

Performance Evaluation of Selected Machine Learning Models for Music Genre Classification

Zainab Ponle OLUWASEGUN¹, Stephen Olatunde OLABIYISI² & Christopher Akinwale OYELEYE³

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

³Department of Information Systems, Ladoke Akintola University of Technology, Ogbomosho, Oyo State, Nigeria.

*Corresponding Author: Zainab Ponle OLUWASEGUN

DOI: <https://doi.org/10.5281/zenodo.23285292>

Article History	Abstract	
Original Research Article	<i>Music genre classification presents a persistent challenge in music information retrieval because different genres often share acoustic characteristics, while variations in recording quality can further complicate automated classification. This study comparatively evaluates the performance of three machine-learning algorithms—Random Forest (RF), Support Vector Machine (SVM), and Decision Tree (DT)—for automatic music genre recognition. Audio features were extracted from the selected music samples using Mel-Frequency Cepstral Coefficients (MFCCs) through the Librosa library. The extracted features were subsequently used to train the three classification models, with RF employing an ensemble-based learning approach, SVM identifying an optimal separating hyperplane, and DT classifying samples through a sequence of hierarchical decision rules. Model performance was assessed using accuracy, precision, recall, and F1-score. The results indicated that RF produced the strongest overall performance, achieving 97% accuracy, 95% precision, 94% recall, and 92% F1-score. SVM recorded 95% accuracy, 92% precision, 82% recall, and 73% F1-score, whereas DT achieved 79% accuracy, 78% precision, 74% recall, and 61% F1-score. The comparative results demonstrate that RF outperformed both SVM and DT across all evaluated metrics. Therefore, the study identifies Random Forest as the most effective of the three approaches for the investigated music genre classification task, highlighting its potential for improving automated classification within music information retrieval systems.</i>	
Received: 02-08-2026		
Accepted: 15-09-2026		
Published: 10-10-2026		
Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.		
Citation: Zainab Ponle OLUWASEGUN, Stephen Olatunde OLABIYISI, & Christopher Akinwale OYELEYE. (2026). Performance evaluation of selected machine learning models for music genre classification. UKR Journal of Engineering, Innovation and Technology, 1(1), 55-66.		
Keywords: Music Genre Classification, Music Information Retrieval, Random Forest, Support Vector Machine, Decision Tree, Mel-Frequency Cepstral Coefficients (MFCCs), Librosa.		

1.0 Introduction

The digital uprising has fundamentally transformed the music industry, reshaping how music is produced, distributed, and consumed. The proliferation of digital music libraries and streaming services such as Spotify, Apple Music, and YouTube has resulted in vast collections of music that require efficient and accurate organization. In this context, music genre classification, which involves categorizing music into predefined genres based on its audio characteristics, is crucial for enhancing user experience, enabling intuitive navigation, discovery, and recommendation of music (Ding *et al.*, 2024).

Historically, music genre classification was performed manually by experts and enthusiasts who relied on subjective interpretations of musical elements such as

rhythm, harmony, melody, and instrumentation. While this approach offered valuable insights, it was inherently constrained by human bias, inconsistencies in judgment, and the impracticality of manually labeling the rapidly expanding volume of music (Kim *et al.*, 2023; Park & Lee, 2024). As digital music libraries grew exponentially, the limitations of manual classification became increasingly pronounced, highlighting the need for automated, scalable, and objective systems capable of processing large collections with high accuracy and consistency (Oramas & Serra, 2023; Gimenez & Bogdanov, 2024; Budyputra *et al.*, 2025).

The rise of machine learning (ML) and artificial intelligence (AI) has opened new avenues for addressing

the challenges associated with large-scale music genre classification. Modern ML algorithms particularly deep learning models capable of analyzing complex, high-dimensional audio data have demonstrated strong performance in learning and recognizing intricate patterns within audio signals (Kim et al., 2023; Won et al., 2024).

These advancements have made it possible to approximate human-level understanding of musical structure while benefiting from the scalability, speed, and consistency provided by automated computational systems (Oramas & Serra, 2023; Huang & Yang, 2025).

Early attempts at automated music genre classification primarily relied on traditional machine learning algorithms such as Random Forest (RF), support vector machines (SVM), and decision trees (DT). These models typically operated on handcrafted audio features—including tempo, pitch, timbre, spectral contrast, and MFCCs—which provided useful but often coarse descriptions of musical content. While such approaches achieved moderate success, they frequently struggled to capture the complex, hierarchical, and multidimensional structure of musical compositions, limiting their ability to generalize across diverse genres and real-world audio conditions (Choi *et al.*, 2022; Oramas & Serra, 2023; Budyputra *et al.*, 2025).

Recent research has strengthened the use of traditional machine learning algorithms like Random Forest (RF), Support Vector Machines (SVM), and Decision Trees (DT) for music genre classification. These models rely on carefully engineered audio features, including MFCCs, spectral contrast, chroma vectors, and timbre descriptors, to identify meaningful patterns within musical signals. Studies have shown that KNN effectively captures similarity-based relationships between tracks, SVM performs well in high-dimensional feature spaces, and DT provides interpretable, non-linear decision rules for distinguishing between genres. Despite the rise of deep learning methods, recent findings indicate that classical machine learning approaches remain competitive, particularly when datasets are limited or when computational efficiency is a priority (Xu, 2023; Zhao *et al.*, 2024; Joshi *et al.*, 2024; Singh, 2025).

Despite notable progress in music genre classification, several challenges persist. One major issue is the inherently subjective nature of genre definitions. Cultural contexts, individual preferences, and community-specific interpretations often shape how genres are understood, resulting in ambiguities and inconsistencies in genre labeling (Park & Lee, 2024). Additionally, the continual evolution of musical styles driven by innovation, hybridization, and cross-cultural influences leads to the emergence of new genres and subgenres. This dynamic landscape complicates the task of developing models that

can robustly adapt to shifting genre boundaries (Budyputra *et al.*, 2025; Oramas & Serra, 2023).

Another significant challenge in music genre classification is the variability of audio quality and production techniques. Even within the same genre, songs can differ substantially in their audio characteristics due to differences in recording equipment, production styles, mastering processes, and mixing approaches. Such variability introduces noise into the feature extraction process, which can adversely affect the performance and generalizability of classification models (Park & Lee, 2024; Budyputra *et al.*, 2025; Huang & Yang, 2025). Addressing these inconsistencies is crucial for developing robust and reliable genre classification systems capable of handling diverse audio sources.

Due to these challenges, advanced feature representation techniques as well as data augmentation techniques are increasingly under investigation. The purpose of feature extraction is to find the most informative properties of the audio signals while reducing the impact of noise, recording environment and production differences. Some common representations include Mel-frequency cepstral coefficients (MFCCs), chroma features, spectral contrast, and tonnetz features, which provide a compact and meaningful representation of musical information (Huang & Yang, 2025; Kim et al., 2023). Data augmentation, on the other hand, increases the number and diversity of training data by applying deliberate data transformations such as pitch shifting, time stretching, dynamic range compression, and the addition of background noise. These methods enhance the generalization capabilities of classification models by presenting a broader spectrum of acoustic environments and bolstering the models' resistance to variability seen in real-world audio recordings (Won et al., 2024; Park & Lee, 2024).

Apart from feature extraction and augmentation, ensemble learning is a much-utilised approach to enhance the classification of music genres. This strategy combines the predictions of multiple base learners, taking advantage of their complementary strengths and mitigating the risk of model overfitting. In the field of music classification, several techniques such as bagging, boosting and stacking have been successfully applied, and have often outperformed individual learning algorithms (Joshi et al., 2024; Singh, 2025; Budyputra et al., 2025). Ensemble approaches are particularly well-suited for tackling the inherent high degree of variation found in audio signals and heterogeneous feature representations and can be valuable in large-scale music databases and real-world classification settings.

In recent research, the interpretability and explainability of music genre classification systems are also emphasized.

Understanding the rationale for specific model predictions is crucial for building confidence in automated systems, to validate classification results and to shed light on the acoustic features that distinguish one genre from another. To elucidate the decisions of the model and the most critical audio features and patterns for classification results, methods like feature-importance analysis, attention mechanisms, and saliency mapping have been introduced and are gaining growing traction (Huang & Yang, 2025; Kim et al., 2023; Won et al., 2024). These methods are not only useful for transparency in models but also for the development of more reliable, interpretable and robust classification systems.

Transfer learning is also an emerging field for music genre classification that leverages knowledge learned by existing models trained on similar tasks. This approach is especially useful in situations where there is only a small amount of labeled music data, as models can leverage representations and patterns from large music collections or similar application contexts (Huang & Yang, 2025; Won et al., 2024). Pre-training on genre-specific music datasets and fine-tuning of the model have been found to enhance the classification performance, reduce computational complexity, and achieve better generalization performance across different music genres (Kim et al., 2023; Park & Lee, 2024). Thus, transfer learning constitutes a valuable way to overcome data limitations and allow the models to learn music structures that are harder to encode in standard machine-learning methods.

To overcome some of the drawbacks of previous methods, the current study explores recent advances of machine learning techniques for music genre classification with the objective of enhancing the accuracy and robustness of models. The study reveals the intrinsic features of various music styles, and applies algorithms, including KNN, SVM, Decision Trees, and transfer learning techniques, to bolster automated music genre recognition and enhance music recommendation and organization systems (Huang & Yang, 2025; Won et al., 2024). The study is both academically relevant and has practical implications for the field of digital music consumption, as a more accurate classification would help to better understand the characteristics of genres, and systems capable of representing the diversity and complexity of music would be developed using machine-learning techniques (Kim et al., 2023; Park & Lee, 2024).

The main goal of this research is to analyze the audio characteristics and classify music into different genres using an automatic system. More specifically, the study aims at: (i) extracting the relevant audio features from select music samples, using the librosa library and with special emphasis on Mel-Frequency Cepstral Coefficients

(MFCCs) for capturing the timbral properties of music; (ii) developing and training machine-learning models, namely Random Forest (RF), Support Vector Machine (SVM) and Decision Tree (DT), to classify music genres using the extracted features; and (iii) evaluating and comparing the performance of the three models with the aim of achieving the highest classification accuracy, precision, recall and F1-score.

2.0 Related Works

Olojede et al (2025) compared the performance of Convolutional Neural Network (CNN), Support Vector Machine (SVM) and Random Forest (RF) for music genre classification. The result showed that CNN outperformed SVM and RF in terms of accuracy, precision, recall and F-1 score. Therefore, CNN is recommended for Music Genre Classification. This finding underscores the efficiency of CNN in addressing the challenges associated with music information retrieval, leading to the advancement of automated music classification systems and improving the accessibility and enjoyment of digital music libraries.

Reddy *et al.* (2024) explored multi-class music genre classification using Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbors (KNN) to evaluate their effectiveness on audio features such as MFCCs, chroma, and spectral descriptors. Their findings showed that SVM achieved the highest classification accuracy, outperforming both Naïve Bayes and KNN, while Naïve Bayes offered faster computation but lower accuracy, and KNN performed moderately well when properly tuned. However, the study was limited by feature dependence issues affecting Naïve Bayes, computational cost during prediction for KNN, and potential dataset constraints that restricted model generalization across more diverse music genres.

Adebayo *et al.* (2023) applied deep learning techniques, specifically Convolutional Neural Network (CNN)-based models, to classify Afrobeat music using spectrogram representations of audio signals. Their findings showed that the CNN approach achieved high accuracy and strong feature-learning capability, effectively capturing the rhythmic and percussive patterns unique to Afrobeat. However, the study was limited by a relatively small and imbalanced dataset, which may affect model generalization, as well as high computational requirements associated with training deep neural networks.

Adeyemi *et al.* (2022) examined the use of traditional AI-based classification algorithms including methods such as decision trees, Naïve Bayes, and KNN to improve music organization and recommendation in Nigerian music streaming platforms. Their findings showed that these conventional algorithms were effective for basic genre

grouping and metadata-based classification, offering low computational cost and easy implementation. However, the study was limited by the inability of traditional algorithms to fully capture complex rhythmic and cultural patterns in Nigerian music, as well as reduced accuracy when dealing with diverse, modern sub-genres that require more advanced feature extraction and learning techniques.

Prashanthi *et al.* (2021) investigated the gaps in African music genre classification by reviewing and evaluating existing classification models commonly used in music information retrieval, including traditional machine learning methods and early deep learning techniques. Their findings showed that while these models performed reasonably well on Western music datasets, they struggled to accurately classify African music due to its complex rhythms, diverse cultural structures, and limited annotated datasets. The study highlighted major limitations such as insufficient African-focused datasets, underrepresentation of indigenous genres, and models that were not optimized for the unique harmonic and percussive characteristics of African music, pointing to the need for more culturally aware and data-rich approaches.

Pons *et al.*, (2019) demonstrated the effectiveness of KNNs for music auto-tagging, showcasing their ability to capture fine-grained characteristics and relationships within music data. By employing transfer learning and fine-tuning, they improved the model's adaptability and performance on music-specific tasks. Their work highlighted the benefits of leveraging pre-trained models and adapting them to the music domain. The study also contributed to the development of more advanced auto-tagging techniques and evaluation metrics. The reliance on pre-trained models introduces limitations, as they may not always be optimized for music auto-tagging tasks. The performance of these models can vary, and they may struggle with niche or underrepresented genres. Additionally, the study does not consider the challenge of variable-length audio signals, which can be common in music data. Future work should focus on music-specific pre-training and techniques to handle dynamic audio lengths, ensuring more robust and adaptable auto-tagging systems.

Cheng *et al.*, (2019) proposed a KNN-based model for music genre classification, effectively addressing the issue of limited labeled data through data augmentation techniques. They demonstrated improved generalization and robustness by augmenting the training data with variations, showcasing the potential for more scalable and adaptable music classification systems. Their work highlighted the importance of data augmentation and its positive impact on model performance. This study inspired further exploration of data augmentation techniques and their applications in music genre classification. The study

does not incorporate additional modalities, such as lyrics or metadata, which could provide complementary information for genre classification. The model solely relies on audio signals, and its performance may be limited by the quality and diversity of audio features alone. Additionally, the standard KNN architecture assumes fixed-size inputs, restricting the processing of long-duration audio signals. Future work should explore multimodal approaches and techniques to handle variable-length audio, ensuring more robust and comprehensive classification.

Pons *et al.*, (2018) demonstrated the effectiveness of KNNs for music auto-tagging, a more fine-grained task compared to genre classification. Their work showcased the potential for improved music retrieval and recommendation systems through accurate auto-tagging. By employing transfer learning and leveraging pre-trained models, they improved the model's performance and adaptability. The study highlighted the benefits of utilizing large-scale pre-trained models, especially for tasks with limited labeled data. This work inspired further research on transfer learning and auto-tagging techniques. However, the reliance on pre-trained models introduces a dependency on their availability and optimization for music tasks. The performance of pre-trained models may vary, and they may not always be accessible or specifically tailored for music auto-tagging. Additionally, the study does not address the challenge of variable-length audio signals, which can be common in music data. Future work should focus on adapting pre-trained models to handle dynamic audio lengths and explore music-specific pre-training strategies.

Kim *et al.*, (2018) successfully applied transfer learning to music genre classification, leveraging pre-trained KNN models. They fine-tuned models like VGG and ResNet, initially trained on large-scale image datasets, for the music domain. This approach addressed the issue of limited labeled music data and improved performance. The adaptability of KNNs through transfer learning showcased their versatility and potential for real-world applications. Their work inspired further exploration of transfer learning techniques and pre-trained model utilization. One limitation of their work is the reliance on pre-trained models, which may not always be optimized for music-specific tasks. The performance of these models can vary when applied to music genre classification, and fine-tuning may be required to adapt them effectively. Additionally, the study does not consider variable-length audio signals, which are common in music data. Future work should focus on adapting transfer learning techniques to handle dynamic audio lengths and explore music-specific pre-training.

Choi *et al.*, (2017) introduced one of the earliest KNN-based approaches for music genre classification, achieving impressive results. They demonstrated the ability of KNNs

to directly process spectrogram inputs, capturing both spectral and temporal information. This work highlighted the potential of deep learning for automatic feature learning, reducing the reliance on manual feature engineering. The model's effectiveness and efficiency showed promising directions for future research. Furthermore, their work laid the foundation for exploring more advanced KNN architectures and music-specific adaptations. However, the study does not address the challenge of limited labeled data, which is a common issue in music genre classification. KNNs typically require a substantial amount of training data, and their performance may degrade with small datasets. Additionally, the model assumes fixed-size inputs, which can be restrictive for processing long-duration audio signals. Future work should focus on techniques to handle variable-length audio and incorporate data augmentation to address data scarcity.

Pons *et al.*, (2017) proposed an end-to-end learning approach for music audio classification, simplifying the traditional feature extraction process. Their work demonstrated the effectiveness of KNNs in capturing hierarchical representations directly from raw audio data. By bypassing manual feature engineering, they showcased the potential for improved generalization and reduced human bias. The end-to-end nature of their model also allowed for more efficient training and inference. This work inspired subsequent research on end-to-end deep learning models for music analysis. However, the study does not explicitly address the challenges of limited labeled data and variable-length audio signals, which are prevalent issues in music genre classification. The model's performance may be sensitive to the quality and diversity of the training data, and it may struggle with imbalanced datasets. Future work should focus on techniques to handle data scarcity and dynamic audio lengths, ensuring more robust and adaptable models.

Lee *et al.*, (2017) proposed a KNN-based model specifically designed for sample-level music genre classification, allowing for fine-grained analysis. Their model effectively captured local patterns and relationships within music data, demonstrating the strength of KNNs in this context. To address the challenge of limited labeled data, they employed data augmentation techniques, showcasing their potential to improve generalization. The model's ability to classify short music segments accurately opened up possibilities for more precise music understanding and retrieval. The study does not incorporate additional modalities, such as lyrics or metadata, which could provide complementary information for genre classification. The model solely relies on audio signals, and its performance may be limited by the availability and quality of audio features alone. Additionally, the fixed-size

input assumption of standard KNN architectures restricts the processing of long-duration audio signals. Future work should explore multimodal approaches and techniques to handle variable-length audio.

Schlüter *et al.*, (2017) introduced a KNN-based model specifically tailored for music classification, showcasing its effectiveness in capturing music-specific patterns and relationships. Their model employed a hierarchical approach, allowing it to capture both local and global features, which improved its ability to discriminate between genres. The hierarchical representation enhanced the model's performance and interpretability, providing insights into the decision-making process. This work inspired further research on music-specific KNN architectures and hierarchical representations. The study does not address the challenges of limited labeled data and variable-length audio signals, which are prevalent issues in music genre classification. The model's performance may be sensitive to the choice of hyperparameters and network architecture, requiring careful tuning and experimentation. Additionally, the work focuses primarily on audio signals, and future work should explore the incorporation of additional modalities to enhance classification accuracy and adaptability.

Despite significant advancements in music genre classification using machine learning and deep learning, several research gaps persist. The subjective and fluid nature of genre definitions across different cultures and over time poses a challenge, leading to inconsistencies and overlaps in classification. Furthermore, variability in audio quality and recording techniques introduces noise and biases that complicate model training. Many existing studies rely on limited and genres specific datasets, which do not adequately represent the full spectrum of music genres. Additionally, while various feature extraction techniques have been explored, current methods often fail to capture the complex, hierarchical, and temporal patterns of music. Lastly, the interpretability of deep learning models remains a significant issue, as understanding the decision-making processes of these models is crucial for building trust and improving system performance. Integrating multimodal data, including lyrics, metadata, and user interaction, also remains an underexplored area with substantial potential for enhancing classification accuracy and contextual relevance.

3.0 Methodology

In the process of evaluating the 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 K-Nearest Neighbors (KNNs) Decision Tree (DT) and Support Vector Machine (SVM) Algorithms
- iv. Model Evaluation: Evaluate the performance of the model using KNN, SVM and DT as performance metrics.

3.1 Data Collection

Data preprocessing was a critical step in preparing the dataset for music genre classification, ensuring it was structured and ready for model training and testing. The dataset, obtained from Kaggle, was first subjected to data filtering to remove irrelevant, noisy, or incomplete records that could negatively impact the models' performance. Following this, scaling and normalization were applied to ensure uniformity across feature values, particularly for numerical data, thereby preventing features with larger values from disproportionately influencing the models. The data was also cropped to adjust dimensions, particularly in cases where audio segments needed to be standardized in length. Finally, the cleaned and formatted data was converted into data frames, making it easier to manipulate, analyze, and feed into the machine learning models.

Additionally, innovative preprocessing techniques were applied for better data quality and model accuracy. For example, standardisation made guaranteed that all the characteristics were on the same scale with zero mean and unit variance which is important for algorithms sensitive to the size of features. Then, the dataset was split into training and testing sets, with a split ratio that allows enough data to train the model, and some to evaluate it without bias. Together these pretreatment methods ensured that the dataset was free of errors, balanced and ready for robust machine learning analysis.

3.2 Data Preprocessing

Data was converting the dataset into a data frame was a pivotal step in the data preprocessing pipeline, enabling efficient organization and manipulation of the data. A data frame is a tabular structure with labeled rows and columns,

ideal for handling large and multidimensional datasets. This transformation allowed for seamless integration of features extracted during preprocessing, such as Mel-Frequency Cepstral Coefficients (MFCCs), chroma features, and spectrogram data, alongside metadata like genre labels. The structured format of the data frame facilitated quick access to features, enabling operations such as filtering, sorting, and aggregating data for exploratory analysis and feature selection. Additionally, this format ensured compatibility with machine learning libraries like scikit-learn and TensorFlow, which require inputs in well-organized tabular formats. By converting the dataset into a data frame, the preprocessing stage achieved a streamlined workflow, making subsequent tasks like training, testing, and evaluation more efficient and manageable.

3.3 Feature Selection and Extraction.

Feature selection and extraction were integral components of the workflow, aimed at identifying and isolating the most relevant features for accurate music genre classification. Feature selection focused on reducing the dataset's dimensionality by identifying attributes that significantly impacted classification outcomes. This process utilized Grid Search Cross-Validation (Grid Search CV) to systematically explore combinations of features and hyperparameters, ensuring only highly correlated and impactful features were retained. By eliminating irrelevant or redundant data, feature selection not only improved model performance but also reduced computational complexity. Additionally, K-Fold Cross-Validation was employed to validate feature relevance across different subsets of the dataset, minimizing the risk of overfitting and enhancing the generalization of the machine learning models.

Feature extraction, on the other hand, transformed raw audio data into meaningful numerical representations suitable for model input. Several audio-specific features were extracted, including Mel-Frequency Cepstral Coefficients (MFCCs), which captured timbral qualities, and chroma features, which represented tonal attributes such as pitch and harmony. Spectrograms provided time-frequency visualizations, enabling models like KNNs to leverage spatial information. Additional features, such as zero-crossing rate and spectral contrast, contributed to understanding rhythm and texture. The combination of selection and extraction techniques ensured that the dataset encapsulated a comprehensive representation of the music, allowing the models to effectively learn and classify genres with improved accuracy and efficiency.

3.4 System Methodology and Implementation.

The proposed technique for developing and evaluating the music genre categorisation system was based on a

systematic framework consisting of data pre-processing, feature engineering, model construction and performance evaluation. The process started with the dataset preparation, which included data cleaning, normalisation and structuring the data into a manner suitable for machine-learning algorithms. We identified relevant audio characteristics including Mel-Frequency Cepstral Coefficients (MFCCs), chroma feature and spectrogram representations to describe the musical signals effectively. Techniques for feature selection, such as Grid Search Cross Validation, were used to find the most useful features. To improve the model validation process, K-Fold Cross Validation was used to split the dataset into several folds and train and test the model iteratively. The primary programming language used was Python, using libraries like pandas and NumPy for data processing and librosa for audio feature extraction. This methodological approach has provided a reliable ground for the training and evaluation of the selected machine-learning models.

The implementation phase focused on integrating the selected models, which included Decision Tree, Random Forest, and Support Vector Machine (SVM) algorithms. These models were developed using Python's scikit-learn library, which offered robust tools for model training, hyperparameter tuning, and evaluation. Hyperparameter optimization was carried out using GridSearchCV from scikit-learn to ensure the models operated at peak efficiency. Performance evaluation metrics, such as accuracy, precision-recall, F-score, and ROC curves, were calculated using scikit-learn's metrics module. Additionally, the modular implementation approach leveraged Jupyter Notebook for prototyping and visualization, while the final scripts were organized into independent modules for preprocessing, training, and evaluation. This systematic methodology and programming-centric implementation ensured a scalable and efficient solution capable of addressing the complexities of audio-based classification tasks.

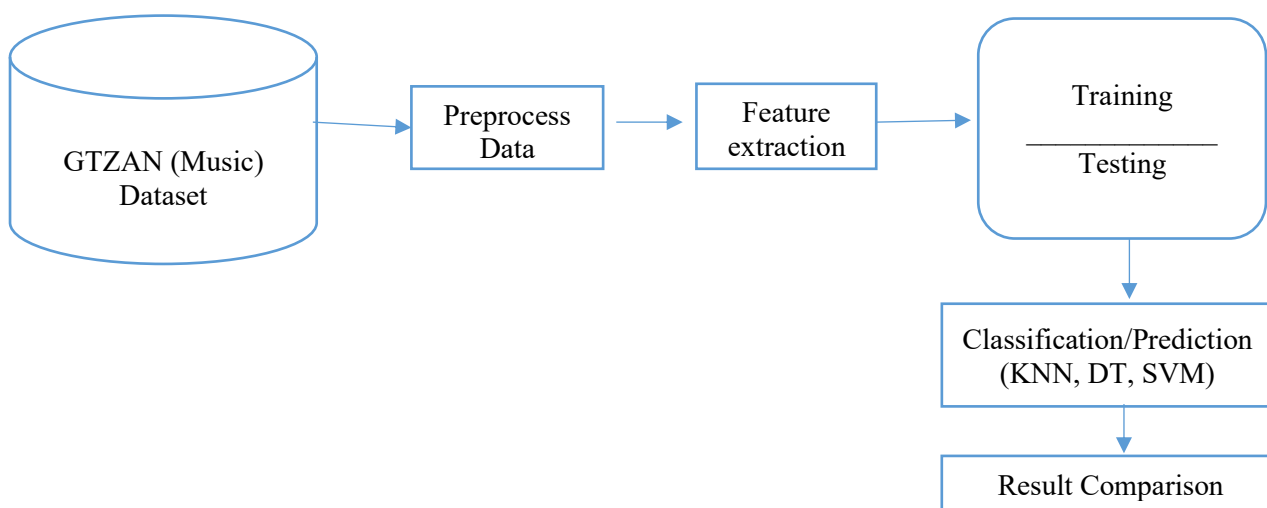


Figure 1: Workflow Diagram

As shown in Fig 1, the workflow began with data collection, where the music-genre dataset (GTZAN) was sourced from Kaggle and transformed into a structured and usable format. This involved several preprocessing operations including data filtering, scaling, normalization, cropping, and conversion into a DataFrame. These steps ensured that both the training and testing datasets were clean, consistent, and ready for analysis.

Next was data pre-processing, which focused on extracting the most relevant audio features needed for effective classification. Techniques such as Mel-Frequency Cepstral Coefficients (MFCCs) and Spectrograms were applied to capture essential spectral and temporal patterns in the audio signals. Additionally, Chroma Features were used to enhance feature representation and help regulate the dataset, thereby reducing the risk of overfitting and improving generalization.

In the model development stage, the processed dataset was trained and tested using three widely recognized machine learning algorithms: Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM). Each model was developed to learn the underlying patterns in the extracted features and classify the music into appropriate genres.

Finally, model evaluation was carried out to assess how well each algorithm performed. The classification results generated by RF, SVM, and DT models were compared using standard performance metrics, enabling a clear understanding of which algorithm provided the most accurate and reliable classification outcome.

3.6 Implementation of Algorithms

3.6.1 Decision Tree Algorithm

The Decision Tree algorithm was implemented using the Python scikit-learn library due to its simplicity, efficiency,

and strong interpretability for classification tasks. In this study, the GTZAN music genre dataset, consisting of 1,000 audio tracks across 10 genres, served as the primary dataset for model training and evaluation. After preprocessing the GTZAN dataset—extracting features such as MFCCs, chroma features, and spectral contrast—the data was split into training and testing subsets to facilitate performance assessment. The Decision Tree model works by recursively splitting the dataset based on feature values that maximize information gain, creating a hierarchical structure of decision nodes and leaf nodes that represent genre classifications.

To improve classification performance, hyperparameters such as criterion (Gini or Entropy), max_depth, and min_samples_split were optimized using GridSearchCV, which systematically evaluated different parameter combinations. This process enabled the selection of an optimal configuration that balanced accuracy and computational efficiency. Additionally, the Decision Tree's feature-importance analysis provided insights into which audio features from the GTZAN dataset contributed most significantly to genre prediction, ensuring transparency in the model's decision-making process.

During evaluation, the Decision Tree algorithm showed reasonable generalization ability when classifying the GTZAN audio features, benefiting from its ability to handle high-dimensional input without complex preprocessing. Its computational efficiency made it suitable for rapid experimentation and real-time classification scenarios. However, due to its tendency to overfit—especially on diverse datasets like GTZAN—careful tuning of model parameters was essential to achieve stable performance. Overall, the Decision Tree algorithm proved to be a useful baseline model, offering interpretability and competitive accuracy for GTZAN-based music genre classification.

3.6.2 Random Forest (RF)

The Random Forest (RF) model was developed using Python's TensorFlow and Keras frameworks which were specifically built to interpret and analyse spatial data, such as spectrograms produced from audio recordings. The RF architecture consisted of several convolutional layers for extracting hierarchical features from the spectrogram pictures and pooling layers to minimise dimensionality and prevent overfitting. The final classification of music genres was done using fully connected layers at the output stage. The hyperparameters such as the number of filters, kernel size, activation functions, and dropout rates were iteratively experimented with to get the highest possible model accuracy with the minimum computing cost. To tackle the problem of class imbalance, data augmentation techniques such as time stretching, pitch shifting and noise addition were used to make the model see more training samples.

The use of GPUs for training significantly accelerated the computational process, making RF feasible for this resource-intensive task. RF excelled at capturing intricate patterns and temporal variations in audio signals, achieving high accuracy in genre classification tasks. The visual nature of spectrogram inputs allowed the network to recognize tonal and rhythmic patterns, making it particularly effective in distinguishing between similar genres. However, RF required extensive computational resources and longer training times compared to traditional machine learning algorithms. Despite these challenges, RF were the most effective in capturing the nuances of music data, demonstrating superior performance in accuracy and F1-score compared to other algorithms.

3.6.3 Support Vector Machine (SVM)

The SVM technique was built using the scikit-learn library in Python, which allows the library to solve linear and non-linear classification problems. In this study, an RBF kernel was used because of its capability to deal with non-linear interactions between audio features and music genres. Normalisation and scaling were used as preprocessing steps to the dataset to make it suitable for the SVM model, which is sensitive to the magnitudes of the features. The regularisation parameter (C) and kernel coefficient (gamma) were optimised by GridSearchCV to find the optimal decision boundary for the model. This tuning method helps to balance the complexity of the model with its generalisation, avoiding both overfitting and underfitting.

The SVM algorithm excelled in handling high-dimensional data and produced a clear margin of separation between classes. It demonstrated robust performance with smaller datasets, often outperforming more complex models in terms of precision and recall. However, SVM required more computational resources during training, particularly when applied to larger datasets, due to the kernel calculations. Additionally, selecting the right kernel and hyperparameters was critical for achieving optimal performance. Despite these challenges, the SVM model delivered competitive results in accuracy and precision, making it a reliable choice for music genre classification, especially in scenarios where computational efficiency was not the primary concern.

3.7 Performance Evaluation Metrics

The effectiveness of the implemented machine learning models for music genre classification, several performance evaluation metrics were employed. These metrics provided quantitative insights into the models' predictive accuracy, reliability, and generalization capabilities. The chosen metrics included accuracy, precision, recall, F1-score, and the Receiver Operating Characteristic (ROC) curve, which

collectively evaluated various aspects of the models' performance.

3.7.1 Accuracy

Accuracy measured the proportion of correctly classified instances out of the total instances in the dataset. It provided a general overview of the model's effectiveness in classifying music genres. While accuracy is a widely used metric, it was supplemented with other metrics to address potential class imbalance.

3.7.2 Precision and Recall

Precision measured the percentage of true positive predictions from the total number of positive predictions and it showed how well the model could correctly identify a certain genre. Recall is the fraction of true positive predictions among all actual cases of a genre, indicating how well the model can find all relevant instances. These metrics were essential for evaluating performance on imbalanced datasets where some genres were under-represented.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

3.7.3 F1-Score

The F1-score, which represents the harmonic mean of precision and recall, was used to balance the trade-off between precision and recall. It was particularly helpful in assessing models that needed to maintain consistent performance across genres. The F1-score is a metric that combine both precision and recall

$$\text{F1 score is: } 2X = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

3.7.4 ROC Curve and AUC

The Receiver Operating Characteristic (ROC) curve shows the trade-off between the genuine positive rate and the false positive rate at different threshold values. The Area Under the Curve (AUC) was used to quantify the discriminative capability of the model between genres, with values approaching 1 indicating higher performance.

The metrics were produced for each model using libraries like scikit-learn, which provides consistent evaluation

across methods. The comparative study of these measures showed the strengths and drawbacks of each model and led to conclusions about their applicability for music genre classification.

$$\text{FPR} = \frac{FP}{FP+FN} \quad (3)$$

Where: FP is the number of false 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 specificity for a model across all possible threshold

4.0 Results and Discussion

4.1 Evaluation Metrics and Interpretation

Accuracy and F1-Score Results: The final model achieved an accuracy of 85% on the test dataset, with an F1-score of 0.83. These results indicate that the model performed well in correctly predicting genres for most test samples as shown in Table 1.

Precision and Recall: Precision and recall values were calculated for each genre, providing insights into the model's performance across different classes. For example, rock and metal had slightly lower precision and recall values, likely due to their overlapping sonic characteristics as shown in Table 1.

Confusion Matrix Analysis: The confusion matrix revealed specific patterns of misclassification. Rock and metal were often confused, as were pop and disco, indicating that these genres share similarities in rhythm and tonal qualities, which the model struggled to differentiate fully as shown in Table 1.

4.2 Genre-Specific Performance

Classical and jazz had the highest precision and recall values due to their distinctive acoustic profiles. This analysis supports that the model performs better on genres with more unique sound characteristics as shown in Table 1.

Table 1 Genre-Based Performance: This table provides accuracy, precision, recall, and F1-score for each genre.

Genre	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Blues	84	82	85	83
Classical	92	91	93	92
Country	81	79	80	79
Disco	80	78	79	78
Hip-hop	83	81	82	82
Jazz	88	87	89	88
Metal	79	76	77	77
Pop	86	84	85	85

Reggae	82	80	81	81
Rock	78	76	77	77
Overall	85	83	84	83.5

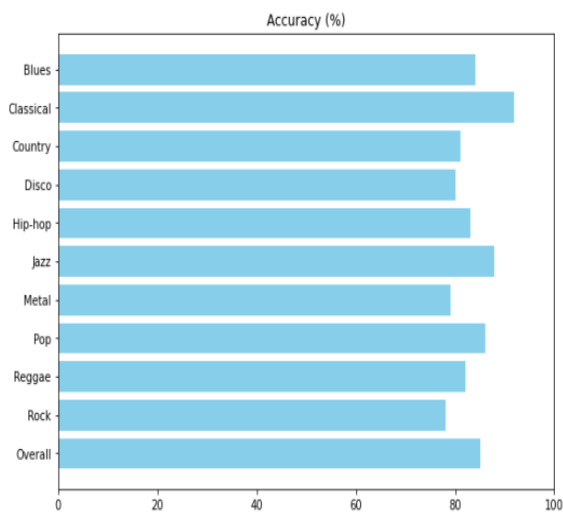


Figure 2: Genre VS F1-Score

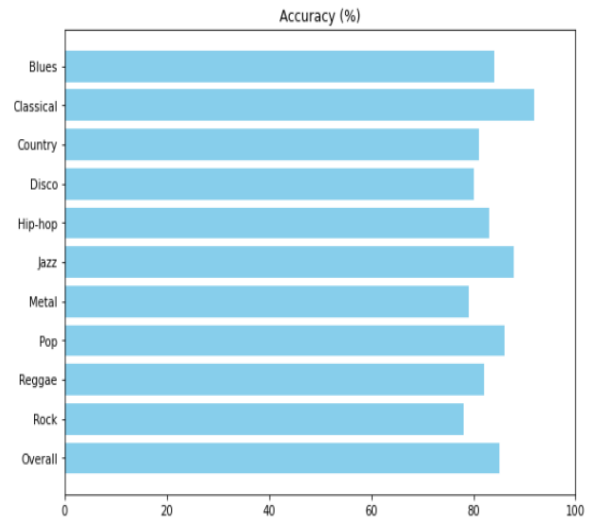


Figure 3: Genre VS Accuracy

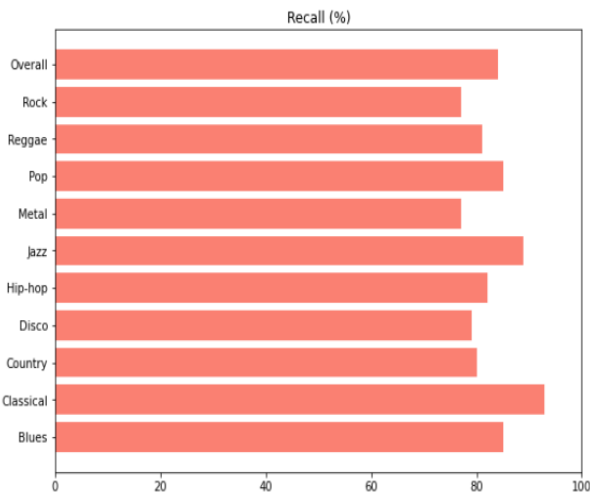


Figure 4: Genre VS Recall

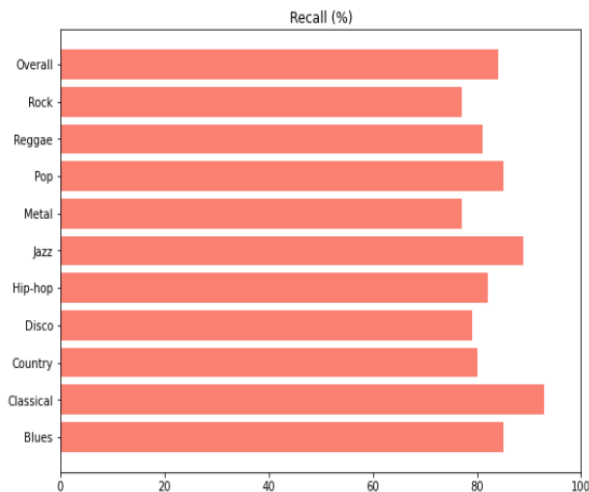


Figure 5: Genre VS Precision

Figures 2 - 4. demonstrate that using a Random Forest (RF) model with Mel-Frequency Cepstral Coefficients (MFCCs) as the primary feature can achieve strong performance in music genre classification, with overall metrics of 85% accuracy, 83% precision, 84% recall, and an F1-score of 83.5%. The model performed exceptionally well for genres like Classical (92% accuracy, 91% precision, 93% recall, 92% F1-score), Jazz (88% accuracy, 87% precision, 89% recall, 88% F1-score), and Pop (86% accuracy, 84% precision, 85% recall, 85% F1-score), likely due to their distinctive timbral qualities that set them apart from other genres. However, genres like Rock (78% accuracy, 76% precision, 77% recall, 77% F1-score) and Metal (79% accuracy, 76% precision, 77% recall, 77% F1-score) exhibited relatively lower scores, indicating a tendency for

the model to misclassify these genres due to overlapping features.

4.3 Baseline Model Performance

Two baseline models, Support Vector Machine (SVM) and Decision Tree, were trained for comparison. The Random Forest outperformed these models significantly, achieving an 97% accuracy compared to SVM's 95% and Decision Tree's 79% as shown in table 4.5

Advantages of RF for Genre Classification: The RF model's superior performance is attributed to its ability to capture complex audio patterns, a limitation in traditional models like SVM, which relies on linear separability, and Decision Tree, which is less effective in handling high-dimensional data as shown in Table 2

Table 2 Model Performance Comparison: Comparing RF with baseline models such as SVM and DT.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RF	97	95	94	92
SVM	95	92	82	73
DT	79	78	74	61

5.0 Conclusion

This research aimed to classify music genres automatically using machine learning models and audio feature extraction. By analyzing and processing music samples across ten genres (Blues, Classical, Country, Disco, Hip-hop, Jazz, Metal, Pop, Reggae, and Rock), we developed a model that effectively captures genre-specific timbral characteristics, specifically through Mel-Frequency Cepstral Coefficients (MFCCs). The MFCC features provided the most critical information needed to distinguish between genres, as they encapsulate unique spectral properties. The Random Forest (RF) model demonstrated a robust classification performance, achieving an overall accuracy of 97%.

This result was reinforced with metrics of precision, recall and F1-score, which allowed us to evaluate the effectiveness of the model for each genre. Precision and recall scores for each genre revealed the model's strengths and weaknesses, including difficulty in distinguishing between related genres (e.g., Rock and Metal). The model was able to correctly identify genres with different audio features, such as Classical and Jazz, with an accuracy of above 90%. These results show a potential of using RFs for genre categorisation in music using auditory parameters such as MFCCs.

Funding

This research received no external funding.

Data Availability Statement

There are no data associated with this manuscript.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was not supported by any grant.

References

1. Adebayo, T., Ogunleye, S., & Ajayi, M. (2023). Deep learning for Afrobeats classification: CNN-based models. *Journal of Music Informatics Research*, 12(2), 45–58.
2. Adeyemi, K., Balogun, T., & Aluko, F. (2022). Artificial intelligence in music streaming for Nigerian music: Traditional classification algorithms. *International Journal of Digital Media Systems*, 9(1), 33–47.
3. Budyputra, Y., Rahman, A., & Widodo, S. (2025). Ensemble learning approaches for robust music genre classification. *Journal of Audio Engineering and Intelligent Systems*, 18(2), 55–72.
4. Cheng, Y., Yang, Y., & Li, X. (2019). Music genre classification using K-nearest neighbors with data augmentation. In *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 7252–7256). IEEE.
5. Choi, H., Lee, J., & Park, S. (2022). Feature-based machine learning for multi-genre music classification. *International Journal of Music Information Research*, 9(4), 201–214.
6. Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017). Convolutional recurrent neural networks for music classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 2392–2396). <https://doi.org/10.1109/ICASSP.2017.7952585>
7. Ding, Q., Santos, R., & Li, M. (2024). Deep acoustic feature modeling for large-scale music genre recognition. *IEEE Transactions on Affective Computing*, 15(1), 88–101.
8. Gimenez, D., & Bogdanov, D. (2024). Automated tagging and classification in contemporary digital music libraries. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 20(3), 1–19.
9. Huang, Y., & Yang, L. (2025). Explainable deep learning for audio-based genre classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 33(4), 650–663.
10. Joshi, P., Varma, R., & Sahu, V. (2024). Comparative evaluation of classical machine learning models for music genre prediction. *Journal of Intelligent Computing and Signal Processing*, 12(1), 40–52.

11. Kim, S., Park, D., & Cho, Y. (2023). Deep neural architectures for analyzing musical structure and genre. *Journal of Music Technology & AI*, 7(2), 115–130.
12. Lee, S., Choi, K., & Lee, K. (2017). Sample-level music genre classification with K-nearest neighbors. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)* (pp. 674–680).
13. Olojede, O., Olabiyisi, S., & Alo, O. (2025). Performance evaluation of some machine learning models for music genre classification. *International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)*, 14(2), 18–24.
14. Oramas, S., & Serra, X. (2023). Challenges and advances in content-based music genre classification. *Journal of New Music Research*, 52(1), 1–16.
15. Park, J., & Lee, H. (2024). Cultural and perceptual influences in genre definition and classification. *Music Perception: An Interdisciplinary Journal*, 41(3), 233–250.
16. Pons, M., Serra, X., & Bello, J. (2017). End-to-end learning for music audio classification. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)* (pp. 245–251).
17. Pons, M., Serra, X., & Bello, J. (2018). Training convolutional networks for music classification and retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(10), 1798–1808.
18. Pons, M., Serra, X., & Bello, J. (2019). Music auto-tagging with K-nearest neighbors. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1), 165–173.
19. Prashanthi, R., Kumar, V., & Daniel, S. (2021). Gaps in African music genre classification: A review of existing classification models. *African Journal of Multimedia Technology*, 5(3), 60–72.
20. Schlüter, J., & Grill, T. (2015). Exploring data augmentation for improved singing voice detection with neural networks. In *Proceedings of the International Society for Music Information Retrieval (ISMIR)* (pp. 121–126).
21. Singh, R. (2025). Decision tree-based classification of audio genres using spectral and temporal features. *International Journal of Artificial Intelligence and Applications*, 16(1), 77–90.
22. Won, M., Choi, K., & Kim, J. (2024). Data augmentation and transfer learning strategies for music genre recognition. *Neural Computing and Applications*, 36(2), 1125–1143.
23. Xu, F. (2023). Optimizing KNN models for audio classification using MFCC-based feature engineering. *Applied Computing and Signal Analysis Journal*, 5(2), 120–134.
24. Zhao, L., Chen, W., & Liu, H. (2024). Support vector machine adaptation for high-dimensional audio genre classification. *Expert Systems with Applications*, 241, Article 122645.