GenreBeats: A Hybrid CNN-LSTM Approach for Music Genre Classification/Detection

Jhansi Vazram Bolla
Professor
Department of Information Technology
Narasaraopeta Engineering College,
Narasaraopeta,
Andhra Pradesh.
jhansi.bolla@gmail.com

Venkata Hari Pavan Kota
Student
Department of Information Technology
Narasaraopeta Engineering College,
Narasaraopeta,
Andhra Pradesh.
haripavankota0909@gmail.com

Dola Sankar Sela
Student
Department of Information Technology
Narasaraopeta Engineering College,
Narasaraopeta,
Andhra Pradesh.
dolasankarsela123@gmail.com

Segu Akash
Student
Department of Information Technology
Narasaraopeta Engineering College,
Narasaraopeta,
Andhra Pradesh.
ckowshik604@gmail.com

Abstract: Using a limited amount of music genres, Music Genre Classification (MGC) model automatically classifies musics into distinct genres based on frequencies and qualities. This subject is very important for retrieval of music information, since it offers a method for arranging, analysing files. Convolutional machine learning methods like KNN, RNN - LSTM, FNN, SVM can be used for MGC. In order to divide the audio files into several genres, these algorithms are trained to identify various musical characteristics. Research indicates that deep learning algorithms like CNN and other outperform traditional machine learning algorithms in a variety of applications. As a result, the CNN algorithm is modified to carry out the music file classification. This uses CNN deep learning algorithms to classify musical genres. Accuracy is used to assess how well the MGC algorithms work. It is used in music production, music education, automated music recommendation systems. With CNN, the MGC task is completed with an accuracy of 97%.

1: Introduction

Music Genre Classification refers to the process of categorizing musical content of various groups based on distinctive musical characteristics. This classification plays a significant role in music information retrieval, with applications in playlist Generation, audio analysis, and music recommendation system. David et al. [1] postulated the method to evaluate current most common feature extraction toolboxes and libraries for an audio feature. It will be necessary to critically, analyze their coverage, efforts, and presentation. There are several toolboxes for extracting audio features, such as “Essentia” which is recommended by D. Bogdanov et al.[2], “Librosa”, which can be found in B. McFee et al.[3], which describes a Python Package for audio and music signal processing, and so on. In this work, we discuss the methods and techniques used to classify music genres.

Furthermore, this paper discusses the issues concerning the music genre classification, focusing on high levels of required accuracy for classification of various genres and applicability of classification algorithms in conditions where the signal quality would be blurred by the distortion and noise. At the end of this work, it will be possible for the reader to understand music genre classification fully and to apply the same when carrying out his/her analysis.

Music, as an art form, merges melodies and sounds to create powerful emotional experiences. It is a universal form of expression that transcends cultural and historical boundaries. Music can be created through a wide array of instruments, including stringed, wind-based, and digital tools. Common genres which are classical, country, disco, hip-hop, jazz, metal, pop, blues, reggae, and rock, each with unique characteristics often tied to particular social, historical, and cultural contexts.

Mel-frequency cepstral co-efficients (MFCC) is categorization of musical instruments is described by R. Loughran et al.[4]. MFCCs stands for the spectrum of audio signals and contains a lot of significant information while keeping compact. The feature set described for music analysis below is a result of learning from previous research and experience. J. H. Jensen et al.,[5] briefly describes in the previous research how the so called MFCC is being calculated. They preserve features of sound such as its tone quality and pitch which are can be effectively used in categorizing musical styles. In recent years, the various fields of Artificial Intelligence have made great strides especially in Machine learning, Deep learning, they have transformed those approaches to music genre classification. Traditional methods such as statistical models and decision trees have given way to more advanced neural network-based techniques. This study presents a unique approach connecting Convolutional Neural Networks (CNNs), Long Short-Term Memory Networks (LSTMs), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). LSTMs are used for capturing intricate features from audio signals, including spectral centroid, MFCCs and Librosa while CNNs are utilized to further refine these features. The outputs from both networks are fed into a dense layer with softmax activation for classification.

Our model was tested on publicly available dataset, achieving high precision, which underscores the effectiveness of using CNNs and LSTMs together for music genre classification. The insights from this study could be beneficial in developing more reliable and accurate algorithms, which in turn, may enhance the performance of music recommendation systems and improve the user experiences in music streaming services.

2: Related Work

In this research, classification of music genre is performed using the approaches that are related to both machine learning, deep learning. Because of development in machine learning, this study return to the traditional approach of analysing instruments in music by relying on automated methods.

C. Chang et al.,[6] applied Convolutional Neural Networks in which there were 5 layers of Convolution for the classification genres of music. The accuracy that they achieved was 83.3%. The size of the hop allowed was 256 and the Fast Fourier Transform applied to 1024 frames. Another work was N. Ndou et al.,[7] where Traditional Machine Learning and Deep Learning approach were adopted. They submitted a well-analysed paper on those approaches. They were able to decide their study with a 92.96% accuracy by k-Nearest Neighbours. They have also applied GTZAN dataset on the classification of music genres. Mentioned research like M. Ashraf et al.,[8] have used the hybrid CNN and a variant of RNN model for achieving up to 89.30% accuracy and N. Ndou et al.,[7] achieved 92.69% using k-Nearest Neighbours. This has a great influence on the new generation of digital technologies, and provides accurate music recommendations and personalized playlist for streaming to make the discovery of users more valuable. Probably, the further development of the approaches and models for music genre classification is regarded as the highly prospective issue from the standpoint of the unceasing evolutions of the various approaches to the machine learning as well as the electric music. This research also looks forward to more accuracy of the genre classification systems in the next couple of year.

3: Problem Statement and Objective

A. Problem Statement

Music plays a vital role in enriching people's lives, with individuals showing distinct preferences based on their personal tastes. However, it is quite fragile because music is rather an individual concept and is rather difficult to categorize according to some specific genres. An accurate automatic classification of the music by genre is required, especially in the use cases like Music Recommender Systems or Music Streaming Services. Doing this manually is time-consuming and requires substantial domain expertise. Moreover, the task is made complex by the overlapping nature of genres and the subjectivity involved in classification.

To address these challenges, deep learning technique, Convolutional Neural Networks (CNNs) is useful for automation in music genre classification (MGC). CNNs are used to identify patterns of audio data, making them ideal for this task. The purpose of this particular study is therefore to design an appropriate model of automatic genre classification using CNN for effective and credible music categorization.

B. Objectives

- To understand various audio formats and their features.
- To extracting the Mel-Frequency Cepstral Coefficients (MFCCs) from processed audio files, which are essential for feature representations.
- This study has extracted the features of the WAV-formatted 30 sec music files that were provided in the dataset so that one re-use those extracted features to train the proposed model and test or evaluate it.
- Here we attempted CNN, LSTM, SVM, KNN to classify the genres of music and then proposed the modified Convolutional Neural Network model, which gives the best accuracy on (GTZAN) dataset for MGC among all the trained models.
- The inquiry has evaluated the proposed Convolutional Neural Network model and compared it to other models and to some well-known published papers.

C. Challenges in Music Genre Classification

The task in music genre classification involves identifying genres of musical piece solely based on its auditory content. While this has useful applications in areas of music recommendation systems, automatic transcription, music streaming services, it is a complex task. Music genres are inherently subjective, and there is significant variability even within a single genre. Additionally, genres often have subgenres or influences from other genres, further complicating the classification the classification process.

A key challenge in music genre classification is that different listeners may categorize the same piece of music into different genres and there is no universal genre taxonomy that everyone agrees upon. This subjectivity adds to the complexity of creating an accurate classification model.

Recent advances in machine learning have provided several methods for genre classification, including statistical models, decision trees, and neural networks. However, the complexity of music genres and the high dimensionality of audio data pose significant challenges for these approaches. CNNs, with their abilities to handle large amounts of data and capture intricate patterns in audio spectrograms, offer a promising result for addressing challenges and increasing accuracy of music genre classification.

4: Methodologies

The methodology section outlines the processes involved in building a robust music genre classification system, from data collection to evaluation of model's performance. This includes data preprocessing techniques, features extraction and application of deep learning algorithms to achieve desired classification accuracy.

A. Data Collection

The dataset analyzed in this study are derived from well-known source that (GTZAN) Dataset. The (GTZAN) dataset consists 1,000 audio tracks of 10 genres, each track being a 30 seconds WAV file, providing a balanced dataset for training purposes. The dataset provides the genre representation that highlighting their distinct pattern and spectral variations, Fig.1 and Fig.2 depict sample wave plots from the datasets.

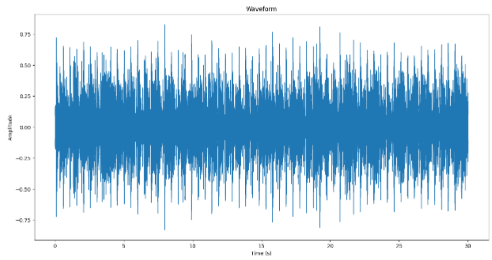

Fig.1: Waveplot of a Genre

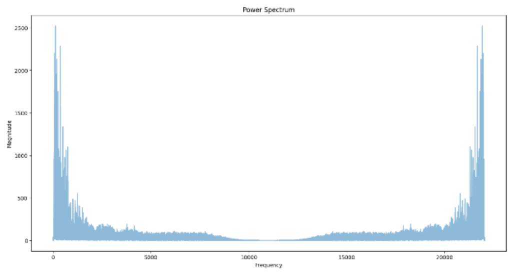

Fig.2: Spectrum of a Genre

B. Data Preprocessing

In the preprocessing phase, each 30-seconds audio file is converted into Mel-spectrograms after being sampled at a rate of 22,050 Hz. The audio signal is first divided into overlapping chunks using a 4- second window with a 2-second overlap, creating 15 chunks per file. This chunking technique helps reduce noise and makes patterns in the audio more visible. Audio segments are then processed using Mel-Frequency Cepstral Coefficients (MFCCs), which capture critical sound features by focusing on perceptual aspects of the audio. This method compresses the audio data, making it less susceptible to noise while maintaining key features that differentiate genres.

Feature extraction plays a vital role in the performance of the model. By isolating the most relevant parts of the data, the model can more easily distinguish between genres. Fig.3 & Fig.4 shows the extracted MFCCs, with the time on x-axis and the MFCC index on the y-axis. The colour in the plot represents the value of coefficients, which helps visualize important aspects of the sound data.

Moreover, techniques like spectral roll-off and zero-crossing rate are applied to better understand the tonal and rhythmic characteristics of audio. The spectral roll-off helps capture the point where most of the signal's energy is concentrated, giving insight into how high-pitched or energetic the track is. The zero-crossing rate, which measures how often the signal crosses the horizontal axis, is useful for distinguishing between percussive genres (like hip-hop) and more melodious ones (like classical or jazz).

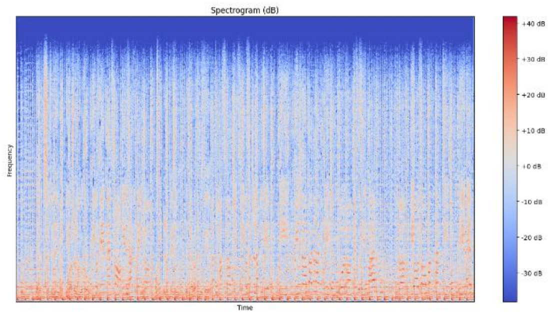

Fig.3: Spectogram of a Genre

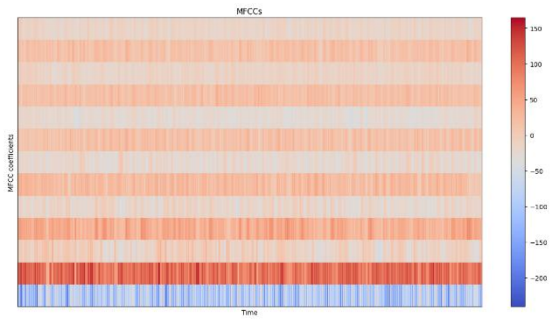

Fig.4: MFCCs coefficients of a Genres after Extraction

C. Feature Engineering

The feature engineering process is critical in improving the test accuracy of the classification task. Several features, including MFCCs, librosa, softmax, Spectral Centroid, Spectral Contrast, Magphase, were extracted to capture different sound attributes. These features represent various aspects of the audio signal such as: MFCC, Librosa, Stft, Spectral Centroid, Spectral Contrast, Magphase, Softmax.

1. **MFCC:** It is a method for compressing a spectrum into a smaller number of coefficients. MFCCs are often used for speech recognition and music information retrieval.
2. **Librosa:** A python package that can analyze and manipulate audio files, and extract key audio features and metrics. Librosa can be used to process audio files to generate MFCCs, as well as other values like magnitude and stft. The **librosa.feature.mfcc** method in Librosa can be used to obtain MFCCs by setting arguments like the no.of frames, hop length, and no.of MFCCs.
3. **Stft:** A short-time Fourier transform (STFT) spectrogram is a time frequency representation a signal that shows how frequency content changes over time. The STFT spectrogram is the squared magnitude of the STFT coefficients. It's a quick and simple way to analyze the time-frequency domain.
4. **Spectral Centroid:** Identifies "center of mass" of spectrum, aiding in distinguishing brighter sounds (like electronic music) from darker sounds (like classical or jazz).
5. **Spectral Contrast:** Measure the contrast in energy across frequency bands, which helps capture the textual variations in genres like jazz or hip-hop.
6. **Magphase:** Extracts magnitude and phase from the Fourier Transform, providing richer information about the signal's frequency content.
7. **Softmax:** It converts raw output scores (logits) into interpretable probabilities for multi-class classification tasks, usually applied at the final layer of a neural network.

D. Classification

Before training, dataset is split into training (70%) and testing (30%) sets. The CNNs architecture designed to work with MFCC features extracted from the audio files. The CNN then learns to classify genres by identifying patterns in the Mel-spectrograms associated with each genre. Fig.5 illustrates architecture of CNN model used in this work.

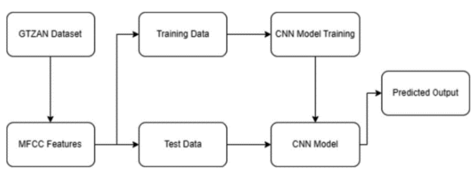

Fig.5: Flowchart of a CNNs Model

E. Architecture of CNN Model

In CNN architecture typically consists several layers stacked sequentially, each playing a vital role in extracting and learning features from input data. The fundamental components include convolutional layers, pooling layers, fully connected layers, and an output layer designed for classification tasks. The basic operation in convolutional layers is represented mathematically by the following formula for 2D convolution:

$$G = [\mathrm{i},\mathrm{j}] \sum_{u=-k}^{k} H[u,v].F[i-u,j-v]$$

Here, $G = [i\,,j]$ is output at position (i,j), representing the result of the convolution. $H = [u,v]$ is the filter or kernel, a matrix of weights applied to input signal F during the convolution operation. The expression $F[i-u,j-v]$ corresponds to the input signal being shifted by (u,v).

Convolutional layers are designed to detect spatial features such as edges, texture, and more complex patterns within the input data. These layers learn a set of filters that perform convolution across the input, capturing local features at different scales. As CNNs has evolved, various improvements have been proposed in research to optimize their performance, where CNNs were used to extract musical patterns from mel-spectrograms of audio signals.

In this project, the CNN architecture consists multiple convolutional layers, each detecting more intricate features as they progress. Pooling layers are used for down-sampling, reducing spatial dimensions while preserving essential information. The output from these layer is flattened and passed through fully connected (dense) layers, which further process high-level features. For multi-class classification tasks, a softmax activation function applied at final output layer to generate probabilities for each class. The softmax formula is as follows:

$$P(yi) = \frac{e^{zi}}{\sum_{j=1}^{k} e^{zj}}$$

In this formula, $P(yi)$ represents the probability of the input in the class i, where zi are class scores, exponentiated to ensure they are positive. These scores are normalized for producing a probability distribution over all classes.

In CNN model used here, the first convolutional layer employs 128 filters of size 3x3, activated using the ReLU function:

$$f(x) = max\,(0, x)$$

This is followed by max-pooling layers, batch normalization, which stabilizes learning process. Additional convolutional layers are applied, including a final layer with 64 filters of size 2x2. The feature map is passed through fully connected layers, including dense layer with 64 units, ReLU activation. Dropout is added for regularization, preventing overfitting during training. The softmax output layer handles the final classification task across 10 genres.

This structured CNN architecture, as illustrated in Fig.7, demonstrates a hierarchical approach to feature extraction, making it suitable for a variety of tasks requiring robust pattern recognition and classification.

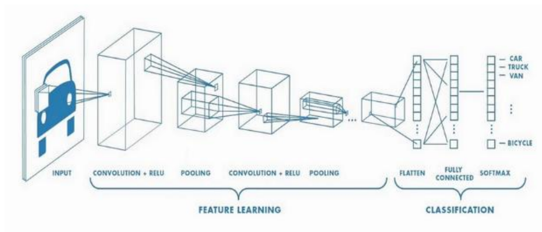

Fig.6: Architecture of CNN Model

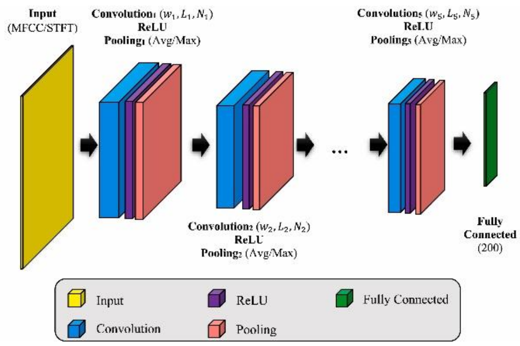

Fig.7: Layers of CNN

F. Architecture of RNN-LSTM Model

This study introduces a sophisticated Recurrent Neural Network (RNN) architecture it is designed specifically for sequential data processing and the extraction for nuanced patterns from audio inputs, enhancing the precision of music genre classification. Each component in the architecture plays a distinct role in identifying temporal and sequential aspects of music, which are key to its structure.

RNNs, particularly suited for tasks involving temporal sequences, are well-equipped to capturing both short-term and long-term dependencies of data, allowing them to music analysis. The study of RNN accepts audio data in the form of sequential feature vectors, where MFCCs serve as primary feature representation. MFCCs encapsulate the most crucial spectral information that reflects the timbral characteristics of music.

The implementation of this architecture specifies the use of recurrent layers with Long short-term memory (LSTM) cells. LSTM layers are a powerful enhancement to traditional RNNs, enabling the network to learn long-term dependencies by mitigating the vanishing gradient problem. The study employs multiple LSTM layers capture to intricate temporal patterns in music. These recurrent layers feed into fully connected layers with Rectified Linear Units (ReLU) for non-linear transformation, enabling model to discern complex relationships in data.

To ensure the model's generalizability and prevent overfitting, several regularization techniques are employed. Dropout layers, set with a probability of 0.3, are strategically placed between fully connected layers. This forces the model to learn a more diverse set of features by randomly dropping some neurons during training. The output layer, which corresponds to the no.of music genres 10, uses a softmax activation functions generates probability distribution across all possible genres. The genre with topmost probability that is selected for model's final prediction. The softmax function is determined as follows:

$$P(yi) = \frac{e^{zi}}{\sum_{j=1}^{k} e^{zj}}$$

Where $\boldsymbol{P(yi)}$ is predicted probability for class i, and zi is input score of that class. The softmax function ensures the output values sum to 1, producing a valid probability distribution across all classes.

The RNN architecture depicted in Fig.8, is combined with multiple LSTM layers, each configured with 128 and 64 units, respectively. These layers are equipped with L2 regularization to enhance generalization. After the recurrent layers, Dense (fully connected) layers with ReLU activations are added, followed by drop-out layers to further counter overfitting. The model's final layer utilizes softmax activation to categorize the input into one of the 10 music genres.

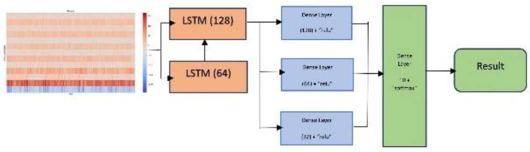

Fig.8: Architecture of RNN-LSTM Model

By leveraging the sequential nature of music data, this architecture effectively captured temporal correlations, proving its potential to significantly improves the accuracy of genre classification tasks.

5: Result and Analysis

The evaluation of various models in this study revealed that the proposed Convolutional Neural Network (CNN) model significantly outperformed other techniques for the task of music genre classification. CNN model achieves remarkable accuracy 96.68% on (GTZAN) dataset, demonstrating a substantial improvement over the base CNN, which provided an accuracy of 85.56%.

Other models such as Support vector machine (SVM), Long short-term memory (LSTM), and k-Nearest neighbors (kNN) achieved. The detailed comparison of results is summarized in Table.1, with the CNN model demonstrating superior classification performance for the task.

Method	Accuracy
CNN	96.68%
RNN-LSTM	89.75%
SVM	85.29%
kNN	88.52%

Table.1: Accuracy on GTZAN Dataset

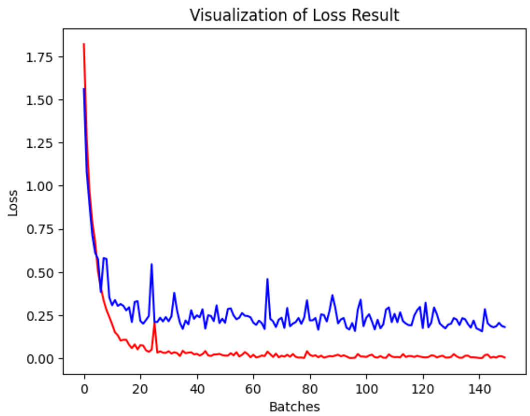

Fig.9: CNN model loss after 150 epochs

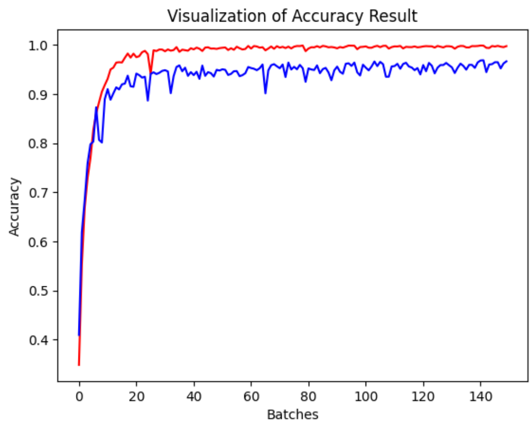

Fig.10: Accuracy of CNN model after 150 epochs

In the implementation of the CNN model for this project, the architecture involved several convolutional layers along increasing filter sizes: 32, 64, 128 and finally 512 filters, each with a kernel size of 4x4 and ReLU activation functions. A drop-out rate of 0.3 is to regularize model and prevents overfitting. The model was trained for a total of 150 epochs using a batch size of 32, yielding test accuracy 96.68% and train accuracy of 99.94%, showcasing its high performance on unseen data.

Similarity, the LSTM model used in this study comprised layers with 128 units, followed by fully connected layers with 64 units. Dropout layers were applied with a rate of 0.5 to improve generalization. This model was trained up-to 250 epochs, achieving train accuracy 99.64% and test accuracy of 89.75%, which, while lower than CNN model, still performed commendably on temporal data.

The performance of other models such as SVM and kNN was also evaluated, with the SVM model yielding a train accuracy of 91.29% and test accuracy of 85.29%, and kNN model providing a train accuracy of 93.35% and test accuracy of 88.52%. These results underline the robustness of CNN model architecture in extracting spatial patterns from audio data, while LSTM models show promise in handling sequential dependencies, albeit with slightly lower classification performance.

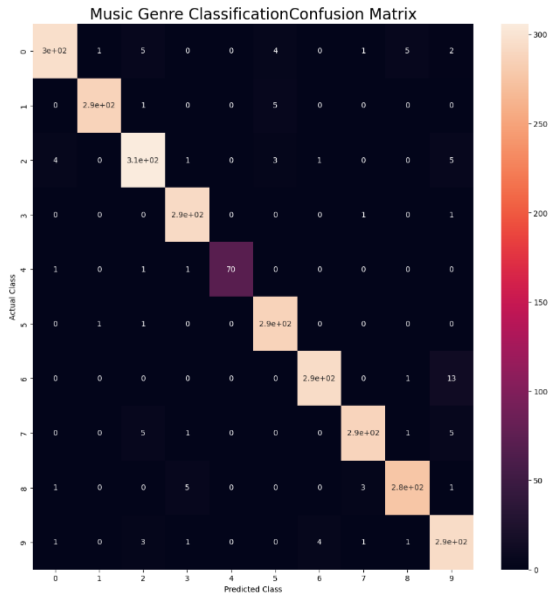

Fig.11: Confusion Matrix

This in particular refers to precision — i.e., the percentage of total number of true positive predictions from the model that is actually correct.

$$Precision = \frac{TP}{TP + FP}$$

The goal of recall is to find the most appropriate items from all the available options. It quantifies the percentage of predictions that came true out of the dataset's actual positive cases. Where TP is true positive and FN is false negative.

$$Recall = \frac{TP}{TP + FN}$$

In the case of classification models used in different fields, F1 Score is a must have metric because it balances precision and recall and how well it can deal with imbalanced datasets. The F1 score was computed using the below formulae:

$$F1 = \frac{2 * Precision * Recall}{Precison + Recall}$$
$$= \frac{2 * TP}{2 * TP + FP + FN}$$

B: Future Scope

The Future Scope of Music Genre Classification holds exciting possibilities across various dimensions, driven by advances in Artificial Intelligence, Deep Learning and Big Data. Some of the key areas are: Improved Model Accuracy with Advanced Architectures, Real-Time Classification and Personalizations, Genre Blending and Sub-genre Detection, Integration with Music Creation Tools, Cross-cultural Genre Classification, Emotional and Sentiment Analysis, Augmenting other audio Recognition application.

6: Conclusion

This study implements a Convolutional Neural Network (CNNs) to achieve 96.68% accuracy in classifying of music genres, outperforming other models and demonstrating its potential for precise genre classification. Each audio file was split into shorter segments, and Mel-frequency cepstral co-efficients (MFCCs) are extracted from segments to represent audio features. By encoding the sound into a compact feature space, MFCCs reduce the complexity of the data, allowing deep learning models to process it more efficiently. The CNN model excels at capturing genre-specific attributes like rhythm and timbre, making it ideal for music genre classification tasks. The model performed best with WAV files, but further research is needed to accommodate other formats like MP3 and FLAC. Though CNNs have proven highly effective for this task, genre classification remains complex due to culture and historical influences. Future research could explore additional audio characteristics like rhythmic patterns or lyrics, and ensemble learning or attention mechanisms may help refine results.

7: References

[1] D. Moffat, D. Ronan, and J. D. Reiss. An evaluation of audio feature extraction tool boxes, doi: 10.13140/RG.2.1.1471.4640, 2015.

[2] D. Bogdanov, N. Wack, E. Gomez Guti ́ errez, S. Gulati, ́ H. Boyer, O. Mayor, G. Roma Trepat, J. Salamon, J. R. Zapata Gonzalez, X. Serra, et al. Essentia: An audio ́ analysis library for music information retrieval, 2013.

[3] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, and O. Nieto. librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference, volume 8, pages 18–25, 2015.

[4] R. Loughran, J. Walker, M. O'Neill, and M. O'Farrell. The use of mel-frequency cepstral coefficients in musical instrument identification. In ICMC, 2008.

[5] J. H. Jensen, M. G. Christensen, M. N. Murthi, and S. H. Jensen. Evaluation of mfcc estimation techniques for music similarity. In 2006 14th European Signal Processing Conference, pages 1–5. IEEE, 2006.

[6] Y.-H. C. P.-C. Chang and C.-N. Kuo. Convolutional neural networks approach for music genre classification. In International Symposium on Computer, Consumer and Control (IS3C), pages 309–403. IEEE, 2020.

[7] N. Ndou, R. Ajoodha, and A. Jadhav. Music genre classification: A review of deep-learning and traditional machine-learning approaches. In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), pages 1–6. IEEE, 2021.

[8] M. Ashraf, F. Abid, I. U. Din, J. Rasheed, M. Yesiltepe, S. F. Yeo, and M. T. Ersoy. A hybrid cnn and rnn variant model for music classification. Applied Sciences, 13(3): 1476, 2023.

[9] S. W. J, P. K. R M, P. K. K and P. J, "Music Genre Classification Using LSTM and CNN", *3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, pp. 205-209, doi: 10.1109/ICPCSN58827.2023.00039, 2023.

[10] N. Choudhury, D. Deka, S. Sarmah and P. Sarma, "Music Genre Classification Using Convolutional Neural Network," *2023 4th International Conference on Computing and Communication Systems (I3CS)*, Shillong, India, pp. 1-5, doi: 10.1109/I3CS58314.2023.10127554, 2023.

[11] S. P. S, M. B S, D. S and K. P, "Utilizing Spectrograms and Deep Learning Techniques for Improved Music Genre Classification," *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, Kollam, India. pp. 1723-1728, doi: 10.1109/ICCPCT61902.2024.10673160, 2024.

[12] M. H. Yimer, Y. Yu, K. Adu, E. Favour, S. M. Liyih and R. A. Patamia, "Music Genre Classification using Deep Neural Networks," *2023 35th Chinese Control and Decision Conference (CCDC)*, Yichang, China, pp. 2384-2391, doi: 10.1109/CCDC58219.2023.10327367, 2023.

[13] T. R. Ragav, D. B and K. A, "Musical Genre Classification Using Support Vector Machines and Convolutional Neural Networks," *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, Hamburg, Germany, pp. 389-393, doi: 10.1109/ICCCMLA58983.2023.10346701, 2023.

[14] S. Si, "A Feature Extraction Intelligent Music Genre Classification Method Based on Deep Learning Technology," *2024 IEEE 2nd International Conference on Control, Electronics and Computer Technology (ICCECT)*, Jilin, China, pp. 1084-1088, doi:

10.1109/ICCECT60629.2024.10545770, 2024.

[15] T. Shaikh and A. Jadhav, Music Genre Classification Using Neural Network, in Proceedings of the International Conference on Automation, Computing and Communication, ICACC, 05 May 2022, Mumbai, India, 2022.

[16] P. Kaushik, S. P. Singh Rathore, K. Chahal, S. Saraf, G. Singh Chauhan and P. Kumar, "RhythmQuest: Unifying Indian Music Classification and Prediction with Hybrid Deep Learning Techniques," *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2024, pp. 1-6, doi: 10.1109/IATMSI60426.2024.10503056, 2024.