www.ijarp.com ISSN 2456-9992 Page: 01-04

International Journal Advanced Research Publications

Review Article

Volume: 02

Issue: 04

MUSIC GENRE RECOMMENDATION USING ML

***Harish U., Mrs. J. Vinitha M.Sc., (Ph.D.)**

Student, Assistant Professor, Department of Artificial Intelligence and Machine Learning, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India.

Article Received: 10 March 2026, Article Revised: 30 March 2026, Published on: 20 April 2026

***Corresponding Author: Harish U.**

Student, Assistant Professor, Department of Artificial Intelligence and Machine Learning, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India.

DOI: https://doi-doi.org/101555/ijarp.2136

ABSTRACT

The landscape of modern music recommendation is currently dominated by proprietary "Black Box" algorithms that excel in accuracy but remain inaccessible for research. This project addresses this critical "Precision Gap" by developing a novel, high-performance music classification engine built on the ideological foundation that signal is truth. By implementing a 3-second segmentation strategy, the model effectively expands the feature space by 10x compared to traditional 30-second averaging techniques. By extracting 58 distinct "Acoustic DNA" parameters, the system analyzes the texture and timbre of audio with a level of fidelity previously reserved for enterprise-grade systems. The resulting model utilizes optimized machine learning kernels to classify genres based purely on acoustic content, surpassing the precision of standard metadata-based models and demonstrating that high-fidelity, transparent music classification is achievable on a local scale.

1. INTRODUCTION

In the digital age, music consumption is overwhelmingly dominated by the sophisticated but opaque algorithms of industry giants like Spotify and Apple Music. While these commercial systems provide highly accurate experiences, they operate as inaccessible "Black Boxes" locked behind corporate firewalls, preventing independent researchers from inspecting the underlying logic.

Consequently, those seeking to study or build their own music information retrieval systems rely on standard open-source models, which often suffer from a severe lack of fidelity known as the "Precision Gap". Traditional models analyze audio files as single, static data points, relying on a broad 30-second statistical average to represent an entire track. This rudimentary approach blurs the audio data, causing the system to miss critical dynamic events such as fleeting guitar solos or sudden bass drops, and rarely exceeds classification accuracy rates of 60 to 70%.

To bridge this gap, this project introduces a High-Fidelity Granular Classification Engine. Rather than accepting the flawed standard of whole-song averaging, this system re-engineers the analysis pipeline by implementing a 3-second micro-segmentation strategy on the standard GTZAN dataset, effectively expanding the data resolution by a factor of ten.

2. SYSTEM ARCHITECTURE AND METHODOLOGY

The EchoVibe system follows a Modular Pipeline Architecture consisting of four primary modules to ensure scalability and easier debugging:

- **Data Preprocessing (Granular Augmentation):** The system applies a windowing function to slice a standard 30-second .wav file into ten non-overlapping 3-second segments. This effectively multiplies the training dataset size from 1,000 files to 10,000 segments.
- **Feature Extraction:** Utilizing the Librosa library, the system extracts 58 unique features per segment, including 20 Mel-Frequency Cepstral Coefficients (MFCCs), Tempo (BPM), Spectral Centroids, and Zero-Crossing Rates.
- **Classification Model:** The primary algorithm is a Support Vector Machine (SVM) with a Radial Basis Function (RBF) Kernel, chosen for its ability to calculate complex boundaries in high-dimensional space. An ensemble Random Forest Classifier is used secondarily for validating results and handling outliers.

3. IMPLEMENTATION AND RESULTS

The implementation was executed within a Python 3.9 environment. The 10,000 extracted segments were randomly divided into an 80/20 train-test split for model evaluation. Rigorous hyperparameter tuning was conducted using Grid Search Cross-Validation to optimize variables such as the SVM's C and gamma parameters.

The granular SVM architecture achieved an overall classification accuracy of 84%. This performance significantly surpasses standard open-source baselines, which typically plateau at 60-65% on the GTZAN dataset, proving that the micro-segmentation strategy successfully resolves the "Precision Gap". Furthermore, the end-to-end process from the moment a user uploads a 30-second track to the generation of the final predicted genre is completed seamlessly in under 2.5 seconds.

4. CONCLUSION AND FUTURE ENHANCEMENTS

The completion of the EchoVibe project marks a significant milestone in local-scale Music Information Retrieval (MIR), proving that high-precision genre classification is achievable without relying on the opaque algorithms of enterprise giants. By implementing a granular analysis pipeline and extracting a 58-dimensional "Acoustic DNA" feature vector, the locally hosted SVM successfully captures fleeting acoustic textures with remarkable fidelity.

Future development will focus on a Deep Learning Migration, specifically integrating Convolutional Neural Networks (CNNs) to analyze Mel-spectrograms, effectively treating sound as highly detailed visual images. By implementing Hybrid Ensemble Modeling and scaling up to train on the massive 100,000-track Free Music Archive (FMA) dataset, the system will evolve into a ubiquitous platform capable of pushing classification accuracy well beyond the 90% commercial threshold.

5. BIBLIOGRAPHY

Book References

1. **Goodfellow, I., Bengio, Y., & Courville, A.** (2016). *Deep Learning*. MIT Press. (Fundamental text for the planned CNN and neural network migration).
2. **Müller, M.** (2015). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer. (Comprehensive guide on audio signal processing and feature extraction).
3. **Bishop, C. M.** (2006). *Pattern Recognition and Machine Learning*. Springer. (Authoritative text providing the mathematical foundation for Support Vector Machines and Radial Basis Function kernels used in this project).
4. **Hastie, T., Tibshirani, R., & Friedman, J.** (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. (Essential reference for understanding the mechanics of ensemble methods like Random Forests).

Digital References

5. **Tzanetakis, G., & Cook, P.** (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5), 293-302. Available at: https://doi.org/10.1109/TSA.2002.800560 (The foundational paper that introduced the GTZAN dataset).
6. **Sturm, B. L.** (2013). The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use. *arXiv preprint*. Available at:

https://arxiv.org/abs/1306.1461 (Crucial research detailing the flaws of 30-second averaging, justifying the granular approach).

7. **Breiman, L.** (2001). Random Forests. *Machine Learning*, 45(1), 5-32. Available at: https://doi.org/10.1023/A:1010933404324 (The original foundational paper for the Random Forest ensemble classifier utilized to prevent overfitting).
8. **Choi, K., Fazekas, G., Sandler, M., & Cho, K.** (2016). Automatic tagging using deep convolutional neural networks. *International Society for Music Information Retrieval Conference (ISMIR)*. (Key reference supporting the future scope of migrating to Mel-spectrogram image analysis via CNNs).
9. **Davis, S., & Mermelstein, P.** (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28(4), 357-366. (The historical academic origin of Mel-Frequency Cepstral Coefficients [MFCCs] used for acoustic DNA extraction).
10. **Kumari, P., Kamath, P., Varun, K., Ganiga, V. D., & Shilpa, K. A.** (2025). Music Genre Classification Using Multilayer Perceptron: A Machine Learning Approach. *2025 9th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*. Available at: https://doi.org/10.1109/CSITSS67709.2025.11294788
11. **Pedregosa, F., et al.** (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. (The official academic citation for the Scikit-learn library powering the SVM and ML pipeline).
12. **Librosa Development Team.** (2024). *Librosa: Audio and Music Signal Analysis in Python (Official Documentation)*. Available at: https://librosa.org/doc/latest/