition corresponding to a simple decision rule based on one feature. These decision rules are organized in a hierarchical tree structure, where each internal node represents a decision based on a specific feature, and each leaf node represents the predicted outcome. According to a study by Quinlan (1986) [33], decision trees are built using a top-down, greedy approach, where at each step, the algorithm chooses the feature that best splits the data, aiming to maximize the homogeneity of the resulting partitions in terms of the target variable. Decision trees are interpretable models that offer insights into the decision-making process and feature importance, making them valuable for understanding complex data relationships. However, they are susceptible to overfitting, particularly when dealing with noisy data or complex decision boundaries [33].

Stochastic Gradient Descent

Stochastic Gradient Descent (SGD) is a widely used optimization algorithm in machine learning for training various models, including neural networks and linear classifiers. It operates by iteratively updating the model parameters to minimize a loss function based on the observed data. Unlike traditional gradient descent, which computes the gradient of the loss function using the entire dataset, SGD computes the gradient using only a randomly selected subset of the data, making it computationally efficient and suitable for large-scale datasets. According to Bottou et al. (2018) [34], this stochastic nature introduces noise into the optimization process but allows for faster convergence and better generalization. SGD updates the model parameters in small steps along the direction of the negative gradient, adjusting the parameters to decrease the loss function. Despite its simplicity, SGD can efficiently navigate complex and high-dimensional parameter spaces, making it a cornerstone of modern machine learning algorithms. However, the convergence of SGD may be affected by factors such as the learning rate schedule and the choice of mini-batch size, requiring careful tuning to achieve optimal performance [34].

K-Nearest Neighbours

K-Nearest Neighbors (KNN) is a straightforward yet effective supervised learning algorithm used for classification and regression tasks in machine learning. It operates by finding the majority class or average value among the K nearest data points to predict the class or value of a new data point. According to Han et al. (2011) [35], KNN is based on the assumption that similar data points tend to belong to the same class or have similar values. The algorithm calculates the distance between the new data point and all existing data points in the training set, typically using metrics like Euclidean distance or Manhattan distance. Then, it selects the K nearest neighbors and assigns the class or value based on their majority or average. KNN is known for its simplicity and intuitiveness, making it easy to understand and implement. However, its performance can be sensitive to the choice of K and the distance metric used, requiring careful parameter tuning. Despite its simplicity, KNN can perform competitively with more complex algorithms, particularly in scenarios with small to medium-sized datasets or when the decision boundary is non-linear. In summary, KNN offers a versatile and interpretable approach to classification and regression tasks, providing a simple yet effective baseline for comparison with other machine learning methods [35].

3.2.2 Deep Learning Classifiers

Multi Layer Perceptron (MLP) Classifier

A Multilayer Perceptron (MLP) is a type of artificial neural network commonly used for supervised learning tasks like classification and regression. It consists of multiple layers of interconnected nodes, including an input layer, one or more hidden layers, and an output layer. Each node in the hidden layers and output layer applies a weighted sum of inputs followed by a non-linear activation function, such as the sigmoid or ReLU function. MLPs are capable of learning complex non-linear relationships between input and output variables, allowing them to capture intricate patterns in the data. According to Haykin (1994) [36], the training of MLPs typically involves forward propagation of input data through the network to generate predictions, followed by backward propagation of errors using techniques like backpropagation to adjust the weights and minimize the difference between predicted and actual outputs. MLPs are known for their ability to approximate any continuous function to arbitrary accuracy, given a sufficiently large number of hidden units and appropriate training data. However, MLPs are prone to overfitting, especially when dealing with high-dimensional or noisy data, necessitating regularization techniques like dropout or weight decay. Despite their computational complexity and sensitivity to hyperparameters, MLPs have demonstrated superior performance in various domains, including image and speech recognition, making them a popular choice in modern machine learning applications [36].

Deep Learning Models

Custom deep learning models refer to neural network architectures tailored to specific tasks or datasets, often designed by researchers or practitioners to address unique challenges or achieve specific objectives in machine learning. These models are typically constructed by combining various layers, activation functions, and connectivity patterns to form a network structure suitable for the given problem domain. According to Goodfellow et al. (2016) [37], custom deep learning models can range from simple architectures with a few layers to complex structures with multiple branches, skip connections, or attention mechanisms, depending on the complexity of the task and the characteristics of the data. Designing custom models often involves experimentation and iterative refinement, guided by domain knowledge and empirical observations. Custom models can offer advantages such as improved performance, efficiency, or interpretability compared to off-the-shelf architectures. However, they may also require careful parameter tuning and validation to ensure robustness and generalization to unseen data. Despite the challenges, custom deep learning models have been successfully applied in various domains, including computer vision, natural language processing, and reinforcement learning, demonstrating their versatility and effectiveness in solving real-world problems [37].

Chapter 4

Dataset

4.1 Dataset for Music Genre Classification

This study relied on the use of the "Bangla Music Dataset" [38]. This dataset provides roughly 1742 songs total, divided into six categories. The groups consist of Adunik Music, Hip-Hop Music, Band Music, Nazrulgeeti (Songs written and composed by Kazi Nazrul Islam), Folk song which is known as 'Palligeeti', Rabindra Sangeet(Songs written and composed by Rabindranath Tagore). Every song has thirty inherent features altogether. RMS, zero crossing, spectral centroid, spectral rolloff, spectral bandwidth, chroma frequency, melspectogram, mfcc, etc. are typically a song's primary characteristics. These characteristics are specific to each song, but they read similarly across every category. Within this extensive dataset lie approximately 1742 songs, thoughtfully organized into six distinct musical genres. These encompass Adunik Music, Hip-Hop Music, Band Music, Nazrulgeeti (compositions by Kazi Nazrul Islam), Folk songs, fondly referred to as 'Palligeeti', and the timeless Rabindra Sangeet (songs by Rabindranath Tagore). Each composition is meticulously characterized by a comprehensive array of thirty distinct musical features. These encompass a myriad of essential elements such as RMS (Root Mean Square), zero crossing rate, spectral centroid, spectral rolloff, spectral bandwidth, chroma frequency, melspectrogram, and the intricate MFCCs (Mel-frequency cepstral coefficients). Despite the inherent diversity and uniqueness inherent in musical expression, these features exhibit remarkable consistency across all genres, thereby providing a rich foundation for detailed analysis, classification, and exploration within the realm of musicology. The amount of songs in each category is displayed in a Bar graph in Figure 4.1. Additionally, Figure 4.2 shows a pie chart of the data dissemination. Table 4.1 shows the total number of data and the data type. Table 4.2 shows the number of songs in each genre categories.

Table 4.1: Dataset Details

Data Type	Numeric
Total Data	1742

Table 4.2: Number of songs for each categories.

Category	**Number of Songs**
Bangla Adunik	283
Bangla Hip-Hop	295
Bangla Band Music	295
Nazrulgeeti	312
Palligeeti	260
Rabindra Sangeet	297
Total	1742

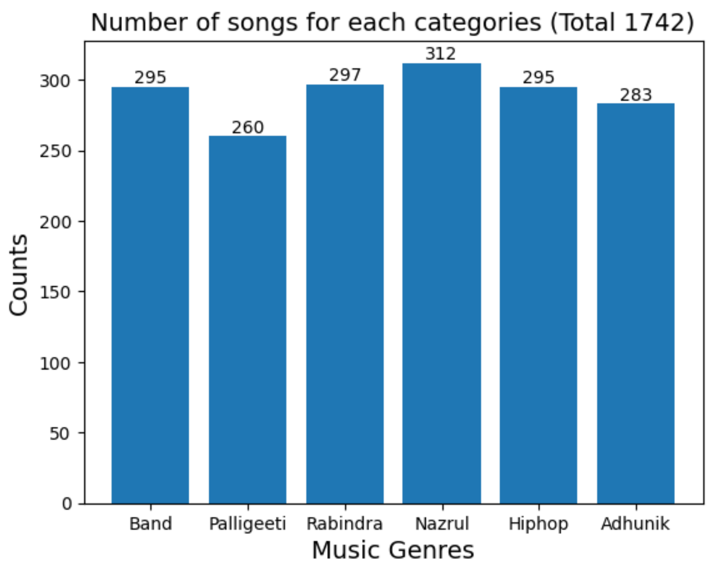

Figure 4.1: Number of songs in each Genres.

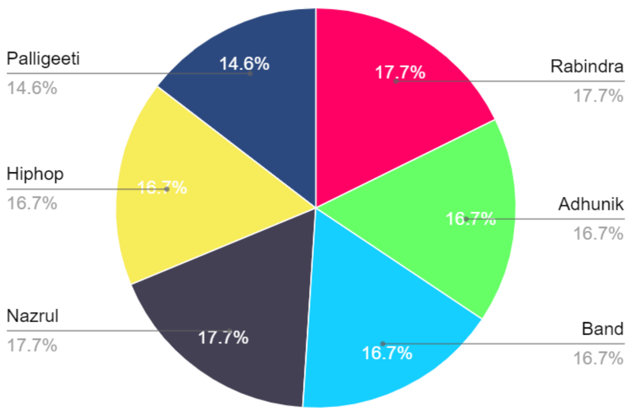

Figure 4.2: Data Distribution of Each Categories.

4.2 Dataset for Music Sentiment Analysis

Due to the absence of a sufficient dataset including relevant audio features for sentiment analysis of Bengali music, we had to depend on the same "Bangla Music Dataset" [38] as our primary source of data.

There is an approximate total of 1742 songs, which are categorised into six distinct groups. Each song possesses a total of thirty audio properties, including rmse, zero crossing, spectral centroid, spectral rolloff, spectral bandwidth, chroma frequency, melspectogram, mfcc, and others. These features usually represent the fundamental characteristics of a song.

Nevertheless, this dataset primarily focuses on categorising Bengali music into several genres and does not include any sentiment categories. Consequently, we found ourselves tasked with the responsibility of meticulously assigning a sentiment category to each and every song in our dataset. This involved carefully evaluating the emotional tone and lyrical content of each track, ensuring that our categorizations accurately reflected the sentiment conveyed. Additionally, we endeavored to maintain consistency and reliability throughout the process, acknowledging the nuanced nature of musical expression and its impact on sentiment interpretation. We meticulously categorised all the songs into five separate sentiment categories, namely: "Amusement", "Romanticism", "Feeling Pumped Up", "Relaxation", and "Sadness".
This categorization was done manually.

The number of songs in each category is depicted as a bar chart in Figure 4.3. Furthermore, Figure 4.4 shows a pie chart of the data distribution of each sentiment categories. Table 4.3 shows the number of songs in each sentiment categories.

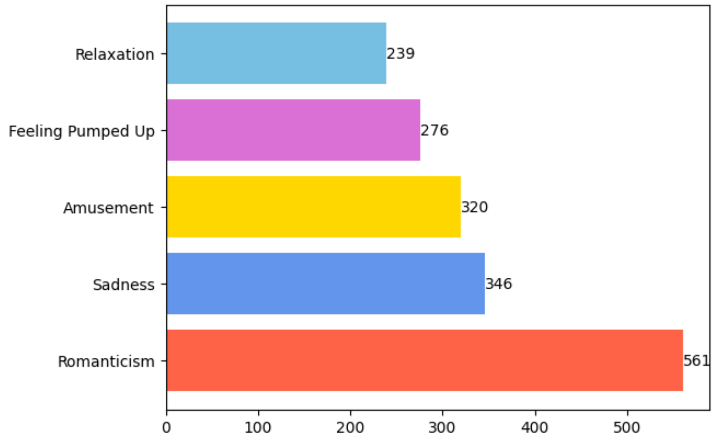

Figure 4.3: Quantity of music in each Sentiment Categories.

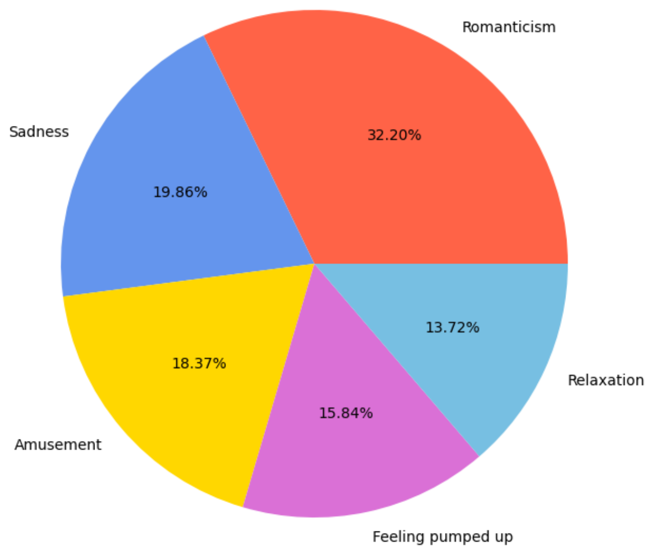

Figure 4.4: Allocation of Songs in Each Sentiment Categories.

Table 4.3: Number of songs for each categories.

Category	**Number of Songs**
Relaxation	239
Feeling Pumped Up	276
Amusement	320
Sadness	346
Romanticism	561
Total	1742

In order to categorise the songs based on sentiment, we thoroughly listened to all the songs in the dataset and classified them with the assistance of knowledgeable music artists. In order to determine the categories, we implemented the following measures:

Romanticism: Refers to the emotional expression of love, devotion, intense desire, commitment, and affection for a person with whom one shares a romantic connection. When a song conveys these sentiments, it is categorised as a romantic song.
Sadness: We categorise sadness as the portrayal of misery, dissatisfaction, sorrow, melancholy, discontent, and gloom in a song.
Amusement: Refers to the classification of a song's sentiment when its lyrics and tunes evoke feelings of merriment, delight, laughter, enjoyment, fun, and entertainment.
Feeling Pumped Up: A song is categorised as "Feeling Pumped Up" if it aims to evoke sensations of thrill, excitement, inspiration, intrigue, exhilaration, and enthusiasm.
Relaxation: Songs that aim to convey a sense of composure, calmness, tranquilly, peacefulness, and pleasure are categorised under this genre.

Following a meticulous procedure of categorising the songs into appropriate sentiment groups, we eventually attained our desired dataset.

Chapter 5

Bengali Music Genre Classification

Our proposed methodology to achieve our desired result in classifying the music genres is divided into two parts as shown in Figure 5.1: data pre-processing and implementation of algorithms. The results are then analysed to evaluate the performance of various classifiers.

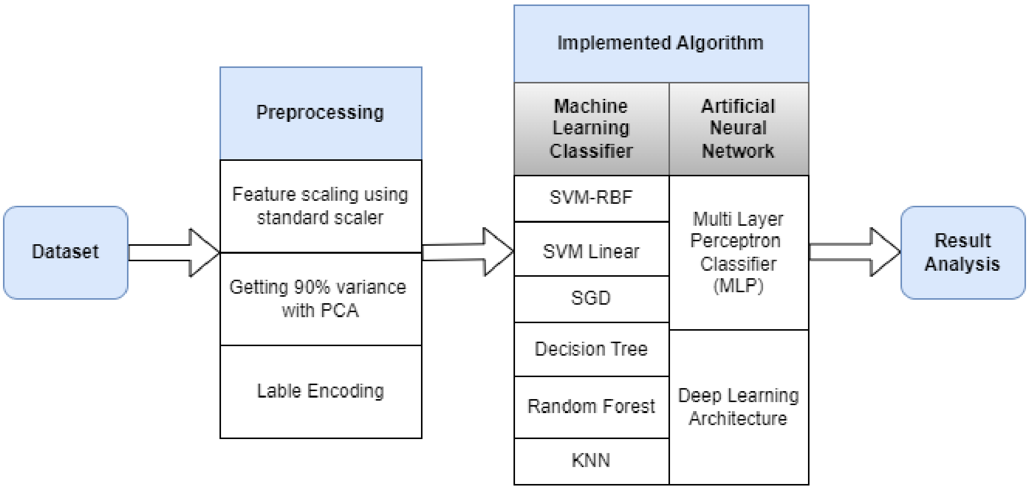

Figure 5.1: Flowchart of Proposed Methodology.

5.1 Technical

5.1.1 Data Preprocessing

Dimentionality Reduction

Our dataset contains a total of twenty-nine features for each song, which is a lot of data to work with. Furthermore, some features have a greater impact than others in obtaining the target classes. That is why Dimentionality Reduction is used to remove features that have little impact on the calculation, making it easier to compute. This also makes classifiers more efficient because it eliminates the need for some unnecessary computations.

Principal Component Analysis is a popular technique for reducing dimension. It is employed in order to reduce the number of features in a dataset while retaining as much information as possible. In our case, we used PCA to calculate the effects of various features. Following that, we discovered that we only needed 16 features to retain 90% of the variance data. As a result, only the most important 16 features were considered, and the rest were eliminated. As a result, the number of computations decreased, and classifiers improved in calculation and performance.

Feature Scaling

Feature Scaling is a critical step in machine learning that involves normalising a dataset's range of features. Real-world datasets frequently contain features with varying magnitude, range, and units, which can lead to inaccurate predictions by models.
Scaling the features to the same scale allows the model to give equal weight to all features and make accurate predictions. Scaling can also help the optimisation process by smoothing the gradient descent flow and allowing algorithms to converge faster. Normalisation and standardisation are two scaling techniques that are used to rescale each feature to have a standard deviation of 1 and a mean of 0[39].

Label Encoding

Label encoding is a vital preprocessing step in machine learning that converts categorical columns into numerical ones, enabling the use of algorithms that only process numerical data. By assigning unique numerical labels to distinct categories, label encoding helps machine learning models to interpret and process categorical data more effectively[40].

5.1.2 Implementation

There are numerous classifiers available today to achieve efficient results in classification problems. To achieve our desired classification results, we used six machine learning classifiers and two neural network-based approaches. Also, the hyperparameters were tuned accordingly.

Machine Learning Algorithms

We tried both the linear kernel and the radial basis function kernel for SVM. C, which denotes the regularisation parameter, was set to 0.025 for the linear kernel. The hyperparameters max depth=10, random state=42, max features=None, and min samples leaf=15 were set for the decision tree. The loss was set to 'modified huber' for SGD. Random forest and K-nearest neighbours were also used.

Artificial Neural Networks

In the case of artificial neural network approaches, we started with a Multi Layer Perceptron classifier. The activation function for MLP was 'relu,' the solver or optimizer was 'adam', the initial learning rate was 0.001, and the maximum iteration was 1000 with early stopping enabled. A sequential model with six dense layers was used for our custom neural network architecture.
The number of neurons was 256, 128, 64, 32, 16, and 6 in that order. The activation function was 'relu' for all layers except the last, which used the 'sigmoid' function. Also, the dropout values for the layers were 0.6, 0.4, 0.3, and 0.2, respectively. The loss was set to sparse categorical cross-entropy during compilation, and the optimizer was 'rmsprop'. In addition, 150 epochs were used with an 8-batch size and early stopping enabled.

5.1.3 Algorithm

Table 5.1: Definition of used variables in the algorithm

Notation	Definition
X_train	This variable contains the features (input variables) from the dataset that will be used for training the model.
X_test	This variable contains a subset of the features from the dataset that will be used for testing the model's performance.
y_train	This variable contains the corresponding target values (labels) for the training dataset.
y_test	This variable contains the corresponding target values for the test dataset
PCA	Principal Component Analysis
reducedData	This variable contains the dataset after the principal component analysis has been done
processedData	This variable contains represents data after using feature scaling on "reducedData"

Algorithm 1: Algorithm

Input: Music Data from Dataset **Output**: Trained Model to Categorize Songs
Pre-Processing:
X ← data
Y ← label
reducedData ← Dimentionality Reduction using PCA(X)
processedData ← Feature Scaling using Standard Scaler(reducedData)
Y ← Label Encoding
X_train, X_test, y_train, y_test ← train-test-split(reducedData, Y)
Training:
for *each Algorithm algo used in the experiment* **do**
    for *each data i in dataset* **do**
        feed data into algorithms
    Model Evaluate:
    compute accuracy$_{algo}$, $f1 - score_{algo}, recall_{algo}, precision_{algo}$

Chapter 6

Sentiment Analysis of Bengali Music

The methodology we propose comprises two fundamental components: data pre-processing and algorithm implementation. Furthermore, the findings are subsequently examined to assess the effectiveness of different classifiers.

6.1 Data Preprocessing

6.1.1 Analyzing the data

Upon generating the dataset, we carefully reviewed the data to detect possible abnormalities and gain a comprehensive understanding of the total dataset. The dataset comprises an overall total of twenty-nine distinct attributes for every individual song. Upon doing a comprehensive assessment, we came across a correlation heatmap depicted in Figure 6.1.

A correlation heatmap is a graphical tool that displays the correlation between multiple variables as a color-coded matrix.

Through analysis of the heatmap, it becomes evident that the sentiments are mostly influenced by several distinct characteristics namely zero crossing, spectral rolloff, chroma frequency etc. The other features have negligible impact on classifying the sentiment groups.

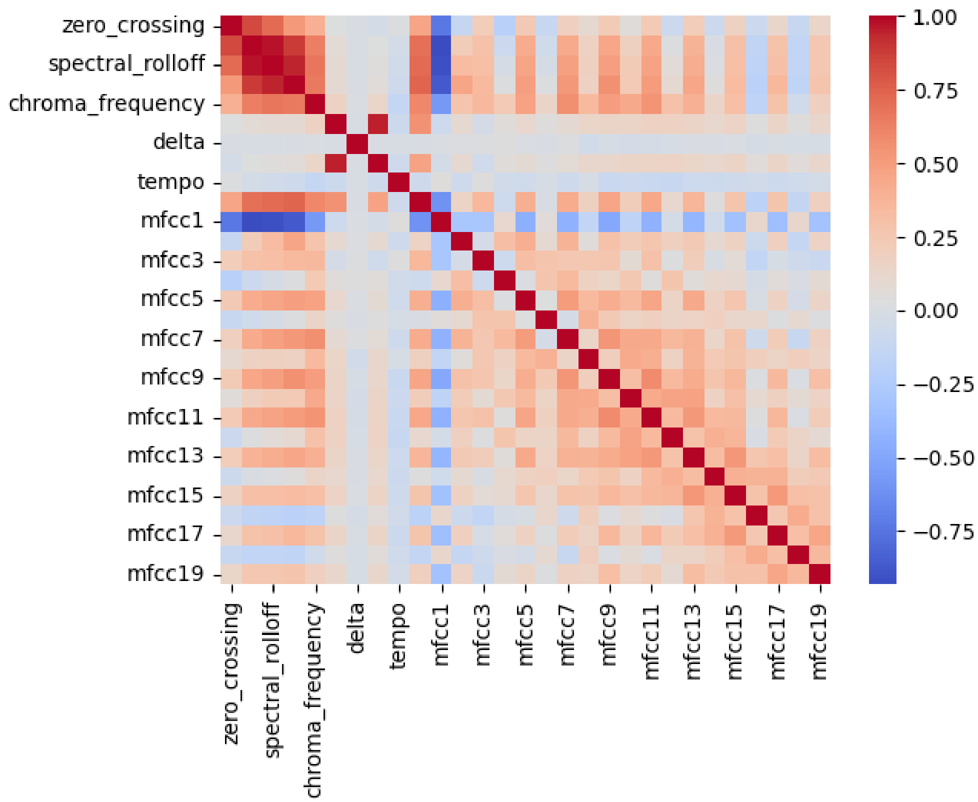

Figure 6.1: Data Correlation Heatmap.

6.1.2 Dimentionality Reduction

Dimensionality Reduction is employed to eliminate audio elements with minimal influence on the classification of target classes, hence simplifying the computation process. This also enhances the efficiency of classifiers by eliminating the requirement for some superfluous computations.

Principal component analysis, or PCA, is a dimensionality reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Accordingly, Principal Component Analysis (PCA) is a widely used method for minimising the number of dimensions in a dataset. It is utilised to decrease the amount of features in a dataset while preserving maximum information. We employed Principal Component Analysis (PCA) to quantify the impacts of different characteristics.

Subsequently, it was determined that a mere 16 attributes were sufficient to preserve 90% of the variation in the data. Consequently, just the most crucial 16 traits were taken into account, while the remaining ones were discarded. Therefore, there was a reduction in the amount of computations, leading to enhanced calculation and performance of classifiers.

6.1.3 Feature Scaling and Label Encoding

Machine learning relies on feature scaling to normalise a dataset's characteristics. Real-world datasets often have features with different magnitude, range, and units, which can cause model errors. By scaling all features to the same scale, the model can assign them equal weight and produce accurate predictions. Scaling can improve optimisation by smoothing gradient descent and accelerating algorithm convergence [39].

Furthermore, an important preprocessing step in machine learning is label encoding, which turns categorical columns into numerical ones. By assigning unique numerical labels to distinct categories, label encoding helps machine learning models to interpret and process categorical data more effectively [40].

6.2 Implementation of algorithms

In the realm of classification tasks, the wide variety of classifiers available offers ample opportunities for researchers and practitioners to select the most suitable methods. Through thoughtful exploration and experimentation, individuals can identify the optimal classifiers to meet their specific needs and achieve desired outcomes effectively and confidently.

Currently, there is a wide range of classifiers that can be used to produce optimal outcomes in classification tasks.

In order to attain our intended classification outcomes, we employed five machine learning classifiers and two techniques based on neural networks..

6.2.1 Machine Learning Algorithms

The machine learning algorithms used in our study are Support Vector Machine (SVM), Random forest, K-nearest neighbors (KNN), Stochastic Gradient Descent(SGD) and Decision Tree.

Since there are different hyperparameters for all the models, we used grid search cv provided by scikit learn to achieve the best combination of hyperparameters.

6.2.2 Articial Neural Networks

Regarding artificial neural network methodologies, our initial step was utilising a Multi Layer Perceptron classifier. We employed grid search cross-validation to optimise the parameters of the MLP classifier.

In addition, our unique neural network design employed a sequential model consisting of five dense layers, including the output layer. The sequence of neuron counts was 128, 64, 32, 16, and 5, respectively in each layer. The activation function employed 'relu' for all levels, with the exception of the final layer that utilised the 'sigmoid' function. The dropout rates for the layers were 0.4, 0.3, and 0.2, respectively. The loss function was specified as 'sparse categorical crossentropy' throughout the compilation process, and the optimizer used was rmsprop.

Furthermore, the training process involved 200 epochs, where each epoch consisted of a batch size of 20. Additionally, early stopping was enabled.

6.2.3 Algorithm for the work flow

Table 6.1: Definition of used variables in the algorithm

Notation	**Definition**
X_train	This variable contains the features (input variables) from the dataset that will be used for training the model
X_test	This variable contains a subset of the features from the dataset that will be used for testing the model's performance
y_train	This variable contains the corresponding target values (labels) for the training dataset
y_test	This variable contains the corresponding target values for the test dataset
PCA	Principal Component Analysis
reducedData	This variable contains the dataset after the principal component analysis has been done
processedData	This variable contains represents data after using feature scaling on "reducedData"

Algorithm 2: Algorithm

Input: Music Data from Dataset **Output**: Trained Model to Analyze Sentiments
Pre-Processing:
X ← data
Y ← label
reducedData ← Dimentionality Reduction using PCA(X)
processedData ← Feature Scaling using Standard Scaler(reducedData)
Y ← Label Encoding
X_train, X_test, y_train, y_test ← train-test-split(reducedData, Y)
Training:
for *each data i in dataset* **do**
    feed data into algorithms
Model Evaluate:
evaluate various models in terms of accuracy, f1-score, recall, precision

Chapter 7

Results & Analysis

7.1 Experimental Tools

We used Anaconda distribution and Python 3.10 as the programming language for our project. Python library functions were used for programming, and sci-kit learn was used to implement machine learning algorithms. Keras library was used for implementing deep neural networks. For our research, we used a computer with 8 gigabytes of RAM, a core i5-1235U CPU, and an RTX 3050 GPU. The GPU was used to train the deep neural network model.

7.2 Analysis of results for Genre Classification

The evaluation criteria of recall, precision, f1-score, and accuracy were determined for each model under consideration. These metrics are standard tools for performance evaluation. We implemented various machine learning classifiers as well as artificial neural network. With regard to machine learning classifiers, we have used Random forest, SVM-RBF, SVM Linear, SGD, Decision tree and KNN. Additionally, in the context of artificial neural network, we have implemented Multi Layer Perceptron classifier and our custom Deep Learning architecture. The accuracy, precision, recall, f1-score across all our models in our analysis has been illustrated in Table 7.1.

When compared with machine learning classifiers and our deep learning architecture, the deep learning architecture has achieved the highest performance which is 83.65%. Among all the implemented machine learning classifiers, it is evident that SVM-RBF excelled with an accuracy of 74.35%. Moreover, SVM-Linear and Random Forest exhib-

Table 7.1: **Performance comparison of implemented models for Genre Classification**

Method	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Random Forest	71.22	71.66	71.66	71.07
SVM-RBF	74.35	74.63	74.37	74.24
SVM-Linear	71.52	71.10	71.47	71.29
SGD	64.78	65.00	64.81	64.18
Decision Tree	61.26	61.59	61.52	61.03
KNN	65.57	70.10	65.41	65.32
MLP	75.61	75.65	75.85	75.48
Deep Neural Network	83.65	83.92	83.27	83.11

ited almost similar levels of accuracy, both recording 71.52% and 71.22%, respectively, closely approaching the accuracy of SVM-RBF. Furthermore, SGD, Decision tree, KNN has achieved the accuracy of 64.78%, 61.26%, 65.57% respectively. The results show significant improvement from other related works [5] [4] both in performance of machine learning classifiers and neural network classifiers. Our revised neural network architecture outperforms all classifiers developed for this project as well as networks developed by other researchers [4] [5]. The accuracy of the neural network architecture proposed by Ahmed et al., [4] was about 77.68% and the Gated Recurrent Unit based model proposed by Sarma et al., [5] was about 80.6% whereas our neural network architecture was able to achieve a outstanding accuracy of 83.65% while also being computationally cost effective. The dimensionality reduction is one of the primary reasons why our proposed method works relatively better. Even though the dataset is not very large, it has an extensive amount of dimensions, making it difficult for classifiers to find enough characteristics for each class. Our dimensionality reduction technique reduced the dimensions, leaving only the most important features. This allows the classifiers to work more efficiently and provide better performance. In addition, our neural network is one layer larger than the model proposed by Ahmed et al., [4], and as our dimension was reduced, it also became more computationally more efficient. It could also handle the data with great effectiveness.

Finally, we attempt to evaluate the models from our machine learning and deep learning classifiers by comparing their performances in classifying each genre. Tables 7.2 to 7.5 shows the accuracy, precision, recall and f1-score respectively for classifying each of the genre classes.

Table 7.2: Accuracy Percentage for Classifying Each Genre

Genre	Palligeeti	Hip-Hop	Nazrul	Rabindra	Adhunik	Band
Random Forest	73.51	66.85	72.72	73.92	69.18	70.33
SVM-RBF	75.21	73.14	72.94	74.81	76.26	73.88
SVM-Linear	74.26	69.76	74.79	71.67	69.54	70.83
SGD	65.81	62.23	66.09	67.36	61.92	63.17
Decision Tree	64.43	56.91	58.79	61.97	59.14	61.92
KNN	69.21	63.87	62.98	66.12	61.76	67.66
MLP	79.29	72.98	73.11	78.21	73.87	76.22
DLM	86.52	79.77	84.91	87.43	81.28	85.82

Table 7.3: Precision Percentage for Classifying Each Genre

Genre	Palligeeti	Hip-Hop	Nazrul	Rabindra	Adhunik	Band
Random Forest	70.78	70.96	74.16	71.44	73.17	73.26
SVM-RBF	71.87	73.42	77.39	76.21	75.56	74.31
SVM-Linear	68.29	70.44	72.51	73.63	69.88	72.94
SGD	61.23	65.87	67.59	64.22	66.18	67.01
Decision Tree	59.12	61.93	60.21	60.74	61.43	62.18
KNN	67.48	69.29	70.71	70.28	70.12	72.07
MLP	73.98	74.51	78.29	76.77	74.63	76.94
DLM	81.12	84.71	84.23	82.95	82.17	83.41

Table 7.4: Recall Percentage for Classifying Each Genre

Genre	Palligeeti	Hip-Hop	Nazrul	Rabindra	Adhunik	Band
Random Forest	70.39	72.02	70.99	71.45	72.12	71.87
SVM-RBF	72.61	74.82	73.98	75.24	75.13	74.26
SVM-Linear	69.82	71.23	72.17	70.95	72.68	71.11
SGD	62.29	65.22	64.59	63.91	65.73	65.12
Decision Tree	60.15	59.83	61.99	61.27	60.62	62.14
KNN	67.29	63.18	67.55	66.42	64.89	65.72
MLP	77.13	74.29	74.97	75.62	76.89	75.02
DLM	85.02	81.71	84.39	81.93	83.74	82.95

Table 7.5: F1-Score Percentage for Classifying Each Genre

Genre	Palligeeti	Hip-Hop	Nazrul	Rabindra	Adhunik	Band
Random Forest	68.92	72.15	69.74	70.29	72.31	70.88
SVM-RBF	71.58	72.98	76.15	73.72	74.96	74.08
SVM-Linear	70.17	70.86	72.43	72.16	71.72	70.51
SGD	62.29	61.54	66.07	65.09	63.28	64.42
Decision Tree	59.41	62.17	60.28	60.89	61.55	60.92
KNN	63.49	64.27	67.04	64.68	66.12	65.77
MLP	76.32	74.92	75.81	74.73	75.12	76.16
DLM	81.78	83.54	81.33	85.01	82.67	82.29

7.2.1 Analysis of results for Sentiment Analysis

Each model used for classification was assessed based on the evaluation metrics of recall, precision, f1-score, and accuracy. These measures are commonly used metrics for assessing performance. We have incorporated a range of machine learning classifiers, along with artificial neural networks. Regarding machine learning classifiers, we have employed Random Forest, SVM with Radial Basis Function (RBF) kernel, SVM with Linear kernel, Stochastic Gradient Descent (SGD), Decision Tree, and K-Nearest Neighbours (KNN). In addition, we have incorporated a Multi Layer Perceptron classifier and our own Deep Learning architecture for the artificial neural network. The accuracy, precision, recall, and f1-score for all of our models in our analysis are presented in Table 7.6.

Table 7.6: **Performance evaluation of implemented models for Sentiment Analysis**

Method	Accuracy	Precision	Recall	F1 Score
SVM-RBF	72.61	72.72	71.16	71.49
SVM-Linear	65.48	66.37	64.18	64.54
Random Forest	67.74	66.37	64.18	64.54
SGD	52.96	50.90	54.55	51.74
Decision Tree	54.35	54.73	52.81	53.42
KNN	63.74	62.32	61.55	60.71
MLP	68.87	68.65	69.55	69.05
Deep Neural Network	76.79	76.35	77.02	76.81

Based on the data shown in Table 7.6, it is evident that the SVM with RBF kernel surpasses all other machine learning classifiers in terms of accuracy, achieving a score of 72.61%. The random forest model has nearly equivalent performance, achieving an accuracy rate of 67.74%. The model with the lowest performance is the SGD model, which has an accuracy of 52.96%. In contrast, deep learning models surpass all other machine learning models in performance. The MLP classifier achieved an accuracy of 68.87%, surpassing all other machine learning classifiers. Furthermore, our custom deep learning model offers the highest efficiency. With a remarkable accuracy of 76.79%, it surpassed and exceeded all other models. Our models aimed to categorise songs into five distinct sentiment categories, which is a notable achievement in the field of sentiment analysis of Bengali music. Additionally, we implemented a unique strategy by utilising the inherent audio properties to determine the sentiment classes of Bengali music. Despite Mahajebin et al.'s [41] achievement of approximately 65% accuracy on comparable tasks, they employed a different methodology. They used lyric-based sentiment analysis and classifying sentiments into four separate groups. Despite the limited size of our dataset, our classifiers demonstrated exceptional performance.

Furthermore, our dataset includes 29 inherent audio characteristics for each of the songs. After conducting principal component analysis, we selected and retained just the 16 most significant features for training the models. In order to gain a more detailed understanding of the significance of these 16 characteristics, we employed our random forest classifier to produce the feature importance. Figure 7.1 demonstrates that zero crossing has the

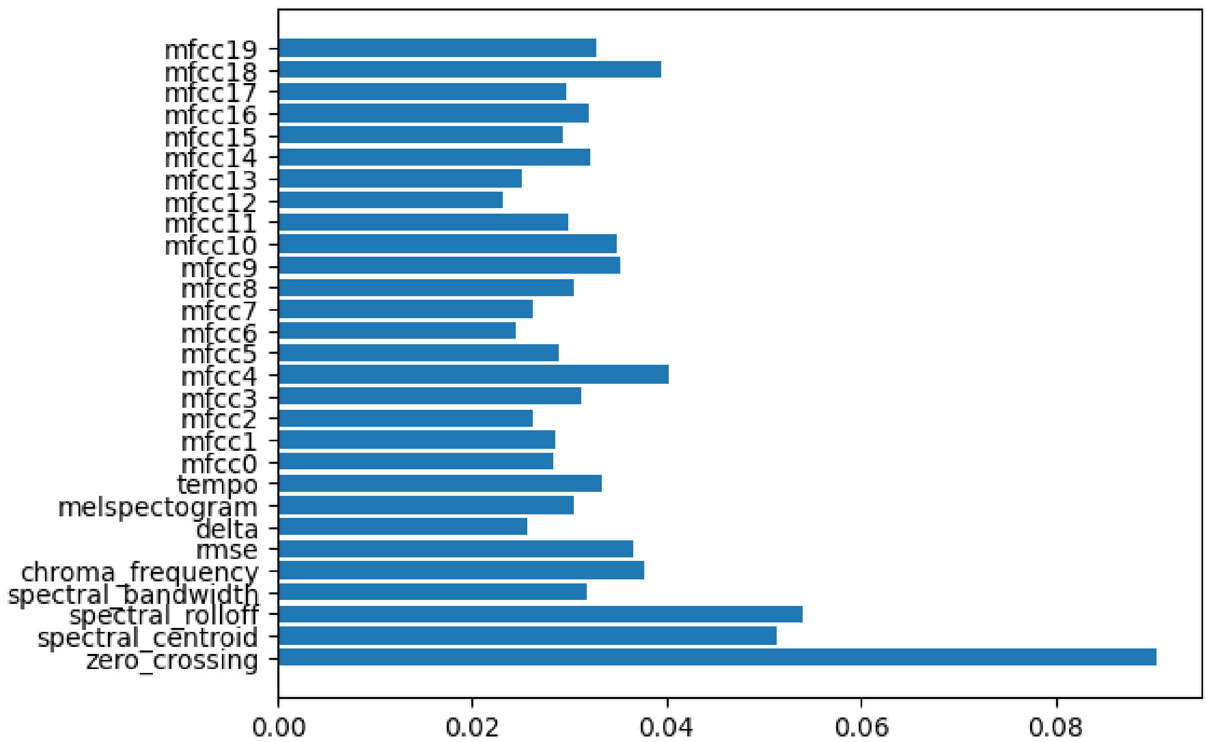

Figure 7.1: Feature Importance Barplot.

greatest influence on song sentiments, followed by spectral rolloff and spectral centroid as mentioned in the heatmap in Figure 6.1. Among the selected 16 features, Mfcc 12 has the lowest impact in sentiment classification.

Tables 7.7 to 7.10 shows the accuracy, precision, recall and f1-score respectively for classifying each of the sentiments.

Table 7.7: Accuracy Percentage for Classifying Sentiment

Genre	Romanticism	Sadness	Amusement	FPU	Relaxation
Random Forest	67.18	70.02	65.91	66.47	68.85
SVM-RBF	74.32	70.89	73.67	72.43	70.24
SVM-Linear	67.38	66.19	64.57	66.63	63.12
SGD	55.21	53.67	51.42	55.58	50.81
Decision Tree	52.21	55.29	53.88	56.12	53.95
KNN	61.23	65.05	62.89	61.91	66.22
MLP	67.21	70.14	67.82	69.58	66.93
DLM	75.15	77.12	74.98	79.02	76.24

Table 7.8: Precision Percentage for Classifying Sentiment

Genre	Palligeeti	Hip-Hop	Nazrul	Rabindra	Adhunik	Band
Random Forest	71.46	73.18	71.81	73.95	71.92	72.11
SVM-RBF	67.51	65.98	66.14	68.23	65.72	66.54
SVM-Linear	64.82	65.21	65.97	67.89	65.45	66.94
SGD	52.76	49.62	52.45	51.18	49.23	51.91
Decision Tree	53.49	56.27	53.12	54.91	55.83	54.46
KNN	59.47	61.82	60.19	63.45	63.91	62.06
MLP	66.32	68.91	67.54	69.27	68.09	69.73
DLM	78.02	75.88	77.46	76.91	75.75	76.18

Table 7.9: Recall Percentage for Classifying Sentiment

Genre	Palligeeti	Hip-Hop	Nazrul	Rabindra	Adhunik	Band
Random Forest	69.42	71.79	70.18	72.55	70.91	69.98
SVM-RBF	62.43	61.27	64.72	63.84	65.19	64.91
SVM-Linear	62.59	60.92	63.75	64.26	65.08	63.71
SGD	62.27	64.82	65.14	63.19	63.98	65.77
Decision Tree	50.72	54.29	51.63	54.18	53.09	51.95
KNN	60.12	62.79	60.67	61.98	62.13	60.21
MLP	67.93	68.21	69.79	70.12	69.38	68.74
DLM	74.86	76.52	77.45	78.13	76.94	77.25

Table 7.10: F1-Score Percentage for Classifying Sentiment

Genre	Palligeeti	Hip-Hop	Nazrul	Rabindra	Adhunik	Band
Random Forest	70.19	72.62	70.91	69.73	71.85	72.15
SVM-RBF	63.21	65.17	63.92	65.45	63.78	64.13
SVM-Linear	67.02	64.89	65.21	66.43	65.77	64.98
SGD	49.61	52.12	50.43	50.88	51.92	52.27
Decision Tree	51.18	54.21	53.89	53.13	52.76	54.02
KNN	58.92	60.33	61.29	59.84	61.52	60.27
MLP	66.92	70.11	68.25	69.63	69.82	68.71
DLM	78.12	76.32	75.95	77.23	76.55	77.29

Chapter 8

Future Work

In the coming times, our focus will be on refining the computer algorithms used for sorting and comprehending Bengali music, aiming for improved efficiency and accuracy. Our objective is to enhance their ability to distinguish between different music genres and interpret the associated emotions effectively.

We will start by adjusting the parameters of these algorithms to optimize their performance. Through experimentation, we will explore various configurations to determine the most effective ones for discerning between different types of music and understanding their emotional nuances.

Expanding our Bengali music collection will also be a priority. Utilizing tools like librosa, we will gather a larger selection of songs spanning diverse genres and emotional expressions. This broader dataset will provide our algorithms with more examples to learn from, enhancing their capability to analyze a wide range of Bengali music. To ensure diversity in our collection, we will seek input from music experts. Their guidance will help us curate a well-rounded selection of songs representing various musical styles and emotional tones. Additionally, we will investigate advanced techniques to further refine our algorithms. These methods may involve leveraging pre-existing models and combining different algorithms to improve overall performance and accuracy.

In summary, our future efforts will focus on enhancing the efficiency and accuracy of our algorithms for understanding Bengali music. This will involve adjusting parameters, expanding our dataset, seeking guidance from experts, and exploring advanced techniques to achieve optimal results.

Chapter 9

Conclusion

The classification and categorization of Bangla music genres and sentiments play a pivotal role in the development of dedicated online platforms catering to Bengali music enthusiasts. These platforms aim to provide users with a seamless experience, allowing them to explore and discover music that resonates with their tastes and moods. Through effective organization based on genre and sentiment, users can easily navigate vast music libraries, enhancing their overall engagement and satisfaction.

While significant strides have been made in the field of music classification and sentiment analysis, there is always room for improvement. Current models have demonstrated promising results, but their performance can be further enhanced through iterative refinement and expansion of the underlying datasets. This continual improvement process is essential for ensuring the accuracy and efficiency of genre classification and sentiment analysis in Bengali music.

One approach to improving model performance is the gradual tuning of hyperparameters. Hyperparameters are crucial settings that govern the behavior of machine learning models, and fine-tuning them can have a significant impact on performance. By systematically adjusting these parameters over time, researchers can optimize model performance to achieve better results in genre classification and sentiment analysis tasks. This iterative tuning process allows for the identification of optimal parameter configurations that maximize model accuracy and efficiency. In addition to hyperparameter tuning, expanding the dataset is another key strategy for improving model performance. A larger and more diverse dataset provides the model with a richer source of information, enabling it to learn more effectively and generalize better to unseen data. By incorporating additional songs with accurate features, researchers can enhance the model's ability to classify music genres and analyze sentiments accurately. This expansion of the dataset not only improves the performance of existing models but also lays the groundwork for the development of more advanced algorithms in the future.

The process of dataset expansion involves collecting and annotating a diverse range of Bengali songs, ensuring that they cover various genres and emotional themes. Careful curation of the dataset is essential to maintain quality and relevance, as well as to minimize biases that may affect model performance. Researchers may leverage crowd-sourcing or collaborate with music experts to gather and annotate the data, ensuring its accuracy and comprehensiveness. Once the expanded dataset is in place, researchers can use it to train and evaluate improved models for genre classification and sentiment analysis. By incorporating a larger and more diverse set of examples, these models can learn more robust representations of Bengali music, leading to better performance on real-world tasks. Additionally, researchers can employ advanced techniques such as transfer learning and ensemble methods to further boost model performance and robustness.

Overall, the classification and categorization of Bangla music genres and sentiments are essential for the development of effective online platforms that cater to the diverse preferences of Bengali music enthusiasts. Through iterative refinement and expansion of datasets, researchers can continue to improve the accuracy and efficiency of genre classification and sentiment analysis algorithms, ultimately enhancing the user experience and promoting the discovery of new and exciting music.

Bibliography

[1] "Practise bangla at all levels of education." [Online; accessed 10-November-2023].

[2] M. A. Al Mamun, I. Kadir, A. S. A. Rabby, and A. Al Azmi, "Bangla music genre classification using neural network," in *2019 8th international conference system modeling and advancement in research trends (SMART)*, pp. 397–403, IEEE, 2019.

[3] R. Hasan, S. Hossain, F. I. Alam, and M. Barua, "Bangla music genre classification using fast and scalable integrated ensemble boosting framework," in *2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pp. 1–6, IEEE, 2021.

[4] T. Ahmed, M. A. Alam, R. R. Paul, M. T. Hasan, and R. Rab, "Machine learning and deep learning techniques for genre classification of bangla music," in *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, pp. 1–6, IEEE, 2022.

[5] M. Sen Sarma and A. Das, "Bmgc: A deep learning approach to classify bengali music genres," in *Proceedings of the 4th International Conference on Networking, Information Systems & Security*, pp. 1–6, 2021.

[6] R. Hossain and A. Al Marouf, "Banglamusicstylo: a stylometric dataset of bangla music lyrics," in *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pp. 1–5, IEEE, 2018.

[7] H. Ali, M. F. Hossain, S. B. Shuvo, and A. Al Marouf, "Banglasenti: A dataset of bangla words for sentiment analysis," in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–4, IEEE, 2020.

[8] M. Mia, P. Das, and A. Habib, "Verse-based emotion analysis of bengali music from lyrics using machine learning and neural network classifiers," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 359–370, 2024.

[9] D. Nath, A. Roy, S. K. Shaw, A. Ghorai, and S. Phani, "Textual lyrics based emotion analysis of bengali songs," in *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 39–44, IEEE, 2020.

[10] A. N. Tusher, S. C. Das, M. L. H. Moeen, M. S. R. Sammy, M. R. S. Sakib, and A. I. Aunik, "Sentiment analysis of bangla song comments: A machine learning approach," 2023.

[11] D. Nath and S. Phani, "Mood analysis of bengali songs using deep neural networks," in *Information and Communication Technology for Competitive Strategies (ICTCS 2020) Intelligent Strategies for ICT*, pp. 1103–1113, Springer, 2021.

[12] S. Ghosh and M. O. F. Riad, "Attention-based cnn-bigru for bengali music emotion classification," *Indonesian Journal of Computer Science*, vol. 11, no. 3, 2022.

[13] Z. Cataltepe, Y. Yaslan, and A. Sonmez, "Music genre classification using midi and audio features," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–8, 2007.

[14] S. Lippens, J.-P. Martens, and T. De Mulder, "A comparison of human and automatic musical genre classification," in *2004 IEEE international conference on acoustics, speech, and signal processing*, vol. 4, pp. iv–iv, IEEE, 2004.

[15] M. Haggblade, Y. Hong, and K. Kao, "Music genre classification," *Department of Computer Science, Stanford University*, 2011.

[16] C. McKay and I. Fujinaga, "Musical genre classification: Is it worth pursuing and how can it be improved?," in *ISMIR*, pp. 101–106, 2006.

[17] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 282–289, 2003.

[18] A. Meng, P. Ahrendt, J. Larsen, and L. K. Hansen, "Temporal feature integration for music genre classification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 5, pp. 1654–1664, 2007.

[19] T. Li and M. Ogihara, "Music genre classification with taxonomy," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 5, pp. v–197, IEEE, 2005.

[20] S. Oramas, F. Barbieri, O. Nieto Caballero, and X. Serra, "Multimodal deep learning for music genre classification," *Transactions of the International Society for Music Information Retrieval. 2018; 1 (1): 4-21.*, 2018.

[21] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.

[22] L. Nanni, Y. M. Costa, A. Lumini, M. Y. Kim, and S. R. Baek, "Combining visual and acoustic features for music genre classification," *Expert Systems with Applications*, vol. 45, pp. 108–117, 2016.

[23] F. Gouyon, F. Pachet, O. Delerue, *et al.*, "On the use of zero-crossing rate for an application of classification of percussive sounds," in *Proceedings of the COST G-6 conference on Digital Audio Effects (DAFX-00), Verona, Italy*, vol. 5, p. 16, 2000.

[24] E. Priya, P. S. Reshma, S. Sashaank, *et al.*, "Temporal and spectral features based gender recognition from audio signals," in *2022 International conference on communication, computing and internet of things (IC3IoT)*, pp. 1–5, IEEE, 2022.

[25] U. Ayvaz, H. Gürüler, F. Khan, N. Ahmed, T. Whangbo, and A. Bobomirzaevich, "Automatic speaker recognition using mel-frequency cepstral coefficients through machine learning," *CMC-computers materials & continua*, vol. 71, no. 3, 2022.

[26] J. T. Abraham, A. N. Khan, and A. Shahina, "A deep learning approach for robust speaker identification using chroma energy normalized statistics and mel frequency cepstral coefficients," *International Journal of Speech Technology*, vol. 26, no. 3, pp. 579–587, 2023.

[27] J. M. Ha, H. B. Shin, J. F. Joung, W. J. Chung, J.-E. Jeong, S. Kim, S. H. Hur, S.-Y. Bae, J.-Y. Kim, J. Y. Lee, *et al.*, "Rational molecular design of azaacene-based narrowband green-emitting fluorophores: Modulation of spectral bandwidth and vibronic transitions," *ACS Applied Materials & Interfaces*, vol. 13, no. 22, pp. 26227–26236, 2021.

[28] C. Cox, W. Trojak, T. Dzanic, F. Witherden, and A. Jameson, "Accuracy, stability, and performance comparison between the spectral difference and flux reconstruction schemes (2020)," *arXiv preprint arXiv:2009.06072*.

[29] M. F. Asmussen, J. Liniger, and H. C. Pedersen, "Fault detection and diagnosis methods for fluid power pitch system components—a review," *Energies*, vol. 14, no. 5, p. 1305, 2021.

[30] M. Ortíz, P. Vicente, E. Iáñez, E. Montiel, and J. M. Azorín, "Assessing footwear comfort by electroencephalography analysis," *IEEE Access*, vol. 9, pp. 134259–134269, 2021.

[31] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," vol. 2049, pp. 249–257, 09 2001.

[32] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, Oct 2001.

[33] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, pp. 81–106, Mar 1986.

[34] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Review*, vol. 60, no. 2, pp. 223–311, 2018.

[35] J. P. Jiawei Han, Micheline Kamber, *Data Mining: Concepts and Techniques, 3rd Edition.* Morgan Kaufmann, 2011.

[36] S. Haykin, *Neural Networks: A Comprehensive Foundation.* USA: Prentice Hall PTR, 2nd ed., 1998.

[37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning.* MIT Press, 2016. `http://www.deeplearningbook.org`.

[38] A. A. Mamun, "Bangla music dataset," 2020.

[39] "Introduction to feature scaling: Normalizing and standardization." [Online; accessed 7-November-2023].

[40] "Label encoding in python – 2024." [Online; accessed 7-November-2023].

[41] M. Mahajebin, M. R. A. Rashid, and N. Mansoor, "Mood classification of bangla songs based on lyrics," in *Inventive Communication and Computational Technologies* (G. Ranganathan, G. A. Papakostas, and Á. Rocha, eds.), (Singapore), pp. 585–597, Springer Nature Singapore, 2023.

Music Genre Classification and Sentiment Analysis of Bengali Music based on various inherent features

By

Atika Humayra

Student ID: **2030021**

Md Maruf Kamran Sohag

Student ID: **2130278**

Spring, 2024

The student modified the internship final report as per the recommendation made by his or her academic supervisor and/or panel members during final viva, and the department can use this version for achieving.

Signature of the Supervisor

Dr. Mohammed Anwer

Professor

Department of Physical Sciences

School of Engineering, Technology & Sciences

Independent University, Bangladesh

Signature of the Co-Supervisor

Dr. Mahady Hasan

Associate Professor

Department of Computer Science & Engineering

School of Engineering, Technology & Sciences

Independent University, Bangladesh