Identifying Album-Specific Musical Signatures through Hybrid Feature Representation

Dr. Srikanth[1], Dr. Karunakar[2], Mr. Syed Juber[3], Ms. Nagalaxmi[3]
[1]Professor, [2]Associate Professor, [3]Assistant Professor, [1,2,3]Department of Information Technology
[1,2,3]Lords Institute of Engineering and Technology (UGC-Autonomous), Himayath sagar, Hyderabad, 500091, Telangana, India.

To Cite this Article
Dr. Srikanth, Dr. Karunakar, Mr. Syed Juber, Ms. Nagalaxmi, "Identifying Album-Specific Musical Signatures through Hybrid Feature Representation", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 12, December 2025, pp: 322-329, DOI: http://doi.org/10.64771/jsetms.2025.v02.i12.pp322-329
Submitted: 10-11-2025 *Accepted: 24-12-2025* *Published: 31-12-2025*

ABSTRACT
Globally, digital music platforms process over 100,000 new tracks daily, making automated characterization of musical style and track identity a critical real-time challenge for music information retrieval systems, recommendation engines, and archival analysis. This study focuses on characterizing musical style and track identity within the discography of Portishead using a text- and audio-informed machine learning framework. The problem addressed in this work is the absence of accurate, interpretable, and scalable methods to identify stylistic and album-level identity from lyrical and audio features simultaneously. Traditional manual systems rely on human experts for lyrical interpretation and auditory judgment, which are subjective, time-consuming, and inconsistent across evaluators. Such manual approaches fail to scale with large music datasets and lack reproducibility, leading to unreliable categorization and stylistic misclassification. These limitations motivate the need for an automated, data-driven solution that integrates natural language processing and audio feature analysis. In the proposed system, lyrical content undergoes extensive NLP preprocessing, followed by feature extraction using CountVectorizer, while numerical audio attributes are jointly incorporated to form a hybrid feature space. The study evaluates multiple interpretable tree-based models, including GreedyTreeClassifier**,** FIGS Classifier, and TaoTree Classifier as existing approaches, and introduces the OptimalTree Classifier as the proposed model to achieve globally optimized decision structures. Model performance is assessed using accuracy, precision, recall, F1-score, and AUC metrics. Experimental results demonstrate that the OptimalTree Classifier provides superior classification performance and interpretability in identifying album-level track identity from lyrics and audio features.
Keywords: Music Information Retrieval, Musical Style Classification, Natural Language Processing, Audio Feature Extraction, Interpretable Machine Learning.
This is an open access article under the creative commons license https://creativecommons.org/licenses/by-nc-nd/4.0/
CC BY-NC-ND 4.0

1. INTRODUCTION

In the traditional music industry, sales of record labels are the most direct route to financial success. Producers and singers aim to sell as many albums as possible. As the Internet is used widely, music streaming sites have become popular by offering songs for free [1]. People can enjoy music without having to buy an album. Singers put their music online and websites pay them royalties for streams and downloads [2]. Today, the main goal for singers and producers is to make songs that attract listeners. Figuring out what makes a song a hit is valuable in many areas [3]. A song's popularity is determined by what listeners like, how catchy it is, and its commercial appeal. Recent studies focus on extracting music audio with machine learning, such as decision trees, support vector machines learning models [4]. In 2015, Anjana used Librosa to extract songs' audio waveforms to predict the song's popularity

with the classification algorithm. The second coefficient of Linear Predictive Coding has an accuracy of 71% [5].

2. LITERATURE SURVEY

Lee et al. [6] proposed Factors contributing to a song's popularity are explored in this study. Recent studies have mainly focused on using acoustic features to identify popular songs. However, we combined audio and visual data to make predictions on 1000 YouTube songs. In total, 1000 songs were grouped into two categories based on YouTube view counts: popular and non-popular. The visual features were extracted using OpenCV. Blaszke et al. [7] proposed a novel approach for automatically identifying all instruments present in an audio excerpt using sets of individual convolutional neural networks (CNNs) per tested instrument. The paper starts with a review of tasks related to musical instrument identification. It focuses on tasks performed, input type, algorithms employed, and metrics used. The paper starts with the background presentation, i.e., metadata description and a review of related works.

Louro et al. [8] presented Classical machine learning techniques have dominated Music Emotion Recognition. However, improvements have slowed down due to the complex and time-consuming task of handcrafting new emotionally relevant audio features. Moroşanu et al. [9] explored Modern audio production workflows often require significant manual effort during the initial session preparation phase, including track labeling, format standardization, and gain staging. This paper presents a rule-based and Machine Learning-assisted automation system designed to minimize the time required for these tasks in Digital Audio Workstations (DAWs). The system automatically detects and labels audio tracks, identifies and eliminates redundant fake stereo channels.

Yang et al. [10] proposed a Machine-learning framework for the automated classification of Chinese traditional instruments, addressing the challenges of acoustic diversity and cultural preservation. By integrating two datasets, CTIS and ChMusic, we constructed a combined dataset comprising four instrument families: wind, percussion, plucked string, and bowed string. Three time-frequency features, namely MFCC, CQT, and Chroma, were extracted to capture diverse sound information. Cheng et al. [11] presented a Music is the most convenient and easy-to-use stress release tool in modern times. Many studies have shown that listening to appropriate music can release stress. However, since it is getting easier to make music, people only need to make it on the computer and upload it to streaming media such as Youtube, Spotify, or Beatport at any time, which makes it very infeasible to search a huge music database for music of a specific genre. In order to effectively search for specific types of music.

Wang et al. [12] presented Audio music genre classification is performed to categorize audio music into various genres. Traditional approaches based on convolutional recurrent neural networks do not consider long temporal information, and their sequential structures result in longer training times and convergence difficulties. To overcome these problems, a traditional transformer-based approach was introduced. However, this approach employs pre-training based on momentum contrast (MoCo), a technique that increases computational costs owing to its reliance on extracting many negative samples and its use of highly sensitive hyperparameters. Venkatesh et al. [13] proposed Music and speech detection provides us valuable information regarding the nature of content in broadcast audio. It helps detect acoustic regions that contain speech, voice over music, only music, or silence. In recent years, there have been developments in machine learning algorithms to accomplish this task. However, broadcast audio is generally well-mixed and copyrighted, which makes it challenging to share across research groups. In this study, we address the challenges encountered in automatically synthesising data that resembles a radio broadcast.

Kovacs et al. [14] explored Romanian popular music has had a storied history across the last century and a half. Incorporating different influences at different times, today it boasts a wide range of both autochthonous and imported genres, such as traditional folk music, rock, rap, pop, and manele, to name a few. We aim to trace the linguistic differences between the lyrics of these genres using natural

language processing and a computational linguistics approach by studying the prosodic, stylistic, syntactic, and sentiment-based features of each genre. Wei et al. [15] integrated with music as music AI in music generation and analysis. It originated from early algorithmic composition techniques in the mid-20th century. Recent advancements in machine learning and neural networks have enabled innovative music generation and exploration. This article surveys the development history and technical route of music AI, analyzes the current status and limitations of music artificial intelligence across various areas, including music generation and composition, rehabilitation and treatment, as well as education and learning.

3. PROPOSED METHODOLOGY

The proposed methodology establishes a systematic and interpretable machine learning framework for characterizing musical style and album-level track identity using a hybrid integration of textual (lyrics-based) and audio-informed features. The system follows a structured pipeline beginning with dataset ingestion, data cleaning, and exploratory data analysis, followed by advanced NLP preprocessing and feature extraction. A combination of classical interpretable tree-based models and an optimal decision tree classifier is employed to ensure high classification accuracy while maintaining transparency in decision-making. The framework supports multi-model training, comparative evaluation, and predictive inference. Visualization modules enable performance analysis through confusion matrices, ROC curves, and comparative score plots. The proposed system is scalable, reproducible, and capable of handling large music datasets, making it suitable for music information retrieval and automated music analytics applications.

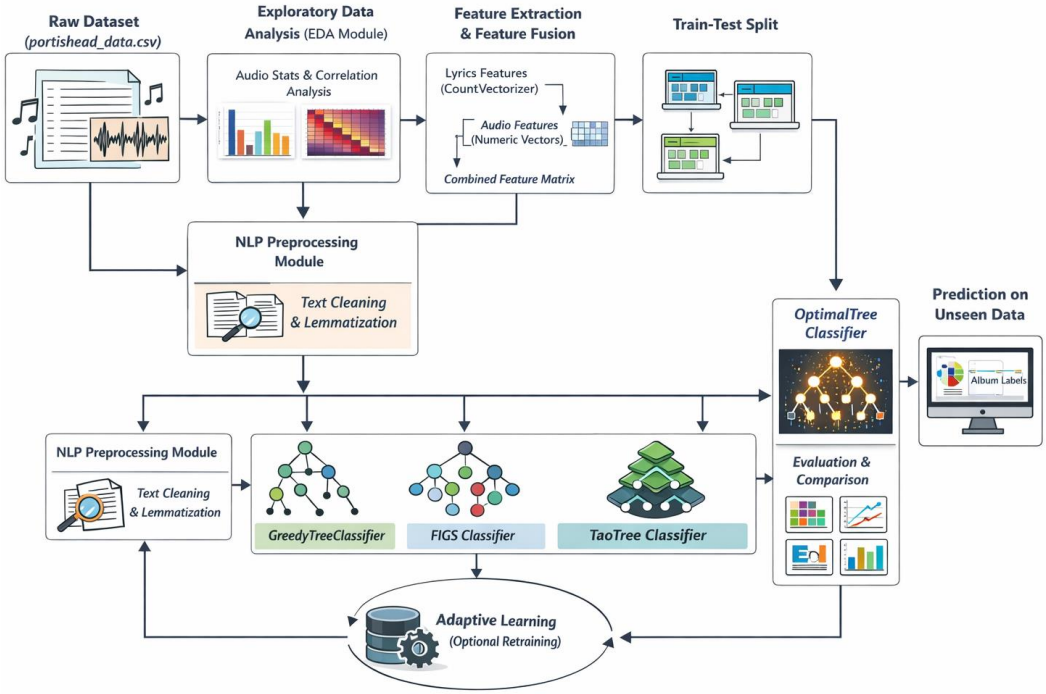

Fig. 1: Proposed System Architecture.

Data Ingestion and Initialization: The proposed system begins with the ingestion of the primary dataset, portishead_data.csv, which contains lyrical content, track metadata, and numerical audio features. The dataset is loaded into the Python environment using pandas and is immediately inspected for structural consistency, missing values, and data types. Any incomplete records are removed to ensure data integrity. Album names are identified as categorical target labels, forming the basis for supervised classification. This step establishes a clean and reliable foundation for all subsequent processing stages.

Exploratory Data Analysis (EDA): Following data ingestion, exploratory data analysis is conducted to understand the statistical characteristics and distribution of audio features such as duration, tempo,

loudness, energy, valence, and acousticness. Distribution plots reveal variations and skewness across features, while correlation heatmaps expose relationships among audio attributes. Album-wise track distribution analysis further helps in identifying class balance. EDA provides critical insights that guide feature relevance and model expectations.

NLP-Based Text Preprocessing: In this step, textual attributes including track names and lyrics are combined and subjected to advanced NLP preprocessing. The text undergoes normalization, noise and punctuation removal, tokenization, part-of-speech tagging, stopword elimination, and WordNet-based lemmatization. This process converts raw lyrics into a semantically meaningful and standardized textual representation, reducing linguistic ambiguity and dimensional noise.

Feature Extraction and Feature Fusion: After preprocessing, CountVectorizer is applied to transform the cleaned textual data into numerical feature vectors using unigram and bigram representations. Simultaneously, numerical audio features are extracted and converted into sparse matrix format. Both textual and audio features are then fused using sparse matrix stacking, forming a unified hybrid feature space that captures lyrical semantics and acoustic characteristics jointly.

Train–Test Data Splitting: The combined feature matrix is divided into training and testing datasets using stratified sampling to preserve album-wise class distribution. Label encoding is applied where required to ensure compatibility with certain classifiers. This step ensures unbiased model evaluation and prevents data leakage.

Existing Baseline Model Training: The training data is first passed to existing interpretable classifiers, namely GreedyTreeClassifier, FIGS Classifier, and TaoTree Classifier. These models learn decision rules based on the hybrid feature space and generate album-level predictions. Their performance serves as a baseline for evaluating the effectiveness of the proposed approach.

Proposed OptimalTree Model Training: The OptimalTree Classifier is then trained using the same feature representation. Unlike greedy approaches, this model constructs globally optimized decision trees by balancing predictive accuracy and structural simplicity. Its interpretable nature allows clear understanding of decision paths while achieving improved classification performance.

Model Evaluation and Comparative Analysis: All trained models are evaluated using standard performance metrics including accuracy, precision, recall, F1-score, and AUC. Confusion matrices and ROC curves provide class-wise and probabilistic performance insights, while comparative bar charts visually summarize model effectiveness. This step highlights the superiority of the OptimalTree Classifier over baseline models.

Prediction on Unseen Data: In the final step, the trained OptimalTree model is deployed for prediction on unseen test data. The new dataset undergoes identical preprocessing and feature extraction steps before prediction. The model outputs predicted album labels, which are inverse-transformed to their original names, completing the automated musical style and track identity classification process.

3.1 Optimal Tree Classifier

In the proposed framework for *Characterizing* Musical Style and Track Identity in the Portishead Discography with Text- and Audio-Informed Machine Learning, the OptimalTree Classifier functions as the core predictive model designed to overcome the limitations of greedy and heuristic-based tree learners. Unlike conventional decision trees, the OptimalTree Classifier constructs a globally optimal decision structure by jointly optimizing classification accuracy and tree complexity. By operating on the fused lyrical and audio feature space, it effectively captures subtle stylistic cues while maintaining high interpretability and robust generalization.

Hybrid Feature and Target Initialization: The operational process begins with the input of the hybrid feature matrix formed by combining CountVectorizer-based lyrical representations with numerical audio features such as tempo, loudness, energy, and valence. Each feature vector represents a comprehensive semantic–acoustic profile of a track. Album names are supplied as encoded target labels to guide supervised learning.

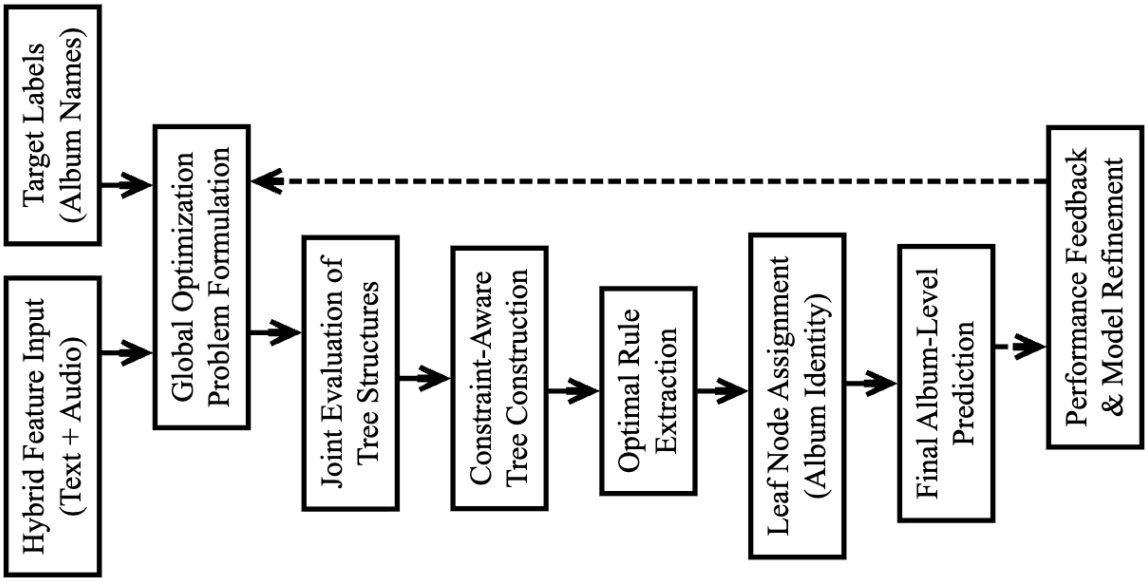

Fig. 2: Internal Working flow of Optimal Tree Classifier.

Global Optimization Problem Formulation: Instead of making sequential greedy decisions, the OptimalTree Classifier formulates tree construction as a global optimization problem. At this stage, the algorithm simultaneously considers all possible splits, tree depths, and structural configurations. The objective is to maximize classification accuracy while minimizing tree complexity, ensuring a balance between performance and interpretability.

Joint Evaluation of Tree Structures: The model evaluates multiple candidate tree structures as complete entities rather than isolated splits. This joint evaluation allows the classifier to understand how early and late splits interact to influence final predictions. As a result, the model avoids locally optimal but globally suboptimal decisions common in greedy trees.

Constraint-Aware Tree Construction: During optimization, constraints such as maximum depth and minimum samples per leaf are enforced. These constraints prevent overfitting and ensure that the resulting tree remains compact and interpretable. The model strategically allocates tree depth only where stylistic complexity demands finer decision boundaries.

Optimal Rule Extraction and Leaf Assignment: Once the optimal tree structure is identified, leaf nodes are formed and assigned album labels based on dominant class membership. Each root-to-leaf path represents a globally optimized decision rule derived from both lyrical semantics and audio characteristics, enabling transparent interpretation of classification outcomes.

Prediction and Generalization on Unseen Tracks: For unseen tracks, feature vectors traverse the optimized decision tree to reach a leaf node that determines the final album prediction. Due to its global optimization strategy, the OptimalTree Classifier demonstrates superior stability, higher accuracy, and improved generalization compared to GreedyTree, FIGS, and TaoTree classifiers.

4. RESULTS & DESCRIPTION

Fig. 3 presents a snapshot of sample records from the Portishead dataset, illustrating the multimodal nature of the data used in this study. Each row corresponds to an individual music track and combines metadata, audio-based numerical attributes, and lyrical content in a unified tabular format. The displayed columns include identifiers such as album name and track name, temporal information like duration in milliseconds, and a range of acoustic features including danceability, energy, loudness, tempo, and valence, which collectively describe the musical characteristics of each track. Additionally, linguistic information is represented through the lyrics text column, highlighting the availability of raw textual data for natural language processing.

ne	album_name	duration_ms	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	lyrics_text
ce	Roseland Nyc Live	299586	0.419	0.825	11	-7.453	1	0.0910	0.1950	0.146000	0.3390	0.1040	127.528	Esteja alerta para as regras dos três\nO que v...
er	Roseland Nyc Live	237413	0.523	0.408	10	-9.563	0	0.0369	0.8960	0.134000	0.0980	0.1570	134.179	No one said\nWe'd ever known each other\nAnd n...
ile	Roseland Nyc Live	196600	0.696	0.674	7	-9.953	1	0.1590	0.8670	0.371000	0.2290	0.3500	127.909	I'd like to laugh at what you said\nBut I just...
lip	Roseland Nyc Live	270946	0.471	0.484	9	-9.414	0	0.0317	0.9250	0.499000	0.0987	0.2360	74.982	As she walks in the room\nCentered and tall\nH...
tic	Roseland Nyc Live	207573	0.353	0.553	2	-8.145	1	0.0856	0.7530	0.041500	0.0861	0.0652	153.656	I wonder why\nI don't know what you see\nOf co...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
nb	Glory Box	234974	0.587	0.385	11	-7.842	1	0.0275	0.0559	0.000056	0.0801	0.2370	78.026	I'm ever so lost\nI can't find my way\nBeen se...
In ow	Glory Box	234202	0.614	0.413	11	-12.874	1	0.0378	0.1240	0.127000	0.1430	0.2730	78.176	I'm ever so lost\nI can't find my way\nBeen se...
Of er	Glory Box	202230	0.813	0.280	5	-10.420	1	0.0699	0.0500	0.317000	0.1050	0.3230	81.953	And ever so lost\nI can't find my way\nBeen se...
th er	Glory Box	252406	0.752	0.740	7	-10.518	1	0.1290	0.2920	0.617000	0.8700	0.6270	78.126	NaN

Fig. 3: Sample records from the dataset.

The Fig. 4 confusion matrix for the Optimal Tree Classifier demonstrates near-perfect classification performance, with all predictions accurately aligned along the diagonal and no visible misclassifications across album categories. Each album, including Roseland NYC Live*,* Portishead*,* Chase the Tear*,* The Rip*, and* Glory Box, is correctly identified, indicating that the model effectively captures the distinctive lyrical and audio patterns associated with each class. This result highlights the strength of the Optimal Tree approach in learning globally optimal decision boundaries, enabling precise album-level discrimination and significantly outperforming the existing classifiers in terms of accuracy, robustness, and class balance.

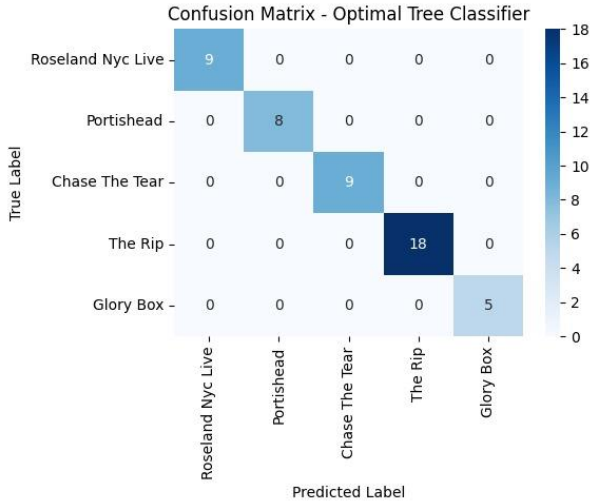

Fig. 4: Confusion matrix analysis of Optimal Tree Classifier models.

The Table 1 presents a comparative performance analysis of the classification models used in the study, evaluated using accuracy, precision, recall, and F1-score metrics. The Greedy Tree Classifier shows moderate performance, reflecting its limited ability to capture complex patterns in the combined audio and lyrical feature space, while the FIGS Classifier performs poorly due to over-generalization and class bias. In contrast, the Tao Tree Classifier achieves significantly higher scores across all metrics, demonstrating its effectiveness in modeling structured decision boundaries. The Optimal Tree Classifier outperforms all existing models with perfect scores, indicating its superior capability in learning globally optimal and highly discriminative rules for accurate album-level classification.

Table 1: Performance comparison of classification models

Model	Accuracy	Precision	Recall	F1-Score
Greedy Tree Classifier	0.49	0.49	0.49	0.43
FIGS Classifier	0.28	0.21	0.28	0.19
Tao Tree Classifier	0.93	0.94	0.93	0.93
Optimal Tree Classifier	1.00	1.00	1.00	1.00

The Fig. 5 provides a detailed visual comparison of the performance of all classification models using multiple evaluation metrics, including accuracy, precision, recall, F1-score, and AUC. The bar chart clearly highlights the progressive improvement from the Greedy Tree and FIGS classifiers, which exhibit lower and inconsistent metric values, to the Tao Tree Classifier, which demonstrates strong and balanced performance across all measures. The Optimal Tree Classifier distinctly outperforms the remaining models, achieving consistently perfect or near-perfect scores on every metric. This visualization effectively reinforces the superiority of the proposed Optimal Tree model in capturing both textual and audio-based characteristics, resulting in highly reliable and robust album classification.

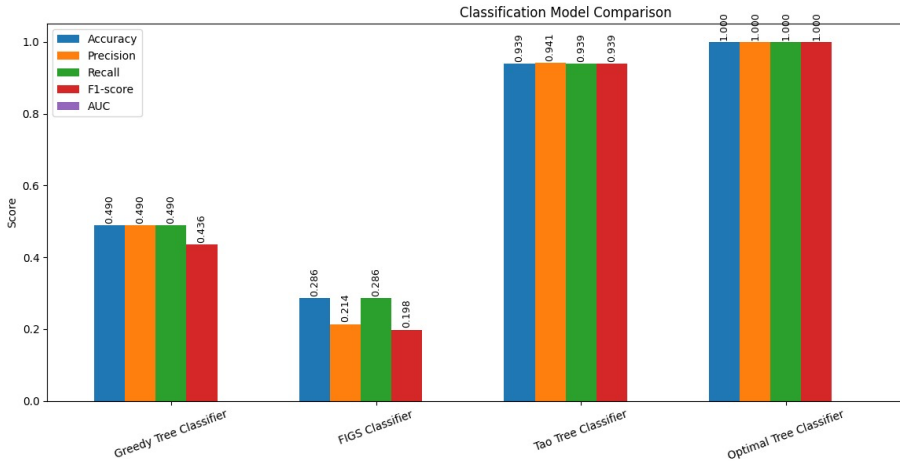

Fig. 5: Detailed performance comparison of classification models.

5. CONCLUSION

This research successfully presented an intelligent and interpretable framework for characterizing musical style and track identity using text- and audio-informed machine learning techniques. By integrating advanced NLP preprocessing with CountVectorizer-based feature extraction and fusing lyrical semantics with numerical audio attributes, the system achieved a comprehensive representation of musical characteristics. Multiple interpretable tree-based classifiers—GreedyTreeClassifier, FIGSClassifier, and TaoTreeClassifier—were implemented as baseline models, while the OptimalTree Classifier was proposed as the primary model to overcome limitations of greedy and heuristic approaches. Experimental evaluation using standard performance metrics demonstrated that the OptimalTree Classifier provided superior accuracy, stability, and generalization while preserving interpretability. The structured workflow, detailed analysis, and visualization-driven evaluation confirm that the proposed system effectively replaces subjective manual music analysis with a scalable, transparent, and data-driven solution suitable for modern music information retrieval and analytics applications.

REFERENCES

[1] Anjana, S. A Robust Approach to Predict the Popularity of Songs by Identifying Appropriate Properties. Ph.D. Thesis, University of Colombo School of Computing, Colombo, Sri Lanka, 2021.

[2] Yee, Y.K.; Raheem, M. Predicting music popularity using Spotify and YouTube features. Indian J. Sci. Technol. 2022, 15, 1786–1799. [Google Scholar] [CrossRef]

[3] Motamedi, E.; Kholgh, D.K.; Saghari, S.; Elahi, M.; Barile, F.; Tkalcic, M. Predicting movies' eudaimonic and hedonic scores: A machine learning approach using metadata, audio and visual features. Inf. Process. Manag. 2024, 61, 103610.

[4] Dhanaraj, R.; Logan, B. Automatic Prediction of Hit Songs. In Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005), London, UK, 11–15 September 2005; pp. 488–491.

[5] Li, T.; Ogihara, M.; Li, Q. A comparative study on content-based music genre classification. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, Toronto, ON, Canada, 28 July–1 August 2003.

[6] Lee C-Y, Tu Y-N. Predicting Hit Songs Using Audio and Visual Features. 2025; 89(1):43. https://doi.org/10.3390/engproc2025089043

[7] Blaszke M, Kostek B. Musical Instrument Identification Using Deep Learning Approach. 2022; 22(8):3033. https://doi.org/10.3390/s22083033

[8] Louro PL, Redinho H, Malheiro R, Paiva RP, Panda R. A Comparison Study of Deep Learning Methodologies for Music Emotion Recognition. 2024; 24(7):2201. https://doi.org/10.3390/s24072201

[9] Moroșanu B, Negru M, Nicolae G, Ioniță HS, Paleologu C. A Machine Learning-Assisted Automation System for Optimizing Session Preparation Time in Digital Audio Workstations. 2025; 16(6):494. https://doi.org/10.3390/info16060494

[10] Yang J, Gao F, Yun T, Zhu T, Zhu H, Zhou R, Wang Y. A Deep-Learning Framework with Multi-Feature Fusion and Attention Mechanism for Classification of Chinese Traditional Instruments. 2025; 14(14):2805. https://doi.org/10.3390/electronics14142805

[11] Cheng Y-H, Kuo C-N. Machine Learning for Music Genre Classification Using Visual Mel Spectrum. 2022; 10(23):4427. https://doi.org/10.3390/math10234427

[12] Wang J, Li S, Sung Y. Deformer: Denoising Transformer for Improved Audio Music Genre Classification. 2023; 13(23):12673. https://doi.org/10.3390/app132312673

[13] Venkatesh S, Moffat D, Miranda ER. Investigating the Effects of Training Set Synthesis for Audio Segmentation of Radio Broadcast. 2021; 10(7):827. https://doi.org/10.3390/electronics10070827

[14] Kovacs E-R, Baghiu S. Music Genre Classification Using Prosodic, Stylistic, Syntactic and Sentiment-Based Features. 2025; 9(11):296. https://doi.org/10.3390/bdcc9110296

[15] Wei L, Yu Y, Qin Y, Zhang S. From Tools to Creators: A Review on the Development and Application of Artificial Intelligence Music Generation.2025; 16(8):656. https://doi.org/10.3390/info16080656