JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MUSIC GENRE CLASSIFIERUSING MACHINE LEARNING

MAHEK SHARMA, Siddid Kaul, Ishika Singh
STUDENT
AMITY UNIVERSITY

Abstract*—Machine learning (ML) and artificial intelligence (AI) have emerged as transformative technologies across numerous sectors, including music analysis. Music genre classification, a complex task within music information retrieval, involves extracting significant audio features to accurately categorize music. This research investigates various ML algorithms for music genre classification, evaluating their performance on the GTZAN dataset. The study encompasses traditional classifiers, convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models. Results indicate accuracies from 40% to 95%, with CNN-RNN hybrid models, specifically convolution-recurrent neural networks (CRNNs), demonstrating superior performance. Additionally, ensemble methods like AdaBoost were found to enhance classification accuracy. This work contributes to the field by exploring effective music genre classification strategies, particularly emphasizing parallel and ensemble techniques.

Keywords *– Convolution-Recurrent Neural Networks, Ensemble Learning, Music Genre Classification, Music Information Retrieval.*

I. INTRODUCTION

Music, the artistic arrangement of sounds, encompasses diverse genres like classical, folk, rock, pop, electronic, and hip-hop. Evolving perceptions of music have spurred new genres and sub-genres. Technological advancements have transformed how we consume music. Within the field of artificial intelligence (AI), music genre classification – a subset of music information retrieval (MIR) – presents a complex challenge. It employs machine learning (ML) algorithms to categorize music audio files based on stylistic features. Successful classification enables applications in music streaming services (e.g., Spotify), identification apps (e.g., Shazam), and smart assistants for personalized music experiences.

This research investigates various ML algorithms for music genre classification, comparing their performance and exploring the effectiveness of ensemble techniques. Extensive prior work exists, including models utilizing convolutional neural networks (CNNs), recurrent neural networks (RNNs), hybrid approaches, and more [1-10].

Naive Bayes, K-Nearest Neighbors, Decision Tree, and Logistic Regression. Each of these algorithms brings its unique strengths, allowing the system to harness their combined predictive power.

Understanding that accessibility is key to widespread utilization, we have also developed an web application with an intuitive Graphical User Interface(GUI). This platform enables users toeasily input specific parameters related to their heart health and receive instant predictions regarding the state of their heart: healthy or not. [9] With this system, our objective is to bridge the gap between sophisticated ML models and everyday users, democratizing access to state-of-the-art heart disease prediction tools. By making early diagnosis more accessible and efficient, we hopeto contribute to a future where heart disease isdetected in its nascent stages, leading to betteroutcomes and improved quality of life for patients globally.

II. Materials and Methods

Dataset

This study utilizes the GTZAN dataset, a publicly available collection of 1000 audio files (30 seconds each) representing 10 musical genres (blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock). It can be downloaded from Kaggle.

Feature Extraction

We employ librosa [12], a Python library for audio analysis, to extract the following features from each audio file:

Zero-crossing rate: Rate at which the audio signal transitions between positive and negative values.

Chroma STFT: Normalized energy for each chroma bin in a chromagram derived from the waveform.

Spectral centroid: Indicates the frequency spectrum's center of mass.

Spectral bandwidth: Frequency range within each frame of the spectrogram.

Spectral roll-off: Frequency below which a specified percentage (default 85%) of the spectrum's energy is contained.

MFCCs (Mel-Frequency Cepstral Coefficients): 20 MFCCs were used to represent the short-term power spectrum of the audio, based on a mel scale transformation (see Equation 1).

$$\text{Mel}(f) = 2595 \log\left(1 + \frac{f}{700}\right)$$

Equation 1

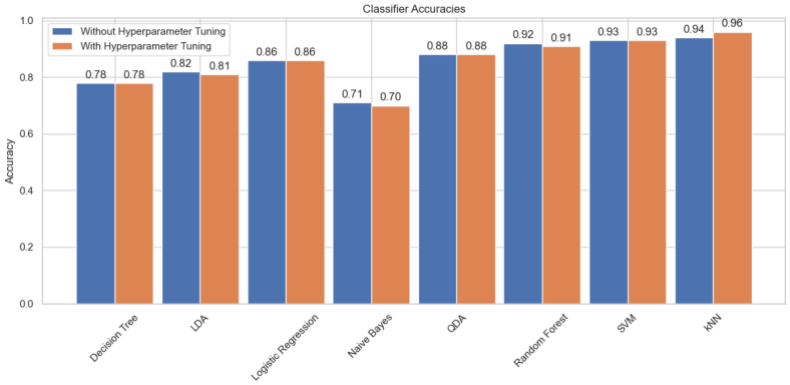


Table 1 Our Proposed Models

We established a baseline performance by training an initial model on the dataset. Subsequently, we employed the following models to improve accuracy:

Logistic regression

Logistic regression is a linear classification model employed for binary tasks. It directly estimates the probability of an input belonging to a specific class using the sigmoid function. Coefficients associated with each feature are linearly combined to compute log-odds, which are subsequently transformed into probabilities bythe logistic function.

Decision Tree

Decision trees are hierarchical, non-parametric models used for classification and regression tasks. They partition the feature space into regions using a series of decision rules applied at each internal node. Branches represent the possible outcomes of a decision, and leaf nodes denote class labels or predicted values. Decision trees offer interpretability and can handle both numericaland categorical features.

Support Vector Machines

(SVMs) are powerful kernel-based classifiers well- suited for music genre classification. They excel at handling complex feature spaces by constructing hyperplanes that optimally separate classes in a high- dimensional representation. SVMs maximize the margin between support vectors (data points nearest to the hyperplane), enhancing generalization. Classification of new data points is based on their position relative to the decision hyperplane.

Gaussian Naïve Bayes (NB)

Naive Bayes is a probabilistic classifier that leverages Bayes' Theorem under the assumption of feature independence. While this assumption is often unrealistic, the algorithm frequently performs well in practice, particularly for text classification and tasks involving high-dimensional data. It estimates the probability of an input belonging to each class, assigning it to the class with the highest calculated probability. Variants like Gaussian Naive Bayes and Multinomial Naive Bayes are tailored for continuous and discrete features, respectively.

k-Nearest Neighbour Classifier (kNN)

The k-Nearest Neighbors (kNN) algorithm is a non- parametric instance-based classifier used in music genre classification. Musical pieces are represented as points in a feature space defined by attributes like tempo, rhythm, pitch, etc. Classification involves identifying the k nearest neighbors of a data point and assigning it the majority class label among these neighbors. The choice of k impacts the bias-variance trade-off, with smaller k leading to more complex decision boundaries and largerk favoring smoother boundaries.

kNN's adaptability to complex relationships between musical features makes it well-suited for genre classification, where decision boundaries are often non- linear. Its non-parametric nature offers robustness when data distributions are unknown. However, careful choice of the distance metric and potential use of dimensionality reduction techniques are important due tokNN's sensitivity to the curse of dimensionality.

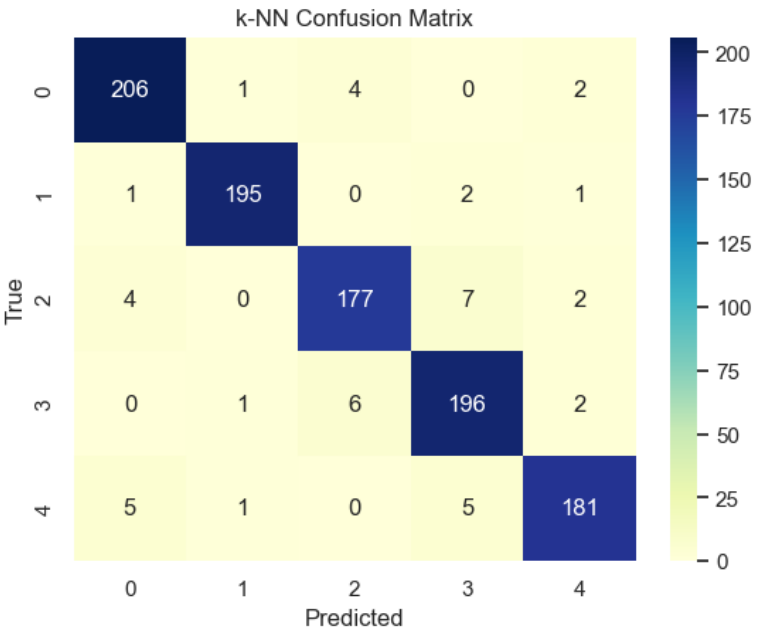


Table 2

Random Forest Classifier

The Random Forest classifier is an ensemble method that leverages multiple decision trees to achieve robust performance in music genre classification. It excels at handling high-dimensional data and complex decision boundaries. Each tree is trained on a bootstrapped sample of the data with a random subset of features. Classification employs a majority vote among the individual tree predictions. This approach mitigates overfitting and offers resilience to noise and outliers in the data. Additionally, Random Forests provide implicit feature importance analysis, aiding the interpretation of relevant musical attributes for genre discrimination.

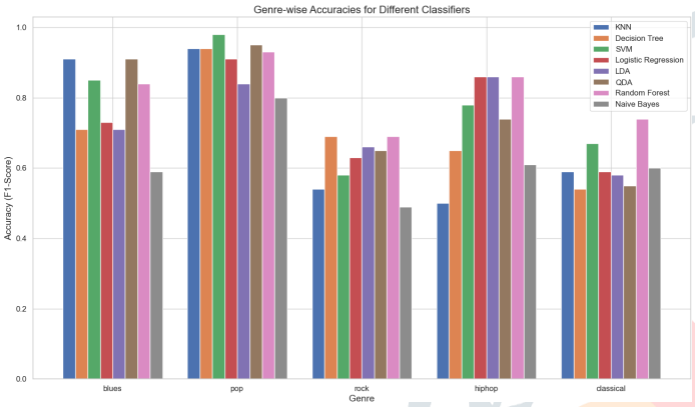


Table 3

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a dimensionality reduction technique commonly used for music genre classification. LDA seeks linear combinations of features that maximize class separation while minimizing intra- class variance. It projects the feature space onto a lower-dimensional subspace by computing eigenvectors of the between-class and within-class scatter matrices. This facilitates visualization and can improve classification accuracy.

LDA's interpretability and its assumption of Gaussian class distributions are potential advantages. However, LDA is most effective when classes are linearly separable. For complex or non-linearly separable data, consider alternative classifiers or pre-processing techniques.

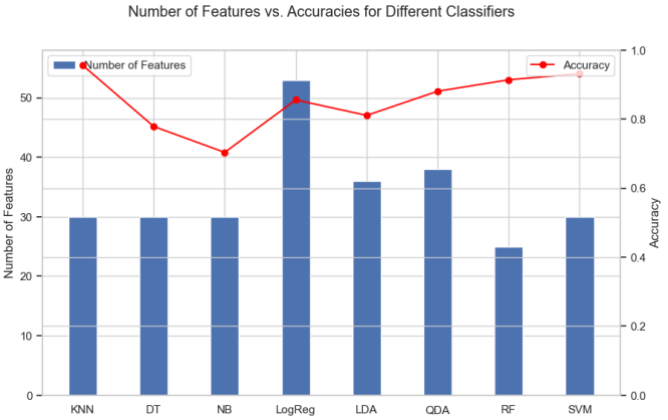
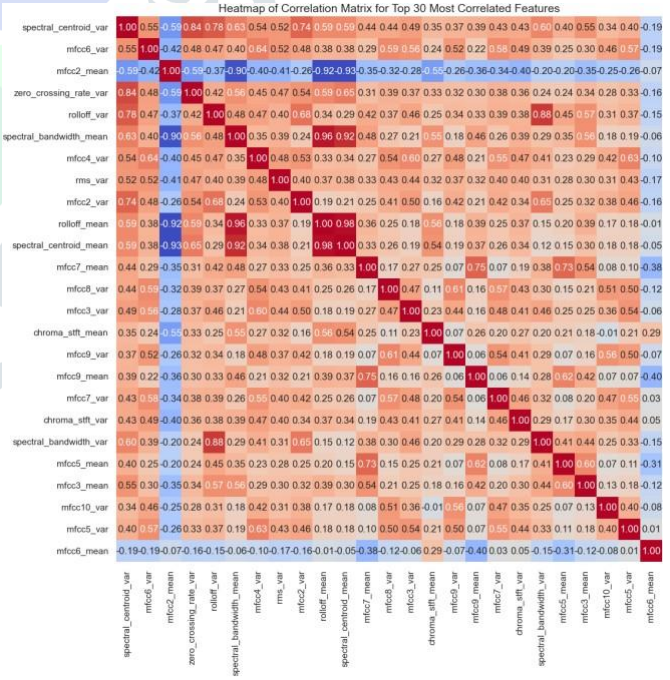


Table 4 Quadratic Discriminant Analysis (QDA)

Quadratic Discriminant Analysis (QDA) is a classification method often used for music genre classification. Unlike Linear Discriminant Analysis (LDA), which assumes equal covariance matrices across classes, QDA models a separate covariance matrix for each class. This flexibility allows QDA to capture complex class distributions, making it well-suited to music data with nonlinear relationships between features and genres. QDA can model nonlinear decision boundaries, enhancing its accuracy in distinguishing genres where linear classifiers may struggle. Additionally, QDA offers insights into the covariance structure of each genre, aiding interpretability.



Table

RESULT

This analysis delves into the effectiveness of various machine learning algorithms for music genre classification using the GTZAN dataset. This dataset comprises 1000 audio excerpts (30 seconds each) belonging to 10 distinct genres – blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, and rock. The provided visualizations offer a comparative perspective on the performance of different algorithms and the characteristics that differentiate musical genres.

The first image portrays a bar graph showcasing the accuracy (measured by F1-score) of various algorithms (KNN, Decision Tree, SVM, Logistic Regression, LDA, QDA, Random Forest, and Naïve Bayes) for each genre. While the overall average accuracy across genres hovers around 70%, a closer look reveals significant differences in algorithm performance. Random Forest emerges as the strongest contender, exhibiting consistently high accuracy across all genres. This suggests its ability to effectively capture the complex relationships between various audio features and genre classification. Conversely, Naïve Bayes consistently delivers the lowest accuracy, potentially due to its underlying assumption of feature independence, which may not hold true for complex audio data. Support Vector Machines (SVMs) and K-Nearest Neighbors (KNN) also demonstrate promising performance, achieving moderate to high accuracy across most genres. These algorithms offer a good balance between accuracy and computational efficiency.

The second Image, a heatmap, visualizes the correlations between the top 30 most relevant features extracted from the audio data in the GTZAN dataset. These features include spectral properties (MFCCs, spectral centroid, rolloff, bandwidth) and temporal characteristics.
Analyzing how features correlate with the genre classification accuracy observed in the first image provides insights into their role in genre differentiation. Both visualizations combined provide valuable guidance for practical implementation.

For real-world applications, the chosen algorithm should strike a balance between accuracy, computational efficiency, and interpretability. Random Forest, while boasting the highest accuracy, can be computationally expensive for large datasets. Support Vector Machines or K-Nearest Neighbors might be more suitable choices in such scenarios. Additionally, while Random Forest offers a powerful model, its internal workings can be less interpretable compared to simpler models like SVMs or KNNs. If understanding the rationale behind genre classification is crucial, these latter algorithms might bepreferable.

CONCLUSION

In the ever-evolving landscape of music, the classification of genres is a complex task that intersects technology, human perception, and artistic expression. The analysis undertaken underscores the pivotal role of machine learning in this domain, particularly in the context of genre classification. Among the array of algorithms explored, Random Forest stands out as a beacon of accuracy. Its ability to aggregate the predictions of multiple decision trees renders it adept at discerning intricate patterns within music data, resulting in robust classification outcomes.

However, the pursuit of the most suitable algorithm extends beyond mere accuracy. Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) emerge as compelling alternatives, offering a delicate equilibrium between accuracy and computational efficiency. SVM, with its prowess in delineating complex decision boundaries, and KNN, with its simplicity and adaptability, present viable options for applications where a balance between performance and computational resources is crucial.

Yet, the choice of algorithm cannot be divorced from the contextual nuances of the application. Practical implementation mandates a holistic consideration of various factors. Firstly, the specific accuracy requirements of the task at hand play a pivotal role. In scenarios where precision is paramount, the nuanced predictive capabilities of Random Forest may outweigh its computational demands.
Conversely, for applications where real-time processing or resource constraints loom large, the efficiency of SVM or KNN may tilt the scales in their favor.

Furthermore, the interpretability of the chosen model emerges as a salient consideration. While Random Forest may excel in predictive accuracy, its ensemble nature may obfuscate the interpretability of individual decisions. In contrast, SVM and KNN offer more transparent decision-making processes, facilitating a deeper understanding of the underlying classification mechanisms—an invaluable asset in applications where interpretability is paramount.

In essence, the efficacy of machine learning for music genre classification is indisputable. However, the optimal choice of algorithm hinges on a nuanced evaluation of various factors, including accuracy, efficiency, and interpretability. As technology continues to evolve and musical landscapes evolve in tandem, the quest for the perfect harmony between algorithmic prowess and practical exigencies remains perpetual.

REFERENCES

[1] Annesi, P., Basili, R., Gitto, R., Moschitti, A., & Petitti, R. (2007).Audio Feature Engineering for Automatic Music
Genre Classification. RIAO, 702–711

[2] Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011).The million song dataset in Proceedings of the
12th International Society for Music Information Retrieval Conference. Miami, October, 24, 591–596.

[3] Chang, K. K., Jang, J.-S. R., & Iliopoulos, C. S. (2010). Music GenreClassification via Compressive Sampling. ISMIR,
387–392.

[4] Chathuranga, D., & Jayaratne, L. (2013). Automatic music genreclassification of audio signals with machine learning
approaches. GSTF Journal on Computing (JoC), 3, 1–12
.

[5] Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017). Convolutionalrecurrent neural networks for music classification.
2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2392–2396.

[6] Feng, L., Liu, S., & Yao, J. (2017). Music genre classification withparalleling recurrent convolutional neural network.
ArXiv Preprint ArXiv:1712.08370

[7] Jeong, I.-Y., & Lee, K. (2016). Learning Temporal Features Using aDeep Neural Network and its Application to
Music Genre Classification. Ismir, 434–440.

[8] Jothilakshmi, S., & Kathiresan, N. (2012). Automatic music genreclassification for indian music. Proc. Int. Conf.
Software Computer App.

[9] Lidy, T., & Schindler, A. (2016). Parallel convolutional neuralnetworks for music genre and mood classification.
MIREX2016, 3

[10] IJEBSS e-ISSN: 2980-4108 p-ISSN: 2980-4272 320
IJEBSS Vol. 1 No. 04, April 2023, pages: 308-320

[11] 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, 8, 18–25.

[12] Pelchat, N., & Gelowitz, C. M. (2020). Neural network music genreclassification. Canadian Journal of Electrical and
Computer Engineering, 43(3), 170–173.