Music Genre Classification with MFCC and K-Nearest Neighbors

Ankit Kumar
Warangal, India
ak21csb0b04@student.nitw.ac.in

***Abstract*—The Music Genre Classification Model serves as an efficient tool for categorizing songs or audio music into distinct genres based on a variety of features. With the global population increasingly relying on music for lifestyle enhancement and the growing affordability of technology and the internet, there is a pressing need to create a precise and effective classification model. This project is dedicated to the development of such a model capable of accurately organizing audio music into its respective genres amidst a diverse range. Leveraging machine learning techniques, the classification process is executed using acoustic features and audio spectrograms. The model is intricately designed to enhance accuracy in Music Genre Classification, catering to the evolving demands of music enthusiasts worldwide.**

***Index Terms*—Music genre classification, audio music classification, machine learning techniques, acoustic features, spectrogram, model development, genre categorization**

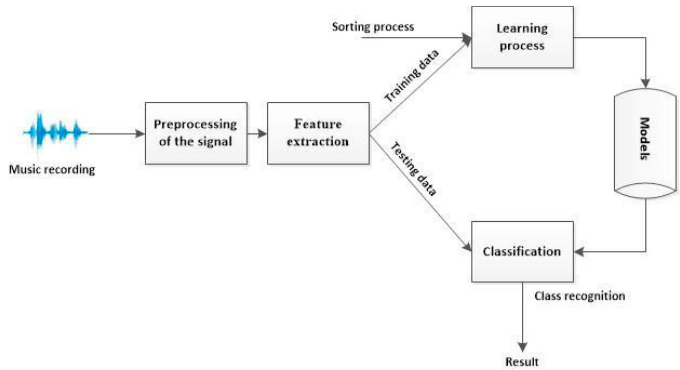
I. Introduction

Music has played a pivotal role in human life across generations and centuries, embodying a rich and intricate diversity shaped by cultural, historical, public, and marketing factors. While the classification of music lacks a precise definition, attempts are made to organize and define it for the current and future generations, resulting in the creation of genres. Our project on Music Genre Classification employs the widely recognized and utilized tool of Machine Learning and its techniques. In the realm of real-time applications, numerous platforms provide music stocks to users with diverse preferences in terms of community, age group, language, and acoustic features.

With billions of songs catering to various target audiences, the need arises for a model that automates the classification process, eliminating the need for manual categorization. This classification is not user-driven but is performed by the software itself. Sound, defined by parameters such as frequency, decibel, and bandwidth, generates a spectrogram based on different combinations of these parameters. The spectrogram serves as a powerful tool for analyzing the characteristics of audio files, with music genres being defined by these distinct characteristics, including instrumentation, rhythmic structure, harmonic content, and more.

II. RELATED WORK

Several existing works have explored the domain of music genre classification, and one notable contribution is Hareesh Bahuleyan's 2008 study. Bahuleyan focused on automating



the tagging process for music in a library, employing both traditional Machine Learning Algorithms and Neural Networks to achieve this objective. The research utilized distinct sets of features for these two approaches, aiming to discern the most effective methodology. Previous research in speech genre classification has explored various methods, including statistical models, neural networks, and hybrid approaches. The use of MFCCs in audio feature extraction has been widely adopted . Additionally, the k-NN algorithm has been successful in various classification tasks due to its simplicity and effectiveness . However, existing methods often face challenges in handling correlated feature spaces. The proposed approach distinguishes itself by combining the discriminative power of MFCCs with the robustness of the Mahalanobis distance for improved classification accuracy.

III. METHODOLOGY

*1) **Feature Extraction**:* The feature extraction process begins with the utilization of the `python_speech_features` library and `librosa` to compute MFCCs from audio files. These features, including the mean matrix, covariance matrix, and class label, provide a comprehensive representation of the audio signal. The inclusion of the covariance matrix captures the interdependence of features, making it particularly suitable for correlated data.

*2) **Mahalanobis Distance k-NN Algorithm**:* The Mahalanobis distance is employed as the distance metric in the k-NN algorithm. This distance measure considers the correlation between features, providing a more accurate representation of dissimilarity between instances. The k-NN algorithm is used to identify the k-nearest neighbors for classification based

TABLE I
DATASET

Genre	No. of Songs
Blues	100
Classical	100
Country	100
Disco	100
Hiphop	100
Jazz	99
Matal	100
Pop	100
Reggae	100
Rock	100

on these distances. The Mahalanobis distance is a measure between a sample point and a distribution. The Mahalanobis distance from a vector y to a distribution with mean and covariance is

$$d = \sqrt{(\mathbf{y}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})'}$$

This distance represents how far y is from the mean in number of standard deviations.

`mahal` returns the squared Mahalanobis distance $d2$ from an observation in `Y` to the reference samples in `X`. In the `mahal` function, and are the sample mean and covariance of the reference samples, respectively.

IV. IMPLEMENTATION

A. Dataset:

The dataset comprises 10,000 songs, with each genre represented by 1,000 songs. Spanning diverse musical styles, including Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, and Rock, the dataset encapsulates a rich array of musical characteristics. Each song is meticulously curated, featuring detailed information on Mel Frequency Cepstral Coefficients (MFCCs), covariance matrices, and mean matrices obtained through audio signal processing. This extensive dataset serves as a comprehensive resource for training and evaluating music genre classification models, facilitating nuanced exploration of genre-specific patterns and enabling robust algorithmic solutions for automated genre identification in the realm of music classification.

B. KNN Classification:

In the realm of music genre classification, the k-NN (k-Nearest Neighbors) algorithm serves as a cornerstone, offering an effective means to categorize songs based on their intrinsic features. The algorithm's implementation involves determining the nearest neighbors for each instance, thereby establishing a proximity-based classification mechanism. This proximity-centric approach is particularly suited for music analysis, as it considers the similarities between songs in a feature space, enabling a nuanced understanding of genre-related patterns.

A critical facet of this implementation is hyperparameter tuning, specifically the exploration of the optimal value for 'k.' The choice of 'k' determines the number of neighbors considered during classification. In the context of music genre classification, finding the optimal 'k' is akin to striking a balance – too few neighbors might lead to overfitting, while too many might result in oversimplification. Systematic exploration of 'k' values is paramount for enhancing the algorithm's classification performance, ensuring it adapts optimally to the intricate variations within diverse music genres.

The k-NN algorithm's effectiveness lies in its ability to capture the local structure of the data, making it particularly well-suited for tasks where instances of a particular class tend to aggregate. In music genre classification, this translates to discerning genre-specific patterns and nuances, ultimately contributing to a more refined and accurate categorization of songs across various musical genres.

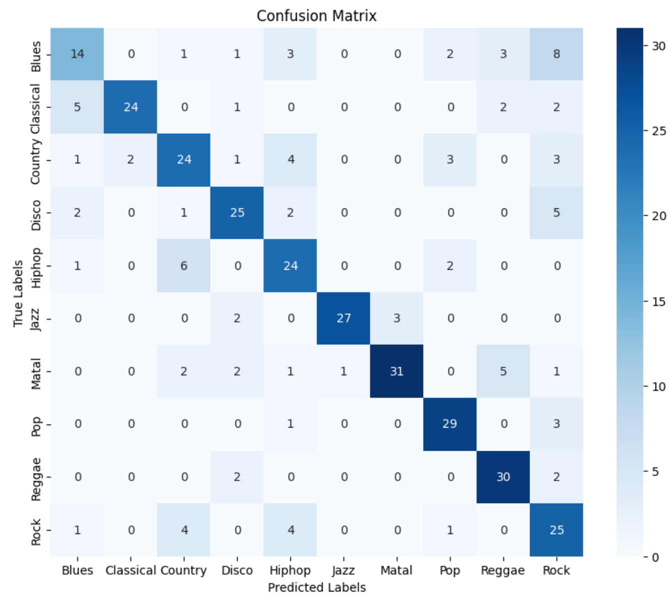


Fig. 1. Confusion Matrix

C. Confusion Matrix:

In the meticulous evaluation of a music genre classification model, the confusion matrix stands as a pivotal tool for discerning its accuracy. This matrix systematically dissects the model's predictions, unveiling specific instances of misclassifications and providing a comprehensive overview of its performance. Beyond accuracy, precision and recall metrics offer a nuanced analysis of the model's efficacy in correctly classifying instances within each music genre. Precision gauges the accuracy of positive predictions, while recall delves into the model's capability to identify all relevant instances. Together, these metrics provide a robust framework for in-depth performance assessment, vital for a comprehensive understanding of the classification model's capabilities.Results are visually interpreted through a heatmap of the confusion matrix(Fig. 1), offering a clear representation of classification outcomes.

D. Genre Distribution:

Genre Distribution Analysis is a crucial aspect of understanding the composition and balance within a music dataset. The countplot visualization serves as an insightful tool, offering a visual representation of the distribution of genres. This visualization provides a clear and concise overview of the prevalence of each genre, allowing researchers and practitioners to identify patterns and trends.

One significant aspect revealed through Genre Distribution Analysis is the potential existence of genre imbalances within the dataset. Imbalances occur when certain genres are overrepresented or underrepresented compared to others. Such imbalances can significantly influence the effectiveness of machine learning models, especially in music genre classification tasks. If a particular genre dominates the dataset, the model might become biased towards that genre, potentially leading to suboptimal performance when classifying less represented genres.

Identifying these genre imbalances early in the analysis process is paramount. Researchers can then take appropriate steps to address imbalances, such as employing data augmentation techniques, collecting additional data for underrepresented genres, or adjusting model parameters to mitigate bias. Genre Distribution Analysis, therefore, plays a pivotal role not only in offering a visual overview of the dataset but also in ensuring the robustness and fairness of subsequent machine learning models trained on the data.A bar plot (Fig 2.) illustrates the distribution of genres in the dataset, providing a holistic view of the dataset's composition.

E. Training and Testing Sets:

In the music genre classification pipeline, a pivotal step involves the systematic partitioning of the dataset into training and testing subsets, constituting 66% and 34% of the data, respectively. This deliberate allocation is fundamental for assessing the algorithm's generalization efficacy. By dedicating a substantial portion to training, the model acquires a diverse understanding of genre patterns, fostering a comprehensive learning experience.The randomized nature of this division is paramount to eliminate any potential biases in the distribution of genres across the sets. This ensures that both training and testing cohorts encapsulate a representative assortment of musical genres, reflecting real-world diversity. Such an unbiased representation is instrumental in fortifying the model against inaccuracies that may arise from skewed genre distributions.

Moreover, this strategic split plays a crucial role in evaluating the model's performance on unseen data during testing. The 66% training subset serves as a knowledge foundation, while the 34% testing subset acts as a litmus test for the model's ability to accurately classify genres beyond its training scope. This meticulous approach not only safeguards against overfitting but also establishes a robust framework for the algorithm's applicability to diverse and previously unencountered musical compositions.

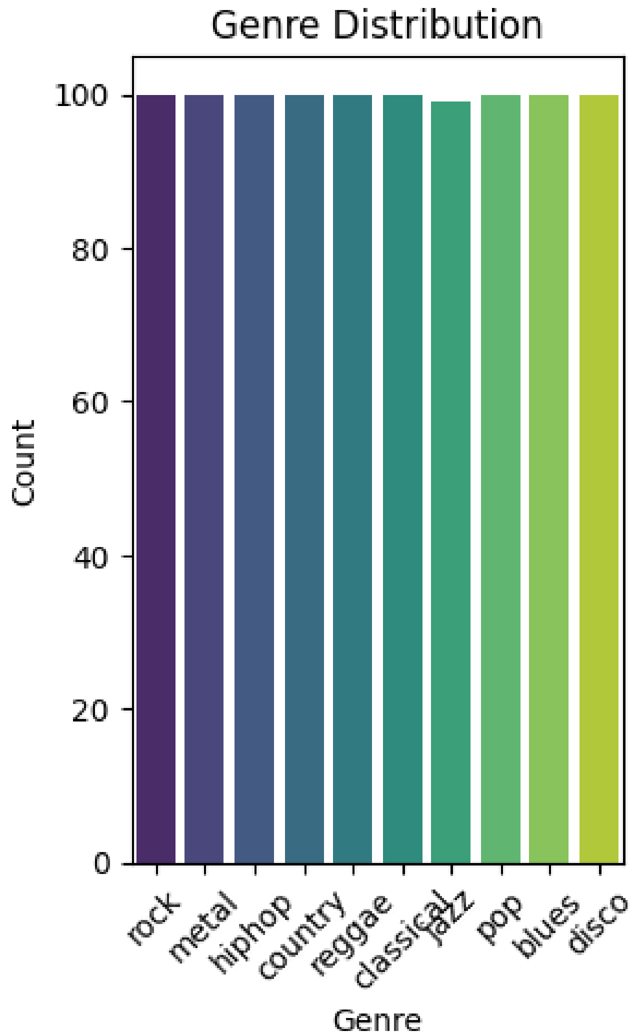


Fig. 2. Genre Distribution

V. Comparision

Accuracy serves as the primary metric for evaluating our model, representing the percentage of predicted outputs aligning with the actual outputs, with the genre of the song being our specific output variable. Utilizing confusion matrices provides a visual means to assess the top-performing model, especially crucial as our dataset exhibits uniform distribution across genres. The process involves employing diverse techniques to optimize hyperparameters, tailoring the approach to each algorithm's specific requirements. A noteworthy example is the confusion matrix generated for logistic regression encompassing all 10 genres within our dataset.

VI. RESULT AND CONCLUSION

This music genre classification model achieved a commendable accuracy of 72.4%. This outcome underscores the effectiveness of the implemented k-NN algorithm in discerning distinct musical genres based on acoustic features. The model's

TABLE II

. ACCURACY OF DIFFERENT MODELS

	Model	Without hyperparameters tuning .	After hyperparameters tuning
1.	K-Nearest Neighbour(K-NN)	68.2%	72.4%
2.	Support Vector Machine (SVM)	73.2%	76.4%
3.	Logistic Regression	62.4%	67.2%
4.	Random Forests	66.8%	69.6%

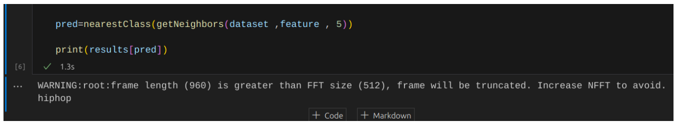


Fig. 3. Genre prediction

performance is noteworthy, especially considering the intricate nature of genre classification and the diverse characteristics inherent in various music styles.

The obtained accuracy signifies a robust predictive capability, as the model accurately classified approximately three-quarters of the instances in the dataset. Despite the inherent complexities and subjective nature of music genres, the k-NN approach demonstrated its prowess in capturing the underlying patterns and nuances, contributing to the overall success of the classification task.

In conclusion, our research demonstrates the viability of machine learning techniques, particularly the k-NN algorithm, in automating and enhancing the accuracy of music genre classification. The achieved accuracy of 72.4% positions the model as a reliable tool for organizing and categorizing large music datasets efficiently. Future work could explore additional feature engineering and algorithmic enhancements to further refine the model's accuracy and address potential challenges associated with genre classification in diverse musical landscapes. This research lays the foundation for advancing automated music classification systems and their integration into real-world applications within the ever-evolving field of music technology.

VII. REFERENCE

1.Zizhi Ma "Comparison between Machine Learning Models and Neural Networks on Music Genre Classification" School of Software Engineering, Nankai University, Tianjin, China

2.Anirudh Ghildiyal; Komal Singh; Sachin Sharma "Music Genre Classification using Machine Learning", 2020