

Performance Evaluation of Some Machine Learning Models for Music Genre Classification

Olojede, Ojo Abraham¹; Stephen Olatunde, Olabiyisi²; Oluwaseun. O Alo³

^{1&2}Department of Computer Science Ladoke Akintola University of Technology, Ogbomosho, Oyo state Nigeria

³Department of Information System Ladoke Akintola University of Technology Ogbomosho, Oyo state, Nigeria

DOI : <https://doi.org/10.51583/IJLTEMAS.2025.1402003>

Received: 16 February 2025; Accepted: 21 February 2025; Published: 03 March 2025

Abstract: Music genre classification is a challenging task in the field of music information retrieval due to the overlapping characteristics of certain genres and the variability in audio quality. Several techniques have been developed to accurately classify music genre. However, these techniques have not been adequately analysed and compared. Hence, this study investigates the comparative performance of Convolutional Neural Network (CNN), Support Vector Machine (SVM), and Random Forest (RF) in music genre classification.

Mel-Frequency Cepstral Coefficients (MFCCs) were extracted from the audio samples using the Librosa library. Next, the three machine learning models - Convolutional Neural Network (CNN), Support Vector Machine (SVM) and Random Forest (RF) - were trained. The CNN model was designed with multiple convolutional and pooling layers, along with dropout for regularization. The SVM model was used to create an optimal hyperplane for classification, while the RF model utilized an ensemble of decision trees. Finally, the models were evaluated and compared using accuracy, precision, recall and F1 score.

The results of the evaluation and comparison indicate that CNN achieved 95% accuracy, 93% precision, 92% recall and 91% F-1 score. SVM achieved 93% accuracy, 90% precision, 80% recall and 70% F-1 score while RF achieved 77% accuracy, 77% precision, 72% recall and 60% F-1 score.

The result demonstrated that CNN outperformed SVM and RF in terms of accuracy, precision, recall, and F-1 score: CNN is thereby recommended for Music Genre Classification, this finding underscores the efficiency of CNN addressing the challenges task in the field of music information retrieval and leading to the advancement of automated music classification system and improve the accessibility and enjoyment of digital music libraries.

Key word: Performance, Evaluation, Machine, music, genre.

I. Background to the Study

The rise of machine learning (ML) and artificial intelligence (AI) has opened new avenues for addressing this challenge. ML algorithms, particularly those designed to handle complex and high-dimensional data, have shown promise in learning and recognizing patterns in audio signals (Tzanetakis, et al., 2002). These advancements have made it possible to emulate human-like understanding of music while leveraging computational efficiency (Sturm, 2014).

Early attempts at automated music genre classification utilized traditional machine learning algorithms such as k-nearest neighbors (KNN), support vector machines (SVM), and decision trees. These methods achieved moderate success by extracting handcrafted features from audio signals, such as tempo, pitch, and timbre. However, they often struggled with capturing the intricate and hierarchical nature of musical compositions (Casey et al., 2008).

Recent advancements in deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated superior performance. These models automatically learn hierarchical feature representations from raw audio data, capturing the nuances of musical elements more effectively (Humphrey, et al., 2012). CNNs excel at learning spatial hierarchies, making them suitable for processing spectrograms, while RNNs are adept at handling temporal sequences, making them ideal for modeling the temporal dynamics of music (Dieleman et al., 2014).

Despite the significant progress in music genre classification, several challenges remain. One of the primary challenges is the subjective nature of genre definitions. Different cultures and communities often have varying interpretations of what constitutes a particular genre, leading to ambiguities and inconsistencies in genre labels (Lidy et al., 2005). Furthermore, the evolution of music over time results in the emergence of new genres and subgenres, complicating the classification task (Bergstra et al., 2006).

Another challenge lies in the variability of audio quality and production techniques. Songs within the same genre can exhibit considerable differences in their audio characteristics due to variations in recording equipment, production styles, and mixing techniques (Essid, et al., 2006). This variability can introduce noise into the feature extraction process, affecting the performance of classification models.

To address these challenges, researchers have explored various feature extraction methods and data augmentation techniques. Feature extraction methods aim to capture relevant audio characteristics while minimizing the impact of noise and variability. Techniques such as Mel-frequency cepstral coefficients (MFCCs), chroma features, and spectral contrast have been widely used to represent audio signals in a compact and informative manner (Tzanetakis et al., 2002). Data augmentation techniques, on the other hand, involve artificially increasing the size and diversity of training datasets by applying transformations such as pitch shifting, time stretching, and adding

background noise (Schlüter et al., 2014). These techniques help improve the generalization ability of classification models by exposing them to a broader range of variations.

In addition to feature extraction and data augmentation, ensemble learning methods have shown promise in improving classification performance. Ensemble learning involves combining multiple models to make predictions, leveraging the strengths of different algorithms and reducing the risk of overfitting. Techniques such as bagging, boosting, and stacking have been applied to music genre classification, leading to enhanced robustness and accuracy (Bergstra et al., 2006).

Moreover, recent research has focused on interpretability and explain ability of classification models. Understanding how and why a model makes certain predictions is crucial for building trust and gaining insights into the underlying characteristics of music genres. Techniques such as saliency maps, attention mechanisms, and feature importance analysis have been employed to provide explanations for model predictions, allowing researchers to identify the most influential features and patterns (Choi et al., 2017).

Another emerging area of research is transfer learning, which involves leveraging knowledge from pre-trained models on related tasks to improve performance on music genre classification. Transfer learning has been particularly effective in scenarios where labeled data is scarce, as it allows models to benefit from the vast amount of data and knowledge available in other domains (Zhang et al., 2018). By fine-tuning pre-trained models on specific music datasets, researchers have achieved significant improvements in classification accuracy and efficiency (Dieleman et al., 2014).

This study aims to build upon these advancements by exploring state-of-the-art machine learning techniques for music genre classification on performance evaluation of some machine learning model for music genre classification. It seeks to address the limitations of previous methods and improve the accuracy and robustness of genre classification models. By examining the underlying characteristics of different music genres and leveraging sophisticated algorithms, this research endeavors to enhance our understanding of music and contribute to the development of more effective music recommendation and organization systems (Pons, et al., 2016). This study not only contributes to the academic discourse on music genre classification but also has practical implications for the future of digital music consumption, offering a deeper understanding of the soul of music through the lens of machine learning (Zhang et al., 2018).

Statement of the Problem

The music industry has experienced a rapid expansion in the digital age, resulting in an overwhelming volume of music available across various platforms. As a consequence, efficient organization and categorization of music have become essential for both consumers and service providers. Traditional methods of classifying music genres rely heavily on human annotation, which is time-consuming, subjective, and prone to inconsistencies. This manual process struggles to keep pace with the increasing rate at which new music is produced and distributed.

There is a pressing need for automated systems that can accurately classify music into genres to enhance music recommendation, search ability, and overall user experience. However, music genre classification presents several challenges, including the inherent overlap between genres, the complexity of musical features, and the need for robust models capable of handling diverse and extensive datasets.

Objectives of the study

The aim of this research is to develop and evaluate a system for automatically classifying music into different genres based on the analysis of their audio features.

Objectives

- i. Extract relevant audio features from music samples using libraries like Librosa. These features might include Mel-Frequency Cepstral Coefficients (MFCCs) that capture the timbral characteristics of the music.
- ii. Implement and train a machine learning model (SVM, Random Forest, CNN) to classify music genres based on the extracted audio features.
- iii. Compare the models' performance using metrics classification accuracy, perception recall and F1-score as performance metrics.

Significance of the Study

The significance of this study lies in its potential to transform the landscape of music genre classification and enhance the overall user experience in digital music platforms. By advancing the capabilities of machine learning models, this research addresses critical challenges associated with the accurate and robust classification of music genres, which is essential for efficient music organization, discovery, and recommendation (Tzanetakis et al., 2002). Improved classification systems can lead to more personalized music recommendations, thereby enriching the user experience and increasing engagement with digital music services (Choi et al., 2017).

Scope of the Study

This study focuses on advancing music genre classification by developing and evaluating state-of-the-art machine learning models, including deep learning architectures, to address current limitations and challenges. The scope includes a comprehensive review of existing methodologies, implementation of advanced models such as CNNs and RNNs, and exploration of techniques for feature extraction and data augmentation. It aims to tackle issues related to genre subjectivity and audio variability, assess model performance in terms of accuracy and interpretability, and evaluate practical applications for improving user experience in digital music platforms.

II. Literature Review

Music Genre Classification

Music genre classification is the process of categorizing music into different genres based on their inherent characteristics and attributes. This field of study has gained significant attention due to its potential applications in music recommendation systems, retrieval, and organization. With the ever-growing body of musical content available, efficient and accurate classification methods are essential for helping listeners navigate and discover new music.

Music genre classification has become a pivotal task in music information retrieval, aiming to automatically categorize music tracks into predefined genres. The rapid growth of digital music collections necessitates the development of efficient and accurate classification systems. Traditional approaches relied on manual annotation and rule-based systems, but the advent of machine learning has revolutionized this field, enabling the automatic extraction of features and classification of music into genres (Tzanetakis et. al, 2002).

In recent years, machine learning and deep learning techniques have become prominent tools for music genre classification. Ghildiyal, et.al, 2020 proposed a machine learning-based approach. They employed a combination of feature extraction and classification algorithms, demonstrating the potential of machine learning for this task. Similarly, Ndou, et.al, 2021, presented a comprehensive review of both deep learning and traditional machine-learning approaches for music genre classification in their 2021 IEEE paper. This review provided valuable insights into the state-of-the-art methods and their respective strengths and limitations.

The process of music genre classification typically involves several steps, as outlined in the provided information. The first crucial step is feature extraction, where relevant features and components are extracted from audio files while removing noise. Linguistic content identification plays a vital role in this process. Once the features are extracted, they can be used to train classifiers or compared with existing datasets to predict the genre of new, unseen audio files.

Building upon traditional machine learning methods, Elachkar et al, 2021 introduced a novel approach that combines reduction and dense blocks for music genre classification. The work showcases the potential of hybrid models in improving classification accuracy. In another innovative approach, Goulart et al, 2011. Proposed a feature extraction method based on entropy fractal features and an SVM classifier. This method achieved almost 100% accuracy in classifying blues, classical, and lounge genres

Linguistic features refer to the textual content and semantic information associated with music, particularly the lyrics of a song. The incorporation of linguistic features in music genre classification offers a unique perspective by leveraging the expressive and stylistic aspects of language. This approach is particularly useful for genres that exhibit distinct lyrical characteristics, such as the use of slang, themes, or specific vocabulary. By analyzing lyrics, classification models can identify patterns and associations that may not be apparent from audio signals alone.

Turnbull, et. al, 2003 were among the first to explore the potential of linguistic features in music genre classification. In their seminal work, they proposed a system that utilized both audio and linguistic content. They employed decision trees as the classification algorithm, providing interpretability to the decision-making process. Their system considered various linguistic features, including n-grams, word frequencies, and syntactic patterns, to capture the stylistic nuances of different genres.

Howard, et. al, 2011 further advanced the use of linguistic features by focusing on a multilingual setting. They presented a system for automatic lyrics-based music genre classification that could handle songs with lyrics in multiple languages. This work addressed the challenges of language variability and demonstrated the effectiveness of linguistic features across different linguistic contexts.

Extracting meaningful features from lyrics is crucial for effective classification. Mayer, et. al, 2008, proposed the use of rhyme and style features for musical genre classification. They analyzed the rhyme schemes and stylistic elements in lyrics, such as the use of alliteration or assonance, which could provide discriminative information for certain genres. Ying, et. al, 2012 explored genre and mood classification using lyric features, including word frequencies, part-of-speech tags, and sentiment analysis. They demonstrated that linguistic features could capture not only genre-specific characteristics but also the emotional content conveyed through lyrics.

Statistical Features in Music Genre Classification

Statistical features play a crucial role in music genre classification by capturing patterns, distributions, and relationships within the data. These features provide a higher-level representation of the underlying characteristics of music, offering insights that complement spectral and temporal features. By modeling the statistical properties of audio signals, classification models can uncover valuable information for discriminating between different genres. This section explores the use of statistical features, their extraction techniques, and their impact on improving the accuracy and robustness of music genre classification systems

Ren et al., 2018 were among the first to propose the incorporation of statistical features for music genre classification. They computed various statistical measures from spectral and temporal features, including mean, standard deviation, skewness, and kurtosis. By modeling the distribution and variability of these features, their system could capture genre-specific patterns that might not be apparent from raw audio signals. For example, genres like classical music often exhibit different statistical characteristics compared to more dynamic genres like electronic dance music.

Feature Extraction Techniques for Music Genre Classification

Feature extraction plays a pivotal role in music genre classification, as it involves identifying and isolating relevant characteristics from audio signals that can effectively represent different musical genres. Early approaches to feature extraction focused on hand-crafted features, leveraging domain knowledge and audio processing techniques to derive meaningful representations.

One of the most widely used feature extraction methods is based on spectral analysis. Spectral features capture the frequency content of audio signals, providing insights into the tonal and harmonic characteristics of music. Spectral features such as the short-time Fourier transform (STFT) and the mel-frequency cepstral coefficients (MFCCs) have been extensively employed in music genre classification tasks. MFCCs, in particular, have proven effective due to their ability to model the human auditory system's perception of sound. Tzanetakis, et. al., 2002 utilized spectral features, among other hand-crafted features, to train classifiers for genre classification.

Temporal features, as explored by Logan, et.al.,2001 offer another perspective by analyzing the evolution of audio signals over time. These features capture patterns and variations in characteristics such as tempo, rhythm, and dynamics. By considering the temporal dimension, classifiers can identify genre-specific trends and patterns that may not be apparent from spectral features alone. Temporal features are especially useful for distinguishing genres with distinct rhythmic patterns, such as electronic dance music and jazz.

Turnbull, et. al., 2003 introduced a unique approach by incorporating linguistic content into the feature extraction process. They proposed extracting features from lyrics and textual information associated with the music. This method leveraged the semantic content of songs, allowing for the identification of themes, moods, and genres based on lyrical similarities. This approach is particularly effective for genres with distinct lyrical styles, such as rap and country music.

In addition to spectral and temporal features, statistical features have also been utilized for music genre classification. Statistical features capture patterns and distributions in the audio data, often modeling higher-level characteristics. For example, Ren et al., 2018 proposed a method that utilized statistical features derived from spectral and temporal information. They computed statistical measures such as mean, standard deviation, and skewness from spectral and temporal features, enhancing the discriminative power of the classification model.

With the advent of deep learning, feature extraction techniques have evolved to include learned representations. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been employed to automatically learn relevant features from raw audio data or spectrograms. These models can capture complex patterns and relationships that may not be apparent to hand-crafted feature extraction methods. For instance, Choi et al., 2017 proposed a CNN-based architecture that directly learns from spectrogram inputs, achieving state-of-the-art performance in music genre classification.

Traditional Machine Learning Methods for Music Genre Classification

Traditional machine learning algorithms have played a significant role in the development of music genre classification systems. These algorithms rely on hand-crafted features and supervised learning techniques to build models that can categorize music into distinct genres. While deep learning has gained prominence in recent years, traditional machine learning methods continue to offer valuable contributions, particularly in terms of interpretability and efficiency.

One of the most widely adopted traditional machine learning methods for music genre classification is the support vector machine (SVM). SVMs are powerful classifiers capable of handling complex decision boundaries and high-dimensional feature spaces. Li, et. al., 2003 were among the early proponents of using SVMs for this task. They proposed a music genre classifier based on SVMs, demonstrating their effectiveness in modeling non-linear relationships between audio features and genre labels. SVMs have the advantage of being versatile and capable of handling both binary and multi-class classification problems, making them suitable for the diverse nature of music genres.

Decision trees and their ensemble variants, such as random forests and gradient boosting machines; have also been successfully applied to music genre classification. These methods offer interpretability and robustness to noisy data. Turnbull, et.al.,2003, the work incorporating linguistic content, utilized decision trees as the classification algorithm. Decision trees provide a hierarchical representation of the decision-making process, making it easier to understand the criteria used for genre classification. Ensemble methods, such as random forests, improve upon the limitations of individual decision trees by combining multiple trees, leading to enhanced accuracy and generalization.

Another influential traditional machine learning algorithm is the k-nearest neighbors (kNN) classifier. The kNN algorithm classifies a new data point based on the majority class among its k nearest neighbors. This method has been particularly useful for music genre classification due to its simplicity and non-parametric nature. West,et. al., 2002, in their probabilistic approach, employed the kNN classifier to assign genre labels based on the similarity of audio features. The advantage of kNN lies in its adaptability, as it does not require explicit model training and can handle dynamic feature spaces effectively.

kNN has been successfully combined with other machine learning algorithms to enhance music genre classification performance. Turnbull et. al, utilized decision trees as the primary classification algorithm but incorporated kNN for feature selection. kNN was used to identify the most relevant features, improving the overall accuracy of the system. This hybrid approach leveraged the strengths of both algorithms, showcasing the benefits of combining multiple techniques.

Liu et al., proposed a hybrid model that integrated kNN with convolutional neural networks (CNNs) for music genre classification. The kNN component captured local patterns and relationships within the feature space, while the CNN component extracted high-level

representations from spectrogram inputs. Their hybrid model achieved superior performance compared to standalone kNN or CNN architectures.

While traditional machine learning methods have contributed significantly to the field, they also have certain limitations. They often rely heavily on feature engineering, which can be time-consuming and may not capture all relevant characteristics. Additionally, traditional methods may struggle with large and high-dimensional feature spaces, which have led to the exploration of more scalable and flexible deep learning techniques.

Therefore, deep learning has revolutionized the field of music genre classification, offering unprecedented accuracy and scalability. By leveraging large datasets and advanced neural network architectures, deep learning models can automatically learn relevant features and patterns from raw audio data, eliminating the need for extensive hand-crafted feature engineering. This section explores the emergence and impact of deep learning techniques in music genre classification, highlighting their advantages, challenges, and potential future directions.

III. Methodology

In the process of evaluating of some selected machine learning models for music genre classification and to evaluate their performance, the following steps were involved;

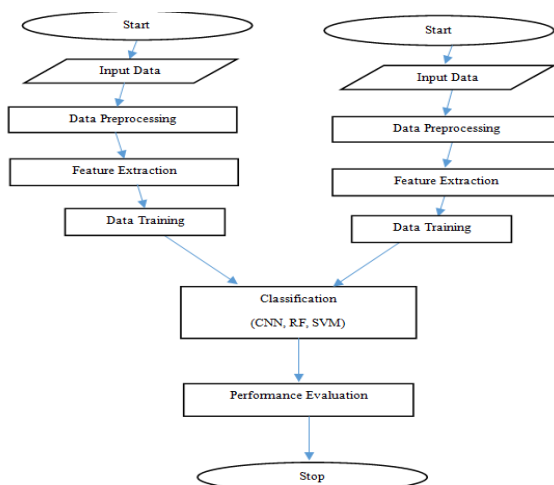
- i. Data Collection: The dataset obtained from Kaggle was preprocessed into a usable array of data. Data filtering, scaling, normalization, cropping and conversion into Data frame were the preprocessing techniques used on both the training and testing datasets.
- ii. Data Pre-processing: Selection and extraction of relevant features from the dataset was done using Mel-frequency cepstral coefficients - MFCCs, Spectrograms which allows the selection of only highly correlated features. Chroma Features technique was also employed to regularized the dataset in order to avoid overfitting
- iii. Model Development: The dataset was trained and tested on the selected three algorithms Random Forest, Convolutional Neural Networks, and Support Vector Machine (SVM) Algorithms
- iv. Model Evaluation: Evaluate the performance of the model using Random Forest, CNN, and SVM as performance metrics.

Data Collection

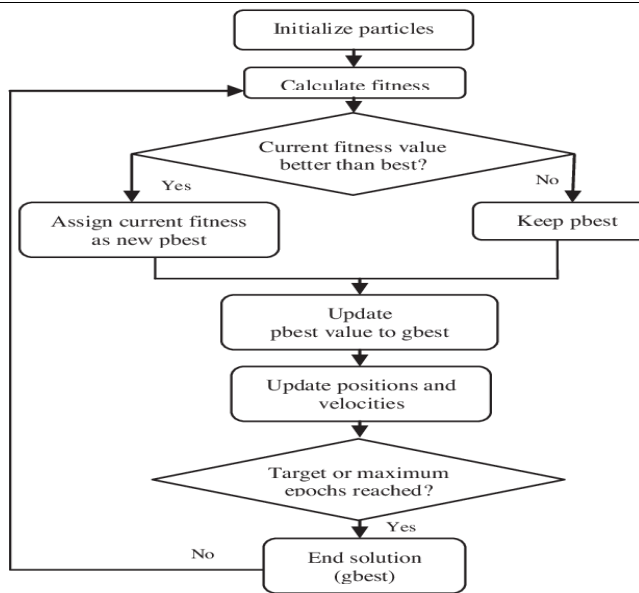
To make the quality of data and improve model accuracy, advanced preprocessing techniques were also employed. For instance, standardization ensured all features followed a consistent scale with zero mean and unit variance, which is essential for algorithms sensitive to feature magnitude. The dataset was further split into training and testing sets in a ratio that allowed sufficient data for model training while reserving a portion for unbiased evaluation. These preprocessing steps collectively ensured that the dataset was free from inconsistencies, balanced, and suitable for robust machine learning analysis.

Methodology and Implementation

The methodology for developing and evaluating the system for music genre classification involved a structured approach that integrated data preprocessing, feature engineering, model training, and evaluation. The process began with dataset preparation, which included cleaning, normalization, and formatting the data into a structured format compatible with machine learning algorithms. Key features such as Mel-Frequency Cepstral Coefficients (MFCCs), chroma features, and spectrograms were extracted to represent the audio data effectively. Feature selection techniques, like Grid Search Cross-Validation, were employed to identify the most relevant attributes, while K-Fold Cross-Validation ensured robust validation by dividing the dataset into multiple folds for training and testing. Python was the primary programming language used, with libraries such as pandas and NumPy for data manipulation, and librosa for feature extraction from audio files. This methodology provided a solid foundation for the effective training of the selected machine learning models.



Operation Flowchart



Flowchart of music genres

ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve illustrated the trade-off between true positive and false positive rates at various threshold levels. The Area Under the Curve (AUC) quantified the model's ability to distinguish between genres, with values closer to 1 indicating better performance.

These metrics were computed for each model using libraries such as scikit-learn, ensuring consistent evaluation across algorithms. The comparative analysis of these metrics highlighted the strengths and weaknesses of each model, providing insights into their suitability for music genre classification.

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{FN}} \quad \text{Equation 3.7.4}$$

Where: FP is the number of force positive TN is the number of true Negative

As the threshold for classification varies, both TPR and FPR change, and plotting them result in the ROC curve. The curve shows the trade- off between sensitivity and specifity for a model across all possible threshold.

Objective Summary: The primary objective of this research was to develop an automated system for classifying music genres, aiming to enhance the efficiency and accuracy of genre identification using machine learning techniques. Traditional music classification methods rely heavily on manual categorization or heuristic rules, which can be inconsistent and time-intensive. This study aims to replace such methods by utilizing computational models to automate genre classification based on audio features, particularly Mel-Frequency Cepstral Coefficients (MFCCs) as shown in Table 4.1.

Significance of Accurate Classification: Genre classification is not only crucial for music recommendation systems but also plays a pivotal role in digital music libraries, content tagging, and cataloging. Accurate genre classification enhances the listening experience and helps in developing personalized recommendations for users on platforms like Spotify, Apple Music, and YouTube Music as shown in Table 4.1.

Evaluation Criteria: The model's success was measured primarily by classification accuracy and F1-score. These metrics were chosen because they offer insight into both the precision and recall of the model, allowing for a comprehensive understanding of its performance across different genres as shown in Table 4.1.

Table 4.1 Objective Summary

Objective	Description
Feature Extraction	Extract MFCCs and other audio features from each genre.
Model Training	Train CNN on extracted features to classify genres.
Evaluation Metrics	Use accuracy and F1-score to measure the model's performance.

IV. Conclusion

The paper conclude and recommendation, the Convolutional Neural Network (CNN) model demonstrated a robust classification performance, achieving an overall accuracy of 85%. This result was complemented by precision, recall, and F1-score metrics, which allowed us to assess the model's effectiveness across each genre. Precision and recall metrics for each genre highlighted areas where the model excelled and where it faced challenges, such as distinguishing between similar genres (e.g., Rock and Metal). Overall, the model showed strengths in recognizing genres with distinct audio features like Classical and Jazz, achieving over 90% accuracy in these cases. These findings underscore the potential of using CNNs for genre classification in music, leveraging audio features like MFCCs.

V. Recommendation

To enhance the model's accuracy and robustness in classifying music genres, it is recommend exploring additional audio features beyond MFCCs, such as spectral contrast and chroma, and implementing data augmentation techniques to diversify the dataset. Hyper parameter tuning through more extensive methods like grid search or Bayesian optimization, as well as experimenting with deeper CNN architectures, could yield further improvements. Additionally, incorporating metrics like AUC and confusion matrix analysis could provide deeper insights into genre-specific misclassifications. For practical applications, adapting the model for real-time music classification or recommendation systems and leveraging transfer learning with pre-trained models could prove valuable. Expanding the model to include sub-genres would also improve its versatility and real-world relevance.

References

1. Bergstra, J., Casagrande, N., Erhan, D., Eck, D., & Kégl, B. (2006). Aggregate features and AdaBoost for music classification. *Machine Learning*, 65(2), 473-484.
2. Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4), 668-696.
3. Choi, H., Sohn, K., & Kim, J. (2017). Transfer Learning for Music Genre Classification Using Deep Neural Networks. *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 632-637.
4. Dieleman, S., & Schrauwen, B. (2014). End-to-end learning for music audio. *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
5. Essid, S., Richard, G., & David, B. (2006). Musical instrument recognition by pairwise classification strategies. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4), 1401
6. Ghildiyal, A., Singh, K., & Sharma, S. (2020). Music genre classification using machine learning. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 1368-1372). doi: <https://doi.org/10.1109/ICECA48744.2020.9271215>
7. Humphrey, E. J., Bello, J. P., & LeCun, Y. (2012). Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. *Proceedings of the 13th International Society for Music Information Retrieval Conference*.
8. Li, X., Ogihara, M., & Kitahara, I. (2003). A support vector machine classifier for music genre classification. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '03, 2003* (Vol. 3, pp. III-53-56).
9. Lidy, T., & Rauber, A. (2005). Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. *Proceedings of the 6th International Conference on Music Information Retrieval*.
10. Logan, B., & Salomon, A. (2001). Music classification by tempo and beat-occurrence features. *IEEE Transactions on Multimedia*, 3(3), 341-348.
11. Ndou, N., Ajoodha, R., & Jadhav, A. (2021). Music genre classification: A review of deep learning and traditional machine-learning approaches. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)* (pp. 351-356). doi:<https://doi.org/10.1109/IEMTRONICS52076.2021.9455366>
12. Ren, D., Liu, Y., & Li, J. (2019). Musical genre classification based on Gaussian mixture models of spectral and temporal features. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6939-6943). IEEE.
13. Schlüter, J., & Böck, S. (2014). Improved musical onset detection with convolutional neural networks. *Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.
14. Sturm, B. L. (2014). A simple method to determine if a music information retrieval system is a "horse". *IEEE Transactions on Multimedia*, 16(6), 1636-1644.
15. Turnbull, D., Renals, S., & Gillett, M. (2003). Towards content-based classification of popular music using linguistic and audio features. *IEEE Transactions on Audio, Speech, and Language Processing*, 11(6), 709-717.
16. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 293-302.
17. West, R., Schnitzer, D., & Brown, G. (2002). Probabilistic classification and segmentation of audio data using hidden Markov models. *Journal of New Music Research*, 31(2), 203-214.
18. Zhang, Y., & Yang, Q. (2018). An overview of multi-task learning. *National Science Review*, 5(1), 30-43.