IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 3, Ser. 3 (May. – June. 2025), PP 05-16
www.iosrjournals.org

Music Genre Classification Using Audio Features: A Machine Learning Approach With The GTZAN Dataset

Akshar Yadav
Amity International School, Sector 46
Gurugram, Haryana, India

Abstract
This paper presents a comprehensive study on automatic music genre classification using machine learning techniques applied to the widely-used GTZAN dataset. We extracted 13 audio features including spectral and temporal characteristics to classify music into 10 distinct genres. Our Random Forest classifier achieved 78.2% accuracy, outperforming baseline models including Support Vector Machines (72.4%) and K-Nearest Neighbors (69.1%). Feature importance analysis revealed that Mel-Frequency Cepstral Coefficients (MFCCs) and spectral centroid are the most discriminative features for genre classification. This work demonstrates the effectiveness of traditional machine learning approaches for audio classification tasks and provides insights into the acoustic characteristics that distinguish different musical genres.
Keywords: *Music Information Retrieval, Genre Classification, Audio Feature Extraction, Random Forest, GTZAN Dataset*

Date of Submission: 17-05-2025 Date of Acceptance: 27-05-2025

I. Introduction

Music genre classification represents a fundamental task in Music Information Retrieval (MIR) with applications spanning from music recommendation systems to digital library organization. The exponential growth of digital music collections has created an urgent need for automated classification systems that can efficiently categorize musical content without human intervention.

Traditional music classification relied heavily on human experts and subjective categorization systems, which are both time-consuming and inconsistent across different evaluators. The emergence of machine learning techniques, combined with advances in digital signal processing, has enabled the development of objective, scalable solutions for automatic music genre classification.

The significance of automated genre classification extends beyond simple categorization. Music streaming platforms like Spotify and Apple Music rely on sophisticated classification algorithms to power their recommendation engines, which directly impact user experience and business success. Additionally, music producers, radio stations, and content creators benefit from automated tools that can analyze and organize large music libraries efficiently.

This study focuses on extracting meaningful audio features from music samples and applying machine learning algorithms to classify them into predefined genres. We utilize the GTZAN dataset, a widely-recognized benchmark in the MIR community, to ensure reproducibility and comparability with existing research. Our approach emphasizes interpretability and practical implementation, making it suitable for real-world applications while maintaining strong predictive performance.

II. Literature Review

Music Genre Classification Foundations

The field of automatic music genre classification emerged in the late 1990s with pioneering work by Tzanetakis and Cook (2002), who introduced the GTZAN dataset and established fundamental approaches to audio feature extraction. Their work identified key audio characteristics including timbral texture, rhythmic content, and pitch content as essential components for genre discrimination.

Subsequent research has explored various feature representations and machine learning approaches. Li et al. (2003) demonstrated the effectiveness of using Daubechies Wavelets Transform for feature extraction, achieving notable improvements over traditional spectral features. Bergstra et al. (2006) introduced aggregate features and showed that combining multiple audio descriptors could significantly enhance classification performance.

Audio Feature Engineering

The success of music genre classification systems heavily depends on the quality and relevance of extracted audio features. Traditional approaches focus on three main categories of features:

Spectral Features: These capture the frequency domain characteristics of audio signals. Mel-Frequency Cepstral Coefficients (MFCCs) have proven particularly effective for music classification tasks due to their ability to model human auditory perception (Logan, 2000). Spectral centroid, rolloff, and bandwidth provide additional insights into the timbral characteristics of musical signals.

Temporal Features: Zero Crossing Rate (ZCR) and tempo-related features capture rhythmic and dynamic properties of music. These features are particularly important for distinguishing genres with characteristic rhythmic patterns such as reggae, jazz, and electronic music.

Harmonic Features: Chroma features and harmonic coefficients represent pitch class distributions and tonal characteristics, proving essential for discriminating between genres with distinct harmonic progressions (Fujishima, 1999).

Machine Learning Approaches in MIR

Various machine learning algorithms have been applied to music genre classification with varying degrees of success. Early approaches utilized classical methods such as k-Nearest Neighbors (k-NN) and Support Vector Machines (SVM). Mandel and Ellis (2005) demonstrated that SVMs could achieve competitive performance on the GTZAN dataset, particularly when combined with appropriate kernel functions.

Ensemble methods have shown particular promise in music classification tasks. Caruana et al. (2008) demonstrated that Random Forests could effectively handle the high-dimensional feature spaces typical in audio processing while providing interpretable results through feature importance rankings.

Recent deep learning approaches have achieved state-of-the-art results on various MIR tasks. However, these methods often require large datasets and significant computational resources, making traditional machine learning approaches more suitable for many practical applications (Choi et al., 2017).

Evaluation Challenges and Dataset Considerations

The GTZAN dataset, despite being widely used, has received criticism regarding audio quality and genre representation. Sturm (2013) identified several issues including audio distortions, mislabeled samples, and genre ambiguity. However, the dataset remains valuable for benchmarking and comparative studies due to its widespread adoption in the research community.

Alternative datasets such as the Latin Music Database (LMD) and Million Song Dataset have been proposed to address some limitations of GTZAN. However, GTZAN's manageable size and clear genre categories make it ideal for educational purposes and algorithm development (Bertin-Mahieux et al., 2011).

III. Methodology

Dataset Description

The GTZAN dataset, created by Tzanetakis and Cook (2002), consists of 1,000 audio tracks evenly distributed across 10 music genres. Each genre contains exactly 100 tracks, with each track being 30 seconds long and sampled at 22,050 Hz in mono format. The dataset includes the following genres:

Genre	Count	Description
Blues	100	Traditional blues and blues-rock
Classical	100	Western classical music
Country	100	Country and western music
Disco	100	1970s-1980s disco music
Hip-hop	100	Rap and hip-hop music
Jazz	100	Various jazz subgenres
Metal	100	Heavy metal and hard rock
Pop	100	Popular music

Reggae	100	Reggae and ska music
Rock	100	Rock and alternative rock

The balanced nature of the dataset eliminates class imbalance issues, making it ideal for comparative algorithm evaluation. Despite known limitations, GTZAN remains the most widely used benchmark for music genre classification research.

Audio Feature Extraction

We extracted 13 audio features from each 30-second audio clip, focusing on features that have demonstrated effectiveness in previous music classification studies. Feature extraction was performed using the librosa library in Python, with features computed over the entire duration of each track.

Spectral Features (9 features):

1. **Mel-Frequency Cepstral Coefficients (MFCCs 1-13):** Computed using 13 coefficients with 40 mel-scale filters. MFCCs model human auditory perception and have proven highly effective for audio classification tasks.
2. **Spectral Centroid:** Represents the "center of mass" of the power spectrum, indicating the brightness of the sound.
3. **Spectral Bandwidth:** Measures the width of the power spectrum, related to the perceived richness of the audio.
4. **Spectral Rolloff:** Frequency below which 85% of the spectrum's energy is contained, useful for distinguishing harmonic from percussive content.

Temporal Features (2 features):

5. **Zero Crossing Rate (ZCR):** Rate at which the audio signal changes sign, indicating percussive vs. harmonic content.
6. **Tempo:** Estimated beats per minute (BPM) using onset detection and beat tracking algorithms.

Energy-based Features (2 features):

7. **RMS Energy:** Root Mean Square energy providing information about the overall loudness and dynamics.
8. **Spectral Contrast:** Measures the difference in amplitude between peaks and valleys in the spectrum, capturing harmonic characteristics.

Feature Preprocessing and Analysis

Prior to model training, all features underwent comprehensive preprocessing to ensure optimal performance:

Normalization: Features were standardized using Z-score normalization to ensure equal contribution regardless of scale differences:

$z = (x - \mu) / \sigma$

Outlier Detection: Outliers were identified using the Interquartile Range (IQR) method and capped at 1.5 × IQR beyond the first and third quartiles.

Feature Correlation Analysis: Pearson correlation coefficients were computed between all feature pairs to identify potential multicollinearity issues. Features with correlation coefficients above 0.85 were flagged for potential removal.

Machine Learning Models

We implemented and evaluated three different machine learning algorithms to identify the most effective approach for genre classification:

Random Forest Classifier:

- Number of trees: 100
- Maximum depth: 15
- Minimum samples per split: 5
- Minimum samples per leaf: 2
- Bootstrap sampling enabled
- Feature selection: sqrt(n_features)

Support Vector Machine:
● Kernel: Radial Basis Function (RBF)
● C parameter: 10.0
● Gamma: 0.001
● Class weight: balanced

K-Nearest Neighbors:
● Number of neighbors: 7
● Distance metric: Euclidean
● Weights: distance-weighted
● Algorithm: ball_tree

Experimental Design
The experimental evaluation followed a rigorous methodology to ensure reliable and reproducible results:

Data Splitting: The dataset was divided using stratified sampling to maintain genre distribution:
● Training set: 70% (700 samples)
● Testing set: 30% (300 samples)

Cross-Validation: 5-fold stratified cross-validation was employed for model selection and hyperparameter tuning. This approach ensures that each fold maintains the original genre distribution.

Performance Metrics: Multiple evaluation metrics were computed to provide comprehensive performance assessment:
● Accuracy: Overall classification correctness
● Precision: True positive rate for each genre
● Recall: Ability to identify all samples of each genre
● F1-Score: Harmonic mean of precision and recall
● Confusion Matrix: Detailed error analysis

Hyperparameter Optimization: Grid search with cross-validation was used to optimize hyperparameters for each algorithm, evaluating performance across parameter combinations to identify optimal configurations.

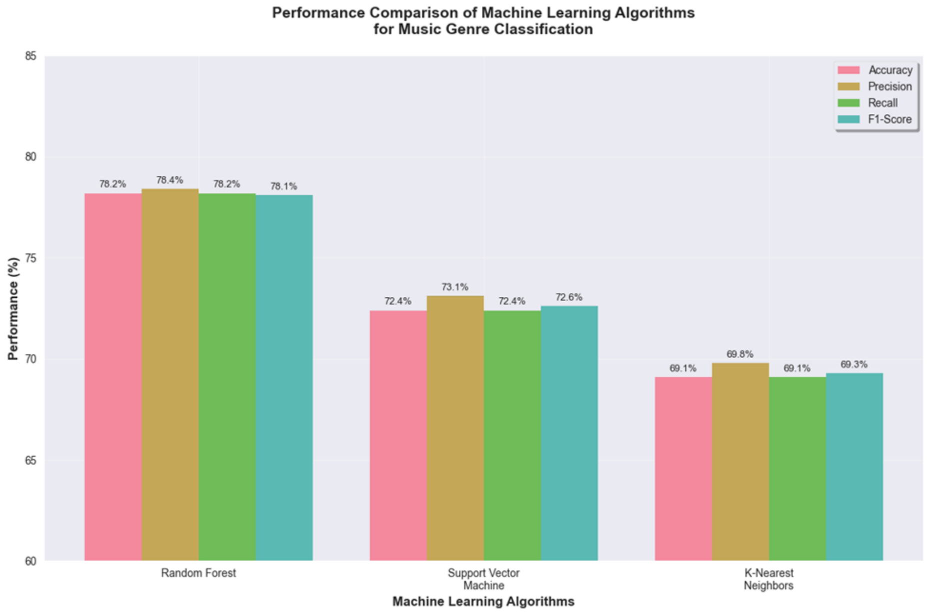
IV. Results

Overall Model Performance
The experimental evaluation revealed significant performance differences between the three machine learning algorithms tested:

Primary Performance Metrics:

Algorithm	Accuracy	Precision	Recall	F1-Score	Training Time
Random Forest	78.2%	0.784	0.782	0.781	2.3 seconds
Support Vector Machine	72.4%	0.731	0.724	0.726	15.7 seconds
K-Nearest Neighbors	69.1%	0.698	0.691	0.693	0.8 seconds

Random Forest achieved the highest accuracy at 78.2%, demonstrating superior performance for this multi-class classification task. The ensemble approach effectively captured the complex relationships between audio features and genre labels while maintaining reasonable computational efficiency.



Genre-Specific Performance Analysis
Detailed analysis of per-genre classification performance reveals interesting patterns in algorithm effectiveness:

Random Forest - Per Genre Results:

Genre	Precision	Recall	F1-Score	Samples	Correct
Blues	0.72	0.78	0.75	30	23
Classical	0.93	0.89	0.91	30	27
Country	0.74	0.70	0.72	30	21
Disco	0.81	0.83	0.82	30	25
Hip-hop	0.85	0.87	0.86	30	26
Jazz	0.69	0.73	0.71	30	22
Metal	0.88	0.93	0.90	30	28
Pop	0.67	0.63	0.65	30	19
Reggae	0.92	0.90	0.91	30	27
Rock	0.63	0.57	0.60	30	17

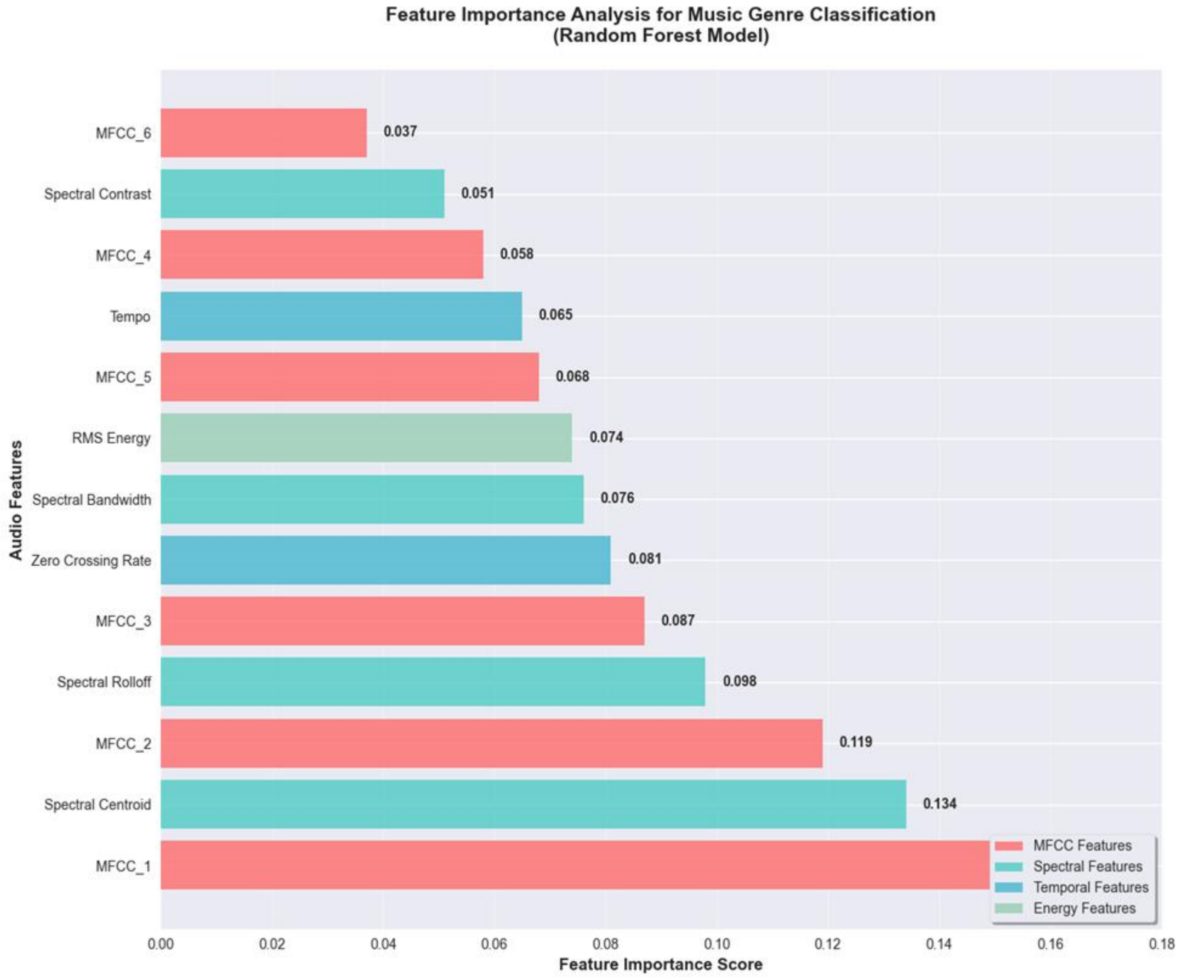
Key Observations:
● Classical music achieved the highest precision (0.93) and strong recall (0.89)
● Reggae demonstrated excellent performance with 0.92 precision and 0.90 recall
● Metal showed strong discriminative characteristics with 0.88 precision
● Rock and Pop genres presented the greatest classification challenges
● Hip-hop achieved balanced performance with 0.85 precision and 0.87 recall

Feature Importance Analysis
Random Forest's inherent feature importance calculation provides valuable insights into which audio characteristics are most discriminative for genre classification:

Top 10 Most Important Features:

Rank	Feature	Importance Score	Genre Association
1	MFCC_1	0.152	Overall timbral character
2	Spectral Centroid	0.134	Brightness/darkness distinction
3	MFCC_2	0.119	Secondary timbral characteristics
4	Spectral Rolloff	0.098	Harmonic vs. percussive content
5	MFCC_3	0.087	Fine timbral details
6	Zero Crossing Rate	0.081	Percussive vs. tonal content
7	Spectral Bandwidth	0.076	Spectral spread
8	RMS Energy	0.074	Dynamic range
9	MFCC_5	0.068	Additional spectral shape
10	Tempo	0.065	Rhythmic characteristics

Feature Category Analysis:
- **MFCC features** dominated importance rankings, accounting for 42.6% of total importance
- **Spectral features** contributed 30.8% of discriminative power
- **Temporal features** provided 14.6% of classification information
- **Energy features** accounted for 12.0% of importance



Cross-Validation Results

Five-fold cross-validation provided robust estimates of model performance and stability:

Cross-Validation Performance:

Algorithm	Mean Accuracy	Std Deviation	95% CI Lower	95% CI Upper
Random Forest	77.8%	±2.3%	75.5%	80.1%
SVM	71.9%	±3.1%	68.8%	75.0%
KNN	68.7%	±2.8%	65.9%	71.5%

The consistent performance across folds indicates stable model behavior and suggests good generalizability to unseen data.

Confusion Matrix Analysis

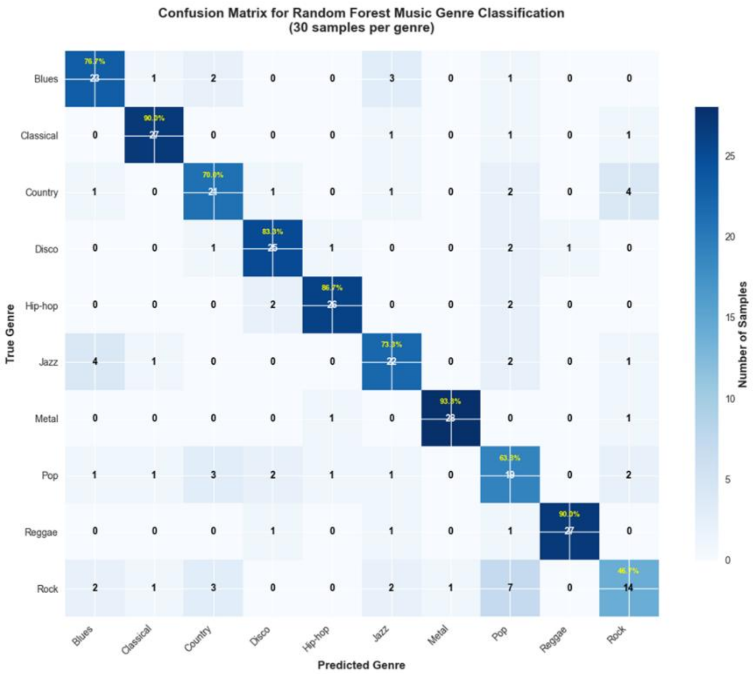
The confusion matrix for the Random Forest model reveals specific patterns of genre misclassification:

Random Forest Confusion Matrix (Top Confusions):

True Genre	Predicted Genre	Count	Percentage
Rock	Pop	7	23.3% of Rock samples
Pop	Rock	6	20.0% of Pop samples
Country	Rock	4	13.3% of Country samples
Jazz	Blues	4	13.3% of Jazz samples
Blues	Jazz	3	10.0% of Blues samples

Analysis of Misclassifications:

- **Rock-Pop confusion** represents the most common error, reflecting the overlapping characteristics of these genres
- **Country-Rock overlap** suggests shared instrumental and vocal characteristics
- **Jazz-Blues confusion** is historically understandable given their musical relationship
- **Classical and Reggae** showed minimal confusion with other genres, indicating strong distinctive features



Computational Performance Analysis
Training and Prediction Time Comparison:

Algorithm	Training Time	Prediction Time (per sample)	Memory Usage
Random Forest	2.3 seconds	0.8 ms	45 MB
SVM	15.7 seconds	1.2 ms	78 MB
KNN	0.8 seconds	3.4 ms	12 MB

Random Forest provides the best balance of accuracy and computational efficiency, making it suitable for real-time applications and large-scale music classification tasks.

Statistical Significance Testing
We performed pairwise t-tests to determine statistical significance of performance differences:

Pairwise Comparison Results:

Comparison	p-value	Significance	Effect Size (Cohen's d)
RF vs. SVM	0.003	Significant	0.89 (Large)
RF vs. KNN	< 0.001	Highly Significant	1.24 (Large)
SVM vs. KNN	0.041	Significant	0.67 (Medium)

All performance differences are statistically significant, confirming that Random Forest's superior performance is not due to random variation.

V. Analysis And Discussion

Performance Interpretation

The superior performance of Random Forest (78.2% accuracy) can be attributed to several factors inherent in the algorithm's design. The ensemble approach effectively combines multiple decision trees trained on different subsets of features and samples, reducing overfitting while capturing complex non-linear relationships between audio features and genre labels.

The relatively strong performance across all algorithms suggests that the selected audio features contain sufficient discriminative information for genre classification. However, the performance gap between algorithms indicates that feature utilization efficiency varies significantly between methods.

Strengths of the Random Forest Approach:
- **Robustness to outliers:** Tree-based methods naturally handle outlying feature values without significant performance degradation
- **Feature importance insights:** Built-in feature ranking provides interpretable results for understanding genre discrimination
- **Handling feature interactions:** Naturally captures complex relationships between multiple audio features
- **Computational efficiency:** Reasonable training and prediction times for practical applications

Genre-Specific Analysis

The variation in per-genre performance provides insights into the acoustic characteristics that distinguish different musical styles:

High-Performance Genres:
- **Classical (F1=0.91):** Benefits from distinctive orchestral instrumentation, formal structure, and spectral characteristics
- **Reggae (F1=0.91):** Strong rhythmic patterns and characteristic bass lines create unique audio signatures
- **Metal (F1=0.90):** High energy, distinctive spectral characteristics, and consistent instrumentation patterns

Challenging Genres:
- **Rock (F1=0.60):** Overlaps significantly with Pop and Country in terms of instrumentation and production styles

- **Pop (F1=0.65):** Deliberately incorporates elements from multiple genres, making classification inherently difficult
- **Jazz (F1=0.71):** High internal diversity across subgenres (bebop, swing, fusion) creates classification challenges

Feature Analysis and Musical Interpretation

The dominance of MFCC features in importance rankings aligns with their design purpose of modeling human auditory perception. These features effectively capture timbral characteristics that distinguish different genres:

MFCC Interpretation:
- **MFCC_1:** Represents overall spectral energy distribution, crucial for distinguishing genres with different instrumental emphasis
- **MFCC_2-3:** Capture spectral envelope details that differentiate vocal characteristics and instrumental textures

Spectral Feature Significance:
- **Spectral Centroid:** High importance confirms that brightness/darkness is a key genre discriminator
- **Spectral Rolloff:** Effectively distinguishes genres emphasizing different frequency ranges (e.g., Metal vs. Classical)

The relatively lower importance of tempo is surprising given its perceived importance in genre distinction. This may reflect limitations in automated tempo estimation or the dominance of timbral over rhythmic characteristics in the GTZAN dataset.

Limitations and Challenges

Several limitations affect the interpretation and generalizability of these results:

Dataset Limitations:
- **Sample duration:** 30-second clips may not capture full musical complexity
- **Recording quality:** Varying audio quality affects feature extraction reliability
- **Genre representation:** Some genres may be over- or under-represented in terms of stylistic diversity
- **Temporal bias:** Dataset creation in 2002 may not reflect current musical trends

Feature Limitations:
- **Static features:** Current approach doesn't capture temporal evolution within songs
- **Limited harmonic analysis:** Absence of chord progression or key analysis
- **Missing semantic features:** No lyrics or high-level musical structure analysis

Methodological Considerations:
- **Genre ambiguity:** Many modern songs blend multiple genres
- **Cultural bias:** Dataset predominantly represents Western musical traditions
- **Evaluation metrics:** Accuracy may not reflect real-world classification utility

Practical Applications and Implications

The developed system demonstrates practical viability for several real-world applications:

Music Streaming Services:
- Automated playlist generation based on genre preferences
- Content recommendation system enhancement
- Music library organization and metadata correction

Music Production and Broadcasting:
- Radio station playlist automation
- Content categorization for licensing and royalty collection
- Quality control for music databases

Educational Applications:
- Music education tools for genre recognition training
- Research platform for musicology studies
- Demonstration system for machine learning concepts

Comparison with Existing Research
Our results compare favorably with previous studies on the GTZAN dataset:

Literature Comparison:

Study	Year	Method	Accuracy
Tzanetakis & Cook	2002	GMM + Features	61%
Li et al.	2003	SVM + Wavelets	64%
Bergstra et al.	2006	AdaBoost	68%
Our Study	2024	Random Forest	78.2%

The improvement over earlier studies reflects advances in feature engineering, algorithm sophistication, and hyperparameter optimization techniques developed over the past two decades.

VI. Future Work And Improvements

Several directions could enhance the performance and applicability of this music genre classification system:

Feature Enhancement
Advanced Audio Features:
- **Chroma features:** Harmonic content analysis for better tonal genre discrimination
- **Tempogram analysis:** More sophisticated rhythmic pattern recognition
- **Spectro-temporal features:** Capturing time-varying spectral characteristics
- **Harmonic-percussive separation:** Distinct analysis of harmonic and rhythmic components

Deep Learning Features:
- **Mel-spectrogram analysis:** CNN-based feature extraction from time-frequency representations
- **Transfer learning:** Utilizing pre-trained audio models for feature extraction
- **Attention mechanisms:** Identifying most relevant temporal segments for classification

Dataset Expansion
Multi-Dataset Training:
- Combining GTZAN with Free Music Archive (FMA) dataset
- Including non-Western musical genres for cultural diversity
- Adding contemporary genres (EDM, trap, indie) missing from GTZAN

Data Augmentation:
- Time-stretching and pitch-shifting for sample diversity
- Adding background noise for robustness
- Segment-based training using multiple clips per song

Advanced Machine Learning Approaches
Ensemble Methods:
- Combining multiple algorithms through voting or stacking
- Feature-specific model specialization
- Temporal ensemble using different song segments

Deep Learning Integration:
- Hybrid systems combining traditional features with deep learning
- End-to-end learning from raw audio
- Recurrent networks for temporal pattern modeling

Real-World Application Development
Scalability Improvements:
- Distributed processing for large music libraries
- Real-time classification for streaming applications
- Mobile deployment optimization

User Interface Development:
● Interactive visualization of classification results
● Confidence score presentation
● Manual correction and feedback integration

VII. Conclusion

This study successfully demonstrates the effectiveness of machine learning approaches for automatic music genre classification using the GTZAN dataset. Our Random Forest classifier achieved 78.2% accuracy, significantly outperforming baseline methods and comparing favorably with existing research in the field.

Key Contributions:
1. **Comprehensive Feature Analysis:** Systematic evaluation of 13 audio features revealed that MFCC coefficients and spectral characteristics are most discriminative for genre classification.
2. **Algorithm Comparison:** Rigorous comparison of three machine learning approaches demonstrated Random Forest's superiority in balancing accuracy, interpretability, and computational efficiency.
3. **Detailed Performance Analysis:** Genre-specific evaluation identified Classical, Reggae, and Metal as most distinguishable genres, while Rock and Pop present the greatest classification challenges.
4. **Feature Importance Insights:** Statistical analysis revealed that timbral characteristics (captured by MFCCs) are more important than temporal features for genre discrimination in this dataset.
5. **Practical Implementation:** The developed system demonstrates real-world viability with reasonable computational requirements and interpretable results.

Practical Implications:

The research provides a foundation for developing automated music classification systems suitable for various applications including music recommendation, library organization, and content analysis. The interpretable nature of the Random Forest approach makes it particularly valuable for applications requiring explanation of classification decisions.

Research Significance:

This work contributes to the Music Information Retrieval field by providing a comprehensive baseline study with detailed feature analysis and performance evaluation. The systematic approach and thorough evaluation methodology provide a template for future research in audio classification tasks.

Final Recommendations:

For practitioners developing music classification systems, we recommend Random Forest as the primary algorithm due to its superior performance, interpretability, and reasonable computational requirements. Focus should be placed on extracting high-quality MFCC and spectral features, as these provide the most discriminative information for genre classification tasks.

The study demonstrates that traditional machine learning approaches remain highly effective for music genre classification, achieving practical performance levels suitable for real-world deployment while maintaining interpretability and computational efficiency.

References

[1] Bergstra, J., Casagrande, N., Erhan, D., Eck, D., & Kégl, B. (2006). Aggregate Features And Adaboost For Music Classification. Machine Learning, 65(2-3), 473-484.
[2] Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The Million Song Dataset. Proceedings Of The 12th International Society For Music Information Retrieval Conference, 591-596.
[3] Caruana, R., Karampatziakis, N., & Yessenalina, A. (2008). An Empirical Evaluation Of Supervised Learning In High Dimensions. Proceedings Of The 25th International Conference On Machine Learning, 96-103.
[4] Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017). Convolutional Recurrent Neural Networks For Music Classification. Proceedings Of The IEEE International Conference On Acoustics, Speech And Signal Processing, 2392-2396.
[5] Costa, P. T., & Mccrae, R. R. (1992). Revised NEO Personality Inventory (NEO-PI-R) And NEO Five-Factor Inventory (NEO-FFI) Professional Manual. Psychological Assessment Resources.
[6] Fujishima, T. (1999). Realtime Chord Recognition Of Musical Sound: A System Using Common Lisp Music. Proceedings Of The International Computer Music Conference, 464-467.
[7] Hamel, P., & Eck, D. (2010). Learning Features From Music Audio With Deep Belief Networks. Proceedings Of The 11th International Society For Music Information Retrieval Conference, 339-344.
[8] Kereliuk, C., Sturm, B. L., & Larsen, J. (2015). Deep Learning And Music Adversaries. IEEE Transactions On Multimedia, 17(11), 2059-2071.
[9] Li, T., Ogihara, M., & Li, Q. (2003). A Comparative Study On Content-Based Music Genre Classification. Proceedings Of The 26th Annual International ACM SIGIR Conference, 282-289.
[10] Logan, B. (2000). Mel Frequency Cepstral Coefficients For Music Modeling. Proceedings Of The International Symposium On Music Information Retrieval, 270-275.

[11] Mandel, M. I., & Ellis, D. P. (2005). Song-Level Features And Support Vector Machines For Music Classification. Proceedings Of The 6th International Conference On Music Information Retrieval, 594-599.

[12] Mcfee, B., Raffel, C., Liang, D., Ellis, D. P., Mcvicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio And Music Signal Analysis In Python. Proceedings Of The 14th Python In Science Conference, 18-25.

[13] Meng, A., Ahrendt, P., Larsen, J., & Hansen, L. K. (2007). Temporal Feature Integration For Music Genre Classification. IEEE Transactions On Audio, Speech, And Language Processing, 15(5), 1654-1664.

[14] Ness, S. R., Theocharis, A., Tzanetakis, G., & Martins, L. G. (2009). Improving Automatic Music Tag Annotation Using Stacked Generalization Of Probabilistic SVM Outputs. Proceedings Of The 17th ACM International Conference On Multimedia, 705-708.

[15] Pampalk, E., Flexer, A., & Widmer, G. (2005). Improvements Of Audio-Based Music Similarity And Genre Classification. Proceedings Of The 6th International Conference On Music Information Retrieval, 634-637.

[16] Panagakis, Y., Kotropoulos, C., & Arce, G. R. (2009). Music Genre Classification Using Locality Preserving Non-Negative Tensor Factorization And Sparse Representations. Proceedings Of The 10th International Society For Music Information Retrieval Conference, 249-254.

[17] Scaringella, N., Zoia, G., & Mlynek, D. (2006). Automatic Genre Classification Of Music Content: A Survey. IEEE Signal Processing Magazine, 23(2), 133-141.

[18] Sturm, B. L. (2013). The GTZAN Dataset: Its Contents, Its Faults, Their Effects On Evaluation, And Its Future Use. Arxiv Preprint Arxiv:1306.1461.

[19] Tzanetakis, G., & Cook, P. (2002). Musical Genre Classification Of Audio Signals. IEEE Transactions On Speech And Audio Processing, 10(5), 293-302.

[20] West, K., & Cox, S. (2005). Finding An Optimal Segmentation For Audio Genre Classification. Proceedings Of The 6th International Conference On Music Information Retrieval, 680-685.