**TRIBHUVAN UNIVERSITY**

**INSTITUTE OF ENGINEERING**

ADVANCED COLLEGE OF ENGINEERING AND MANAGEMENT

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

KALANKI, KATHMANDU



**A Minor Project Final Defense Report On**

**“Music Genre Classification using CNN”**

[CT654]

**Submitted By:**

Subash Limbu (ACE076BCT086)

Sushant Tripathee (ACE076BCT090)

Ujen Lal Shrestha (ACE076BCT094)

**Supervisor:**

Narayan KC

A Minor Project Final report submitted to the department of Electronics and

Computer Engineering in the partial fulfillment of the requirements for degree of Bachelor of Engineering in Computer Engineering

Kathmandu, Nepal

2079/11/02

ADVANCED COLLEGE OF ENGINEERING AND

MANAGEMENT

DEPARTMENT OF COMPUTER AND ELECTRONICS ENGINEERING

APPROVAL LETTER

The undersigned certify that they have read and recommended to the Institute of

Engineering for acceptance, a project report entitled “MUSIC GENRE CLASSIFICATION USING DEEP LEARNING”

Submitted by:

Subash Limbu (ACE076BCT086)

Sushant Tripathee (ACE076BCT090)

Ujen Lal Shrestha (ACE076BCT094)

In partial fulfillment for the degree of Bachelor in Computer Engineering.

.................................. ..................................

Project Supervisor Project Supervisor

..................................

External Examiner

..................................

Er. Laxmi Prasad Bhatt

Academic Project coordinator

Department of Computer and Electronics Engineering

Date:2079/11/02

**ACKNOWLEDGEMENT**

We take this opportunity to express our deepest and sincere gratitude to our Project

Supervisor **Er. Narayan KC** for his insightful advice, motivating suggestions, invaluable guidance, help and support in successful completion of this project and also for his/her constant encouragement and advice throughout our Bachelor's programme.

We express our deep gratitude to **Er. Ajaya Shrestha**, Head of Department of Electronics and Computer Engineering, **Er. Bikash Acharya,** Deputy Head, Department of Electronics and Computer Engineering, **Er. Laxmi Prasad Bhatt,** Academic Project Coordinator, Department of Electronics and Computer Engineering for their regular support, co-operation, and coordination.

The in-time facilities provided by the department throughout the Bachelors program are also equally acknowledgeable.

We would like to convey our thanks to the teaching and non-teaching staff of the Department of Electronics & Communication and Computer Engineering, ACEM for their invaluable help and support till the date. We are also grateful to all our classmates for their help, encouragement and invaluable suggestions.

Finally, yet more importantly, I would like to express our deep appreciation to my grandparents, parents, siblings for their perpetual support and encouragement throughout the Bachelor’s degree period.

Subash Limbu (ACE076BCT086)

Sushant Tripathee (ACE076BCT090)

Ujen Lal Shrestha (ACE076BCT094)

## ABSTRACT

Music genre classification is a fundamental problem in the field of music information retrieval (MIR). In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown great promise in solving this problem. In this project, we propose a music genre classification system using CNNs. The system takes spectrograms of audio signals as inputs and trains a CNN model to classify the genre of the music. The proposed system was evaluated on a dataset called GTZAN consisting of 1,000 audio tracks from 10 different genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock.

To preprocess the audio signals, we used the Librosa library to extract Mel-frequency spectrograms with a window size of 2048 samples and a hop length of 512 samples. The CNN model was then trained on these spectrograms. The pre-trained model was fine-tuned on our dataset for 30 epochs with a batch size of 32 and a learning rate of 0.0001.

Overall, our results demonstrate the effectiveness of CNNs in music genre classification and show the potential for their use in real-world applications such as music recommendation systems. The proposed system can be extended to classify other music genres or even other audio signals such as speech or environmental sounds.

**Keywords:** *GTZAN, machine learning, signal processing, optical music recognition, CNN, music information retrieval*

## TABLE OF CONTENTS

[ABSTRACT III](#_Toc129783396)

[TABLE OF CONTENTS IV](#_Toc129783397)

[LIST OF FIGURES V](#_Toc129783398)

[LIST OF ABBREVIATIONS/ACRONYMS VI](#_Toc129783399)

[CHAPTER 1 1](#_Toc129783400)

[INTRODUCTION 1](#_Toc129783401)

[1.1 Background 1](#_Toc129783402)

[1.2 Motivation 2](#_Toc129783403)

[1.3 Statement of the Problem 3](#_Toc129783404)

[1.4 Project objective 4](#_Toc129783405)

[1.5 Significance of the study 5](#_Toc129783406)

[CHAPTER 2 6](#_Toc129783407)

[LITERATURE REVIEW 6](#_Toc129783408)

[CHAPTER 3 8](#_Toc129783409)

[REQUIREMENT ANALYSIS 8](#_Toc129783410)

[3.1 Hardware Requirements 8](#_Toc129783411)

[3.2 Software Requirements 8](#_Toc129783412)

[3.3 Functional Requirements 8](#_Toc129783413)

[3.4 Non-Functional Requirements 9](#_Toc129783414)

[3.5 Feasibility Study 9](#_Toc129783415)

[CHAPTER 4 11](#_Toc129783416)

[SYSTEM DESIGN AND ARCHITECTURE 11](#_Toc129783417)

[4.1 Block Diagram 12](#_Toc129783418)

[4.2 DFD 13](#_Toc129783419)

[4.3 Use Case Diagram 14](#_Toc129783420)

[CHAPTER 5 15](#_Toc129783421)

[METHODOLOGY 15](#_Toc129783422)

[5.1 Incremental Model 15](#_Toc129783423)

[5.2 CNN Algorithm 16](#_Toc129783424)

[CHAPTER 6 18](#_Toc129783425)

[RESULT AND ANALYSIS 18](#_Toc129783426)

[CHAPTER 7 21](#_Toc129783427)

[CONCLUSION, LIMITATIONS AND FUTURE ENHANCEMENT 21](#_Toc129783428)

[7.1 Limitations 21](#_Toc129783429)

[7.1 Future Enhancements 21](#_Toc129783430)

[REFFERENCES 23](#_Toc129783431)

## 

## LIST OF FIGURES

Figure 4.1: Block diagram of genre classification system.................................................12

Figure 4.2.1: DFD level 0 diagram....................................................................................13

Figure 4.2.2: DFD level 1 diagram....................................................................................13

Figure 4.3: Use Case Diagram...........................................................................................14

Figure 5.1: Incremental Model..........................................................................................15

Figure 5.2: CNN Architecture...........................................................................................17

Figure 6.1: GUI Interface…………............................................................................…..18

Figure 6.2: Confusion Matrix...............……….................................................................18

Figure 6.3: Train vs Test accuracy and error……………….............................................19

## 

## LIST OF ABBREVIATIONS/ACRONYMS

**CNN** Convolution Neural Network

**MIR** Music Information Retrieval

**ML** Machine Learning

**MFCC** Mel-Frequency Cepstral Coefficients

**STFT** Short-Time Fourier Transform

**FTT** Fast Fourier Transform

**ReLU** Rectified Linear Unit

## CHAPTER 1

## INTRODUCTION

## 1.1 Background

Music genre classification using CNNs is a rapidly developing field that has received significant attention from researchers in recent years. The goal of music genre classification is to automatically categorize a piece of music into one or more predefined genres based on its audio features. This is important for various applications, such as music recommendation systems, music search engines, and music streaming services.

CNNs have proven to be highly effective in music genre classification tasks due to their ability to learn and extract complex features from the audio spectrograms of music tracks. The CNN architecture is typically composed of several layers, including convolutional, pooling, and fully connected layers, which enable the network to learn and represent high-level features. Moreover, the ability to perform transfer learning has also contributed to the success of CNNs in music genre classification.

Overall, research in music genre classification using CNNs has shown promising results, and there is much ongoing work in this field. In addition to improving the performance of CNNs, researchers are also exploring new methods and techniques for feature extraction, data augmentation, and model interpretation to better understand and analyze the results. The potential applications of music genre classification using CNNs are vast, and the field is expected to continue to grow and develop in the coming years.

### 1.2 Motivation

The motivation to perform music genre classification using CNNs is to automatically categorize a piece of music into one or more predefined genres based on its audio features. This has several potential applications, including music recommendation systems, music search engines, and music streaming services. By accurately classifying music genres, these systems can provide users with more personalized and relevant recommendations, improve search results, and enhance the overall listening experience.

Moreover, music genre classification using CNNs is also useful for musicologists, researchers, and music enthusiasts who are interested in analyzing and understanding different music genres. With the help of CNNs, they can analyze large collections of music tracks and identify patterns and trends in different genres, such as the common chord progressions, rhythms, and melodic motifs. This can lead to a better understanding of the evolution of different genres and their cultural and historical significance.

Overall, music genre classification using CNNs has the potential to improve music-related applications, enhance the listening experience, and facilitate research and analysis in musicology. Therefore, it is an exciting and rapidly growing field of research in music information retrieval.

### 1.3 Statement of the Problem

To accurately classify a given piece of music into one or more predefined genres based on its audio features involves training a CNN model to learn and extract discriminative features from the audio spectrograms of music tracks and predict the corresponding genre labels.

The main challenges in this task include dealing with variations in music recordings, such as differences in instrumentation, tempo, and sound quality, as well as the presence of multiple genres in a single track. Furthermore, selecting an appropriate CNN architecture, preprocessing techniques, and training parameters can greatly impact the performance of the model.

Therefore, the goal of a music genre classification using CNN project is to develop an accurate and efficient classification model that can handle these challenges and achieve high performance on a given dataset. This can involve experimenting with different CNN architectures, preprocessing methods, and hyperparameters, as well as evaluating the model's performance using appropriate metrics.

### 1.4 Project objective

* To develop a machine learning model that can accurately classify a given piece of music into one or more predefined genres based on its audio features.

### 1.5 Significance of the study

The study of music genre classification using CNN project has several significant implications, including:

* **Improved music recommendation systems:** By accurately classifying music into different genres, music recommendation systems can provide users with more personalized and relevant recommendations, leading to an enhanced listening experience.
* **Enhanced music search engines:** Music search engines can use genre classification to improve search results and help users discover new music based on their preferred genres.
* **Facilitating music analysis and research:** Musicologists and researchers can use music genre classification to analyze and understand different music genres and their cultural and historical significance, leading to a better understanding of music evolution.
* **Advancement in machine learning techniques:** Music genre classification using CNN is a challenging task due to variations in music recordings. Therefore, developing accurate and efficient models can help advance machine learning techniques and improve their performance on other related tasks.
* **Commercial applications:** The accurate classification of music into genres can have significant commercial applications, such as targeted advertising, personalized playlists, and content recommendation, leading to increased revenue for the music industry.

## CHAPTER 2

## LITERATURE REVIEW

There are few pre-existing works on the above-described model. In 2008, Hareesh Bahuleyan worked on Music Genre Classification using Machine Learning techniques. His work is to provide tags to the music automatically to the songs in the library. Both Neural network and traditional Machine Learning Algorithms were made use to reach the goal. These two approaches used different sets of features. After comparing both these approaches, the model with the Convolutional Neural Network (CNN) approach gave the highest accuracy. [1]

In 2002, Tzanetakis G. et al. worked on classifying audio signals into a hierarchy of genres. They believed that characteristics of a song depends on the instrumentation, the rhythmic structures, and the harmonic content of the song. They proposed three main feature sets: timbral texture, the rhythmic content and the pitch content. Using the proposed feature sets, their model can classify almost 61% of the songs into the corresponding genre out of ten music genres correctly. [2]

Again in 2002, Lu L. et al. The work on Content analysis for audio classification and segmentation. They divided their work into two steps. The first step is to discriminate and separate the speech and non-speech audio. Algorithms are developed using K-nearest-neighbor (KNN) and linear spectral pairs-vector quantization (LSP- VQ). The second step is to classify the audio into music and other forms of sounds. [1] [3]

In 2010, Tom LH Li et al. Automatic musical pattern feature extraction using convolutional neural network, where they tried to understand better features that helps in developing more accurate Music Genre Classification model. Their work was to extract musical features of the audio using the Convolutional Neural Network. They proved that CNN is a better tool to read informative features from the varying musical pattern. Dataset GTZAN was considered. [1] [4]

In 1997, Hochreiter and Schmidhuber, introduced LSTM and were refined and popularized by many people in the following work. They work tremendously well on a large variety of problems, and are now widely used. [5]

## CHAPTER 3

## REQUIREMENT ANALYSIS

### 3.1 Hardware Requirements

* **Processor:** Intel Core i3 or higher
* **RAM:** 4 GB DDR or higher
* **Storage:** At least 10 GB of free disk space for storing data and model checkpoints

### 3.2 Software Requirements

* **Operating System**: Windows, macOS, or Linux
* **Python:** Version 3.10 or higher
* **Deep Learning Framework:** TensorFlow or PyTorch
* **Audio Processing Libraries:** Librosa, Essentia, or similar libraries for audio feature extraction
* **Data Science Libraries:** NumPy, pandas, and scikit-learn for data processing and analysis
* **Integrated Development Environment (IDE):** Jupyter Notebook or any Python IDE for coding and experimentation

### 

### 3.3 Functional Requirements

* **Audio Data Preprocessing:** The system should preprocess the audio data to extract relevant features, such as spectrograms, for use in music genre classification.
* **Model Training and Validation**: The system should train and validate a CNN model using the preprocessed audio data and labeled genre information to accurately classify music into different genres.
* **Genre Classification**: The system should be able to classify music tracks into different genres based on their audio features and the trained CNN model.
* **Performance Evaluation:** The system should be able to evaluate the performance of the CNN model using appropriate metrics.

### 3.4 Non-Functional Requirements

* **Accuracy:** The system should achieve high accuracy in music genre classification.
* **Scalability:** The system should be able to handle large amounts of data and scale efficiently to support a large number of users and requests.
* **Speed:** The system should be able to classify music tracks in real-time or near real-time for a seamless user experience.
* **Security:** The system should implement appropriate security measures, such as encryption and access control, to protect user data and ensure privacy.
* **Usability:** The system should be easy to use and navigate, with clear instructions and feedback provided to users.
* **Reliability:** The system should be reliable and available at all times, with minimal downtime or errors.
* **Maintainability:** The system should be easy to maintain and update, with clear documentation and version control practices in place.

### 3.5 Feasibility Study

Here are some key factors to consider in a feasibility study:

* **Technical Feasibility:** The technical feasibility of the project depends on the availability of resources and tools required to develop and deploy the CNN model for music genre classification. This includes hardware such as processors and GPUs, software such as Python, TensorFlow or PyTorch, and audio processing libraries. It is important to ensure that the required tools are available and compatible with the project requirements.
* **Economic Feasibility:** The economic feasibility of the project depends on the cost of developing, deploying, and maintaining the music genre classification system. This includes the cost of hardware, software, data, and resources such as personnel and time. The project's economic feasibility should be assessed against the potential benefits, such as increased revenue, improved user experience, and enhanced brand image.
* **Operational Feasibility:** The operational feasibility of the project depends on the availability of data, personnel, and processes required to develop and maintain the music genre classification system. This includes access to labeled audio datasets for model training and validation, skilled personnel with expertise in machine learning and audio processing, and effective processes for data preprocessing, model training, and deployment.
* **Legal and Ethical Feasibility:** The legal and ethical feasibility of the project depends on compliance with regulations and laws governing data privacy, intellectual property, and ethical standards. This includes obtaining the necessary permissions and consent for using the audio data, ensuring that the system does not violate copyright laws, and adhering to ethical principles such as fairness, transparency, and accountability.

## CHAPTER 4

## SYSTEM DESIGN AND ARCHITECTURE

To classify music into different genres, the system needs a large and diverse dataset of audio tracks. This can be obtained from various sources, such as online music libraries, music streaming services, and user uploads. Once the data is collected, it needs to be preprocessed to extract relevant audio features such as Mel-frequency cepstral coefficients (MFCCs) and spectrograms. Additionally, data augmentation techniques such as pitch shifting, time stretching, and noise addition can be used to increase the size and diversity of the dataset, and improve the generalization performance of the model.

The next step is designing the CNN model architecture. A CNN model is well-suited for music genre classification because it can learn to automatically extract relevant audio features from raw audio data. The architecture typically consists of several convolutional layers, pooling layers, and fully connected layers. Transfer learning can also be used to improve model performance, by using a pre-trained CNN model such as VGG16 or ResNet as a starting point for the genre classification model.

Once the model architecture is designed, it needs to be trained and validated using the preprocessed audio data and labeled genre information. Cross-validation techniques can be used to prevent overfitting and improve the generalization performance of the model.

The genre classification itself involves allowing users to upload their music tracks and classifying them into different genres based on their audio features. The preprocessed audio data is fed into the trained CNN model, which outputs probabilities for each genre label. The system then uses these probabilities to determine the genre label for each uploaded music track. The genre classification results can be displayed to the user via a user interface, along with visualizations of the audio features used for classification such as spectrograms and waveforms.

Finally, the system needs to be deployed and maintained for optimal performance. This involves deploying the music genre classification system on a web server or cloud platform, monitoring the system for errors and performance issues, and updating the system with new data and model improvements as needed.

### 4.1 Block Diagram

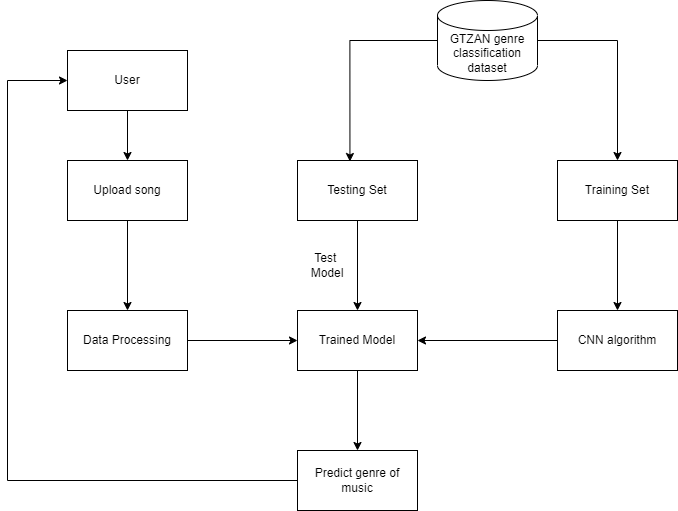


Figure 4.1: Block diagram of genre classification system

### 4.2 DFD

#### 4.2.1 DFD Level 0

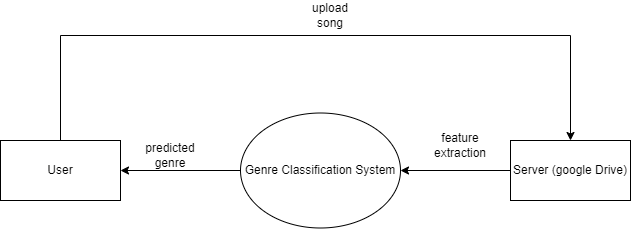


Figure 4.2.1: DFD level 0 diagram

#### 4.2.2 DFD Level 1

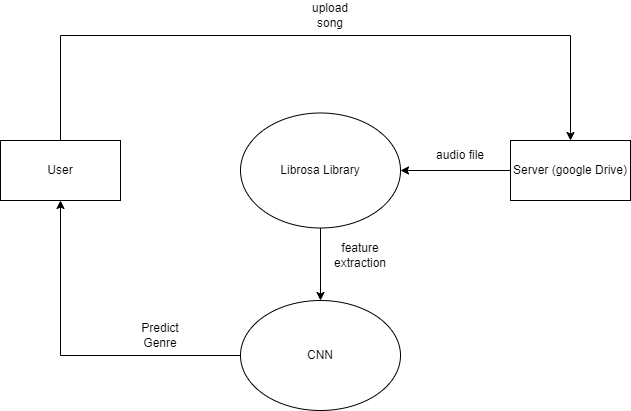


Figure 4.2.2: DFD level 1 diagram

### 4.3 Use Case Diagram

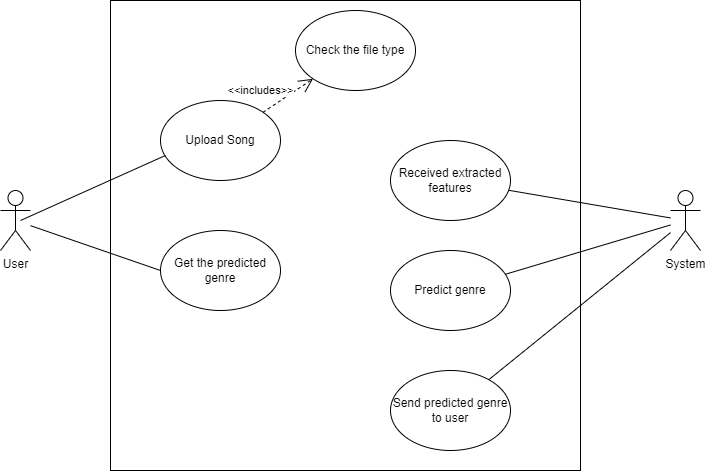


Figure 4.3: Use Case Diagram

## CHAPTER 5

## METHODOLOGY

### 5.1 Incremental Model

An incremental model is an approach to software development that emphasizes building and delivering software in small, incremental stages. In this approach, a small section of the software is developed and delivered in a short period of time, typically weeks or months. This section is then reviewed, tested, and refined before the next section is developed and delivered.

There are several advantages to using the incremental model. First, it allows for early delivery of working software, which can help to build stakeholder confidence and enable the team to get feedback on the software early in the development process. Second, it allows for early identification and resolution of issues, as problems can be caught and addressed at each stage of development. Finally, it allows for a more flexible and adaptable development process, which can help to ensure that the final product meets the needs of the users.

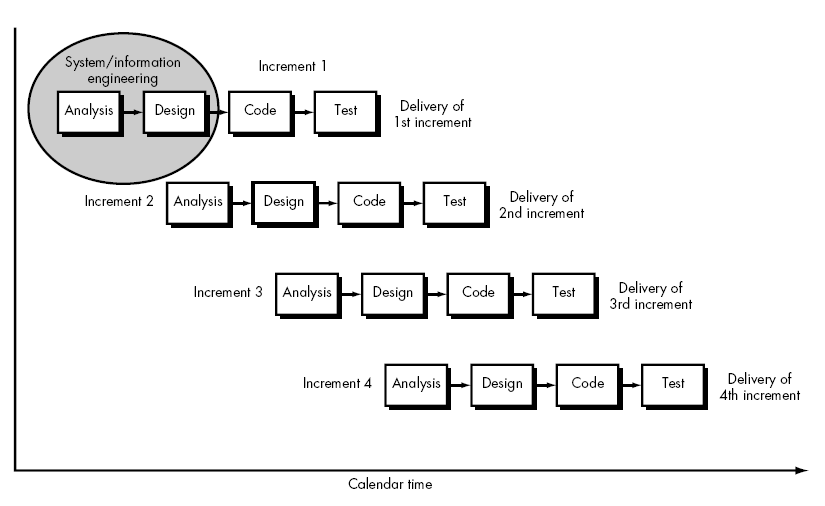


Figure 5.1: Incremental Model

*[Source:* *https://www.researchgate.net/figure/Incremental-Model\_fig3\_311559089 Accessed Dec-12, 2022]*

### 5.2 CNN Algorithm

Convolutional Neural Networks (CNNs) are a type of neural network that have been very successful in image and audio processing tasks. The main idea behind CNNs is to use convolutional layers to extract relevant features from the input data, such as images or audio signals, and then use fully connected layers to perform a classification task.

The working of the CNN algorithm can be explained as follows. First, the input data is fed into a set of convolutional layers. Each convolutional layer applies a set of filters to the input data, which convolve with the input and produce a set of feature maps. These filters are learned through the training process and capture various aspects of the input data, such as edges, shapes, or patterns.

After each convolutional layer, a pooling layer is applied to reduce the spatial dimensionality of the feature maps. The pooling layer down samples the feature maps by taking the maximum or average value of a local neighborhood of pixels or feature activations. This helps to improve computational efficiency and prevent overfitting by reducing the number of parameters in the model.

The output of the final pooling layer is then flattened and fed into a set of fully connected layers, which perform a classification task. The fully connected layers are similar to those used in traditional feedforward neural networks, and learn to map the extracted features to the different classes or labels.

The CNN algorithm is trained using a dataset of labeled data, where the model learns to optimize the weights of the filters and fully connected layers through backpropagation. The optimization process involves adjusting the weights of the model to minimize the classification error between the predicted and actual labels. The model is evaluated on a separate test set to estimate its generalization performance.

Once the model is trained, it can be used for inference on new data to predict their label or class. The input data is first preprocessed to match the input format expected by the CNN model, and then fed into the model. The model outputs a probability distribution over the different classes or labels, and the class with the highest probability is selected as the predicted label.

Overall, the CNN algorithm is a powerful technique for many machine learning tasks, as it can automatically extract relevant features from the input data and learn to map them to the output labels. By using convolutional and pooling layers, the CNN algorithm is able to capture both local and global features of the input data, making it well-suited for complex data processing tasks such as image or audio classification.

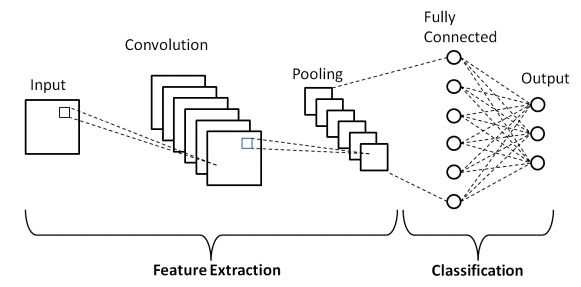


Figure 5.2: CNN Architecture

*[Source:* *https://www.upgrad.com/blog/basic-cnn-architecture/ Accessed Dec-17, 2022]*

## CHAPTER 6

## RESULT AND ANALYSIS

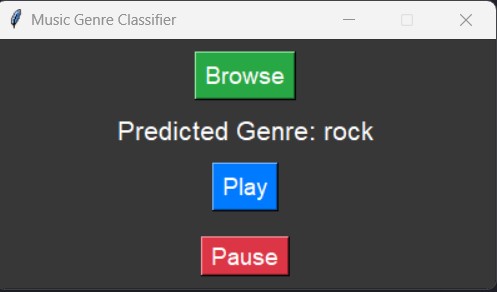


Figure 6.1: GUI Interface

Here in the GUI we can select the song and the GUI displays the genre of the song. There is also a play and pause button for the song.

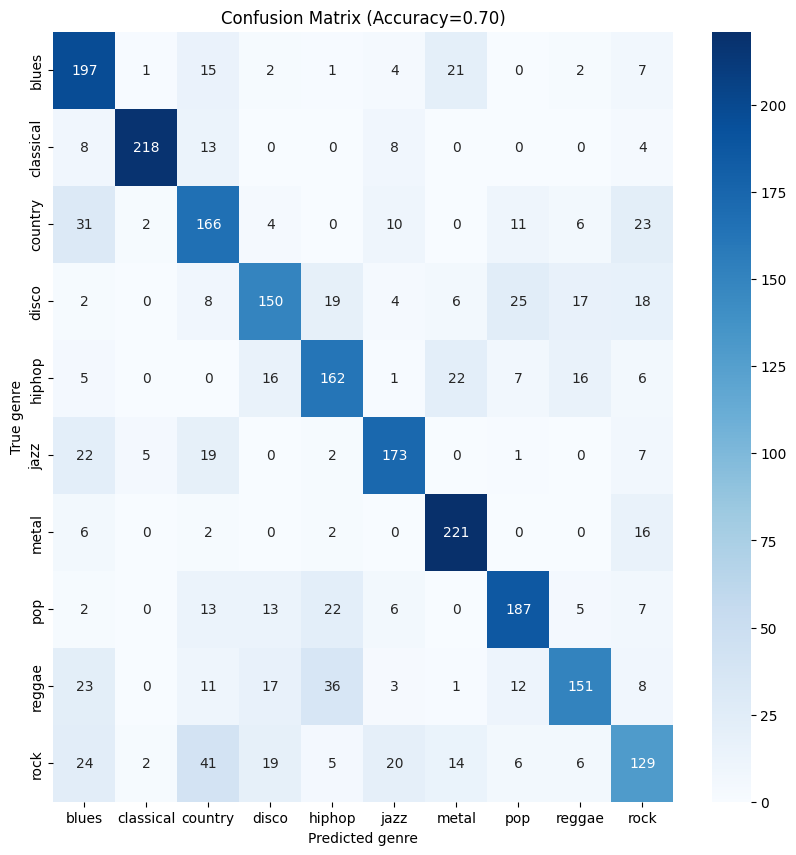


Figure 6.2: Confusion Matrix

Here in the confusion matrix the diagonals show the correctly predicted genres. For blues 197 samples were predicted correctly, for classical 218, country 166 and so on for all the genres.

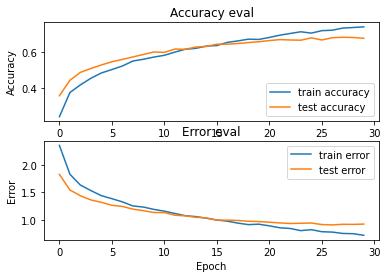


Figure 6.3: Train vs Test accuracy and error

Here:

Testing accuracy = 70.18%.

Training accuracy: 74.82%

Validation accuracy: 70.22%raining accuracy:

Hence, we did prediction on the testing set and among ten it was able to predict seven genres correctly which is according to our testing accuracy. We also did a prediction on our own songs and it was able to predict it’s genre pretty accurately. The difference between the training and validation accuracy is not very large, which suggests that the model does not have overfitting issues. Ideally, the training and validation accuracies should be close to each other.

## 

## CHAPTER 7

## CONCLUSION, LIMITATIONS AND FUTURE ENHANCEMENT

In conclusion, music genre classification using CNNs is a powerful application of machine learning that has shown promising results in accurately categorizing music into different genres. By leveraging the ability of CNNs to automatically extract relevant features from audio signals, we can build models that can classify music genres with high accuracy. With continued research and development, we can build more accurate and robust models for music genre classification that can help us better understand the nuances and characteristics of different genres of music.

### 7.1 Limitations

* **Limited access to high-quality audio data:** Accuracy could be increased if the dataset contained more songs
* **Lack of domain expertise:** Since music genre classification is a complex topic that requires knowledge of both music theory and machine learning we were not able to perform this project at high level
* **Limited Time:** Due to our limited time we couldn’t do more research which would let us learn to optimize our model to increase it’s accuracy
* **Limited computing resources:** We also lack high tech computing resources for accurate depth training and prediction of our model

### 7.2 Future Enhancements

* **Using more sophisticated CNN architectures:** While traditional CNNs can be effective for music genre classification, more advanced architectures such as recurrent or attention-based models may improve performance by better capturing temporal dependencies in the audio signal.
* **Incorporating additional features:** In addition to spectrograms, other features such as lyrics or metadata could be incorporated into the classification model to improve accuracy.
* **Applying transfer learning:** Pre-trained CNN models on larger datasets could be used as feature extractors for the task of music genre classification. This approach can be especially useful when the available training data is limited.
* **Combining multiple models:** Ensemble models that combine the predictions of multiple CNN models trained on different subsets of the data may improve overall classification accuracy.

## REFFERENCES

|  |  |  |
| --- | --- | --- |
| [1] |  | S. S. e. al., "MUSIC GENRE CLASSIFICATION USING MACHINE LEARNING," Prasanna Naik, Bangalore, Karnataka, India, 2021. [Accessed 14 12 2022]. |
| [2] |  | A. V. Oppenheim, "irjmets," Journal of the Acostical Society of America, 12 05 2004. [Online]. Available: https://www.irjmets.com/. [Accessed 17 12 2022]. |
| [3] |  | B. L. a. A. Salomon, "A music similarity function based on signal analysis," Prasanna Naik, Tokyo, Japan, 2001. [Accessed 15 12 2022]. |
| [4] |  | J. J. a. F. Pachet, "Music Genre Classification," *Journal of Modernization in Engineering Technology and Science,* vol. 03, no. 07/May/2021, p. 957, 2021. |
| [5] |  | Oinkina, "Understanding LSTM Networks," *Recurrent Neural Networks,* vol. 05, no. August 27, 2015, p. 43, 2015. [Accessed 16 12 2022]. |
| [6] |  | A. Chowdhry, "clairvoyant," an EXL company, 07 05 2021. [Online]. Available: https://www.clairvoyant.ai/blog/music-genre-classification-using-cnn. [Accessed 17 12 2022]. |
| [7] |  | M. Saint-Félix, "towards data science," Towards Data Science, 20 07 2021. [Online]. Available: https://towardsdatascience.com/music-genre-detection-with-deep-learning. [Accessed 16 12 2022]. |