International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 • Website: www.ijfmr.com • Email: editor@ijfmr.com

Machine Learning Based Music Categorization

Narendra Singh[1], Satyam Singh Rawat[2], Anirudh Ratauri[3], Rohit Singh Rawat[4], Princy Tyagi[5]

[1,2,3,4]Student of Bachelor of Computer Application, Department of Computer Science & Engineering, Himalayan School of Science & Technology, Swami Rama Himalayan University, Dehradun, India
[5]Assistant Professor CSE Department of Computer Science and Engineering, Swami Rama Himalayan university, Dehradun

Abstract:
Music genre categorization is an essential activity in music data retrieval and recommendation systems. This research focuses on classifying music genres using machine learning techniques, specifically the Support Vector Machine. The GTZAN dataset, comprising 10 distinct genres, is utilized for training and evaluation. We took audio features like MFCCs, spectral contrast, and chroma vectors from the GTZAN dataset and used them to train a Support Vector Machine (SVM) model. The categorization model achieved an accuracy of 81.1% across 10 distinct genres in a multi-class setting the research emphasizes the difficulties of genre convergence and the efficacy of machine learning in automating music categorization. Future developments might explore deep learning techniques, like Convolutional Neural Networks, better ways to choose features, and improving data to make music categorization more effective. assignment in music retrieve information

Keywords: Music Genre Categorization, Machine Learning, Support Vector Machine, GTZAN Dataset, Feature Extraction.

Introduction

Music is essential to human culture and recreation. Throughout the years, various music genres have evolved, each characterized by unique attributes such as rhythm, harmony, instrumentation, and vocal styles. The growing number of music streaming platforms has made proper categorization of genres essential for music organizations as well as systems for recommendations and the overall user experience. Traditional genre classification techniques depended on manual annotation by specialists or listeners. This method is time-consuming, subjective, and frequently inconsistent. Advances in machine learning (ML) and artificial Intelligence (AI) has enabled automatic music categorization by extracting and analyzing audio features. [1] ML algorithms can detect patterns in music files and classify them into predefined genres with significant accuracy. This study focuses on implementing a **Support Vector Machine (SVM)** model for multi-class music genre categorization. We use the GTZAN dataset, a widely recognized benchmark dataset that contains **1,000 audio tracks spanning 10 different genres**. The model was trained utilizing extracted data.

We retrieved audio features including Mel-Frequency Cepstral Coefficients, spectral contrast, chroma features, and zero-crossing rate (ZCR) using the GTZAN music collection. The features were utilized for training a Support Vector Machine model. After performing hyperparameter tuning and feature optimization, the model attained a classification accuracy of 81.1%. The study highlights the challenges of genre overlap, where certain genres share similar auditory characteristics, complicating the classification process.

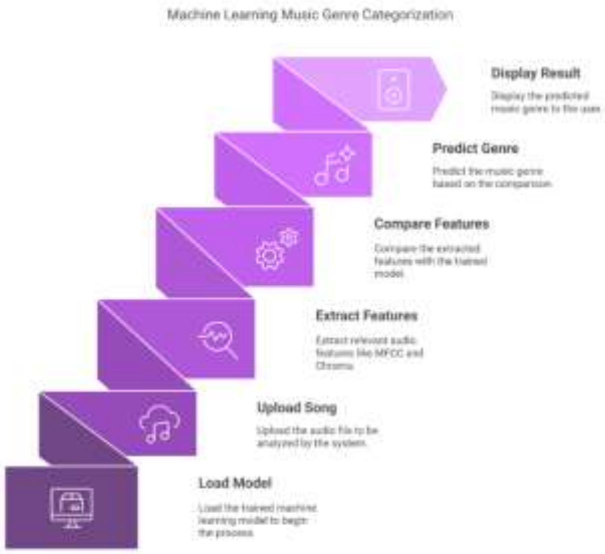

Figure 1: Machine Learning Music Genre Categorization — Process Flow

Future enhancements may incorporate deep learning methodologies, like **convolutional neural networks (CNNs)**, to discern more complex patterns in musical signals. Additionally, incorporating features such as **tempo, pitch, and rhythm analysis** could further enhance classification accuracy.

RELATED WORK

The categorization of music genres has been extensively studied within the realm of Music information retrieval (MIR). Many studies have investigated diverse methodologies for automating the categorization of music into established genres utilizing machine learning and deep learning methodologies.
Tzanetakis and Cook (2002) were among the first to introduce the GTZAN dataset, which has since become a benchmark for genre classification research. [2] Their work showed that using statistical audio feature extraction techniques like Mel-Frequency Cepstral Coefficients (MFCCs), spectral centroid, and zero-crossing rate is effective for telling apart different music genres.
Subsequent research explored different machine learning techniques, including K-Nearest Neighbors (KNN), Decision Trees, Random Forest, and Support Vector Machines (SVM). Among these, SVM has been widely recognized for its ability to handle high-dimensional feature spaces while maintaining strong classification performance. Studies have shown that SVM performs well in genre classification tasks but faces challenges in multi-class classification due to overlapping features among similar genres.
Improvements in deep learning have allowed the use of **Convolutional Neural Networks (CNNs)** and **Recurrent Neural Networks (RNNs)** for classifying spectrograms, leading to significantly better results than traditional machine learning methods.[3] These methodologies regard audio signals as pictures, enabling the model to discern the spatial relationships in sound. Nonetheless, deep learning models necessitate substantial datasets and significant processing resources, and this may not always be practical for small-scale applications.

Several research studies have examined hybrid models that integrate conventional machine learning methods with techniques from deep learning to take advantage of the benefits of both approaches. Researchers have investigated feature selection methodologies, including Principal Component Analysis (PCA) and t-SNE, to improve classification accuracy and decrease computational complexity. [4]
Notwithstanding these developments, music genre classification continues to pose a formidable challenge owing to the subjective essence of genres. overlapping audio characteristics and intra-genre variations. This study builds upon previous research by implementing an SVM-based model on the GTZAN dataset and analyzing its performance in multi-class classification.[5] Further, it explores the impact of feature extraction techniques and hyperparameter tuning on classification accuracy.

METHODOLOGY

This section defines the systematic methodology employed for categorizing music genres via machine learning. The procedure encompasses dataset selection, initial processing, feature extraction, training of models, and evaluation to guarantee an efficient classification pipeline.

Dataset and Preprocessing

This research uses the GTZAN dataset, which contains 1,000 audio samples, Each sample lasts for 30 seconds The samples are uniformly allocated over ten separate music genres, guaranteeing a balanced representation of diverse musical styles.

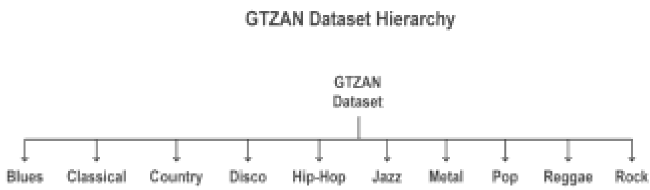

Figure 2: GTZAN Music Categories

To facilitate accurate categorization, we conducted extraction of features on the unprocessed audio recordings. Essential attributes like MFCCs, Chroma, Spectral Centroid, Bandwidth, Zero-Crossing Rate, RMS Energy, and Spectral Rolloff were calculated for each recording. These attributes served as input for the initial training stage of the machine learning model.
Preprocessing techniques were implemented to improve the model's performance. The feature values were normalized with the StandardScaler, and the genre labels were converted into numerical format for classification.

Model Selection and Training

We chose a Support Vector Machine model utilizing a linear kernel for classification because it excels with data rich in features. We trained the model on the complete dataset to obtain genre characteristics. The joblib library thereafter preserved the trained model for future utilization, thus obviating the necessity for retraining.
Instead of using a train-test split from the dataset for testing, we downloaded external audio files to evaluate the model's real-world applicability. These downloaded audio tracks were pre-processed using the same feature extraction techniques before being input into the trained SVM model for classification.

Evaluation Metrics

Various evaluation metrics were employed to evaluate model performance. We evaluated accuracy as the proportion of accurately predicted genres to the total quantity of predictions generated.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision, recall, and F1-score were utilized to evaluate the model's efficacy across several genres. Precision measures the ratio of correctly identified genre labels to the predicted labels, where **TP (True Positives)** denotes correctly categorized genres, **TN (True Negatives)** indicates properly disregarded categories, **FP (False Positives)** refers to incorrectly identified categories, and **FN (False Negatives)** signifies genres that were overlooked.

We utilized precision, recall, and F1-score to evaluate the model's effectiveness across each genre. Precision quantifies to assess the model's efficacy within each genre.

$$\Pr e\, cision = \frac{TP}{TP + FP}$$

Recall measures the model's precision in categorizing genuine genre samples.

$$Recall = \frac{TP}{TP + FN}$$

F1-score, the harmonic median of precision and recall provides an equitable assessment of model efficacy.

$$F1\text{-}Score = 2 \times \frac{\Pr e\, cision \times Recall}{\Pr e\, cision + Recall}$$

We use a confusion matrix to examine incorrect categorization and genre overlaps. It offers insights about the model's capacity to differentiate among related genres, highlighting the areas of the most frequent errors.

RESULT AND DISCUSSION

This research evaluates the efficacy of the Support Vector Machine model in categorizing music genres. We evaluate the model's efficacy for multi-class classification (10 genres) using essential measures like accuracy, precision, recall, F1-score, and confusion matrices. Furthermore, genre-specific misclassifications and possible enhancements are examined.

Multi-Genre Categorization (All 10 Genres)

We trained the model using the GTZAN dataset and evaluated its ability to classify all 10 genres. The classification task introduced complexity due to overlapping characteristics between certain genres.

The final model achieved a total accuracy of 81.10%, demonstrating the efficacy of machine learning in the categorization of music genres.

Table 1: Categorization Report for Multi-Genre Categorization

(10 Genres)

Model Accuracy: 81.10%				
	Precision Score	**True Positive Rate (Recall)**	**F-Measure (F1-Score)**	**Number of Samples**
Blues	0.77	0.85	0.81	100
Rock	0.65	0.63	0.64	100
Reggae	0.74	0.70	0.72	100
Pop	0.94	0.87	0.90	100

Metal	0.92	0.98	0.90	100
Jazz	0.93	0.84	0.88	100
Hip-Hop	0.77	0.78	0.78	100
Disco	0.71	0.75	0.73	100
Country	0.76	0.78	0.77	100
Classical	0.93	0.99	0.96	100
Macro Avg.	**0.81**	**0.81**	**0.81**	**1000**
Weighted Avg.	**0.81**	**0.81**	**0.81**	**1000**

Performance Variations

The model demonstrated **higher accuracy** for genres with unique spectral features, such as **Classical and Metal**, as they have distinct frequency distributions. In contrast, **Rock vs. Metal and Disco vs. Pop** exhibited classification challenges due to their overlapping rhythmic and instrumental characteristics.

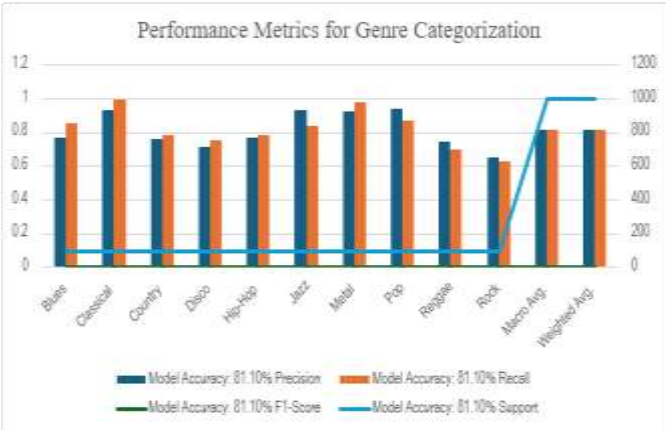

Figure 3: Performance Metrics for Genre Categorization

The chart below illustrates the efficacy of the SVM model for categorization music genres, highlighting the precision, recall, and F1-score for all categories. The findings demonstrate that genres like Classical and Metal attain superior accuracy owing to their unique spectral attributes.

In contrast, genres like Rock and Hip-Hop show lower classification accuracy, likely due to overlapping rhythmic and instrumental features. Misclassification is more common in genres with similar musical properties, such as Rock and Blues or Disco and Pop, where shared instrumentation and tempo variations pose classification challenges.

A more detailed analysis of these misclassifications is provided in the confusion matrix, which illustrates the frequency and nature of incorrect predictions among genres. These findings suggest that refining feature selection techniques and improving preprocessing methods could further enhance classification accuracy

Confusion Matrix Insights:

The confusion matrix analyzes the model's effectiveness by visualizing misclassifications over several genres.

The model achieves high accuracy for Classical, Metal, and Pop, indicating distinct genre characteristics. Blues, Jazz, and Reggae show moderate misclassification, with Blues occasionally predicted as Rock and Country. Hip-Hop and Reggae experience the highest misclassification rates, with Hip-Hop often confused with Reggae due to similar rhythmic structures.

Disco and Pop exhibit notable confusion, as Disco is misclassified as Rock and Pop, reflecting overlapping dance-oriented patterns. Rock has the highest misclassification, frequently confused with Country and Blues, highlighting challenges in distinguishing between these genres.

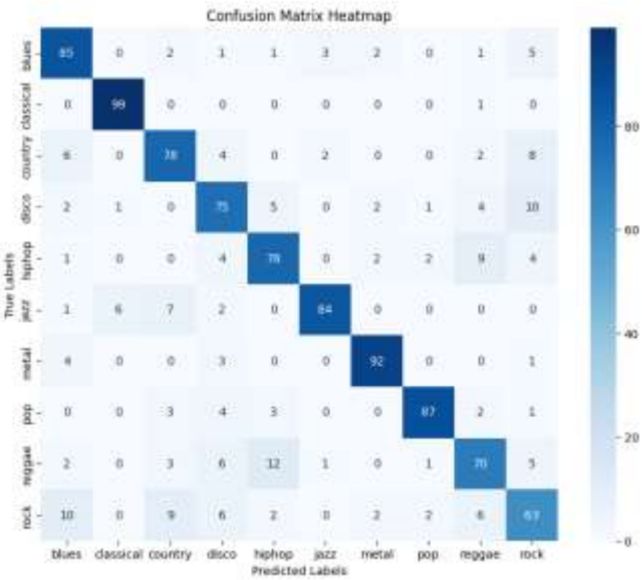

Figure 4: Confusion Matrix for Multi-Genre Categorization

These findings indicate that while the SVM model performs well for genres with distinct spectral features, genres with overlapping rhythmic and instrumental characteristics pose classification challenges.

CONCLUSION AND FUTURE WORK

This research assesses the efficacy of a Support Vector Machine model in categorizing multi-genre music utilizing the GTZAN dataset. The model achieved a comprehensive accuracy of 81.10% for categorizing 10 different genres, illustrating its proficiency in managing intricate categorization tasks.

The confusion matrix and classification reports reveal that misclassifications primarily occur between genres with similar spectral, rhythmic, and instrumental characteristics (e.g., **Rock vs. Metal, Disco vs. Pop**). The model performs well on genres with distinct musical structures, such as **Classical and Metal**, but struggles with overlapping characteristics in other genres.

To enhance classification robustness, future work can explore **deep learning models like CNNs or RNNs**, improved **feature engineering**, and **data augmentation techniques** to better differentiate similar genres.

REFERENCES

1. Bahuleyan, H. (2018). Music genre classification using machine learning techniques. arXiv preprint arXiv:1804.01149.
2. Tzanetakis, G., & Cook, P. (2002). “Musical genre Classification of audio signals.” **IEEE Transactions on Speech and Audio Processing**, 10(5), 293-302

3. Choi, K., Fazekas, G., Sandler, M., & Cho, K. (2017). “Convolutional recurrent neural networks for music Classification.” **2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, 2392-2396.
4. Panagakis, Y., Kotropoulos, C., & Arce, G. R. (2009, August). Music genre classification via sparse representations of auditory temporal modulations. In *2009 17th European Signal Processing Conference* (pp. 1-5). IEEE.
5. SuriyaPrakash, J., & Kiran, S. (2022, November). Obtain better accuracy using music genre classification systemon gtzan dataset. In 2022 IEEE North Karnataka subsection flagship international conference (NKCon) (pp. 1-5). IEEE.