Leveraging Spectrogram Features and Convolutional Neural Networks for Music Genre Classification

Dr. David Raj Micheal
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
davidraj.micheal@vit.ac.in

Ramya V
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology Chennai
Tamil Nadu – 600127
ramya.v2023@vitstudent.ac.in

***Abstract*—Music genre classification is a crucial task in music information retrieval, with applications including recommendation systems, content organization, and media categorization. This study aims to develop an automated music genre classification system using deep learning, specifically utilizing the VGGish model, a pre-trained neural network for audio feature extraction. The system is trained and evaluated on the GTZAN dataset, which contains 1,000 audio files uniformly distributed across ten distinct music genres. The methodology utilizes the VGGish model to extract features by transforming raw audio files into Mel spectrograms, a time-frequency representation that captures essential audio characteristics. Then, a deep convolutional neural network is trained on these extracted features to classify each audio track into its appropriate genre. The study enhances the feature extraction capabilities by fine-tuning the VGGish model on the GTZAN dataset to improve music genre classification. The model's performance is assessed using metrics including accuracy, precision, recall, and F1-score. The study aims to provide a highly accurate and efficient system for classifying music genres, with potential applications in music recommendation engines and other audio analysis platforms.**

***Index Terms*—Deep Learning; Mel spectrograms; Music genre classification; Transfer learning; VGGish**

I. Introduction

Music genre classification is a crucial task in the field of music information retrieval (MIR), which involves automatically categorizing songs into predefined genres based on their audio characteristics. This task has gained significant attention due to its wide-ranging applications in streaming services, recommendation systems, and music organization. The growing volume of digital music available on various platforms has made it essential to develop efficient and accurate systems for genre classification. Traditional approaches to music genre classification relied on manual feature extraction methods, using attributes such as rhythm, timbre, and harmony. However, with the advancements in deep learning, modern techniques now leverage powerful neural networks that can automatically extract complex features from raw audio data, significantly improving classification performance. Spectrograms, which represent the frequency components of an audio signal over time, are commonly used as input features in these systems, as they provide a visual representation of audio that deep learning models can effectively process. In this project, we utilize the VGGish model, a variant of the popular VGG16 architecture fine-tuned for audio tasks, to classify music genres. By converting raw audio files into spectrograms, we aim to leverage the model's ability to learn intricate patterns in the data, thereby achieving improved accuracy in genre prediction. This approach highlights the synergy between feature-rich spectrograms and powerful pre-trained deep learning models for building robust music classification systems.

II. Objective

Explore the potential of deep learning in solving the complex task of music genre classification by identifying relevant patterns and features in audio data. Develop an automated system for music categorization, capable of accurately classifying music into various genres, contributing to more efficient content organization. Contribute to the field of music information retrieval, offering insights into how advanced computational techniques can improve audio data processing and analysis. Improve existing music classification systems, providing more reliable and accurate methods for genre identification, which can be applied to various audio-based applications. Demonstrate the applicability of AI-driven solutions in real world scenarios, such as music recommendation engines and media categorization, enhancing user experience and system efficiency.

III. Related Works

Elbir and Aydin (2020) [1] investigate the use of deep learning for enhancing music genre classification and recommendation systems. They apply Convolutional Neural Networks (CNNs) to analyze spectrograms of audio signals, demonstrating improved accuracy in genre classification compared to traditional methods. Their approach not only refines genre classification but also integrates with collaborative filtering for more effective music recommendations. The study highlights the effectiveness of advanced neural network models in processing complex audio data and providing personalized music suggestions, contributing significantly to the fields of music data analysis and recommendation systems.

Yang, Feng, Wang, Yao, and Luo (2020) [2] propose a parallel recurrent convolutional neural network (RRCNN) approach

for music genre classification tailored for mobile devices. Their method combines convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to leverage both spatial and temporal features of music data effectively. The RRCNN model is designed to be computationally efficient, making it suitable for deployment on mobile platforms. The authors demonstrate that their approach significantly improves classification accuracy while maintaining low computational requirements, making it feasible for real-time genre classification on mobile devices. This work advances the field by providing a robust and efficient solution for genre classification that is well-suited for resource-constrained environments.

Wijaya and Muslikh (2024) [3] present a music genre classification method that combines Bidirectional Long Short-Term Memory (BiLSTM) networks with Mel-Frequency Cepstral Coefficients (MFCCs). Their approach leverages BiLSTM networks to capture both past and future context in sequential audio data, enhancing the model's ability to classify genres accurately. MFCCs are used as input features to represent the audio signal's spectral characteristics. The study shows that the integration of BiLSTM with MFCCs leads to significant improvements in classification performance compared to traditional methods. This work contributes to the field by providing a robust technique for genre classification that effectively utilizes sequential dependencies and spectral features, thereby advancing the accuracy and reliability of music genre classification systems.

Zhang (2021) [4] introduces a music style classification algorithm that combines music feature extraction with deep neural networks (DNNs). The study focuses on extracting meaningful features from music data, which are then fed into a deep neural network for classification. Zhang's approach emphasizes the importance of effective feature extraction in enhancing the performance of deep learning models. The algorithm demonstrates improved accuracy in classifying various music styles by leveraging complex patterns identified through the DNN. The research highlights the effectiveness of integrating sophisticated feature extraction techniques with advanced neural network architectures, advancing the field of music style classification by providing a more accurate and robust solution for categorizing music based on its stylistic elements.

Zhuang, Chen, and Zheng (2020) [5] explore the application of a transformer-based model for music genre classification. The authors utilize the transformer architecture, known for its success in natural language processing, to capture both local and global dependencies within music data. The model processes sequences of audio features and efficiently classifies music genres by learning long-term dependencies. Their results demonstrate that the transformer classifier outperforms traditional machine learning models and some deep learning architectures in terms of classification accuracy. This work highlights the potential of transformer models for music classification tasks, contributing to the field by providing an innovative approach that leverages attention mechanisms for improved performance in genre classification.

Mehta, Gandhi, Thakur, and Kanani (2021) [6] review various music genre classification methods, initially discussing traditional approaches based on handcrafted features like timbre, pitch, and rhythm, which often fail to fully capture the complexity of musical audio. They then focus on deep learning methods, particularly convolutional neural networks (CNNs), which have enhanced classification performance. The authors highlight the importance of mel spectrograms in representing audio data. A key part of the literature review involves transfer learning models, such as VGG16, ResNet50, and InceptionV3, which are pre-trained on large image datasets like ImageNet. These models are fine-tuned to extract deep audio features from log-based mel spectrograms, significantly improving classification accuracy and generalization in music genre tasks, especially with limited labeled data. Transfer learning thus stands out as an efficient solution for audio classification.

Prabhakar and Lee (2023) [7] review the advancements in music genre classification by comparing traditional and modern approaches. Early methods relied on handcrafted features such as rhythm, pitch, and timbre, but these techniques struggled with the complexity of audio signals and often resulted in suboptimal performance. With the rise of deep learning, convolutional neural networks (CNNs) have become a standard for improving classification accuracy by automatically learning relevant features. The authors emphasize the role of transfer learning in modern classification systems, discussing models such as EfficientNet and MobileNet, which are pre-trained on large-scale datasets like ImageNet. These models are adapted to audio tasks using mel spectrograms and other feature representations. The literature highlights how transfer learning reduces computational complexity and enhances performance, making it more efficient for real-time applications. The paper also discusses ensemble methods and hybrid deep learning techniques, offering holistic solutions for improving music genre classification outcomes across diverse datasets.

Jena, Bhoi, Mohapatra, and Bakshi (2023) [8] review various methods for music genre classification, emphasizing the hybrid deep learning approach that combines wavelet analysis with spectrogram analysis. Traditional methods often rely on handcrafted features and struggle with the complexity of musical data. The paper highlights how recent advancements in deep learning, particularly through hybrid models, enhance classification accuracy. The authors discuss the benefits of integrating wavelet transforms with spectrograms, which helps in capturing both time-frequency representations and hierarchical feature extraction. By combining these techniques with deep learning models, such as CNNs, the paper demonstrates improvements in classification performance and robustness. The review focuses on how this hybrid approach leverages the strengths of both wavelet and spectrogram analyses to handle diverse and complex music datasets effectively.

Ashraf, Abid, Din, Rasheed, Yesiltepe, Yeo, and Ersoy (2023) [9] review recent developments in music genre classification, particularly focusing on hybrid models that integrate convolutional neural networks (CNNs) and recurrent neural networks (RNNs). They discuss the limitations of traditional

methods, which rely on handcrafted features and often struggle to capture complex patterns in musical data. The paper highlights how CNNs are effective at extracting spatial features from audio spectrograms, while RNNs are adept at modeling temporal dependencies. By combining these two approaches, the authors demonstrate significant improvements in classification performance and robustness. The review explores various hybrid architectures, their performance metrics, and how these models enhance the accuracy of music genre classification by leveraging both spatial and temporal features of the audio data.

Sharma, Aggarwal, Bhardwaj, Chakrabarti, Chakrabarti, Abawajy, and Mahdin (2021) [10] review methods for classifying Indian classical music, focusing on time-series matching deep learning approaches. They discuss traditional classification methods that use handcrafted features and often struggle with the intricate patterns of Indian classical music. The paper highlights advancements in deep learning, specifically time-series models, that better capture temporal and sequential characteristics of music. The authors examine how models like Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCNs) are used to analyze time-series data effectively. They explore how these models handle the unique aspects of Indian classical music, such as complex rhythmic patterns and melodic variations, thus improving classification accuracy. The review provides insights into how deep learning techniques tailored for time-series analysis offer significant advantages over traditional methods in classifying Indian classical music.

Ceylan, Hardalaç, Kara, and Firat (2021) [11] review automatic music genre classification methods and their connections to music education. They begin by discussing traditional genre classification approaches that rely on handcrafted features and their limitations in accurately categorizing diverse music genres. The paper then focuses on the application of machine learning and deep learning techniques, such as convolutional neural networks (CNNs) and support vector machines (SVMs), to improve classification accuracy. The authors explore how these modern methods handle complex audio features more effectively than traditional approaches. Additionally, they examine the impact of automatic classification systems on music education, highlighting how these technologies can support music educators in analyzing and understanding musical content. The review provides insights into the benefits and challenges of integrating automatic genre classification with educational practices, emphasizing the potential for enhancing both music analysis and pedagogy.

Kostrzewa, Kaminski, and Brzeski (2021) [12] review the search for optimal neural network architectures for music genre classification. They discuss the evolution of classification methods from traditional techniques that rely on handcrafted features to modern approaches utilizing deep learning. The paper provides an overview of various neural network models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures that combine different types of networks. The authors evaluate the strengths and limitations of these models in terms of their ability to handle diverse and complex musical features. They also review recent advancements and innovations aimed at improving classification accuracy, such as deeper network architectures and novel training strategies. The review highlights ongoing challenges in finding the ”perfect” network architecture that can effectively and efficiently classify a wide range of music genres.

Vishnupriya and Meenakshi (2018) [13] review methods for automatic music genre classification, with a focus on convolutional neural networks (CNNs). They discuss traditional classification approaches that rely on handcrafted features, such as spectral and rhythmic characteristics, and their limitations in capturing complex musical patterns. The paper highlights the advantages of using CNNs, which are designed to automatically learn and extract relevant features from raw audio data, particularly from spectrogram representations. The authors examine the effectiveness of CNNs in improving classification accuracy compared to traditional methods. They also discuss various CNN architectures and their performance metrics, demonstrating how these models enhance the ability to classify different music genres by leveraging deep learning techniques. The review emphasizes the significant improvements achieved with CNNs in automatic music genre classification tasks.

Yang and Zhang (2019) [14] review advancements in music genre classification, specifically focusing on the use of duplicated convolutional layers in neural networks. They discuss traditional methods that rely on handcrafted features and their challenges in handling the complexity of music data. The paper introduces an innovative approach where neural networks are enhanced by duplicating convolutional layers to improve feature extraction and classification accuracy. By applying this method, the authors aim to capture more nuanced patterns and representations of musical features. The review highlights how duplicated convolutional layers address issues related to depth and feature learning in neural networks, resulting in more effective and precise music genre classification. The authors evaluate the performance of this approach and compare it with existing methods, demonstrating its potential for advancing classification tasks in the field.

Ghosal and Kolekar (2018) [15] explore advancements in music genre recognition, emphasizing the role of deep neural networks and transfer learning. They analyze the shortcomings of traditional methods that rely on handcrafted features, which often fail to capture the intricate and diverse characteristics of musical data. The paper highlights the advantages of employing deep neural networks, particularly through transfer learning, where pre-trained models are adapted for genre recognition tasks. This approach leverages features learned from large-scale datasets, improving performance on smaller, music-specific datasets. The authors examine various deep learning architectures and their effectiveness, illustrating how transfer learning enhances accuracy and reduces training time. The study provides a comprehensive look at how integrating deep learning with transfer learning addresses challenges in music genre recognition and pushes the boundaries of current techniques.

IV. METHODOLOGY

Dataset

The GTZAN dataset, often referred to as the ”MNIST of sounds,” is a widely used benchmark for music genre classification tasks. It consists of 1,000 audio files, each 30 seconds long, distributed evenly across ten distinct music genres: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, and rock. Each genre contains 100 audio clips, providing a balanced representation of diverse musical styles. The dataset has become a standard in audio analysis, offering a challenging testbed for genre classification models and other music information retrieval applications.

In the time domain representation, certain music genres exhibit distinct patterns that allow them to be easily distinguished from others, as demonstrated in Fig 1. For example, Jazz exhibits unique characteristics compared to genres like Rock and Metal. However, some genres share overlapping characteristics that make classification more challenging; Rock and Metal, for instance, display greater similarity to each other than to Jazz, making them harder to differentiate. Therefore, combining time domain features with additional techniques can help capture more informative characteristics, enhancing the classification process.

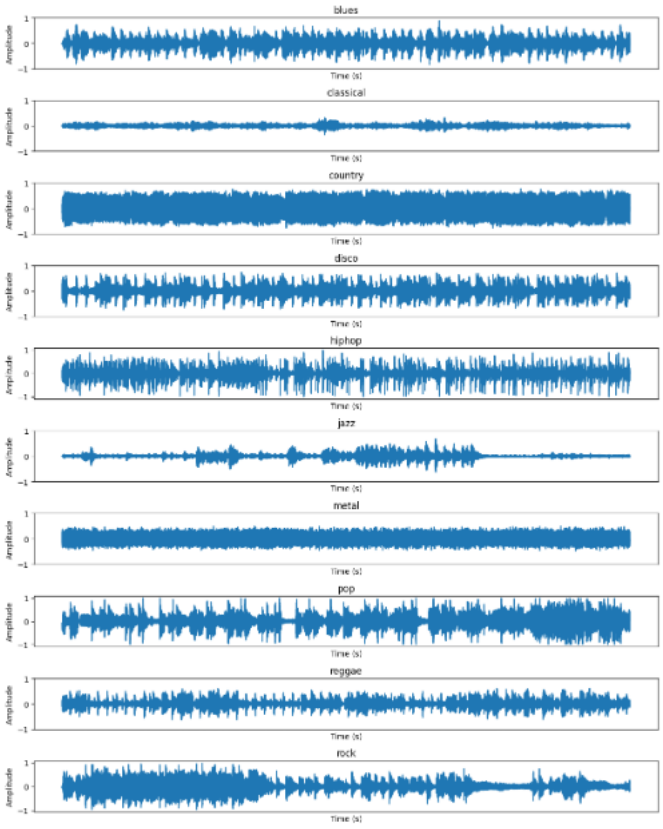

Fig 1 Time domain representation of genres

Preprocessing

Audio Augmentation: In audio classification tasks, a common challenge is acquiring a large, diverse dataset that fully captures real-world variations, which are often difficult to encompass. Audio augmentation techniques address this by expanding the dataset with controlled modifications to existing samples, thereby enhancing the model's generalizability and robustness. By introducing slight modifications, augmentation techniques create varied versions of existing audio files, enabling the model to learn patterns under different conditions, such as tempo, pitch, or background noise variations. Fig 2 illustrates the increased number of samples in each genre after augmentation is applied.

In the context of this music genre classification, audio augmentation is utilized to improve the model's resilience to natural variations in audio data by simulating real-world transformations, including changes in tempo, pitch, or the presence of background noise. These augmentations prevent overfitting by exposing the model to a broader variety of sounds without the need for additional data.

In this study, three key augmentation methods are utilized to enhance the diversity of audio data:

- **Time Stretching:** Time stretching adjusts the speed of audio playback without affecting the pitch. This method, generates a slower or faster version of a song, allowing the model to adapt to tempo variations. Internally, it achieves this by stretching or compressing the waveform along the time axis while preserving pitch, thus maintaining the frequency relationships.
- **Pitch Shifting:** Pitch shifting alters the pitch of an audio sample without changing its speed, helping the model recognize variations in pitch. This is achieved by modifying the waveform's frequency components, simulating changes such as vocal tone shifts or key adjustments in musical compositions.
- **Adding White Noise:** Adding random white noise introduces subtle background interference, enabling the model to recognize audio patterns in the presence of minor noise. This process involves generating and adding random noise to the waveform, thereby slightly modifying the signal's amplitude at various points.

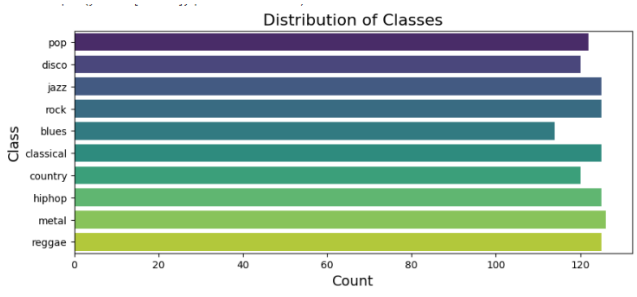

Fig 2 Genre distribution after audio augmentation

Feature Extraction Using VGGish: Feature extraction is a crucial step in transforming raw audio signals into a structured representation that captures the unique characteristics of each genre. VGGish, a pre-trained audio feature extraction model developed by Google, is highly effective for audio classification tasks as it generates embeddings based on the audio content, capturing both timbral and temporal patterns. Leveraging VGGish enables the model to use features that have already been fine-tuned for audio perception tasks, significantly improving the model's performance.

By utilizing a pre-trained model like VGGish, which has been trained on large-scale audio datasets, the classification model benefits from high-level representations of audio, including rhythmic patterns, tonalities, and textures. This allows the neural network to focus on learning the specific nuances of music genres, without the need to extract these basic audio features from scratch. The process of feature extraction using VGGish involves several stages:

- **Audio Processing:** VGGish expects audio to be in a specific format: mono audio with a sample rate of 16 kHz and a fixed length. The librosa library is used to load and process the audio to meet these requirements, which includes converting stereo audio to mono when necessary. Additionally, Mel-spectrograms are computed from the audio waveform. A Mel-spectrogram is a time-frequency representation of the audio that provides a compact and informative summary of the spectral properties of sound, which is crucial for the model to analyze audio. Fig. 3 shows the Mel-spectrogram representation of an audio clip, highlighting the frequency and temporal characteristics. This transformation from the raw waveform to Mel-spectrogram helps capture both the harmonic and rhythmic components of the audio.
- **Feature Embeddings:** Once the audio is preprocessed and transformed into a Mel-spectrogram, it is passed through the VGGish model. Internally, VGGish processes the Mel-spectrogram through its convolutional layers, capturing hierarchical audio features similar to those recognized by the human auditory system. This process transforms the raw audio into embeddings that capture key patterns in both frequency and time domains, making it easier for the model to learn musical features relevant to genre classification.
- **Embedding Vector Output:** VGGish outputs a set of embeddings, typically in the form of 128-dimensional vectors, which summarize the essential auditory features of the audio clip. These embeddings represent characteristics such as melody, harmony, rhythm, and timbre. By reducing the high-dimensional waveform data into a more compact form, VGGish's feature extraction simplifies the input for neural network training, while also reducing noise and standardizing the feature space. The use of Mel-spectrograms during the preprocessing stage ensures that the model effectively captures relevant audio features that improve classification performance.

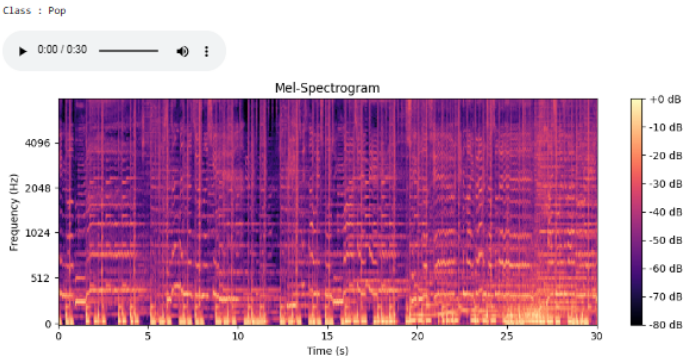

Fig 3 Mel-Spectrogram representation of an audio

Convolutional Neural Network

The Convolutional Neural Network (CNN) utilized for music genre classification is designed to learn hierarchical representations of the extracted audio features. The model receives feature vectors derived from Mel-spectrograms, which represent the time-frequency characteristics of audio signals. These Mel-spectrograms serve as the input to the network, providing a structured representation of the audio clip, capturing critical aspects like rhythm, tonalities, and harmonic structures that are essential for genre classification. Fig. 4 shows the architecture of the proposed CNN.

The network's first layers consist of convolutional filters that automatically learn spatial hierarchies in the input data. Initially, these convolutional layers detect basic patterns such as edges and frequency components. As the data progresses deeper into the network, the successive layers learn more complex patterns, such as musical rhythms or tonal nuances, which are distinct to different music genres. Following each convolutional operation, the output undergoes a non-linear transformation through the ReLU (Rectified Linear Unit) activation function. This introduces non-linearity, enabling the network to model complex relationships within the data and making the model more expressive.

Max pooling layers follow the convolutional layers to reduce the spatial dimensions of the feature maps, retaining the most important features while downsampling the data. This reduces computational complexity and helps prevent overfitting by focusing on the most salient patterns. As the data moves through the network, it passes through several convolutional and pooling layers, progressively abstracting the audio's features.

After the convolutional and pooling stages, the output is flattened and passed through fully connected layers, which perform high-level reasoning based on the extracted features. These fully connected layers synthesize the information captured by the earlier convolutional layers and make the final classification decision. The model ends with a softmax output layer, which produces a probability distribution over the possible music genres. The genre with the highest probability is selected as the model's prediction.

The CNN model is trained using the RMSprop optimizer, which adjusts the learning rate dynamically to stabilize training. This optimizer helps improve convergence by normalizing the gradient, making it effective in preventing issues like vanishing or exploding gradients. The loss function used is categorical cross-entropy, which is well-suited for multi-class classification tasks such as music genre classification. By combining these layers and techniques, the CNN effectively learns to recognize complex patterns in audio and classify music genres with high accuracy.

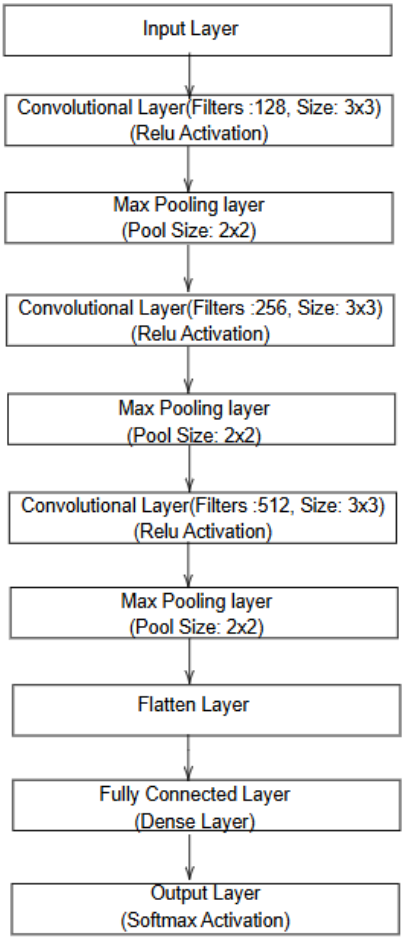

Fig 4 Architecture of CNN

Experimentation and Evaluation

This section, describes the proposed model's experimentation setup and evaluation procedure.

Experimentation Setup: Below is presented a list of hyper-parameters used during model training (Table 1).

Table 1 Model's Hyperparameters

Hyperparameter	Value
Learning Rate	0.001
RMSprop - rho	0.9
Epochs	30
Batch Size	32
Dropout Rate	0.6

Evaluation Measure: When assessing models for multi-class classification, various important metrics are considered to thoroughly gauge their performance. In addition to standard metrics like accuracy and loss, precision, recall, F1 score, and confusion matrices are also analyzed to better understand the model's classification abilities.

- **Accuracy:** Accuracy tells us the percentage of correctly classified samples out of the total. It gives us a general idea of how accurate the model is in its predictions. Accuracy is computed using (1).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{1}$$

- **Loss:** loss measures the discrepancy between predicted and actual labels. It shows us how well the model is doing during training, with lower values indicating better performance. Loss is computed using categorical cross-entropy.
- **Precision:** The ratio of the number of correctly predicted positive cases to all cases that are predicted as positive indicates precision. It shows how well the model avoids incorrect positive predictions. Precision is computed using (2).

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

- **Recall:** The percentage of positively confirmed true cases among all positive test results is known as recall, or sensitivity, or true positive rate. It illustrates how effectively the model captures each positive case. Recall is computed using (3)

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

- **F1 Score:** F1 score is a way for looking at a model's performance by considering both precision and recall. The model performance is certain from the aspect of both false positives and false negatives. F1-Score is computed using (4).

$$\text{F1 Score} = 2 * \frac{\text{Precision*Recall}}{\text{Precision+Recall}} \tag{4}$$

RESULTS

Fig 5. shows the optimal epochs for accuracy and loss of the model.

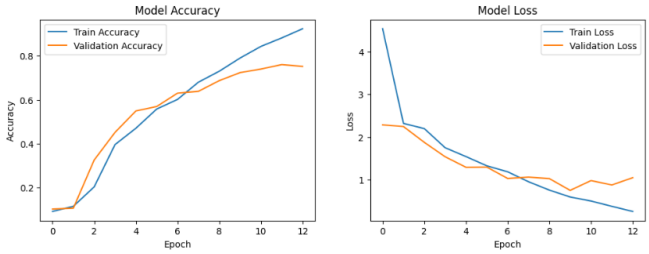

Fig 5 Accuracy and Loss graph of the model

Table 2 shows the Classification metrics achieved by the proposed CNN model

Table 2 Classification metrics

Genre	Precision	Recall	F1 Score
Blues	0.80	0.52	0.63
Classical	0.92	0.92	0.92
Country	0.61	0.71	0.65
Disco	0.64	0.58	0.61
Hip-hop	0.82	0.92	0.87
Jazz	0.80	0.96	0.87
Metal	0.78	0.84	0.81
Pop	0.68	0.52	0.59
Reggae	0.64	0.72	0.68
Rock	0.54	0.52	0.53

Table 3 shows the Performance metrics achieved by the proposed CNN model.

Table 3 Performance Metrics

Model	Accuracy	Loss
CNN	0.7236	0.7544

Fig 6 shows the Confusion matrix of the model

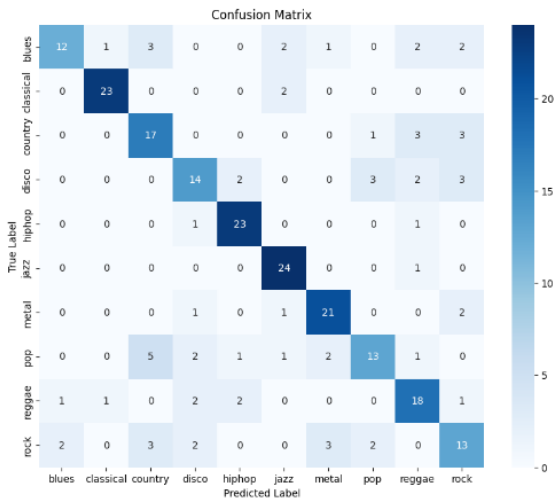

Fig 6 Confusion Matrix

V. CONCLUSION

In conclusion, this study focused on music genre classification using deep learning, specifically employing Mel-spectrograms for feature extraction and a Convolutional Neural Network (CNN) for classification. To enhance the dataset, audio augmentation techniques like time stretching, pitch shifting, and adding white noise were applied, improving the model's robustness and generalization. Feature extraction was carried out using VGGish, a pre-trained model that captured high-level audio representations. These embeddings were then fed into the CNN model, which learned both low-level and high-level features of the music. The model, trained with the RMSprop optimizer and categorical cross-entropy loss, demonstrated effective genre classification. Overall, the project showcases the potential of deep learning models in audio classification tasks. Future work can focus on further improving the model's performance through more advanced techniques in feature extraction, augmentation, and hyperparameter optimization.

REFERENCES

[1] Elbir, A., Aydin, N. (2020). Music genre classification and music recommendation by using deep learning. Electronics Letters, 56(12), 627-629.

[2] Yang, R., Feng, L., Wang, H., Yao, J., Luo, S. (2020). Parallel recurrent convolutional neural networks-based music genre classification method for mobile devices. IEEE Access, 8, 19629-19637.

[3] Wijaya, N. N., Muslikh, A. R. (2024). Music-genre classification using Bidirectional long short-term memory and mel-frequency cepstral coefficients. Journal of Computing Theories and Applications, 1(3), 243-256.

[4] Zhang, K. (2021). Music style classification algorithm based on music feature extraction and deep neural network. Wireless Communications and Mobile Computing, 2021(1), 9298654.

[5] Zhuang, Y., Chen, Y., Zheng, J. (2020, June). Music genre classification with transformer classifier. In Proceedings of the 2020 4th international conference on digital signal processing (pp. 155-159).

[6] Mehta, J., Gandhi, D., Thakur, G., Kanani, P. (2021, April). Music genre classification using transfer learning on log-based mel spectrogram. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1101-1107). IEEE.

[7] Prabhakar, S. K., Lee, S. W. (2023). Holistic approaches to music genre classification using efficient transfer and deep learning techniques. Expert Systems with Applications, 211, 118636.

[8] Jena, K. K., Bhoi, S. K., Mohapatra, S., Bakshi, S. (2023). A hybrid deep learning approach for classification of music genres using wavelet and spectrogram analysis. Neural Computing and Applications, 35(15), 11223-11248.

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

[10] Sharma, A. K., Aggarwal, G., Bhardwaj, S., Chakrabarti, P., Chakrabarti, T., Abawajy, J. H., ... Mahdin, H. (2021). Classification of Indian classical music with time-series matching deep learning approach. IEEE access, 9, 102041-102052.

[11] Ceylan, H. C., Hardalaç, N., Kara, A. C., Firat, H. (2021). Automatic music genre classification and its relation with music education. World Journal of Education, 11(2), 36-45.

[12] Kostrzewa, D., Kaminski, P., Brzeski, R. (2021, June). Music genre classification: looking for the perfect network. In International Conference on Computational Science (pp. 55-67). Cham: Springer International Publishing.

[13] 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.

[14] Yang, H., Zhang, W. Q. (2019, September). Music Genre Classification Using Duplicated Convolutional Layers in Neural Networks. In Interspeech (pp. 3382-3386).
[15] Ghosal, D., Kolekar, M. H. (2018, September). Music Genre Recognition Using Deep Neural Networks and Transfer Learning. In Interspeech (pp. 2087-2091).