Bangla Music Genre Classification using Deep Learning and Machine Learning Techniques

BY

Khairul Islam Ope
ID: 202-15-3839
AND

Md. Mostafizur Rahman
ID: 202-15-3850

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms. Sharmin Akter
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Ms. Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Bangla Music Genre Classification using Deep Learning and Machine Learning Techniques", submitted by Khairul Islam Ope and Md. Mostafizur Rahman to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 13-07-2024.

BOARD OF EXAMINERS

Dr. S.M Aminul Haque
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Board Chairman

13.07.24

Ms. Nazmun Nessa Moon
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

13.7.24

Mr. Deawan Rakin Ahmed Remal
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 2

13.07.2024

Dr. Abu Sayed Md. Mostafizur Rahaman
Professor
Department of CSE
Jahangirnagar University

External Examiner

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Ms. Sharmin Akter, Lecturer (Senior Scale), Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Ms. Sharmin Akter
Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Co-Supervised by:

For

Ms. Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Khairul Islam Ope

Khairul Islam Ope
ID: 202-15-3839
Department of CSE
Daffodil International University

Md. Mostafizur Rahman

Md. Mostafizur Rahman
ID: 202-15-3850
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Ms. Sharmin Akter**, **Senior Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Artificial Intelligence and Machine Learning*" to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. Sheak Rashed Haider Noori, Head of the Department of CSE,** for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

The rapid growth of digital music libraries and the volume of Bangla music has increased the need for efficient and automatic methods to categorize Bangla musical genres. This study provides a thorough examination of the development and implementation of a new Bangla Music Genre Classification model. The proposed model uses advanced machine learning techniques to analyze key elements of Bangla music compositions, allowing for accurate and efficient genre identification. The approach in this study extracts essential audio features including tempo, spectral centroid, chroma frequency, spectral rolloff, RMSE, spectral bandwidth and MFCC, from a complex dataset covering various Bangla music genres. A modern deep learning algorithm, specifically a convolutional neural network, is trained on carefully annotated data and uses these features as input. The research highlights the importance of tailoring feature extraction methods and model designs to the unique qualities of Bangla music. The model aims to improve the accuracy and cultural relevance of automated genre identification by addressing the challenges associated with the diversity within Bangla music. The model is thoroughly tested using benchmark datasets and its performance is compared with existing genre classification methods. The outcomes show how adaptable and effective the suggested method is, especially for Bangla music. This research provides a comprehensive overview of creating and applying a model for classifying Bangla music genres, offering a reliable and culturally aware solution for accurate genre classification in Bangla music. It not only addresses the technical challenges of automated genre identification but also recognizes the unique cultural aspects inherent in Bangla music.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
Table of contents	vi-viii
List of Figures	ix
List of Tables	x
CHAPTER	
CHAPTER 1: INTRODUCTION	**1-6**
1.1 Overview	1
1.2 Background and Present State	1-2
1.3 Problem Statement	2-3
1.4 Objectives	3
1.5 Scope and Limitations	4
1.6 Report Organization	5-6
1.7 Summary	6

CHAPTER 2: LITERATURE REVIEW	**7-18**
2.1 Overview	7
2.2 Related Works	7-12
2.3 Comparison between existing works	13-16
2.4 Open Issues	17
2.5 Summary	17
CHAPTER 3: METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION	**18-26**
3.1 Overview	18
3.2 Proposed Methodology/ System Design	18-21
3.3 Hardware/ Software Requirement	22
3.4 Project Management and Financial Analysis	23
3.5 Summary	24
CHAPTER 4: IMPLEMENTATION	**25-33**
4.1 Overview	25
4.2 Train Model/ Prototype Design	25-30
4.3 System Testing/ Model Evaluation	31-32
4.4 Summary	33

CHAPTER 5: RESULT AND ANALYSIS	**34-45**
5.1 Overview	34
5.2 Experimental/ Simulation Result	34-43
5.3 Performance/ Comparative Analysis	43-45
5.4 Summary	45
CHAPTER 6: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	**46-49**
6.1 Impact on Life	46
6.2 Impact on Society & Environment	46-47
6.3 Ethical Aspects	47
6.4 Sustainability Plan	48
6.5 Summary	49
CHAPTER 7: CONCLUSION AND FUTURE WORK	**50-52**
7.1 Conclusions	50
7.2 Further Suggested Works	51
7.3 Limitations/ Conflict of Interests	52
REFERENCES	**53-54**
APPENDIX	**55-60**

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.2.1: Flowchart of our proposed methodology.	19
Figure 3.2.2: Distribution of data Across Classes.	19
Figure 3.2.3: Plot Histogram of each Data.	20
Figure 3.2.4: Feature Importance.	20
Figure 4.2.1: Architecture of SVMG (source google).	26
Figure 4.2.2: Visualization of the working phase of KNN Algorithm (source google).	27
Figure 4.2.3: Working process of Random Forest (source google).	28
Figure 4.2.4: The structure of a CNN, consisting of convolutional, pooling, and fully-connected layers (source google).	29
Figure 5.2.1: Training and validation accuracy and loss over the epochs for CNN model.	40
Figure 5.2.2: CM of K nearest Neighbour.	41
Figure 5.2.3: CM of Support Vector Machine.	41
Figure 5.2.4: CM of Random Forest Classifier.	42
Figure 5.2.5: CM of Gradient Boosting Classifier.	42
Figure 5.2.6: CM of CNN.	43
Figure 5.3.1: Accuracy comparison among Machine Learning Models and CNN.	43

LIST OF TABLES

TABLES	**PAGE NO**
Table 2.3.1: Comparative analysis with previous work.	13
Table 3.4.1: Estimated Cost for Our Proposed Model.	23
Table 5.2.1: Accuracy for Classification of Machine Leaning Techniques and CNN Model.	34
Table 5.2.2: Precision, Recall, F1-Score, and Support for Machine Learning Models and CNN	36

CHAPTER 1
INTRODUCTION

1.1 Overview

The process of assigning genre names to musical works involves examining the sounds or lyrics. This classification can be applied in a variety of ways to manage and organize music suggestions, advertisements, and streaming audio services, given the exponential increase of music data repositories. Bangladeshi music encompasses a broad range of genres. Much work has been done on classifying music in English or other languages, but not so much on Bangla music. Our classification model is expected to benefit the Bengali music industry.

1.2 Background and Present State

In the age of streaming music digitally, the classification of musical genres plays a pivotal role in enhancing user experiences, aiding in music discovery, and facilitating personalized recommendations. While extensive research has been conducted in classifying music genres in various languages, the realm of Bengali music remains relatively unexplored in the domain of machine learning and artificial intelligence. Bangladesh, with its rich musical heritage spanning diverse genres such as adhunik, hip-hop, bangla band music, Palligeeti, Nazrul Geeti, and Rabindra Sangeet, presents a special chance for exploration.

This research endeavors to address this gap by developing an intelligent model for categorizing genres in Bengali music utilizing CNN [29] alongside traditional machine learning algorithms such as SVM [32], KNN [34], Random Forest [36], and Gradient Boosting [38]. The motivation stems from the potential transformative impact such a classification system could have on the Bangla music industry and the listening experiences of its enthusiasts. By comprehensively understanding and categorizing Bangla music genres [39], this research aims to not only enrich listeners' musical journeys but also empower musicians and content creators.

The main goal of this study is to build a robust classification model that can efficiently categorize Bengali music genres. To accomplish this goal, we will create a unique dataset containing a wide variety of Bengali musical genres. Through deep exploration of Bangla music and leveraging both traditional and advanced machine learning techniques, including random forest, SVM, KNN, gradient boosting, and CNNs, the model will be trained to recognize the unique characteristics and differences between each genre. Importantly, the effectiveness of the model will be assessed on its ability to accurately classify new and previously unheard songs, ensuring its adaptability and reliability in real-world scenarios. The scope of this project extends beyond mere classification, it delves into the cultural significance and historical evolution of Bengali music, emphasizing its diverse mix of traditions and influences. By combining conventional machine learning methods with cutting-edge deep learning techniques, this study attempts to close the gap between traditional music scholarship and modern computational methods, paving the way for innovative applications in music analysis and recommendation systems.

The goal of this study's output is to advance the expanding field of automatically classified musical genres, particularly within the context of Bengali music. Through meticulous data preprocessing, customized model architectures, and parameter fine-tuning for SVM, KNN, Random Forest, Gradient Boosting and CNNs, the project aims to achieve superior accuracy and performance metrics. Ultimately, the success of this project depends on using machine learning to protect and showcase the cultural legacy of Bengali music, making it more enjoyable for listeners everywhere.

1.3 Problem Statement

The task of accurately classifying Bangla music genres poses significant challenges due to the intricate and diverse nature of musical compositions. Traditional methods of genre classification often fall short in addressing the unique characteristics of Bangla music, which includes genres such as Bangla Adhunik, Hip-hop, Nazrul Geeti, Band Music, Rabindra Sangeet, and Palligeeti. Current approaches are limited by insufficient datasets, lack of advanced feature extraction techniques, and inadequate use of state-of-the-art machine learning and deep learning models.

To bridge this gap, this project aims to develop a robust Bangla music genre classification system utilizing advanced machine learning and deep learning techniques. The system will leverage a comprehensive dataset and extract detailed audio features to improve classification accuracy. By comparing various machine learning models, including convolutional neural network, support vector machine, k-nearest neighbor, random forest classifier, and gradient boosting classifier, the project seeks to identify the most effective approach for accurately classifying Bangla music genres. The goal is to ensure reliable performance in real-world applications and contributing to the broader field of music genre classification.

1.4 Objectives

The Bangla music industry is expected to benefit greatly from the classification of music genres, an area that has not yet been thoroughly investigated from the standpoint of machine learning and artificial intelligence. Not only can knowing the many genres of Bangla music improve one's listening experience, but it will also help listeners locate songs that suit their tastes. Since music has a deep emotional connection to people, the project's success will be determined by how much meaning it can bring to both musicians and listeners. Furthermore, by helping people choose their favorite genre, this genre classification technique helps guarantee that they find comfort in the music they listen to. It can also be a useful tool for suggesting songs for particular events, which helps people put together the ideal playlist. Furthermore, because this technology can effectively classify songs, it will improve the whole experience of listening to music on platforms like as YouTube, Spotify, Gaana, and others.

Our objective is to develop an intelligent model for classifying Bangla music genres. In order to create a unique dataset, we are delving deeply into the realm of Bangla music and concentrating just on it. We will take key audio information out of this dataset in order to teach our model about various musical genres. Additionally, the music we deal with might be classified into multiple genres. Our primary goal is to demonstrate how effectively our model comprehends Bangla music by ensuring that it performs well with new songs that it hasn't heard previously. Our goal is to finish the assignment in the allotted time.

1.5 Scope and Limitations

This research effort uses cutting-edge machine learning and artificial intelligence approaches to tackle the long-standing problem of accurately defining Bengali music genres. The issue at hand is that, despite a great deal of study being done on the classification of music genres in different languages, there hasn't been much inquiry and advancement in this particular area. Bengali music offers a distinct but mostly uncharted field in the field of machine learning and artificial intelligence. Its rich and varied legacy includes genres like Bangla Hip-Hop and Rabindra Sangeet.

This project covers a wide range of topics related to the problem. The primary objective is to accurately classify different genres of Bengali music. Advanced computer techniques like CNN, as well as more traditional approaches like support vector machines, k-nearest neighbors, random forests, and gradient boosting, are being used to achieve this. In order to create a robust model that can precisely classify Bengali music genres, we're using a hybrid technique. People will find it simpler to find new music they enjoy and receive suggestions based on their preferences as a result.

Building a comprehensive dataset with a large variety of various Bengali music genres is an essential component of this work. We need this data in order to train and test our classification model and ensure that it performs effectively in practical settings. But classifying music isn't the only goal of the project. We are also interested in finding more about the significance of Bengali music in culture and its past. We want to develop innovative approaches to music understanding and recommendation by fusing traditional music expertise with contemporary computer techniques.

A critical component of our research is testing the performance of our categorization model. We aim to test its ability to correctly categorize music that we are familiar with as well as ones that we have never heard of. We employ metrics like confusion matrices, accuracy, and precision to do this. These enable us to assess how well our model classifies music. In the end, we want to improve the experience of music listening for listeners and support artists and musicians in their endeavors. This research project is an attempt to comprehend and address the challenges of employing intelligent computer techniques for Bengali music genre classification. Through meticulous dataset construction, cultural analysis, and technique testing, our goal is to enhance the automated music genre classification capabilities of computers.

1.6 Report Organization

The report layout encompasses six key sections: Introduction, Background, Research Methodology, Implementation, Experimental Results, and Impact on Society. Each section contributes to a comprehensive understanding of the research, from contextualization and methodology to findings and societal implications. Through concise descriptions, the layout guides readers through the report's structure, facilitating clear comprehension and navigation.

Chapter 1: Introduction

This section lays the groundwork for our report by discussing the importance of music genre classification in today's digital music landscape. We highlight the lack of research in classifying Bengali music genres and outline our objectives and methods.

Chapter 2: Literature Review

Here, we delve into the reasons behind our research, emphasizing its potential impact on the Bengali music industry and user experiences. We explore the cultural richness of Bengali music genres, providing context for our study.

Chapter 3: Methodology

This section outlines the processes we followed during the research process, from preparing data to implementing algorithms. We explain our use of a CNN architecture and feature engineering techniques, along with the significance of tools like the Librosa, python library, MFCC and other features.

Chapter 4: Implementation

The implementation of the Bangla music genre classification system involved preprocessing the Kaggle dataset, extracting features, and training several models including KNN, SVM, Random Forest, Gradient Boosting, and CNN. The models were evaluated using metrics such as accuracy, precision, recall, and F1-score, with the CNN model achieving the best performance. Key libraries like TensorFlow, Keras, Scikit-Learn, and Librosa were utilized to ensure effective implementation.

Chapter 5: Result and Analysis

Here, we present the outcomes of our experiments, including data collection, feature extraction, and model performance analysis. We discuss the datasets used and evaluate the effectiveness of our CNN-based model for Bengali music genre classification.

Chapter 6: Impact on Society, Environment and Sustainability

This part looks at our research's wider societal ramifications. We discuss how our model can enhance music discovery and user experiences, as well as its potential benefits for the Bangla music industry and digital platforms.

Chapter 7: Conclusion and Future Work

In this final section, we summarize our key findings, conclude on the effectiveness of our model, and offer recommendations for future research directions. We also discuss the wider implications for music genre classification and cultural preservation efforts.

1.7 Summary

Using a CNN-based algorithm, this project attempts to categorize Bengali music by genre. Research and development can benefit greatly from the model and dataset that are produced. The primary goal is to demonstrate the effectiveness of our model's comprehension of Bangla music by ensuring that it performs well with previously unheard songs. Furthermore, this concept of genre categorization may be of great use to individuals in identifying their favorite genre, guaranteeing that they find comfort in the music they listen to. Our target is to achieve great accuracy in Bengali music genre categorization by using a CNN architecture, utilizing sophisticated feature engineering and generating a dataset.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

The goal of a literature review is to communicate comprehension of the current debates and research relevant to a particular subject or field of study in the form of a written report. By performing a literature review, we may deepen our understanding of our specific topic. A paper's related work section is a crucial element. As writers, we often have to help readers understand the context of the subject matter by connecting the dots between our current work and past works. It is very important to identify research gaps since they can direct future studies or close gaps in the body of existing literature.

2.2 Related Works

Towkir Ahmed and M. Arfayet Alam (2022) [1] used a deep neural network and five machine learning techniques to categorize the Bangla music dataset. For the deep neural network employed in the experiment, they created a sequential model. The authors have improved the performance of training supervised learning models by utilizing feature scaling and dimensionality reduction. The Bangla Music Dataset is the name of the dataset they used. The accuracy of their deep learning model is 77.68 percent. Their experiment's shortcoming is that their dataset was insufficiently large, which prevented their machine learning models from outperforming the deep learning model in terms of efficiency.

Hasib and Tanzim (2022) [2] suggest a model named BMNet-5 which is intended to determine the Bangla music genre by listening to it.. They attempted to assess BMNet-5's efficacy using a dataset of musical compositions in Bangla and a variety of performance measures. Additionally, explainable AI analysis of genre predictions is performed using SHAP. With an accuracy of 90.32%, their suggested model, BMNet-5, categorized 1742 Bengali songs into 6 genres. They were unable to see the unique feature rankings according to how effectively they functioned under different deformations. If the training set of the model could contain additional audio features, it may perform better.

The Sequence Model from the Keras toolkit is the neural network model that Al Mamun, Kadir and Rabby (2019) [3] suggested. They produced the dataset on their own. The Librosa library assisted in the extraction of several frequency and time domain properties from an audio stream, yielding numerical values as a result. The proposed NN Model has a 74% accuracy rate. Although the study uses two to three hundred songs per genre, more information on the diversity, distribution, and any biases in the dataset might improve the findings. If the model could account for more audio features, it might perform better.

Bhowmik and Chowdhury's (2019) [4] main goal was to categorize songs based on certain attributes. They suggested a novel machine learning approach. A dataset with twelve hundred Bangla songs served as the foundation for their work. The training set was classified and choices were derived using the four chosen classifiers. The continuity feature is a new feature that they unveiled. SVM offered the best accuracy. To increase classification accuracy, future studies may examine new features or sophisticated methods like deep learning. They can also look into classifying more than two genres in multi-class classification.

Parv and Dharaskar (2023) [5] attempted to look at the use of deep learning when it comes to the categorization of musical genres. Digital signal processing techniques were utilized to extract the acoustic qualities of music. They modeled and performed data analysis using the GTZAN dataset. Utilizing the Librosa Python Library, feature extraction and data preprocessing were carried out. When developing a model, each layer of a 2-dimensional CNN is created using the Keras library. Future studies will concentrate on improving the functioning of the system and apply transfer learning as a paradigm. This paper's weak point is its minimal explanation of CNN architecture options and feature engineering.

Because convolutional neural networks can extract relevant characteristics from audio input, they have been shown to be an excellent method for classifying music genres by Tarannum Shaikh and Ashish Jadhav (2022) [6]. This study presents a CNN model that attains an accuracy of up to 92.65%. The model optimizes performance using dropout and Adam's optimizer in addition to using spectrogram images as input. Building and successfully implementing the CNN model depend heavily on the Tensorflow framework. In later development, the initialization of weights and the incorporation of the full song instead of only the 30 second slice categorization could be altered.

A CNN-based genre classifier, offered by Nitin and Deka (2023) [7], is more accurate than previous neural network-based classifiers. For categorization, GTZAN was employed. The Python Librosa package is used to extract the feature vector following the splitting of the data samples. The suggested work uses supervised learning as a classification method. Accuracy results for training, validation, and testing are 97.43%, 78.64%, and 80.50%, respectively. The discrepancy in accuracy between training and validation data points to possible overfitting. Future research will examine how various musical genres affect the human brain.

CNN-RNN models were utilized by Xundong Ma (2023) [8] to categorize musical genres. The GTZAN dataset was used for data testing and training. The Librosa package in Python is used to carry out the full procedure in the background. TensorFlow and Keras are used to create the ANN model after audio features are extracted from the sample files using Librosa. With an accuracy of 80%, their model also clearly shows that CNN is the model that performed the best. The CNN-RNN model option would integrate the benefits of both models. Data visualizations would add a powerful element to the study.

Haitao Peng and Siwei Tan (2023) [9] provide a unique approach which utilizes ANN and CNN to solve the issue of effectively identifying music genres. They use MFCC to preprocess the data after utilizing the GTZAN dataset. The purpose of the study is to compare the performance of four artificial and convolutional neural network based on various feature inputs. ANN2 performs better than other models tested, achieving an amazing accuracy of 94.1. They come to the conclusion that CNN typically outperforms ANN when there are less features feeding into the neural network. Their next research will concentrate on analyzing models in greater detail and adding more layers to increase accuracy and efficiency.

To include the acoustic data, Abhilash Nandy and Manish Agrawal (2020) [10] created a CNN based feature extractor for spectrograms. For the lyrics, they created a CNN based feature extractor. Based on the fused feature vector that is produced, they categorize the music recording. Using the concatenated vector, a neural network-based classifier has performed the final classification. This paper makes use of The FMA Dataset. CNN and HAN(s) are the two types of algorithms they used. Three modules made up their model: a feature fusion module that translates the input to the appropriate musical genre, a convolutional neural network module for the spectrogram and a hierarchical attention module for the lyrics. They may have looked for greater accuracy in this work by using more algorithms.

A convolutional recurrent neural network was employed by Keunwoo Choi and Mark Sandler (2017) [11] for the purpose of music tagging. CRNNs use recurrent neural networks for temporal summarization of the collected features while convolutional neural networks are used for local feature extraction. In this work, the Million Song Dataset is employed. The top 50 tags were predicated by the networks during training. The authors of this article changed the number of parameters to produce comparison that was both memory-controlled and computation-controlled, which in turn changed the size of the networks.

Partha Ghosh and Soham Mahapatra (2023) [12] focused on figuring out whether group methods work better for categorizing musical genres than conventional algorithms. SVC was their baseline model, and it had an accuracy of 46%. They specified the models that followed to have an accuracy higher than that. In the test dataset, ANN scored 67%, whereas CNN scored 77%. This work makes use of several open-source datasets, including Million Song, FMA, CTZAN, and others. Based on the test results, the ultimate accuracy in this paper is 73%. This study employs a variety of methods. However, they are not very accurate. This is an issue, and they should work on making this paper more accurate.

By employing machine learning classifiers, Nandkishor Narkhede and Manish Dadhich (2023) [13] have developed a novel, rigorous approach to the categorization and identification of music genres. In this instance, the GTZAN dataset served as the source of input audio signals. They create a fresh, analytical viewpoint in order to choose the best machine-learning algorithm. They obtained the CNN findings, which showed a 98.9% accuracy level for the system and the accuracy is 99.9% from KNN. Future studies should expand the dataset to include novel, not tested stripes from throughout the globe. Certain stripes may behave differently with colored classifiers on their structure.

Qiuqiang Kong, Xiaohui Feng and Yanxiong Li (2014) [14] have developed an automated genre categorization system that may be applied to a number of fascinating challenges, including music suggestion. This study employs a variety of algorithms. ‘Classifying musical genres with particle swarm optimization and stacked ensemble’ is the first model in this work, while ‘Music genre using spectrogram’ is the second. GTZAN is the dataset that was used. Their model's average accuracy for this publication is 57.5%. In order to obtain better measurements, their future work will focus on fine-tuning the parameters to boost accuracy.

Classifying music genres is the first and most important stage in creating a powerful recommendation engine, according to Priyadharshini Jayadurga and Pratiba (2022) [15]. This study examined the two convolutional neural network deep learning models, Alex-net and Res-net, for the categorization of musical genres using training are employed in this study to categorize the data. The GTZAN datasets were used to evaluate the previously described models. The findings revealed that the res-net model performed less than Alex-net. The Alex-net model, which predicts the music genre just using the spectrogram as an input, has the greatest accuracy.

According to Ting Bian and Minglai Yang (2023) [16], accurately classifying music genres by automated categorization is a common pattern recognition challenge, and several methods have been put forth to achieve this goal. The GTZAN dataset is used in this work. Pattern recognition techniques were extensively employed as baseline approaches in this article. Sometimes the STFT spectrogram can outperform the Mel spectrogram. Overall, it's a well-behaved characteristic. The LGNet-based automated categorization system performs satisfactorily overall. Some shortcomings exist in this study. When it comes to handling drivers, the current approaches are inadequate.

By utilizing convolutional neural networks, Vishnupriya and Meenakshi (2018) [17] have automatically classified music genres. Certain drawbacks, such as computing costs and feature engineering complexity are found while using machine learning approaches by the authors. The most often utilized datasets in the proposed system are the Million Song Dataset and GTZAN. The process of classifying music genres consists of two fundamental steps: feature extraction and categorization. The suggested model makes advantage of supervised learning. One thousand tracks with ten labels are entered. Examining CNN's performance on bigger, more varied datasets and merging CNN with different modalities are research gaps.

Open set recognition has been investigated by Zhanlin Liu and Julien De Mori [18] for the purpose of classifying musical genres. Beyond conventional closed-set situations, OSR seeks to find genres, both well-known and unfamiliar. This research offers an OSR-based method for classifying musical genres and assesses its performance using the GTZAN and FMA datasets. This paper's research gaps include improving model interpretability, going beyond binary prediction, and optimizing the accuracy trade-off. This work aims to enhance the performance and generalizability of OSR. This paper's primary contribution is the clarification of the classification scheme for music.

In order to overcome the shortcomings of previous convolutional neural network structures used for the job, Caifeng Liu, Lin Feng (2021) [19] have introduced an innovative Bottom-up Broadcast Neural Network architecture for the classification of musical genres. Mel-spectrograms are used as the input for the proposed BBNN, which is assessed on three benchmark datasets: GTZAN, Ballroom, and Extended Ballroom. Notably, BBNN outperforms the most recent models. The study highlights difficulties, such as building strong CNN models with little labeled data, and makes recommendations for future research, such as investigating new acoustic variables and utilizing novel distance metric techniques to enhance the calculation of genre similarity.

Jash Mehta and Deep Gandhi (2021) [20] have proposed a log-based MEL spectrogram for the categorization of musical genres utilizing various algorithms including transfer learning. Almost all techniques are feature-based, with few exceptions, and Even while several models use spectrogram techniques, the transfer learning approach continues to be unusual. In this article, Resnet34 achieved the highest accuracy. As a result, the authors see a need to strengthen the conclusions by conducting further tests and improving on the stipulated accuracy.

2.3 Comparison between existing works

Table 2.3.1. Comparative analysis with previous work

SL No	**Author Name**	**Used Algorithm**	**Best Accuracy with Algorithm**
1.	T. Ahmed, M. A. Alam, R. R. Paul [1]	SVM-Linear, K-Nearest Neighbors, SVM-RBF, Stochastic Gradient Descent, random forest, Deep Neural Networks	Deep Neural Networks = 77.68%
2.	Hasib, K.M., Tanzim, A., Shin, J., Faruk, K.O. [2]	Logistic Regression, Shallow Neural Network, CNN, Proposed BMNet-5 Model	BMNet-5 = 90.32%.
3.	Al Mamun, M.A., Kadir [3]	Linear Regression, k-NN, Logistic Regression, SVM, Proposed NN Model	Neural network = 74%
4.	Bhowmik, A. and Chowdhury, A.E. [4]	Decision Tree, Support Vector Machine, Naive Bayes, and K Nearest Neighbors	SVM = 93%
5.	Bhowmik, A. and Chowdhury, A.E. [4]	Logistic regression, random forest, gradient boosting, support vectors, Dense Neural Network, CNN	CNN=76%
6.	Rawat, Parv & Dharaskar [5]	Stochastic Gradient Descent, Pelchat, Gelowitz, Naïve Bayes, Despois, CNN	CNN= 92.65%

7.	Choudhury, Nitin & Deka [7]	CNNs, LSTMs	CNN=80.50%
8.	Ma, X. [8]	CNN, ANN, Random Forest Classifier, KNN	CNN=80%
9.	Li, Z., Peng [9]	Convolutional Neural Network, Artificial Neural Network	ANN=94.1%
10.	Agrawal, M. & Nandy, A. [10]	Convolutional Neural Network, Hierarchical Attention Networks	CNN= 76.2%
11.	Choi, K., Fazekas, G., Sandler, M. [11]	CNNs, CRNN	CRNN=87%
12.	Ghosh, P., Mahapatra, S., Jana, S. [12]	Support Vector Classifier, Logistic Regression, Convolutional Neural Networks, Convolution-Recurrent Neural Network, Artificial Neural Network, g Ensemble technique of Ada Boost.	CRNN=73%
13.	Narkhede, N., Mathur, S., Bhaskar [13]	Support vector machine, k-nearest neighbor, Convolutional Neural Network and Probabilistic Neural Network.	KNN = 99.9%
14.	Kong, Q., Feng, X., & Li [14]	KNN, Decision trees, Random forest, Naive Bayes and SVM.	SVM= 67.2%

15.	Shana, J & Jayadurga [15]	Alex-net model, k-nearest neighbors, Gaussians model	Alex-net model =80.54%%
16.	Bian, T., & Yang [16]	Random forests, Naive Bayes classifiers and support vector machines, CNN, KNN, ResNet.FCN, LGNet-2MLP ,LGNet-4MLP	LGNet-4MLP=82.71%
17.	Vishnupriya, S., & Meenakshi, K. [17]	Convolution Neural Networks, Mel Frequency Cepstral Coefficients, Deep Neural Network	CNN =76%
18.	Liu, K. [18]	OSR, Support Vector Machines, Nearest Neigh bour classifiers and Sparse Representation	OSR=72.4%
19.	Feng, L. [19]	Bottom-up Broadcast Neural Network	BBNN=97.2%
20.	Deep & Thakur [20]	AlexNet, VGG, ResNets, Resent34, Resent 50, VGG16	Resnet34=79%

[1] used the “Bangla Music Dataset” with 1742 songs of 6 categories. [2] and [3] generated their own datasets. [4] worked with a dataset of 1200 Bangla songs. [5] [13] [14] [15] [19] [20] used the GTZAN dataset. [6] and [7] utilized datasets like GTZAN and MusicNet. [10] worked with the FMA (Free Music Archive) Dataset. [11] utilized the Million Song dataset. [12] used multiple datasets like GTZAN, FMA, Million Song, etc. [16] worked on the GTZAN dataset with 1000 tracks. [17] used the GTZAN and Million Song Dataset.[18] utilized the GTZAN and FMA datasets. [19] utilizes GTZAN, Ballroom, and Extended Ballroom. Most papers used the GTZAN dataset (10 genres, 1000 samples) or self-

generated datasets. Some studies explored larger and more diverse datasets like FMA (millions of songs) and Million Song Dataset. Various studies employed algorithms like SVM, k-nearest neighbors, Naïve Bayes, Decision trees, Random forest and more. CNNs and RNNs, were prominently used across several works. Sequential models from the Keras library and CNN-RNN combinations were explored. Specific models like BMNet-5, CRNN, CNN-RNN, and HAN(s) were proposed in certain papers. Some studies, like [16], experimented with multiple techniques, including bidirectional LSTM, FCN-LSTM, and residual networks. Common features included mfcc, chroma features, zero crossing rate, spectral centroid and spectrogram images. Some works included additional features like continuity features or lyrics analysis.

[13] achieved remarkable accuracy with CNNs but used a small dataset. [10] combined CNNs with Hierarchical Attention Network for lyrics and spectrograms, but the accuracy was lower than expected. [12] compared various algorithms and found CRNN to be the most accurate, while ensemble techniques further improved performance. [15] compared AlexNet and ResNet CNNs, with AlexNet achieving higher accuracy. [18] introduced OSR for music genre classification, opening up new research avenues.

ML algorithms like random forest, SVM and k-nearest neighbors were used in earlier studies, achieving accuracies around 70%. CNN and convolutional RNN dominated recent works, achieving accuracies up to 98.9%. CNNs consistently outperformed other models, with the highest accuracy achieved by LGNet at 82.71%. Combining CNNs with other modalities like RNNs or lyrics analysis showed promising results in some studies.

Some studies, like [10], acknowledged challenges like imbalanced data in the FMA dataset. Issues related to feature extraction from audio signals, including ZCR, MFCC, Spectral Centroid, etc., were discussed in multiple works. [7], [11], and [18] specifically addressed the complexities of large and diverse datasets, computational costs, and optimizing model performance. Several works pointed out dataset limitations, such as size, diversity, and potential biases. Feature engineering and model architecture choices were identified as weaknesses in some studies. Many authors highlighted the need for improving model accuracy, exploring additional features, and incorporating advanced techniques like transfer learning, ensemble methods, or hybrid models. Visualization and interpretability were suggested in a few papers to provide a more intuitive understanding of results. The collective findings highlight ongoing efforts to enhance classification accuracy, explore diverse datasets, and develop more robust and efficient models for music genre classification.

2.4 Open Issues

The project on classifying Bangla music genres using machine learning models faces several open issues. Firstly, despite the dataset being relatively comprehensive, the diversity and representation of different Bangla music genres may still be limited. Ensuring a balanced and extensive dataset is crucial for improving model training and generalization. Additionally, the current feature extraction techniques, although effective, may not capture all the nuances and characteristics of Bangla music genres. Exploring additional or alternative audio features could potentially enhance classification performance. Another challenge is the models' performance on completely unseen and diverse music samples. Ensuring robust generalization to new data remains an ongoing concern. Furthermore, training advanced deep learning models such as CNNs requires significant computational power, making access to high-performance hardware and managing computational resources efficiently a continuing challenge. Optimal hyperparameter settings are crucial for the performance of machine learning models. However, the process of hyperparameter tuning is often time-consuming and computationally expensive, requiring further exploration and automation. As Bangla music continues to evolve, new genres may emerge that are not covered in the current classification system. Updating the system to accommodate new genres while maintaining high accuracy presents an ongoing issue. Ensuring that the dataset is used ethically and that the classification system does not reinforce cultural insensitivities is a critical and ongoing concern. Finally, evaluating the system's performance against human music experts to understand the gaps and limitations in automated classification remains an important area for future exploration.

2.5 Summary

These research projects contribute to the improvement of music genre classification by investigating various approaches, algorithms, and datasets. Each study outlines strengths, limits, and prospective topics for further research, with the goal of improving music genre classification systems' accuracy and efficiency.

CHAPTER 3

METHODOLOGY/ REQUIREMENT ANALYSIS & DESIGN SPECIFICATION

3.1 Overview

The methodology for this project involves several key steps to classify Bangla music genres using machine learning models. Initially, a comprehensive dataset of Bangla music tracks is collected, ensuring a diverse representation of various genres. Each track undergoes a preprocessing phase, which includes feature extraction to capture relevant audio characteristics. Different ML techniques, including KNN, SVM, Random forest Classifier, Gradient boosting classifier, and CNN, are then implemented to classify the music tracks.

Using the preprocessed dataset, the models are trained and verified, and their efficacy is assessed using performance measures including accuracy, precision, recall, and F1-score. The performance of each model is optimized by hyperparameter adjustment. Because of its deep learning capabilities, the CNN goes through additional training stages to optimize its parameters and improve the accuracy of genre categorization. The whole approach seeks to create a reliable classification system that can recognize and classify different genres of Bangla music.

3.2 Proposed Methodology

Gathering appropriately categorized music datasets by genre, preprocessing the data, and extracting feature vectors from the preprocessed data are the steps in our proposed work's methodology for the suggested study. Next is the development and classification of the model, followed by the model's implementation. The flow diagram for our project is given below (Figure 3.2.1):

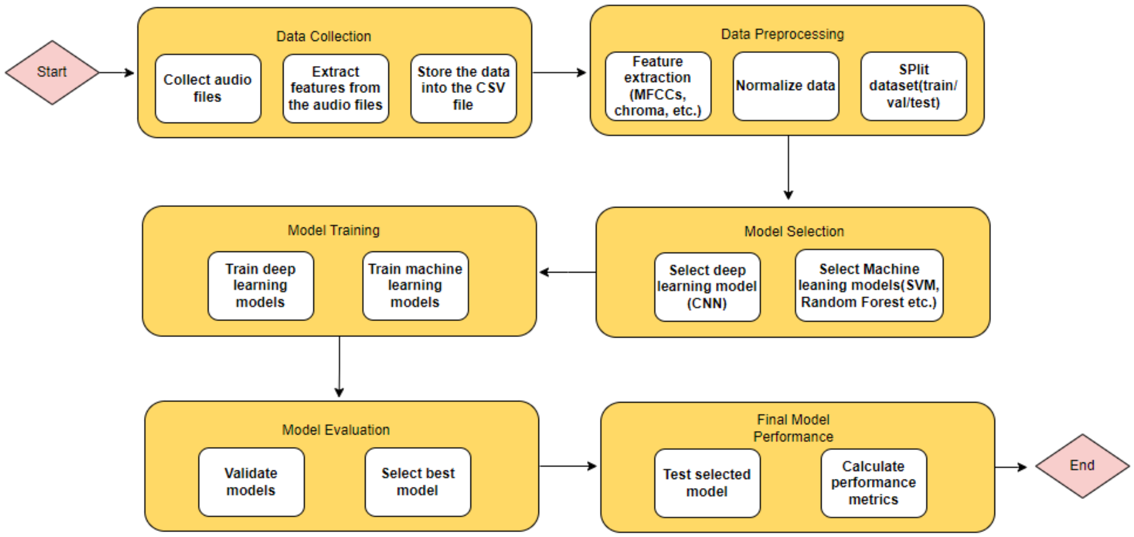

Figure 3.2.1. Flowchart of our proposed methodology.

Data Collection: We will be using the Bangla Music Dataset[24], which is publicly available on Kaggle for using on research, especially for music genre classification purpose. AFIF AL MAMUN built it by extracting different time & frequency domain features of MP3 files. Lots of Bangla music of different genres has been published in the last four years. So, we have extracted the data from those audio files and update the existing csv file. 1742 Bangla songs that are tagged belong to six types in this dataset: Bangla adhunik, bangla hip-hop, Nazrul Geeti, bangla band music, Rabindra sangeet and Palligeeti. The following properties are included in every entry: tempo, rmse, spectral bandwidth, mfcc, delta, spectral centroid, spectral rolloff, zero crossing and chroma frequency. Here the dataset link_ datasets/afifaniks/bangla-music-dataset

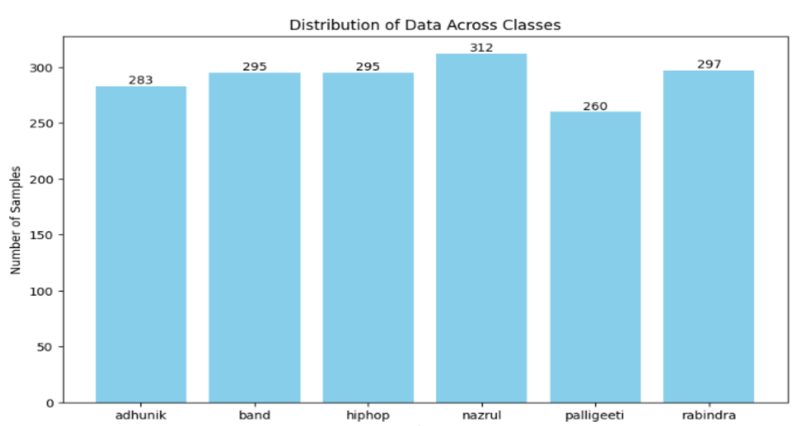

Figure 3.2.2: Distribution of data Across Classes

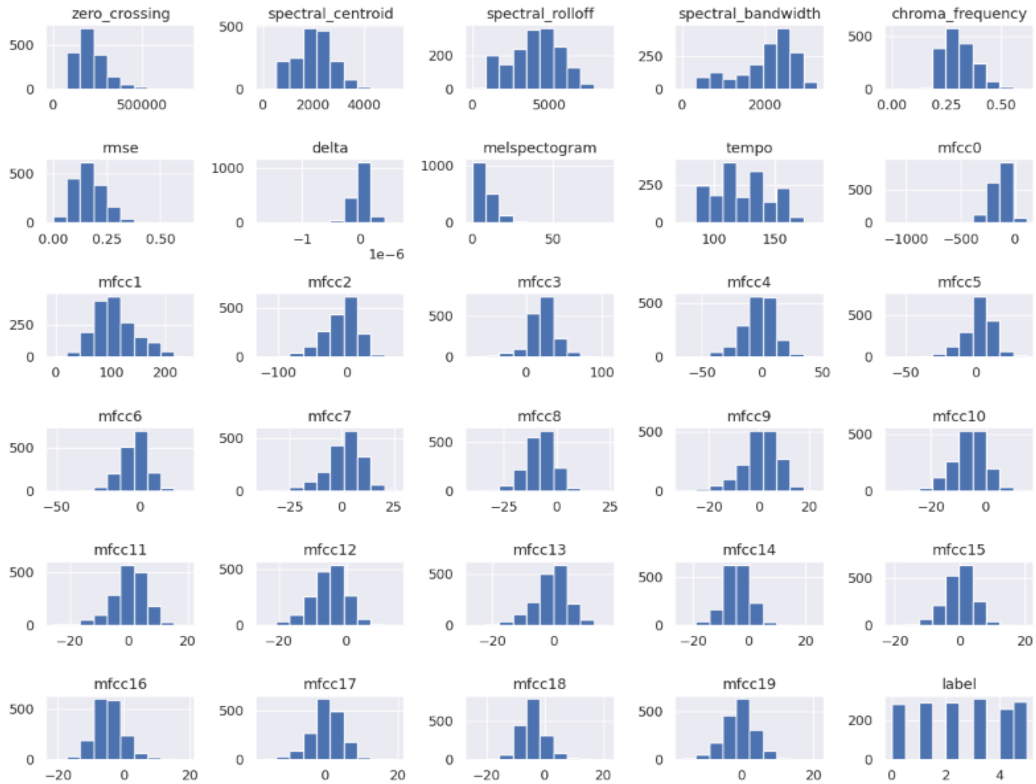

Figure 3.2.3: Plot Histogram of each Data

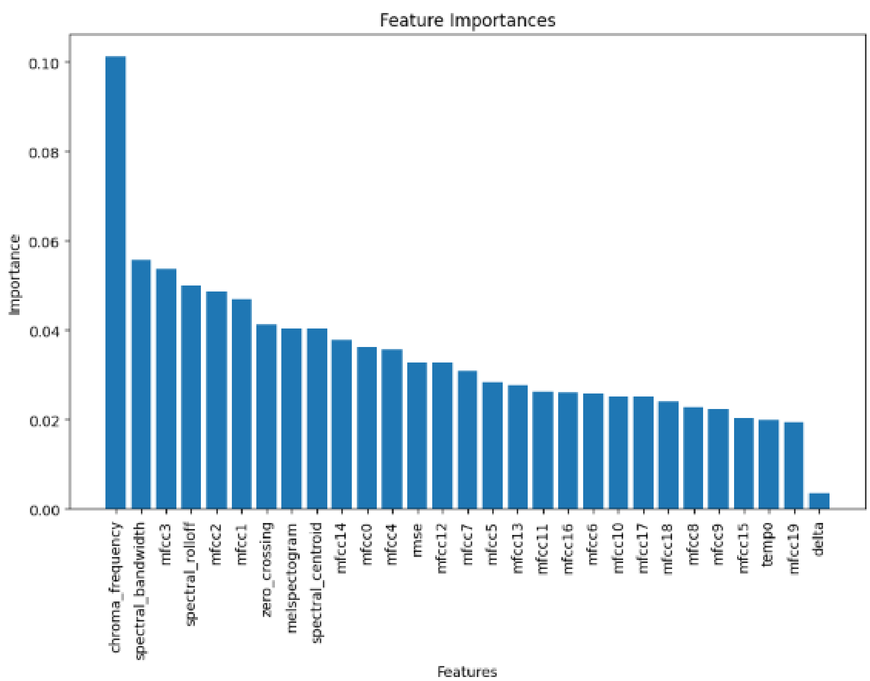

Figure 3.2.4: Feature Importance

Feature Extraction: We extracted a comprehensive set of features that are critical for classification. These features include tempo, rmsse, spectral bandwidth, MFCC, delta, spectral centroid, spectral rolloff, zero crossing rate, and chroma frequencies. These characteristics were extracted using the Librosa package, which allowed us to produce a large dataset for model training.

Data Splitting: We divided the dataset into training and validation sets using the scikit-learn train_test_split function in order to prepare it for training and assessment. A random seed (random_state=22) was used to ensure that the split is repeatable. We put aside 75% of the data for training and 25% for testing.

Model Training: For the classification task, we trained both traditional machine learning models and a deep learning model. The machine learning models include KNN, SVM, random forest, and gradient boosting classifier. These models were chosen for their robustness and effectiveness in classification tasks.

For the deep learning model, we designed a CNN specifically for audio classification. Convolutional layers are used in CNN design for feature extraction, pooling layers are used to reduce dimensionality, and dense layers are used for classification. We employed regularization techniques like dropout to prevent overfitting. The model was assembled using the Adam optimizer. In order to dynamically modify the learning rate based on validation loss, we also created a 'ReduceLROnPlateau' function. By processing the training data in batches of size 80 over 200 epochs, the CNN model was trained using the fit technique. Validation data was used to monitor performance and adjust the learning rate as needed.

Classification and Evaluation: Metrics including accuracy, precision, recall, and F1-score were employed to assess the effectiveness of our models. We employed cross-validation and confusion matrices to provide a robust assessment. By contrasting the outcomes of the CNN with those of conventional machine learning models, we determined the most effective approach for Bengali music genre classification. This thorough assessment contributes to ensuring that the chosen model is robust and accurate, able to classify data reliably in practical applications.

3.3 Hardware/ Software Requirement

The Implementing an effective Bengali music genre classification system using advanced ML and deep learning techniques necessitates a comprehensive set of requirements across various domains, including data acquisition, software, hardware, and development tools.

A. Software Requirements:

- **Programming Languages:** Python for implementing machine learning models, data preprocessing, and feature extraction.
- **Libraries and Frameworks:**
- **Librosa:** For audio processing and feature extraction.
- **Scikit-learn:** For traditional ML algorithms (SVM, KNN, Random Forest, Gradient Boosting).
- **Keras and TensorFlow/PyTorch:** For developing and training deep learning models, particularly CNNs.
- **NumPy and Pandas:** For data manipulation and analysis.
- **Matplotlib and Seaborn:** For data visualization and plotting.
- **Development Environment:** An Integrated Development Environment (IDE) such as Jupyter Notebook, PyCharm, Google Colab or Kagle.

B. Hardware Requirements:

- **Computing Power:** High-performance computing resources, including GPUs to accelerate the training of deep learning models. Cloud-based solutions such as Google Colab is used for scalable computing resources.
- **Storage:** Sufficient storage to accommodate large datasets, including raw audio files and extracted features. This includes both local and cloud storage solutions.

3.4 Project Management and Financial Analysis

Project Management:

This project's technique uses machine learning models to categorize Bangla music genres through a number of crucial processes. Initially, a comprehensive dataset of Bangla music tracks is collected, ensuring a diverse representation of various genres. Each track undergoes a preprocessing phase, which includes feature extraction to capture relevant audio characteristics. Different machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest Classifier, Gradient Boosting Classifier, and Convolutional Neural Networks (CNN), are then implemented to classify the music tracks.

Finance:

Table 3.4.1: Cost for Our Model.

SN	Components	Cost (BDT)
01.	Hardware/Infrastructure	45000-50000
02.	Visiting Stakeholders	2000-3000
03.	Software and Tools	1500-2000
04.	Data Collection and Processing	3000-4000
06.	Documentation and Report Writing	1500-2000
07.	Contingency (10% of total)	5500-6000
Total Cost:		58,500-67,000/-

3.5 Summary

This project employs a structured methodology to classify Bangla music genres using machine learning techniques. The process begins with data collection from a publicly available Bangla music dataset, followed by preprocessing steps and feature extraction to obtain relevant audio features. Multiple machine learning models, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest Classifier, Gradient Boosting Classifier, and Convolutional Neural Networks (CNN), are then trained and validated on this dataset. Performance metrics like accuracy, precision, recall, and F1-score are used to evaluate and compare the models. Hyperparameter tuning is applied to optimize each model's performance. The CNN model undergoes additional training phases to further improve its classification accuracy. This comprehensive approach aims to develop an effective and accurate system for classifying Bangla music genres.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

The implementation of the Bangla music genre classification project involved several key steps. First, the Bangla Music Dataset from Kaggle, containing 1742 songs across six genres, was used. Preprocessing steps are performed on the audio files. Data augmentation techniques were applied to increase dataset diversity. Feature extraction involved obtaining various audio features like tempo, MFCC, and spectral properties. Multiple machine learning models, including KNN, SVM, Random Forest, Gradient Boosting, and CNN, were trained and evaluated using an 75-25 train-test split. Performance was assessed through accuracy, precision, recall, and F1-score metrics. Hyperparameter tuning was performed to optimize the models. The CNN model demonstrated the highest accuracy and robustness, becoming the final chosen model. Python, along with libraries like TensorFlow, Keras, Scikit-Learn, and LibROSA, facilitated the implementation. This structured approach ensured the development of a reliable Bangla music genre classification system.

4.2 Train Model/ Prototype Design

Librosa python Library: Librosa [22] is a Python library that helps in sound and music analysis, allowing programmers to create algorithms for analyzing sound and music document designs. Automatic speech recognition is the main application of this Python library for speech analysis and music, when dealing with audio data, for example in the music field. The library can handle both simple and complex sound and music processing tasks, and is easy to use. Librosa uses various character processing techniques to break down components and aid in the visualization of tonal characters. Fast Fourier transform techniques are used by the Librosa program to analyze the audio input in the time domain, and then internally modify the spectrogram in the frequency domain.

Support Vector Machine: Regression using a highly effective machine learning approach, anomaly detection, and linear and nonlinear classification is the Support Vector Machine [32]. SVMs are known for their adaptability and are employed in many different fields, including as handwriting recognition and spam filtering, text and facial recognition, gene expression analysis, picture classification, and anomaly detection. Their adaptability and efficiency are demonstrated by their proficiency with handling high-dimensional data and capturing nonlinear correlations. The SVM algorithm is very good at finding the optimal separating hyperplane between classes in the target feature space. This unique property contributes to its effectiveness in a wide range of applications.

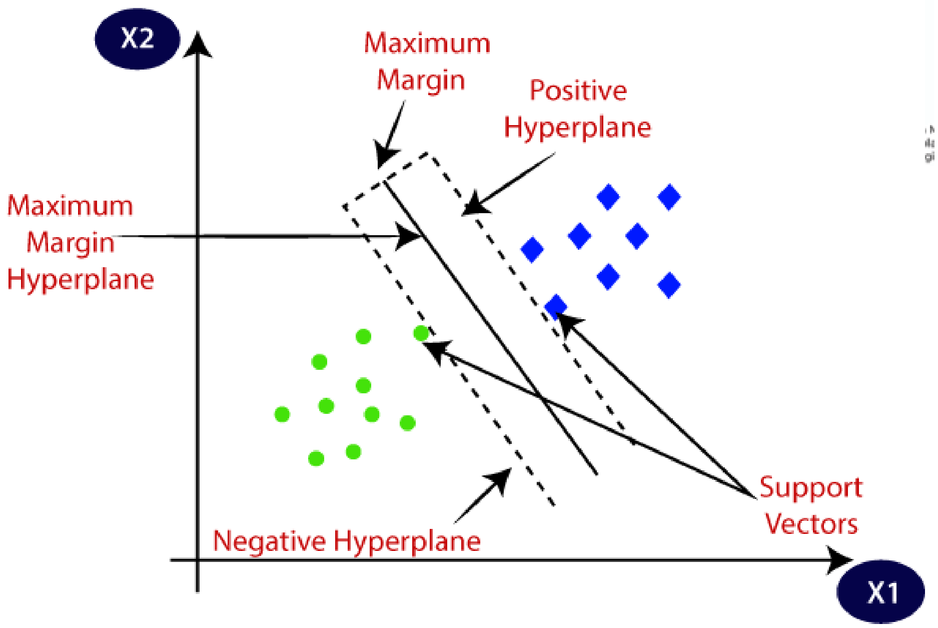

Figure 4.2.1: Architecture of SVM [33].

K-Nearest Neighbors: The K-NN algorithm [34] is founded on the idea of learning via supervision and is a simple machine learning technique. Based on the presumption that the newly acquired data and the current data are comparable, it places new data into the category that most closely matches the existing data. The K-NN method classifies new data points by comparing them with the current dataset, while retaining all the data it has access to. This makes it easy to immediately classify newly discovered data into relevant categories. K-NN is mostly used for classification issues, while it is capable of handling regression jobs as well. The fact that K-NN is referred regarded as a non-parametric technique. That is, one that makes no assumptions about the underlying data structure, must be noted. Lazy learning algorithms, like K-NN, wait to learn from the set of training data until the classification step. Instead, it employs a dataset that has been kept for use in the training phase to categorize newly added data points during the classification phase. In summary, K-NN relies on data similarity rather than assumptions about the distribution of the dataset, making it a flexible tool for classification problems.

.

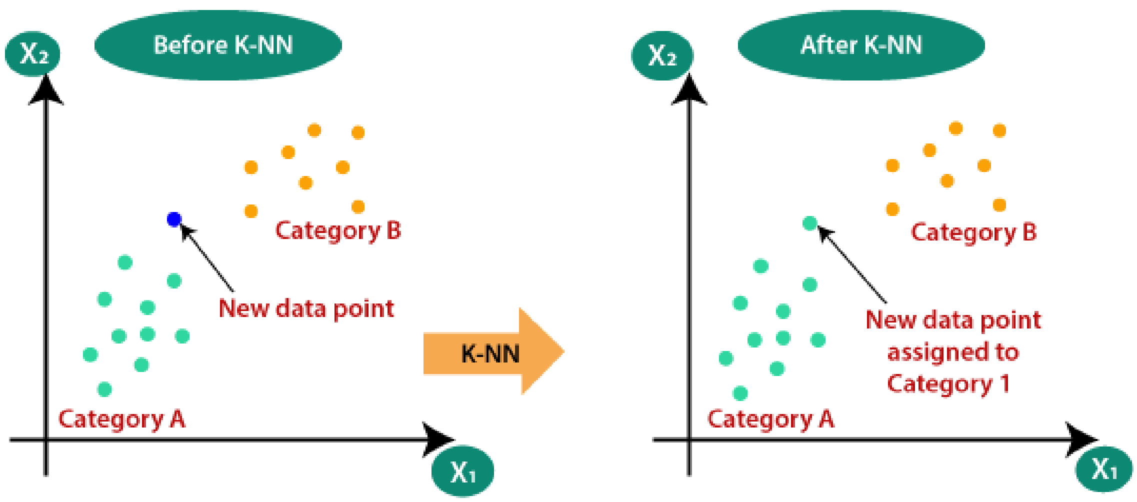

Figure 4.2.2: Visualization of the working phase of KNN Algorithm [35].

Random Forest:

For classification problems, one well-liked ensemble learning method is the Random Forest classifier [36]. During the training phase, it generates a number of decision trees, and the final class is chosen by taking the majority vote of the classes that these trees predicted. When compared to use a single decision tree, this method improves accuracy and lessens overfitting. A randomly chosen subset of the training data is used to train each tree in the forest, with features being selected at random for each split. By increasing the variation amongst the trees, this randomization strengthens the model's resilience and expands its applicability. The Random Forest classifier's capacity to handle big datasets with lots of dimensions is one of its main advantages. Despite its complexity, the algorithm is relatively straightforward to use and adapt, making it a popular choice for various classification problems.

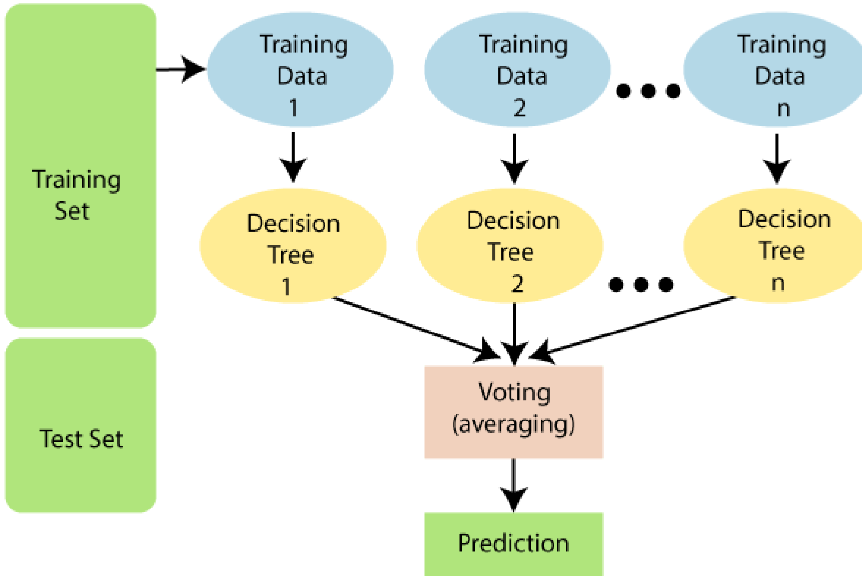

Figure 4.2.3: Working process of Random Forest [37].

Gradient Boosting:

A commonly used machine learning method in regression and classification applications is gradient boosting [38]. It involves building a robust predictive model by integrating multiple weak learners, usually decision trees. First, a base model is trained on the data, and then subsequent models are gradually improved by adapting to the errors of previous models. To successfully learn from the mistakes of its predecessors, each new model aims to minimize these mistakes. The basic idea of gradient boosting is to use a gradient that indicates the direction and progress required at each step. By iteratively minimizing these gradients, the algorithm gradually builds a stronger model that can accurately predict the target variable. Overall, gradient boosting is an iterative optimization technique that combines and leverages the knowledge of multiple weak learners to create highly accurate predictive models.

Convolutional neural networks (CNNs):

A popular kind of artificial neural network for object identification and classification tasks is the CNN. They are classified as deep learning because they are designed to train computers to recognize patterns in data. CNNs do exceptionally well while processing three-dimensional inputs, in contrast to standard neural networks. CNNs consist primarily of convolutional layers, which perform most of the processing. These layers use filters, or kernels, to extract information through a process called convolution. This convolution produces a feature map that shows interesting sections of the element. A pooling layer, which lowers the dimensionality of the feature map and boosts the network's efficiency, comes after the convolutional layer. Max pooling and average pooling are popular pooling techniques that aggregate data from nearby pixels.

Ultimately, a completely linked layer establishes connections between every neuron in one layer and every other layer, allowing the network to make predictions based on data from the previous layer. To create a probability distribution over all potential classes, the output of a fully linked layer is sometimes subjected to a softmax function. CNNs have revolutionized the area of computer vision with their capacity to do tasks including object detection, categorization, face recognition, and picture identification. They are applicable to many other businesses, such as marketing, healthcare, retail, and automotive. During their development, numerous CNN designs have been published, each suited to a specific set of datasets and applications. In conclusion, CNNs are widely used in modern technology and are useful tools for processing visual input.

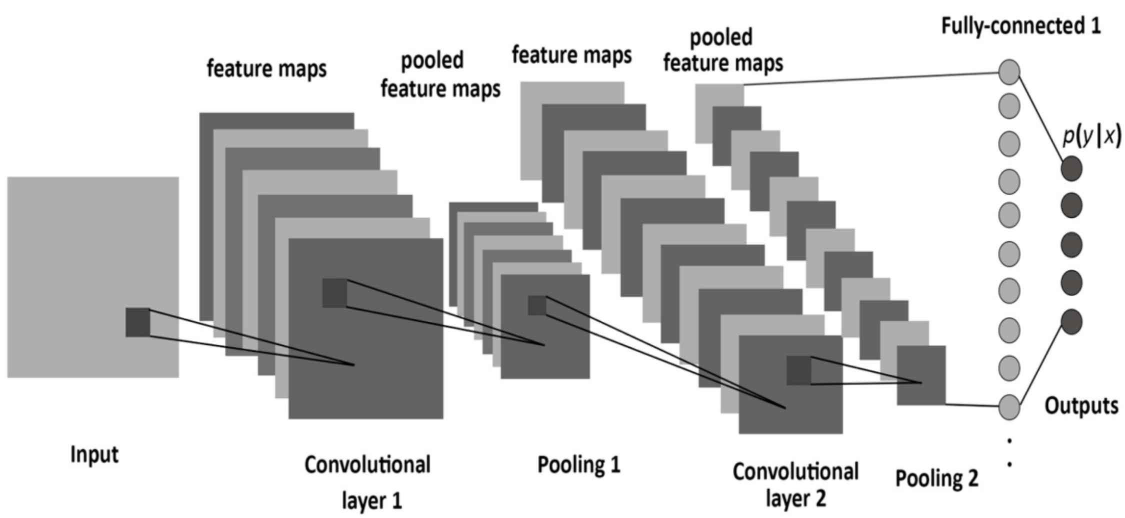

Figure 4.2.4: The structure of a CNN, consisting of convolutional, pooling, and fully-connected layers [40].

Training:

This step involves creating and training machine learning models. The dataset was divided into sets for training and validation. Several models were trained and evaluated for their classification accuracy, including Convolutional Neural Networks (CNN) and traditional machine learning algorithms such as K nearest Neighbour, Support Vector Machines (SVM), Gradient Boosting Classifier algorithm and Random Forest.

X_pca and y are the datasets that we are splitting into preparation and testing subsets using sklearn's train_test_split function. In machine learning, this method is essential for evaluating display execution. Test_size=0.25 selects 25% of the data for testing and reserves the remaining 75% for preparation. The reproducibility of the portion is ensured by the random_state=22 setting.

Next, training and testing are done using machine learning techniques such as KNN, SVM, Random forest classifier, and Gradient boosting classifier. Using Keras, a neural network model is trained via the `train_model} function. Firstly, a learning rate of 0.001 is used to initialize the Adam optimizer. Next, this optimizer is used to build the model, with "accuracy" serving as the training measure and "sparse_categorical_crossentropy" as the loss function.

Furthermore, depending on validation loss, a 'ReduceLROnPlateau' callback is integrated to dynamically adjust the learning rate during training. In order to ensure effective optimization, this callback keeps an eye on the validation loss measure and, if no progress is shown after five consecutive epochs, lowers the learning rate by a factor of 0.2..

The model is subsequently trained using the 'fit' method, which processes the training data in batches of size 80 over 200 epochs. The model's performance during training is verified using the validation data, which consists of X_test and y_test. The 'reduce_lr' callback is included as a parameter to adjust the learning rate dynamically, thereby optimizing training dynamics and enhancing the model's convergence towards optimal performance.

4.3 System Testing/ Model Evaluation

The evaluation of the machine learning and deep learning models for the classification of Bangla music genres involved several key metrics: accuracy, precision, recall, and F1-score. These metrics provided a comprehensive understanding of each model's performance.

Accuracy: The number of accurate forecasts to all predictions is the ratio of accuracy [28]. The model is assessed using this most basic of metrics. Accuracy can have a downside, though. On an unbalanced dataset, it is unable to operate effectively. Let us say that a model determines that the bulk of the data fits into the main class label. It produces results with improved precision. Still, the model performs poorly overall and struggles to categorize on small class labels. Here is the formula provided by:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN)$$

The results showed that K nearest Neighbour achieved a training accuracy of 68% and a model accuracy of 61.5%, Support Vector Machine achieved a training accuracy of 72.4% and a model accuracy of 67.4%, Random Forest Classifier achieved a training accuracy of 99.9% and a model accuracy of 67.7%, Gradient Boosting Classifier achieved a training accuracy of 99.2% and a model accuracy of 67.2%, and CNN achieved a training accuracy of 92.49% and a model accuracy of 74.91%.

Precision, Recall, and F1-score:

Precision: The ratio of genuine positives to the total of true positives and false positives is known as precision [28]. In essence, it examines the optimistic forecasts. The disadvantage of Precision is that it ignores both true and false negatives. The equation is provided by:

$$\text{Precision} = TP/(TP+FP)$$

Recall: The ratio of true positives to the total of true positives and false negatives is known as recall [28]. In essence, it examines how many accurate positive samples there are. Recall's disadvantage is that it frequently raises the false positive rate. The following formula is provided:

$$\text{Recall} = TP/(TP+FN)$$

F1-score: The harmonic mean of recall and accuracy is the F1 score [28]. It has been observed that in the accuracy-recall trade-off, recall falls as precision rises and vice versa. The F1 score aims to balance recall and accuracy. The following formula is provided.:

$$\text{F1 score} = (2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$$

K nearest Neighbour showed variable performance, with high recall in the Hiphop genre (0.81) but lower precision in the Palligeeti genre (0.67), resulting in an overall accuracy of 61%. Support Vector Machine demonstrated higher precision and recall, particularly in the Band genre (precision of 0.83), with an overall accuracy of 67%. Random Forest Classifier showed strong performance in the Nazrul genre (recall of 0.82), achieving an overall accuracy of 68%. Gradient Boosting Classifier maintained consistent precision and recall across genres, resulting in an overall accuracy of 67%. CNN outperformed other models with the highest recall in the Nazrul genre (0.92) and high precision in the Band genre (0.82), achieving the highest overall accuracy of 75%.

Confusion Matrix:

An N x N matrix, where N is the number of target classes, is called a confusion matrix [28]. It shows the quantity of both the current and anticipated outputs. The following are some terms that are used in the matrix:

- True Positives: TP is another name for it. It is the result where both the expected and actual numbers are YES.
- True Negatives: Another name for it is TN. It is the result if there is a difference between the expected and actual numbers.
- False Positives: FP is another name for it. This is the result where the expected answer is YES but the actual value is NO.
- False Negatives: FN is another name for it. This is the result where the expected answer is NO but the actual value is YES.

4.4 Summary

The implementation of this project utilized high-performance hardware, including robust CPUs and GPUs, to support the computational demands of deep learning models. Transfer learning was applied using pre-trained CNN architectures, fine-tuned for classifying Bangla music genres. Various machine learning algorithms, including K nearest Neighbour, Support Vector Machine, Random Forest, Gradient Boosting, and CNN, were implemented using Python and relevant libraries such as TensorFlow and PyTorch. The models were trained, evaluated, and compared based on their performance metrics to identify the most effective approach.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

The results of the project showed varying levels of accuracy among the different models tested for classifying Bangla music genres. The CNN outperformed traditional machine learning techniques, achieving the highest accuracy of 75.91%. Detailed analysis of precision, recall, and F1-score for each genre indicated that CNNs provided superior classification performance across most categories. Traditional models like SVM and Random forest also performed reasonably well, but did not match the CNN's overall effectiveness. The analysis confirmed that deep learning approaches, particularly CNNs, are more adept at capturing complex audio features, leading to better genre classification.

5.2 Experimental/ Simulation Result

Table 5.2.1: Accuracy for Classification of Machine Leaning Techniques and CNN Model

Architecture	**Training Accuracy**	**Model Accuracy**
K nearest Neighbour	**68 %**	**61.5%**
Support Vector Machine	**72.4%**	**67.4%**
Random Forest Classifier	**99.9%**	**67.7%**
Gradient Boosting Classifier	**99.2%**	**67.2%**
CNN	**92.49%**	**74.91%**

The experiment's findings show how well different deep learning and machine learning models categorize different genres of Bangla music. To assess each model and determine how effective each strategy was, the test set's training accuracy and overall model accuracy were taken into consideration. A model accuracy of 61.5% and a training accuracy of 68% were attained with the K-Nearest Neighbour (KNN) technique. The model performed moderately on the training set, but its accuracy dropped on the test set, suggesting that it had trouble generalizing to new data. A superior training accuracy of 72.4% and a model accuracy of 67.4% were demonstrated by the Support Vector Machine (SVM). The fact that SVM outperforms KNN indicates that it can better grasp the intricate patterns seen in the music data, even though test and training results still differ noticeably. With nearly flawless training accuracy of 99.9% on the training set, the Random Forest Classifier showed remarkable performance. On the other hand, overfitting—the model doing remarkably well on the training data but failing to generalize well to fresh, unknown data—was indicated by the test set's 67.7% model accuracy. Likewise, the Gradient Boosting Classifier demonstrated a model accuracy of 67.2% but a high training accuracy of 99.2%. This shows that although the model performed very well on the test set due to overfitting, it also learnt the training data quite well. In this experiment, the Convolutional Neural Network (CNN) proved to be the most efficient and well-rounded model. On the test set, it produced a substantially better model accuracy of 74.91%, compared to a training accuracy of 92.49%. The CNN was able to generalize better and adjust to the heterogeneity in the Bangla music genres, as seen by the comparatively lower gap between the training and test accuracies. In conclusion, the CNN outperformed conventional machine learning models in terms of both training and test accuracies, whereas models such as KNN, SVM, Random Forest, and Gradient Boosting displayed differing degrees of performance. This shows that since deep learning models can identify complicated patterns in the data, especially CNNs, are better suited for challenging audio categorization tasks.

TABLE 5.2.2: Precision, Recall, F1-Score, and Support for Machine Learning Models and CNN.

K nearest Neighbour				
	Precision	**Recall**	**F1-Score**	**Support**
Adhunik	0.51	0.71	0.59	82
Band	0.70	0.62	0.66	74
Hiphop	0.66	0.81	0.72	67
Nazrul	0.61	0.78	0.68	72
Palligeeti	0.67	0.28	0.40	64
Rabindra	0.64	0.47	0.54	77
Accuracy			0.61	436
Macro Avg	0.63	0.61	0.60	436
Weighted Avg	0.63	0.61	0.60	436

Support Vector Machine				
	Precision	**Recall**	**F1-Score**	**Support**
Adhunik	0.64	0.71	0.67	82
Band	0.83	0.78	0.81	74
Hiphop	0.70	0.79	0.74	67
Nazrul	0.66	0.57	0.61	72
Palligeeti	0.56	0.55	0.56	64
Rabindra	0.64	0.64	0.64	77
Accuracy			0.67	436
Macro Avg	0.67	0.67	0.67	436
Weighted Avg	0.67	0.67	0.67	436

Random Forest				
	Precision	**Recall**	**F1-Score**	**Support**
Adhunik	0.66	0.61	0.63	82
Band	0.73	0.69	0.71	74
Hiphop	0.58	0.78	0.67	67
Nazrul	0.76	0.82	0.79	72
Palligeeti	0.69	0.48	0.57	64
Rabindra	0.67	0.68	0.67	77
Accuracy			0.68	436
Macro Avg	0.68	0.68	0.67	436
Weighted Avg	0.68	0.68	0.67	436

Gradient Boosting Classifier				
	Precision	**Recall**	**F1-Score**	**Support**
Adhunik	0.66	0.65	0.65	82
Band	0.74	0.69	0.71	74
Hiphop	0.66	0.73	0.70	67
Nazrul	0.75	0.83	0.79	72
Palligeeti	0.60	0.47	0.53	64
Rabindra	0.60	0.65	0.63	77
Accuracy			0.67	436
Macro Avg	0.67	0.67	0.67	436
Weighted Avg	0.67	0.67	0.67	436

CNN				
	Precision	**Recall**	**F1-Score**	**Support**
Adhunik	0.71	0.68	0.70	82
Band	0.82	0.78	0.80	74
Hiphop	0.72	0.87	0.79	67
Nazrul	0.81	0.92	0.86	72
Palligeeti	0.80	0.55	0.65	64
Rabindra	0.72	0.75	0.73	77
Accuracy			0.76	436
Macro Avg	0.76	0.76	0.75	436
Weighted Avg	0.76	0.76	0.76	436

To categorize Bangla music genres, a Convolutional Neural Network (CNN) and a number of machine learning models were assessed. The evaluation's primary measures, precision, recall, F1-score, and support, are listed in Table 4.2.2. These metrics aid in evaluating how well each model performs in various genres. With an accuracy of 61% overall, the K-Nearest Neighbour (KNN) model demonstrated a decent level of performance. There were differences in precision, recall, and F1-score amongst the genres; the Hiphop genre had the best F1-score, at 0.72, while Palligeeti had the lowest, at 0.40. The model struggled with Palligeeti but did better with Hiphop and Nazrul, suggesting that it might not be as useful for other musical genres. The total accuracy was increased to 67% using the Support Vector Machine (SVM) model. It achieved high precision and recall for the Band genre with an F1-score of 0.81, while the Palligeeti genre had the lowest F1-score of 0.56. SVM showed more balanced performance across genres compared to KNN, making it a more reliable choice for music genre classification. The Random Forest model also achieved an overall accuracy of 68%. It performed well in the Nazrul genre with an F1-score of 0.79, but had lower performance for the Palligeeti genre with an F1-score of 0.57. This model demonstrated strong performance for some genres but showed inconsistency, especially for genres with fewer samples like Palligeeti. Similar to Random Forest, the Gradient Boosting Classifier achieved an overall accuracy of 67%. It had the best performance for Nazrul with an F1-score of 0.79, while Palligeeti again had lower performance with an F1-score of 0.53. The model showed effective performance for some genres, indicating its potential but also its limitations in handling certain genres consistently. The CNN model outperformed all other models with an overall accuracy of 74.91%. It achieved high precision, recall, and F1-scores across most genres. For instance, the Nazrul genre had an F1-score of 0.86, and even the Palligeeti genre, which was challenging for other models, had an improved F1-score of 0.65. All genres showed strong and balanced performance from the CNN model, demonstrating its better capacity to identify and categorize intricate patterns in the music data. In conclusion, the analysis showed that the CNN model achieves the best accuracy and balanced performance metrics across all genres, making it the most effective method for categorizing Bangla music genres. Conventional machine learning models such as SVM, Gradient Boosting Classifier, Random Forest, KNN, and others demonstrated differing degrees of effectiveness; some were good in some genres but not consistently. The results imply that CNNs are more appropriate for this classification task because to their sophisticated feature extraction capabilities.

Training and Validation Accuracy and loss of CNN Model:

The illustration below depicts the CNN model's accuracy and loss throughout training and validation. In accuracy plot ‘train’ accuracy stands for train accuracy and ‘test’ stands for validation accuracy. In error plot ‘train’ stands for train loss and ‘test’ stands for validation loss.

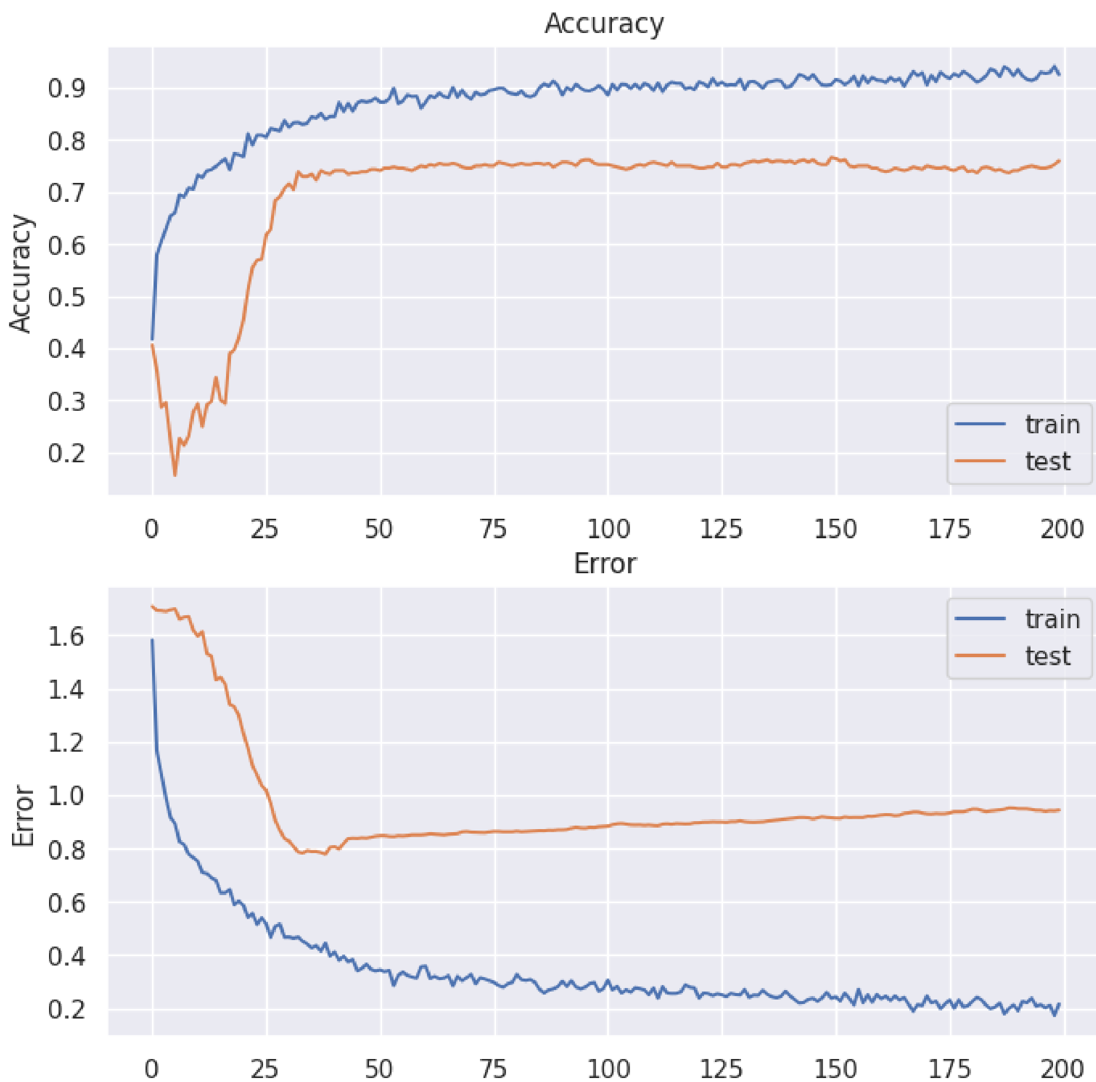

Figure 4.2.1: Training and validation accuracy and loss over the epochs for CNN model.

The plots shows the accuracy of the training and test datasets over epochs, indicating that the training accuracy improves steadily while the test accuracy is also increasing continuously. The bottom plot shows the error for the training and test datasets over epochs, where the training and testing error decreases consistently.

Confusion Matrix for all the models:

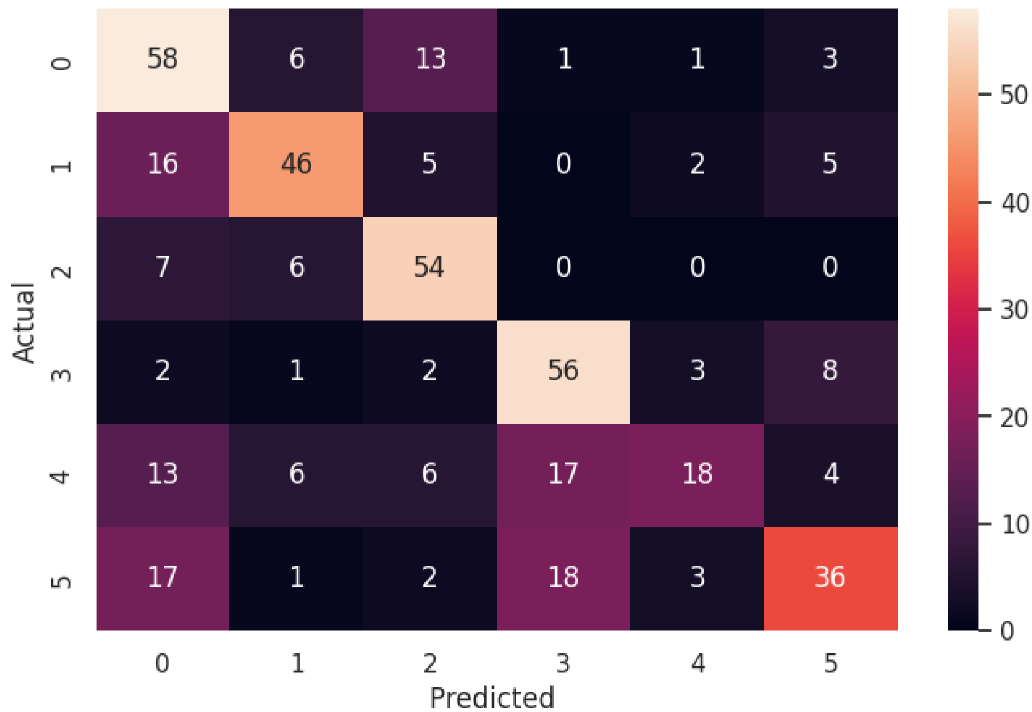

Figure 4.2.2: CM of K-nearest Neighbour

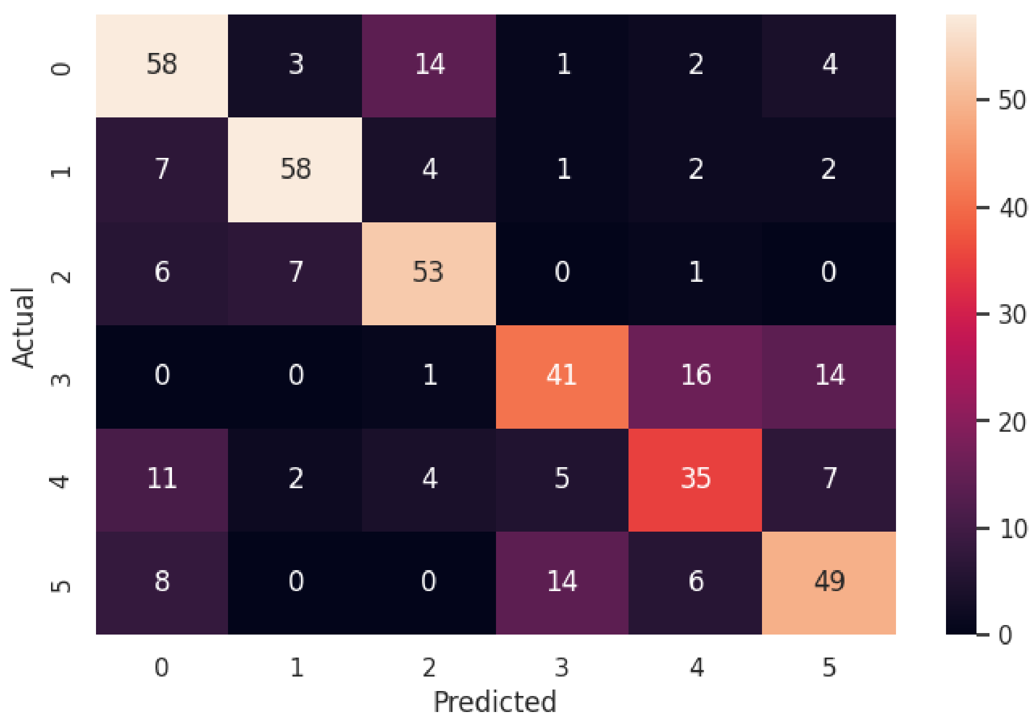

Figure 4.2.3: CM of Support Vector Machine

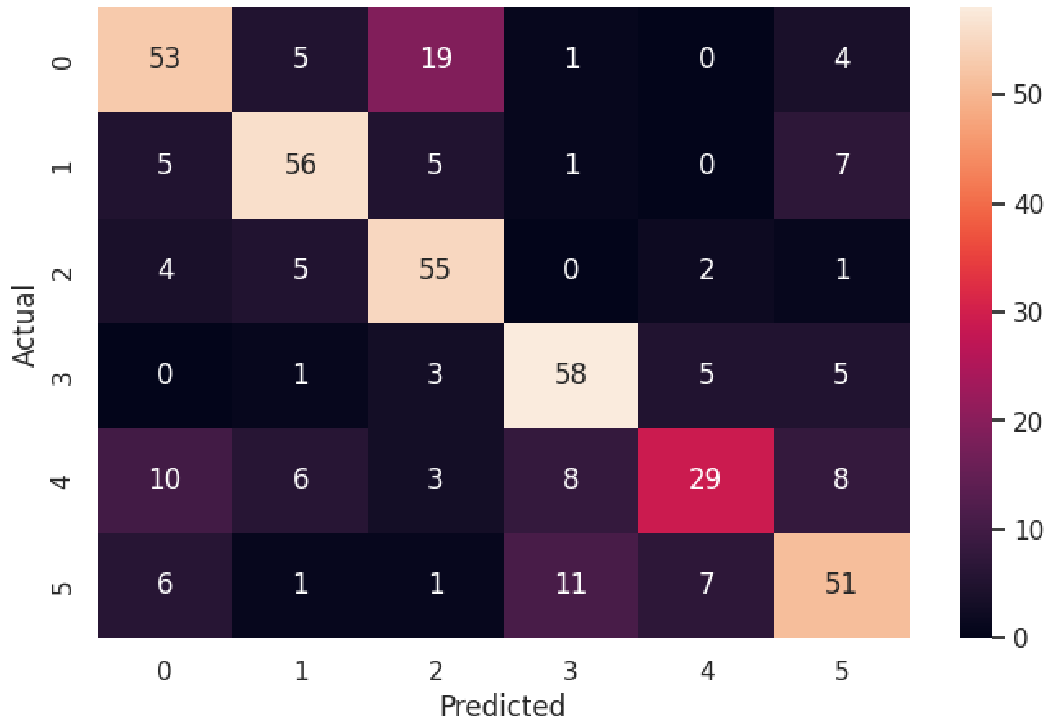

Figure 4.2.4: CM of Random Forest Classifier

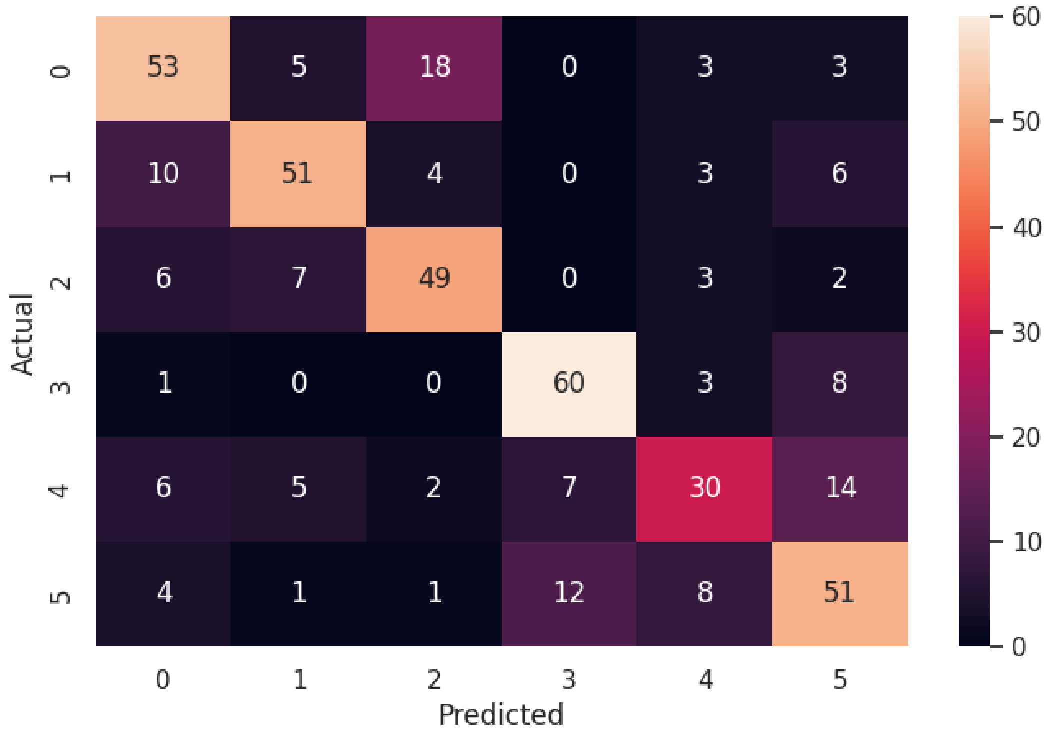

Figure 4.2.5: CM of Gradient Boosting Classifier

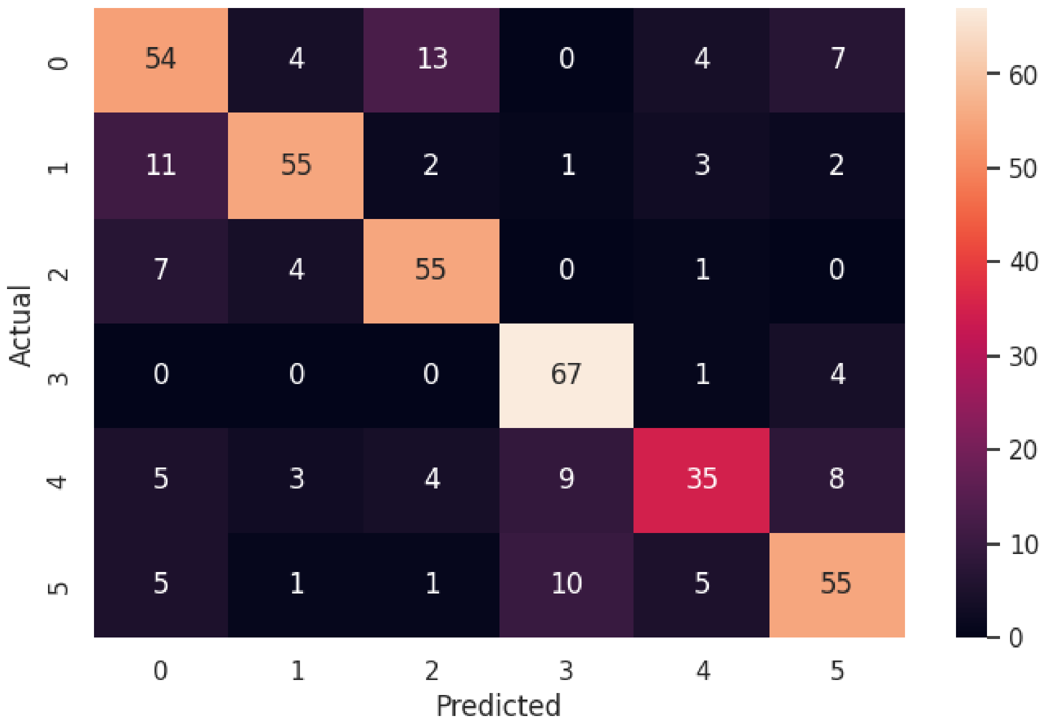

Figure 4.2.6: CM of CNN

5.3 Performance/ Comparative Analysis

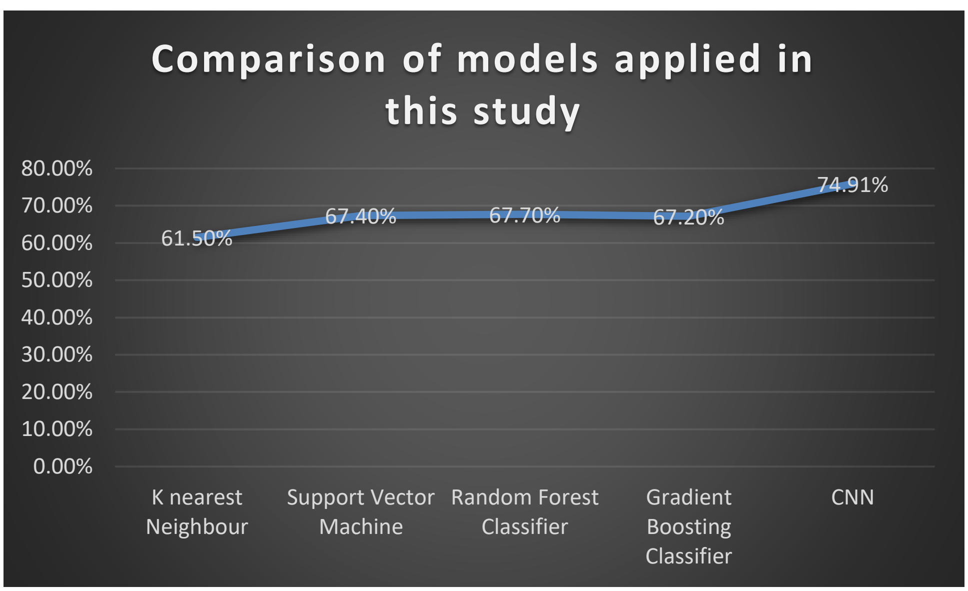

Figure 5.3.1: Accuracy comparison among Machine Learning Models and CNN.

The examination of several CNN and machine learning models for the categorization of musical genres offers insightful information about their usefulness and performance indicators. The study assessed many models, including CNN, SVM, Gradient Boosting Classifier, Random Forest Classifier, and KNN, highlighting their advantages, disadvantages, and implications for further research.

Training accuracy of 68% and testing accuracy of 61.5% were attained using the KNN model. For genres like Hiphop and Nazrul, it performed rather well in terms of precision and recall; nevertheless, it had severe difficulties with genres like Palligeeti, which led to poor F1-scores. The model's mediocre performance suggests that it can't handle large, varied, and complicated music collections, which might provide inconsistent categorization outcomes. With a testing accuracy of 674 and a training accuracy of 72.4%, the SVM model outperformed the KNN. SVM demonstrated relatively balanced precision and recall across different genres, suggesting a better generalization capability. However, the performance is still not ideal for practical applications, particularly for genres with less distinct features, which might cause confusion in classification.

The Random Forest model showed excellent training accuracy at 99.9% but significantly lower testing accuracy at 67.7%. This large discrepancy indicates severe overfitting, where With training data, the model performs remarkably well, but poorly with unknown data. Despite reasonable precision and recall scores, the model’s inability to generalize limits its effectiveness for real-world music genre classification. Similarly, the Gradient Boosting Classifier achieved high training accuracy at 99.2% but lower testing accuracy at 67.2%. The overfitting problem persists, suggesting that while the model captures patterns within the training data efficiently, it struggles with new, unseen data. The balanced precision, recall, and F1-scores across genres indicate potential, but further tuning is required to improve its generalizability.

In terms of training accuracy (92.49%) and testing accuracy (74.91%), the CNN model fared better than other machine learning models. The CNN's superior precision, recall, and F1-scores for most genres, particularly Nazrul and Hiphop, highlight its robust feature extraction and classification capabilities. The hierarchical learning structure of CNNs allows them to capture complex patterns in the music data, making them highly effective for this task. The findings suggest that while traditional machine learning models can provide a baseline, CNNs offer significantly better performance for music genre classification. The high accuracy and balanced performance metrics of CNNs make them suitable for practical applications. Addressing overfitting issues, enhancing data diversity, and improving model interpretability are key areas for future research. These advancements will contribute to developing more accurate and reliable music classification systems, benefiting applications in music streaming services, automated playlist generation, and music information retrieval.

5.4 Summary

The project evaluated various machine learning models for classifying Bangla music genres. The Convolutional Neural Network (CNN) achieved the highest model accuracy at 75.91%, outperforming traditional models like Support Vector Machine, Random Forest, and Gradient Boosting. Analysis of metrics such as precision, recall, and F1-score highlighted the CNN's superior performance in capturing audio features and accurately classifying music genres. This demonstrates the effectiveness of deep learning approaches in handling complex audio data compared to traditional machine learning methods.

CHAPTER 6

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The research on Bangla music genre classification using machine learning and deep learning techniques has significant implications for various aspects of life, particularly in the fields of music and technology. By accurately classifying Bangla music genres, this project enhances music discovery and recommendation systems, providing users with personalized music experiences. For artists and music producers, the model can offer insights into genre-specific trends and preferences, aiding in the creation of music that resonates with target audiences.

Furthermore, this research contributes to the broader field of audio analysis and machine learning, demonstrating the effectiveness of data augmentation and feature extraction techniques in improving model performance. It paves the way for further advancements in automated music classification, which can be applied to other languages and music styles, thereby enriching the global music ecosystem. Ultimately, this work fosters cultural appreciation and preservation by making diverse genres of Bangla music more accessible to listeners worldwide.

6.2 Impact on Society & Environment

Developing an intelligent model for Bengali music genre classification extends far beyond technical advancements; it holds significant implications for society as a whole. By delving into the rich heritage of Bengali music and leveraging advanced machine learning techniques, this research seeks to enhance the music consumption experience for enthusiasts and professionals alike.

One of the most notable impacts lies in the enrichment of listeners' musical journeys. Through accurate genre classification, listeners can discover a diverse array of Bengali music genres that resonate with their unique tastes and preferences. This not only fosters a deeper appreciation for Bengali music but also promotes cultural diversity and understanding.

Major music platforms stand to benefit from this technology as well, as efficient song classification enhances the overall music consumption experience for users worldwide. Overall, Creating an intelligent model for Bengali music genre classification has broad societal impacts. It enriches cultural experiences, promotes diversity, and makes Bengali music more accessible and enjoyable for people worldwide.

The impact on the environment is multifaceted. Firstly, by developing an intelligent model for Bengali music genre classification, the project aims to enhance the digital music consumption experience. This could potentially reduce the demand for physical media such as CDs and DVDs, consequently lowering the production and distribution-related carbon footprint associated with traditional music formats.

Moreover, the classification model, once implemented on music streaming platforms, can lead to more accurate recommendations, thereby potentially reducing unnecessary energy consumption associated with streaming irrelevant content. By aiding users in discovering music that aligns with their preferences, the model could contribute to optimizing energy usage in digital music delivery systems.

Furthermore, the project's emphasis on leveraging machine learning and AI to understand and categorize Bengali music genres can lead to the development of more efficient music recommendation algorithms. These algorithms have the potential to promote the discovery of diverse music genres, fostering cultural exchange and appreciation. In turn, this could contribute to a more sustainable music industry by supporting a broader range of artists and genres, rather than favoring mainstream or commercially-driven content. Overall, while the direct impact of the project on the environment may seem indirect, its potential to optimize digital music consumption, promote diverse music genres, and improve recommendation systems can collectively contribute to a more sustainable and culturally rich music ecosystem.

6.3 Ethical Aspects

Ensuring ethical practices in developing and using the Bengali music genre classification model is crucial. We need to be transparent about how we collect and use data, respecting people's privacy and getting their consent. It's also important to recognize and fix any biases in the data or algorithms to avoid unfairness. We should consider how the model might affect cultural representation and make sure it doesn't ignore less popular genres. Instead, it should celebrate diversity in Bengali music. Everyone should have access to the benefits of the model, no matter their background. Frequent performance evaluations of the model can assist us in identifying and resolving any issues that may arise over time. By being transparent, fair, and mindful of cultural sensitivity, we can develop and use the Bengali music genre classification model in a way that benefits everyone involved.

6.4 Sustainability Plan

To ensure the long-term success and impact of the Bengali music genre classification project, a comprehensive Sustainability Plan has been devised. This plan encompasses various strategies aimed at maintaining the project's relevance, effectiveness, and ethical integrity over time. Central to the sustainability of the project is the ongoing maintenance of the dataset. Regular updates and additions of new songs and genres will be essential to keep the dataset reflective of the dynamic nature of Bengali music. This continuous data maintenance process will also involve ensuring data quality and accuracy through thorough validation and verification procedures.

Furthermore, the optimization of the classification model will be a continual effort. Adapting the model to evolving music trends and technological advancements requires constant refinement of algorithms, adjustment of parameters, and incorporation of new features. Community engagement will play a vital role in sustaining the project's momentum. Building a supportive community of stakeholders, including musicians, researchers, and music enthusiasts, will foster collaboration, generate valuable insights, and garner support for the project's objectives.

Ethical considerations are fundamental to the sustainability of the project. Upholding principles of privacy, fairness, and cultural sensitivity through regular ethical audits and assessments will ensure that the project remains aligned with ethical guidelines and societal values. Collaboration and partnerships with organizations, institutions, and industry stakeholders will amplify the project's impact and reach. Sharing resources, expertise, and data through strategic partnerships will enhance the project's capabilities and broaden its scope of influence.

Effective management of funding and resources is essential for the project's sustainability. Securing sustainable funding sources, efficient resource allocation, and prudent budget planning will ensure the project's long-term viability and impact. By implementing these sustainability strategies, the Bengali music genre classification project aims to thrive, evolve, and make enduring contributions to the music industry and research community, enriching the cultural landscape of Bengali music for years to come.

6.5 Summary

The development of an intelligent model for Bengali music genre classification enriches cultural experiences by promoting diversity and understanding. It enhances music consumption practices through personalized recommendations, potentially reducing environmental impacts associated with traditional media and optimizing energy use in digital platforms. By fostering a sustainable music industry that values cultural diversity and artistic expression, the project contributes to a more enriched societal and environmental landscape.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusions

This study used a combination of deep learning techniques and conventional machine learning algorithms to examine the categorization of Bangla music genres. A dataset comprising 1742 Bangla songs from six different genres, adhunik, hip-hop, band music, Rabindra sangeet, and Palligeeti was used in the study. The process comprised a number of crucial processes, including the collection of audio data, feature extraction, model training, classification, and assessment.

Comprehensive feature extraction from audio files included metrics like tempo, spectral bandwidth, MFCC, delta, spectral centroid, spectral rolloff, zero crossing rate, and chroma frequency, which served as inputs to the classification models. Various models were trained, including traditional ML algorithms and CNN. With a training accuracy of 92.49% and a testing accuracy of 74.91%, CNN performed the best, demonstrating its greater capacity to identify intricate patterns in the audio data.

F1-score, recall, precision, and recall were among the measures used to assess the models' performance for each music genre. CNN outperformed other models, demonstrating better generalization and higher accuracy in classifying Bangla music genres. The results were validated using statistical techniques, which demonstrated the robustness and dependability of the experimental data and confirmed the relevance of the findings.

In conclusion, the study finds that deep learning, specifically CNNs, significantly outperforms traditional ML algorithms in the task of Bangla music genre classification. The use of advanced data augmentation and feature extraction techniques contributed to the high performance of the CNN model. This research underscores the potential of deep learning in the field of music analysis and automated classification systems. Future work should focus on refining these models, enhancing their interpretability, and addressing challenges related to data diversity and potential overfitting. The insights gained from this research can be leveraged to develop more advanced and accurate music classification systems, benefiting both academic research and practical applications in the music industry.

7.2 Further Suggested Works

The results of this study on the genre classification of Bangla music have several significant ramifications for future research. First off, the promise of deep learning techniques in audio classification problems is shown by CNNs' proven dominance over conventional ML systems. In order to further increase classification accuracy, future research can expand on this by investigating more complex and specialized CNN architectures or other deep learning models like RNNs and transformers.

Feature extraction played a critical role in our study. Future studies might concentrate on using end-to-end learning techniques, in which the model learns to extract pertinent characteristics straight from the unprocessed audio input, or integrating more audio features. This could potentially simplify the pipeline and improve performance.

The study also highlights the importance of ethical considerations, particularly in ensuring data privacy and confidentiality. As research progresses, it will be crucial to continue adhering to ethical standards, especially when dealing with more sensitive or personal audio data.

Examining transfer learning and domain adaption strategies is another direction for further study, which could allow models trained on one genre of music or language to be adapted for use in another with minimal retraining. This could make the model more versatile and applicable to a broader range of music classification tasks.

Lastly, our findings highlight the necessity of ongoing development in deep learning models' interpretability and transparency. Future studies could focus on developing methods to better understand and visualize how these models make their classification decisions, which is essential for gaining trust and increased use in practical applications.

7.3 Limitations/ Conflict of Interests

While developing a Bangla music genre classification model marks a significant advancement, several limitations should be noted for future research and application. The dataset of 1742 songs across six genres, while comprehensive, may lack diversity, potentially limiting the model's ability to generalize. Feature extraction relied on standard metrics like tempo and mfcc, which, while effective, may not fully capture the nuances of Bengali music. Models like Random forest and Gradient boosting showed high training accuracies (up to 99.9%) but lower testing accuracies (around 67%), indicating overfitting and highlighting the need for robust regularization and larger datasets for better generalizability. The CNN model, with a training accuracy of 92.49% and testing accuracy of 74.91%, faces challenges in interpretability and transparency. Ethical considerations around data privacy and cultural representation are crucial, alongside computational challenges in deploying deep learning models efficiently. Achieving consistent high accuracy across all genres, particularly for less represented genres like Palligeeti, remains a challenge. Addressing these limitations through enhanced dataset diversity, improved feature engineering, and ethical data practices will be crucial for advancing the model's reliability and applicability.

REFERENCES

[1] 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," 2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, 2022, pp. 1-6, doi: 10.1109/ICAEEE54957.2022.9836434.

[2] Hasib, K.M., Tanzim, A., Shin, J., Faruk, K.O., Al Mahmud, J. and Mridha, M.F., 2022. Bmnet-5: A novel approach of neural network to classify the genre of bengali music based on audio features. IEEE Access, 10, pp.108545-108563.

[3] Al Mamun, M.A., Kadir, I., Rabby, A.S.A. and Al Azmi, A., 2019, November. Bangla music genre classification using neural network. In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) (pp. 397-403). IEEE.

[4] Bhowmik, A. and Chowdhury, A.E., 2019. Genre of Bangla music: a machine classification learning approach. AIUB Journal of Science and Engineering (AJSE), 18(2), pp.66-72.

[5] Rawat, Parv & Dharaskar, Krushna & Nandanwar, Anshula & Dhawale, Krupali Rupesh. (2023). MUSIC GENRE CLASSIFICATION USING DEEP LEARNING. 2582-5208. 10.56726/IRJMETS34036.

[6] Music Genre Classification Using Neural Network Tarannum Shaikh, Ashish Jadhav ITM Web Conf. 44 03016 (2022) DOI: 10.1051/itmconf/2022440301.

[7] Choudhury, Nitin & Deka, Deepjyoti & Sarmah, Satyajit & Sarma, Parismita. (2023). Music Genre Classification Using Convolutional Neural Network. 1-5. 10.1109/I3CS58314.2023.10127554.

[8] Ma, X., 2023. Music Genre Classification Based on Machine Learning Methods. Highlights in Science, Engineering and Technology, 34, pp.168-175.

[9] Li, Z., Peng, H., Tan, S. and Zhu, F., 2023, September. Music classification with convolutional and artificial neural network. In Journal of Physics: Conference Series (Vol. 2580, No. 1, p. 012059). IOP Publishing.

[10] Agrawal, M., & Nandy, A. (2020). A Novel Multimodal Music Genre Classifier using Hierarchical Attention and Convolutional Neural Network. arXiv preprint arXiv:2011.11970.

[11] Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017, March). Convolutional recurrent neural networks for music classification. In 2017 IEEE International conference on acoustics, speech and signal processing (ICASSP) (pp. 2392-2396). IEEE.

[12] Ghosh, P., Mahapatra, S., Jana, S., & Jha, R. K. (2023). A Study on Music Genre Classification using Machine Learning. International Journal of Engineering Business and Social Science, 1(04), 308-320.

[13] Narkhede, N., Mathur, S., Bhaskar, A. A., Hiran, K. K., Dadhich, M., & Kalla, M. (2023, August). A New Methodical Perspective for Classification and Recognition of Music Genre Using Machine Learning Classifiers. In 2023 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC) (pp. 94-99). IEEE.

[14] Kong, Q., Feng, X., & Li, Y. (2014). Music genre classification using convolutional neural network. Proc. Int. Soc. Music Inform. Retrieval (ISMIR).

[15] Shana, J & Jayadurga, Priyadharshini & K R, Pratiba & R, Prakash. (2022). JUXTAPOSITION BETWEEN TWO CONVOLUTIONAL NEURAL NETWORK FOR MUSIC GENRE CLASSIFICATION.

[16] Liu, Z., Bian, T., & Yang, M. (2023). Locally Activated Gated Neural Network for Automatic Music Genre Classification. Applied Sciences, 13(8), 5010.

[17] Vishnupriya, S., & Meenakshi, K. (2018, January). Automatic music genre classification using convolution neural network. In 2018 international conference on computer communication and informatics (ICCCI) (pp. 1-4). IEEE.

[18] Liu, K., DeMori, J., & Abayomi, K. (2022). Open Set Recognition For Music Genre Classification. arXiv preprint arXiv:2209.07548.

[19] Liu, C., Feng, L., Liu, G., Wang, H., & Liu, S. (2021). Bottom-up broadcast neural network for music genre classification. Multimedia Tools and Applications, 80, 7313-7331.

[20] Mehta, Jash & Gandhi, Deep & Thakur, Govind & Kanani, Pratik. (2021). Music Genre Classification

using Transfer Learning on log-based MEL Spectrogram. 1101-1107. 10.1109/ICCMC51019.2021.9418035

[21] Simplilearn is one of the world's leading providers of online training for Digital Marketing, Cloud Computing, Project Management, Data Science, IT, Software Development, and many other emerging technologies.

[22] McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E. and Nieto, O., 2015, July. librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference (Vol. 8, pp. 18-25).

[23] Min Xu; et al. (2004).In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh (eds.). Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer. ISBN 978-3-540-23985-7.

[24] Bangla Music Dataset, available at <<https://www.kaggle.com/datasets/afifaniks/bangla-music-dataset>>, last accessed on 06-07-2024 at 06:00 PM

[25] GTZAN Dataset - Music Genre Classification, available at <<https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification>>, last accessed on 06-07-2019 at 07:00 PM

[26] Google Colab, available at <<https://colab.research.google.com/>>, last accessed on 06-07-2024 at 07:30 PM

[27] Model Evaluation- GeeksforGeeks, available at <<https://www.geeksforgeeks.org/machine-learning-model-evaluation/>>, last accessed on 06-07-2024 at 08:30 PM

[28] Analytics Vidhya, available at <<https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/>>, last accessed on 06-06-2024 at 09:00 PM

[29] Support Vector Machine- GeeksforGeeks, available at <<https://www.geeksforgeeks.org/support-vector-machine-algorithm/>>, last accessed on 06-07-2024 at 9:30 PM

[30] Support Vector Machine (SVM) Algorithm - Javatpoint, available at <<https://images.app.goo.gl/XqoX5k5D2gLTC7nLA>>, last accessed on 06-06-2024 at 12:00 PM

[31] K-Nearest Neighbor(KNN) Algorithm- GeeksforGeeks, available at <<https://www.geeksforgeeks.org/k-nearest-neighbours/>>, last accessed on 06-07-2024 at 10:00 PM

[32] K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint, available at <<https://images.app.goo.gl/3xxX5cxXzSXxsudS8>>, last accessed on 07-07-2024 at 9:00 PM

[33] Random Forest Algorithm- Javapoint, available at <<https://www.javatpoint.com/machine-learning-random-forest-algorithm>>, last accessed on 09-07-2024 at 10:00 PM

[34] Machine Learning Random Forest Algorithm - Javatpoint, available at <<https://images.app.goo.gl/RPYooVr322csVRva8>>, last accessed on 10-07-2024 at 11:00 PM

[35] Gradient Boosting in ML- GeeksforGeeks, available at <<https://www.geeksforgeeks.org/ml-gradient-boosting/>>, last accessed on 11-07-2024 at 06:00 PM

[36] Music of Bangladesh- Wikipedia, available at <<https://en.wikipedia.org/wiki/Music_of_Bangladesh>>, last accessed on 12-07-2024 at 11:00 PM

[37] Albelwi S, Mahmood A. A Framework for Designing the Architectures of Deep Comvolutional Neural Networks. *Entropy*. 2017; 19(6):242. https://doi.org/10.3390/e19060242

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:

Bangla Music Genre Classification using Deep Learning and Machine Learning Techniques

Student ID: 202-15-3839

202-15-3850

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to identify a **music genre classification** problem for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish this goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**

CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**
CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project demonstrates **fundamental engineering (K3)** principles by employing machine learning algorithms and diverse feature extraction methods for music genre classification tasks. The project **demonstrates specialist knowledge (K4)** by leveraging advanced deep learning models, particularly Convolutional Neural Networks (CNNs), and applying the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance, enhancing the accuracy of classifying Bangla music genres, crucial for applications in music recommendation and information retrieval.	**Page no:** [25-29] **Section:** [4.2] **Page no:** [18-21] **Section:** [3.2] **Page no:** [1-2] **Section:** [1.2]

				The project applies **engineering practice and design (K5)** through the structured process of experiments and systematic evaluation of machine learning models. The project addresses **engineering practice and technology (K6)** by employing advanced CNN model to classify Bangla music genres.	**Page no:** [18-21] **Section:** [3.2(B)] **Page no:** [30-32] **Section:** [4.3]
				This project ensures **K8 (Research Literature)** by synthesizing insights from recent studies to advance Bangla music genre classification using deep learning, showcasing a understanding of current methodologies in music analysis and machine learning.	**Page no:** [7-16] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in Bangla music genre classification, including the limitations of traditional methods and the complexities of integrating deep learning techniques like CNNs. Through comparative analysis, it confronts challenges in handling diverse audio features and genre distinctions, offering insights for refining music classification methodologies.	**Page no:** [4] **Section:** [1.5] **Page no:** [16-17] **Section:** [2.4] **Page no:** [52] **Section:** [7.3]

3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting the use of Convolutional Neural Networks (CNNs) as the chosen significant solution for enhancing Bangla music genre classification compared to multiple machine learning approaches.	**Page no:** [34-45] **Section:** [5.2, 5.3]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting musicology and cultural studies by advancing the classification of Bangla music genres. This contributes to the preservation and appreciation of cultural heritage, showcasing **EP-4**.	**Page no:** [4, 7-17] **Section:** [1.5, 2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A
6.	**EP6: Extends of stakeholders involved and conflicting**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, feature extraction, and model implementation, ensuring a holistic solution to complex challenges in Bengali music genre classification, which ensures **EP-7**.	**Page no:** [18-21, 25-32] **Section:** [3.2, 4.2, 4.3]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, machine learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in Bengali music genre classification through feature extraction and CNNs.	**Page no:** [22] **Section:** [3.3]
2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project enhances societal well-being by advancing music genre classification techniques, enriching cultural experiences and promoting diversity in Bengali music. It also supports environmental sustainability through efficient use of computational resources and upholding ethical standards in data privacy.	**Page no:** [46-49] **Section:** [6.1, 6.2, 6.4]

5.	**EA-5: Familiarity**	Yes		This project extends current research by exploring innovative methodologies in Bengali music genre classification using deep learning techniques, demonstrated through detailed experimentation and comparative analysis. It offers fresh perspectives and insights into the application of advanced machine learning models for enhancing music classification accuracy and effectiveness.	**Page no:** [12-16, 18-21] **Section:** [2.3, 3.2]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by using thorough project management practices and financial oversight, ensuring careful planning, efficient use of resources, and accurate budget estimation to make the most of resources at every stage of the research effort.	**Page no:** [23] **Section:** [3.4]
2	CO9	Yes	The project follows ethical principles by keeping data private, getting consent from those involved, and clearly documenting the research process. This ensures that the knowledge gained is shared responsibly and benefits society. By using advanced machine learning and deep learning to classify Bangla music genres, the project respects ethical standards and promotes cultural sensitivity and fairness which comply with **CO9**.	**Page no:** [47] **Section:** [6.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project shows a strong commitment to continuous learning **(CO12)** and adapting to new technology. This is clear from the detailed data collection, careful statistical analysis, well-planned methods, and thorough testing. By continually updating and improving techniques, the project ensures that classifying Bangla music genres stays accurate and effective.	**Page no:** [18-22, 25-32, 34-43] **Section:** [3.2, 3.3, 4.2, 4.3, 5.2]

Bangla Music Genre Classification using Deep Learning and Machine Learning Techniques.pdf

ORIGINALITY REPORT

25%	19%	14%	14%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Daffodil International University Student Paper	3%
3	www.geeksforgeeks.org Internet Source	2%
4	www.mdpi.com Internet Source	1%
5	Md. Nabil Rahman Khan, Khandaker Iffat Jahan Tuli, Most. Sadia Salsabil, S.M. Raiyan Reza et al. "Bengali Music Genre Classification Using WaveNet", 2023 26th International Conference on Computer and Information Technology (ICCIT), 2023 Publication	1%
6	"Data Science and Applications", Springer Science and Business Media LLC, 2024 Publication	1%
7	dspace.bracu.ac.bd	