

Classifico: Music Genre Prediction System using CNN

Aayush Juhkar¹, Yukta Wagh², Kaushik Nakashe³,
Sahil Bane⁴

^{1,2,3,4}Student, Electronics and Computer Science Engineering, Fr. Conceicao Rodrigues College of Engineering

Abstract

The rapid growth of the music streaming industry has led to an overwhelming amount of music available, making it increasingly challenging for users to find specific genres. Recognizing the significance of quickly classifying musical genres in today's society, we introduce Classifico, a web-based music genre classifier powered by a machine learning model. Classifico simplifies the process by allowing users to input audio and receive genre predictions through the trained model. The web page interface provides an intuitive platform for users to interact with the classifier and obtain accurate genre classifications for their music. Using advanced machine learning techniques, Classifico analyses the audio features of the input, such as rhythm, melody, and instrumentation, to make genre predictions. The trained model has been extensively trained on a diverse dataset encompassing various genres, enabling it to provide reliable and accurate results. By leveraging machine learning, Classifico empowers music enthusiasts, professionals, and casual listeners alike to efficiently categorize and explore different musical genres. This web-based music genre classifier opens up new possibilities for discovering music and enhances the overall music listening experience.

Keywords: MFCC, Genre Recognition, Spectrograms, Feature Extraction, Classifico

1. Introduction

The music industry has experienced significant transformations, expanded its customer base and diversified the market for various musical genres. Music has the power to unite people and provide insights into different cultures. To cater to the demands of music enthusiasts, it is crucial to effectively categorize music into genres. However, manually classifying music and ranking it can be a time-consuming and labour-intensive task for listeners. Moreover, individuals often have preferences for specific genres, making it challenging to explore and discover music outside their comfort zone. To address these challenges, innovative solutions have emerged to streamline the process of music genre classification. Machine learning techniques have been employed to automate the categorization of music based on its audio features, such as tempo, pitch, instrumentation, and rhythm. These algorithms can analyse vast amounts of data, enabling efficient and accurate genre classification. By leveraging these advancements, listeners can benefit from automated music genre classification systems. These systems provide personalized recommendations based on users' listening habits and preferences, introducing them to new genres and expanding their musical horizons. Such platforms empower users to explore a wide range of

music effortlessly and discover hidden gems across various genres. Ultimately, the automation of music genre classification not only saves time and effort for listeners but also promotes musical diversity, fosters cross-cultural understanding, and enriches the overall music listening experience.

2. Motivation

This research project was chosen to delve into the field of machine learning and neural networks, specifically in the context of music genre classification. The objective is to study and analyse different algorithms that can effectively classify music genres based on audio files, using Convolutional Neural Networks (CNN). By developing a model that can accurately classify music genres, this project aims to contribute to the improvement of music streaming platforms. These platforms can utilize the genre classification model to create curated playlists tailored to the preferences of their audience. This personalized approach to music recommendations enhances the user experience by providing genre-specific playlists that align with individual tastes and preferences. The application of CNN in music genre classification is a promising approach due to its ability to extract meaningful features from audio data. By training the CNN model on a dataset containing diverse music genres, it can learn to identify distinctive patterns and characteristics associated with each genre. This knowledge can then be utilized to classify unknown audio files into their respective genres. The successful implementation of a robust music genre classification system can greatly benefit music streaming platforms. It allows for more accurate organization and categorization of their vast music libraries, enabling users to easily discover and explore music that aligns with their preferences. This, in turn, enhances user engagement, satisfaction, and retention on these platforms. Overall, this project aims to contribute to the advancement of music streaming platforms by leveraging machine learning techniques, particularly CNN, to provide genre-specific music playlists and improve the overall music listening experience for users.

3. Problem Statement

To develop a web application which can predict the genre of music with good accuracy.

4. Objectives of the Project

1. Extraction of meta data: The first step is to extract relevant meta data from the audio tracks, such as duration, artist, album, and other track-specific information. This meta data will provide additional context and input for the genre prediction algorithm.
2. Genre prediction: Using machine learning techniques, specifically a Convolutional Neural Network (CNN) model, we aim to train the model on a large dataset of audio tracks with labeled genres. The model will learn patterns and features in the audio data that are indicative of specific genres. Once trained, it will be able to predict the genre of unseen audio tracks with good accuracy.
3. Accuracy improvement: Ensuring a high level of accuracy in genre prediction is crucial for the success of the web application. We will focus on fine-tuning the CNN model, optimizing its hyperparameters, and performing rigorous testing and evaluation to achieve the desired level of accuracy. This may involve data augmentation techniques, model architecture modifications, and training on diverse and representative audio datasets.
4. Clean and interactive UI: In addition to accurate genre prediction, the web application will prioritize providing users with a clean and intuitive user interface (UI). The UI will allow users to easily upload

audio tracks, view the predicted genre, and interact with other features of the application. Attention will be given to user experience design principles, ensuring a seamless and enjoyable user journey. By combining these objectives, our aim is to develop a web application that not only accurately predicts the genre of music but also provides users with an engaging and user-friendly experience.

5. Review of Literature

In this chapter, we explore various research papers that have investigated the application of machine learning in music genre classification. We discuss their methodologies and key findings.

The research paper [1] focuses on using the GTZAN dataset, a widely used dataset in the field of music genre classification. The authors of this paper developed multiple models to tackle the classification task. Their proposed model incorporates various inputs from different models, including the audio mel-spectrogram. These inputs are then fed into a Convolutional Neural Network (CNN). The results of their study indicate that the combination of different inputs and the CNN model yielded the highest accuracy in classifying music genres. They achieved an accuracy of 91% across various genres using machine learning techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Multilayer Perceptron (MLP), and decision trees. However, the authors also observed that certain genres were more distinctive and easier to identify accurately, while others posed challenges. For example, country and rock genres were sometimes confused with other genres, while traditional and blues genres were relatively easier to distinguish. Overall, this research paper highlights the effectiveness of using machine learning approaches, particularly CNNs, for music genre classification. It also emphasizes the importance of considering the distinctiveness of genres and the inherent challenges in accurately classifying certain genres.

The research project [2] was conducted in three phases: 'phase A', 'phase B', and 'phase C'. Each phase had its own significance, contributing to the existing body of work in music genre classification. Additionally, the study aimed to compare the accuracy of machine learning and deep learning models in performing classification tasks. After training multiple classifiers, the k-Nearest Neighbors (kNN) model achieved the highest accuracy of 92.69%. It is worth noting that the training time for kNN was relatively short, taking only 78 milliseconds. The researchers attributed the higher accuracy of kNN compared to previous studies to the function set with a duration of 3 seconds, which provided more training data. Interestingly, the study also found that input features lasting 3 seconds may be more accurate compared to features lasting 30 seconds. Apart from kNN, other notable performances were observed by linear logistic regression and support vector machine (SVM), scoring 81.00% and 80.80% accuracy, respectively. However, the convolutional neural network (CNN) implementations in the study exhibited relatively poor accuracy, with the most accurate CNN implementation achieving only 72.40%. Based on their findings, the researchers concluded that automatic music genre classification is indeed possible. Furthermore, they highlighted that traditional machine learning models tend to outperform deep learning approaches in this specific task. These findings contribute to our understanding of the performance and capabilities of different classification models in music genre classification.

Phase A Results

Classifier	Accuracy	Training Time (s)	Hyperparameters
Linear Logistic Regression	81.00%	25.2500	maximum number of iterations for LogitBoost=500 number of trees = 1000 kernel degree=3, tolerance=0.001, epsilon for loss function=0.1, used polynomial kernel: $\gamma u^v + coef_0$, and did not normalize number of hidden layers= number of hidden classes, learning rate=0.3, training time=500 epochs, validation threshold=20 number of neighbours=1, using absolute error for cross-validation, and applied linear search algorithm used normal distribution for numeric attributes and supervised discretization
Random Forests	75.70%	18.0800	
Support Vector Machines	75.40%	3.8200	
Multilayer Perceptron	75.20%	27.480	
k-Nearest Neighbour	72.80%	0.0100	
naïve Bayes	53.20%	0.5600	

Table 1: Phase A Results [2]

Phase B Results

Classifier	Accuracy	Training Time (s)	Hyperparameters
Support Vector Machines	80.80%	0.3000	radial basis function kernel, tolerance=0.001, and regularization=0.17
Multilayer Perceptron	77.30%	0.2300	hidden layers=2, learning rate=0.02, activation=ReLU, max iterations=200, solver=adam, and tolerance=0.0001
Logistic Regression	75.80%	0.0800	solver=newton-cg and max iterations=500
Random Forests	72.40%	61.080	split function=gini, number of trees = 100, and max depth 100
k-Nearest Neighbour	69.70%	0.0110	k=7 with manhattan distance metric, weighting=distance
naïve Bayes	54.50%	0.0019	Gaussian naïve Bayes with smoothing

Table 2: Phase B Results [2]

Phase C Results

Classifier	Accuracy	Training Time (s)	Hyperparameters
k-Nearest Neighbours	92.69%	0.0780	nearest neighbours=1
Multilayer Perceptron	81.73%	60.620	activation=ReLU solver lbfgs
Random Forests	80.28%	52.890	number of trees=1000, max depth=10, $\alpha = e^{-5}$, and hidden layer sizes=(5000,10)
Support Vector Machines	74.72%	3.8720	decision function shape=ovo
Logistic Regression	67.52%	3.6720	penaty=12, multi class=multinomial

Table 3: Phase C Results [2]

In the research paper titled "Music Genre Classification Using Deep Learning" published in 2021 [3], the authors focused on comparing various deep learning techniques for the task of music genre classification. Their project aimed to develop a classifier capable of predicting the genre of audio files. The GTZAN music genre classification dataset was utilized for their experimentation, where they explored different ensemble models to improve genre differentiation. The authors employed three key components in their approach: (1) time and frequency domain analysis, (2) feature extraction techniques, and (3) neural network modelling. These components were integral in designing effective models for music genre

classification. Through their experiments, the authors observed that deep learning methodologies yielded higher accuracy in classifying music genres, achieving approximately 90% accuracy. In comparison, previous models based on spectrogram techniques demonstrated lower accuracy levels, typically around 65-70%. This finding highlights the superiority of deep learning models in accurately categorizing music genres. Overall, the research paper sheds light on the effectiveness of deep learning techniques for music genre classification, showcasing significant improvements in accuracy compared to traditional spectrogram-based approaches. This research contributes to the advancement of music classification systems and provides valuable insights for the development of more accurate genre prediction models.

Model Evaluation using Loss and Accuracy

Model Number	Test Loss	Test Accuracy
1	0.5499	0.8950
2	0.3923	0.9166
3	0.3480	0.9176
4	1.6516	0.9264

Table 4: Model Evaluation [3]

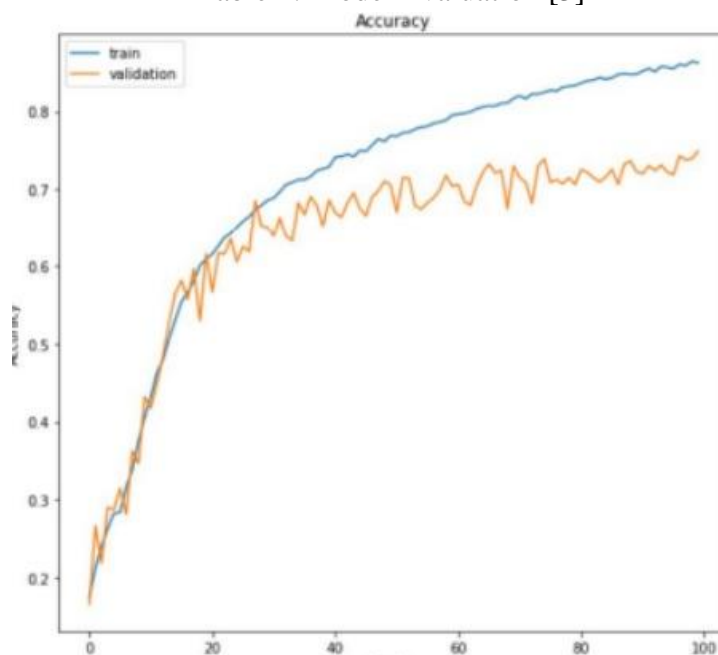


Figure 1: Accuracy Curve [3]

For individuals who are new to music or less familiar with different music genres, searching for a specific style of music through streaming platforms can be time-consuming and inefficient. This is especially challenging for musicians who are seeking a particular genre of music, as they may need to listen to numerous tracks for an extended period to accurately judge the genre. The process can become tiresome and lead to difficulty in making genre judgments, resulting in wasted time and effort [5]. To address this issue and save time, a music genre classifier can be a valuable tool. In a research study [4], the authors proposed a convolutional neural network (CNN) for music genre classification. Their model achieved a promising accuracy rate of 83.3% in classifying music genres. This outcome supports the potential of using CNNs for future work in music genre classification. To further enhance the model's accuracy and

functionality, the authors plan to make improvements such as integrating streaming media and web crawlers. This integration would allow for a more comprehensive and robust music genre classification system. The ultimate goal is to develop a combined CNN architecture that provides a complete solution, enabling both music beginners and musicians to save time and increase efficiency in finding and categorizing music according to specific genres [4]. By leveraging a music genre classifier, individuals can quickly and accurately identify and explore music within their preferred genres, streamlining the process and enhancing their overall music listening experience.

6. Proposed Methodology

6.1 Introduction

In the ever-evolving landscape of the music industry, significant transformations have occurred, shaping the way music is consumed and shared. With a constantly expanding customer base and a diverse range of musical genres, it has become crucial to categorize music efficiently. Manual categorization is time-consuming and challenging due to the vast number of genres and the individual preferences of listeners. To address this issue, our project harnesses the capabilities of machine learning models and Mel-Spectrum cepstral coefficients (MFCCs) to accurately predict the genre of audio tracks.

By utilizing machine learning algorithms, our project automates the genre classification process, reducing the burden on listeners and music professionals. The incorporation of MFCCs allows for a detailed analysis of audio signals, capturing essential features related to timbre, rhythm, and tonality that contribute to genre characteristics. Through this approach, we aim to provide a more efficient and accurate solution for genre classification.

The significance of genre classification extends beyond convenience. Music has the power to bring people together and act as a gateway to understanding foreign cultures. By categorizing music into genres, we facilitate the discovery and exploration of diverse musical styles, allowing listeners to connect with different cultures and broaden their musical horizons.

Furthermore, music streaming services like Spotify recognize the importance of catering to the specific preferences of audiophiles. By generating genre-specific playlists, these services offer personalized recommendations that align with users' tastes. Our project aligns with this objective by providing a reliable and automated method to classify music genres, thereby enhancing the quality and relevance of genre-based playlists.

6.2 Implementation

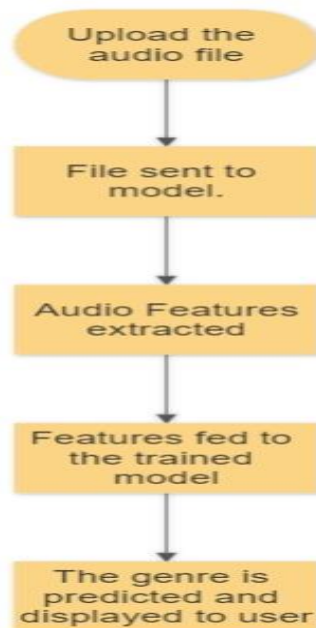


Figure 2: Process Flow Diagram

The depicted process flow diagram of Classifico showcases the steps involved in categorizing music genres using a Keras-based algorithm with a Django backend. The flow begins with the user logging onto the website and subsequently uploading a 30-second audio file. Upon clicking the upload button, the audio file is sent to the backend for further processing. The backend comprises a convolutional neural network (CNN) model that has been trained using the GTZAN dataset, utilizing the scikit-learn library in Python. To enhance accuracy, both the provided dataset audio files and additional self-generated datasets were employed. The MFCC (Mel Frequency Cepstral Coefficients) values were extracted from the audio files, serving as features for prediction. The extracted MFCC values are obtained using the librosa library in Python, which specializes in audio analysis. These values are then fed into the trained CNN model, which performs the genre classification prediction. The predicted output, representing the genre of the uploaded audio file, is displayed on the screen for the user to view. It's important to note that currently, the algorithm supports only .wav file formats. However, future plans include expanding the support to other file formats, thereby accommodating a wider range of audio files for genre classification. Overall, the process flow of “Classifico” demonstrates a well-defined workflow, starting from user login and audio file upload to preprocessing, prediction using a trained CNN model, and displaying the predicted genre output. The inclusion of MFCC feature extraction, the utilization of the scikit-learn and librosa libraries, and the support for expanding file format compatibility contribute to the accuracy and usability of the application.

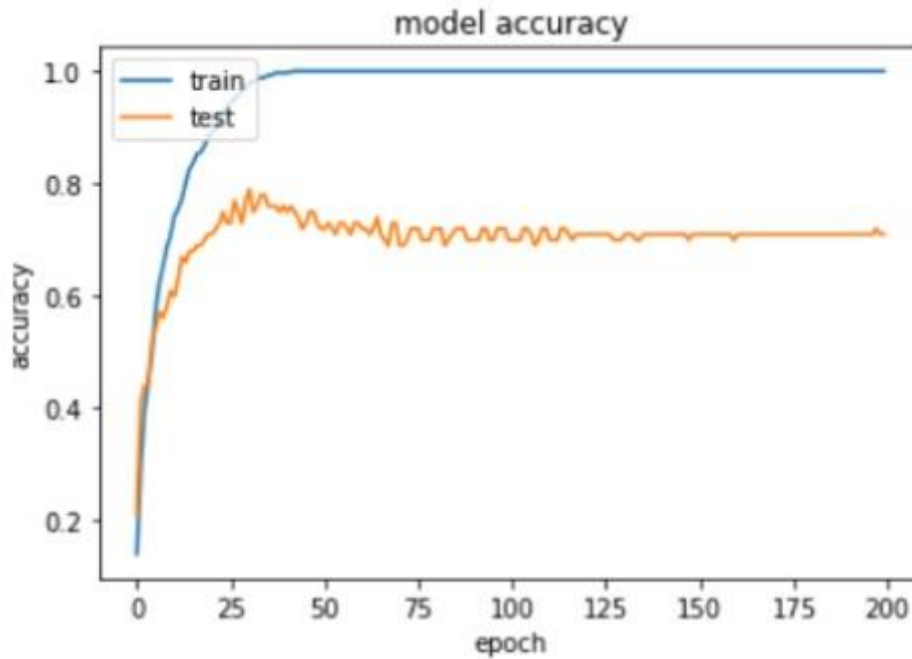


Figure 3: Accuracy curve (training vs testing)

The accuracy curve in the figure shows the performance of the model with respect to the number of epochs. The current model achieves a test accuracy of 76% with a total loss of 1.73%. There is room for improvement by adjusting the weights and activation function to further enhance the accuracy.

6.2.a CNN Model

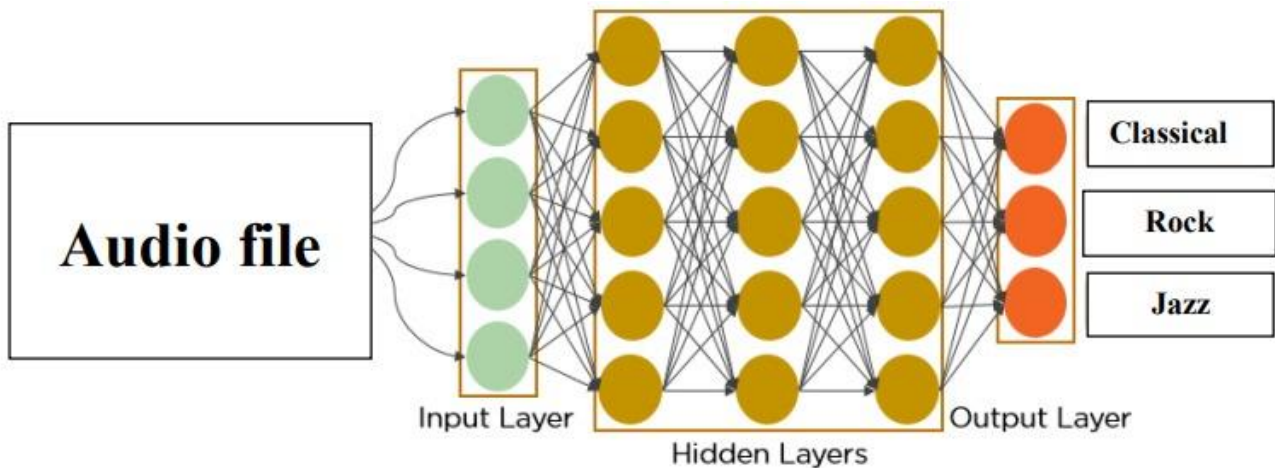


Figure 4: CNN Model

The CNN model used in the architecture comprises four dense layers with 512, 256, and 64 units, respectively, and a final output layer with 10 units representing each music genre. The model is compiled with the Adam optimizer, sparse categorical cross-entropy loss, and accuracy metric. It is trained on the training set for 50 epochs.

To address overfitting, the code implements a regularized ANN with L2 regularization and dropout layers. Each dense layer in the architecture is followed by a dropout layer with a dropout rate of 0.3. The model

is compiled and trained on the training set for 100 epochs. For the convolutional neural network (CNN) model, there are three convolutional layers with 32 filters each. These layers are followed by max pooling and batch normalization layers. The output of the convolutional layers is flattened and passed to a dense layer with 64 units. A dropout layer with a dropout rate of 0.3 is applied before the final output layer, which has 10 units representing the different music genres. The activation function used for the output layer is softmax.

7. Results

7.1 App Snapshots



Figure 5: Web UI

Figure 5 showcases the intuitive web user interface (UI) of Classifico, developed using HTML, CSS, and JavaScript. The UI design focuses on simplicity and user-friendliness, providing a seamless experience for users interacting with the application. With an emphasis on usability, the UI incorporates a provision to upload audio files, allowing users to input the audio tracks they wish to classify.

To facilitate seamless hosting and backend functionality, Classifico is implemented using the Django web framework. Django offers a robust and scalable platform for web application development, enabling efficient handling of user requests, data processing, and model predictions. By leveraging Django's capabilities, Classifico can provide a smooth and responsive experience to users accessing the application.

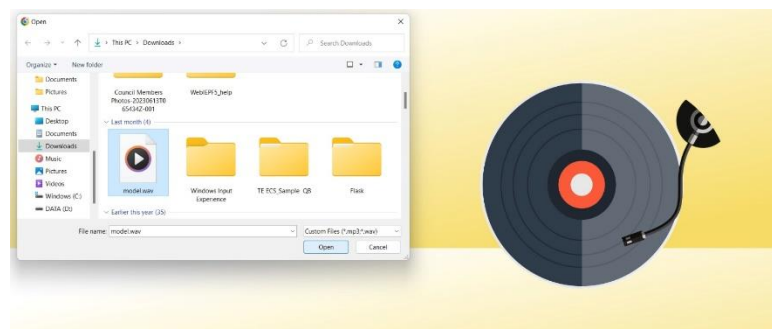


Figure 6: Audio Selection Window

The user can easily feed the song that needs to be classified into the website displayed in the above figure. The website provides a user-friendly interface that allows users to upload their desired song for genre classification.

To further enhance the user experience, the website can include features such as drag-and-drop functionality, allowing users to simply drag and drop their audio files directly onto the designated area in the UI. This intuitive method of uploading the song streamlines the process and eliminates the need for manual file selection.



Figure 7: Prediction

The predicted output is prominently displayed within the UI, as depicted in Figure 7. This allows users to quickly and easily view the genre classification results for the uploaded audio file. The predicted output provides valuable information, giving users immediate insights into the genre to which the audio track belongs.

To enhance the clarity and comprehensibility of the predicted output, additional details may be included, such as the confidence level or probability score associated with the predicted genre. This can provide users with an indication of the model's confidence in its classification result. By presenting this information, users can better understand the reliability of the genre prediction and make informed decisions based on the provided output.

8. Applications

1. **Music Mood and Contextualization:** In addition to genre classification, analyzing the mood or emotional characteristics of songs can further enhance music recommendation systems. By considering the emotional context of songs, music streaming platforms can curate playlists that align with users' current moods or activities, such as creating playlists for relaxation, workout sessions, or studying.
2. **Genre-Based Music Events and Festivals:** Genre classification can facilitate the organization and curation of music events and festivals. By accurately classifying songs into genres, event organizers can create diverse and engaging lineups that cater to specific music preferences. This allows attendees to discover and enjoy performances within their preferred genres, enhancing their overall event experience.
3. **Music Genre Tagging and Metadata Enhancement:** Genre classification can be used to automatically tag and enhance the metadata of music tracks within digital libraries. By assigning accurate genre tags to songs, music platforms and libraries can improve search and browsing functionalities, making it easier for users to discover and organize their music collections based on genres.

4. Music Licensing and Royalty Distribution: Genre classification can play a role in music licensing and royalty distribution processes. By accurately classifying songs into specific genres, music platforms and copyright organizations can ensure fair and appropriate distribution of royalties to artists based on their genre-specific contributions.

Overall, genre classification in the context of music streaming platforms has broad applications ranging from music recommendation and research to marketing, radio broadcasting, event curation, metadata enhancement, and licensing. Continual advancements in genre classification techniques and their integration with other music analysis approaches have the potential to revolutionize the way we discover, enjoy, and interact with music in various domains and industries.

9. Conclusion

In conclusion, the project focused on improving the accuracy of genre classification and playlist curation on a music streaming platform. By fine-tuning parameters, refining the dataset, and exploring different algorithms, the model can achieve higher accuracy in genre classification. The iterative process of training and testing helps determine the optimal number of layers in the model architecture. Additionally, integrating the model with a music streaming platform offers the opportunity to enhance the user experience by automatically classifying genres and curating personalized playlists.

10. Future work

As part of the future scope, incorporating user feedback and contextual information can further enhance the personalization of recommendations. Multimodal analysis, including lyrics and album artwork, can provide a more comprehensive understanding of genres. Implementing explainable AI techniques can foster user trust by providing insights into the model's recommendations. Cross-domain genre classification and collaborative filtering techniques enable the expansion of recommendation systems beyond music, making them more versatile and diverse.

These advancements pave the way for a more accurate, personalized, and engaging music streaming experience. By continuously refining and integrating these approaches, the project contributes to the advancement of genre classification and playlist curation, ultimately benefiting both users and music streaming platforms.

References

1. A. Ghidiyal, K. Singh and S. Sharma, "Music Genre Classification using Machine," IEEE, 2020.
2. N. Ndou, R. Ajoodha and A. Jadhav, "Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning," IEEE 2021.
3. Y.-H. Cheng, P.-c. Chang and C.-N. Kuo, "Convolutional Neural Networks Approach for Music Genre Classification," IEEE, 2022.
4. N. Parab, S. Das, G. Goda and A. Naik, "Music Genre Classification Using Deep Learning," International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 10, 2021.
5. P. Devaki; A. Sivanandan; Raj Senthil Kumar; Mohsin Zahoor Peer, "Music Genre Classification and Isolation," IEEE, 2022