International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 09, Issue – 04, April 2024, PP – 18-33

Genre Classification and Musical Features Analysis

Ozlem Kilickaya

Department of Computer Science, University of the People, Pasadena, USA

Abstract : Genre classification in music is a fundamental task in machine learning research, with implications for music recommendation systems, content organization, and music analysis. This study delves into the process of genre classification and analysis of musical features extracted from audio files. Leveraging the GTZAN dataset, a widely used resource in music genre recognition research, this study demonstrates the extraction and exploration of audio features, including but not limited to spectral centroid, chroma features, and tempo. Three machine learning models—Support Vector Machines (SVM), Random Forest, and XGBoost—are employed for genre classification, with meticulous attention given to hyperparameter tuning and proper train/test set partitioning to prevent data leakage. Notably, precautions are taken to ensure that excerpts from the same song do not overlap between training and testing sets, a common pitfall observed in similar studies. By looking closely at misclassified genre pairs and giving detailed explanations of classification results, the study explores the subtleties of genre labeling and questions how humans detect musical traits are different between genres. This research contributes to advancing understanding in music genre classification and underscores the importance of rigorous methodology in machine learning-based music analysis.

Keywords: Genre classification, Music analysis, Audio feature extraction, Machine learning, Hyperparameter tuning, Genre recognition, Music recommendation systems.

I. INTRODUCTION

Genre classification in music is a fundamental task with broad implications in various domains, including music recommendation systems, content organization, and cultural studies. The automatic categorization of music into distinct genres not only aids in music discovery, but also offers valuable insights into the underlying characteristics of musical compositions.

This paper focuses on the intricate process of genre classification and the analysis of musical features extracted from audio files. Leveraging the extensive GTZAN dataset, a widely recognized benchmark in music genre recognition, this study delves into the extraction and exploration of audio features. These features encompass a diverse range of attributes, such as spectral centroid, chroma features, and tempo, collectively capturing the essence of musical pieces [1].

At the heart of this investigation are three prominent machine learning models: Support Vector Machines (SVM), Random Forest, and XGBoost. Through rigorous experimentation, these models' predictive capabilities are harnessed to classify music into predefined genres. Notably, meticulous hyperparameter tuning and prudent train/test set partitioning are employed to mitigate the risk of data leakage and ensure robust model performance [2].

This study critically addresses the careful handling of excerpts from the same song to prevent overlap between training and testing sets, a common pitfall that can compromise classification integrity [3]. By meticulously managing these nuances, this research aims to provide a comprehensive understanding of genre classification and shed light on the underlying challenges and considerations in music analysis.

Furthermore, this analysis extends beyond classification accuracy to encompass the interpretation of misclassified genre pairs. Conventional assumptions about the distinctiveness of musical characteristics across genres are challenged, and the nuanced relationship between genre labels and underlying musical features is explored.

This research aims to contribute to the advancement of knowledge in music genre classification, highlighting the significance of robust methodology in machine learning-based music analysis. By unraveling the complexities of genre classification and musical feature analysis, this study aims to pave the way for more sophisticated and nuanced approaches to understanding music.

II. RELATED WORK

Musical genre classification has been a focal point of research in the field of music information retrieval (MIR), leveraging various machine learning techniques and audio signal processing methods. Researchers laid the groundwork for genre classification of audio signals, pioneering the application of machine learning in this

domain [4]. Since then, numerous studies have further advanced the field of machine learning, exploring different methodologies and datasets [26].

Zhang and Wang (2018) conducted a comprehensive survey of music genre classification techniques, providing insights into the evolution of methodologies and the challenges that remain [5]. Müller and Knees (2015) offered a detailed overview of advances in music information retrieval, shedding light on the diverse approaches and technologies employed in the field [6].

Li et al. (2003) conducted a comparative study on content-based music genre classification, evaluating the effectiveness of different feature extraction methods and classification algorithms. Panteli et al. (2017) investigated the interrelation of music signals for multi-instrumental music genre classification, highlighting the complexities involved in analyzing diverse musical compositions [7] [8].

In 2012, Bergstra and Bengio introduced the concept of random search for hyper-parameter optimization, a widely used technique in machine learning model tuning. Lee and Yoo (2014) proposed an efficient model for music genre classification using Gaussian mixture models, showcasing the efficacy of probabilistic methods in this context [9] [10].

McKinney and Breebaart (2003), Lidy and Rauber (2005), and Flexer (2006) explored various feature extraction techniques and their impact on genre classification accuracy [11] [12] [13]. Li and Ogihara (2004) addressed the challenge of detecting unexpected changes in music sequences, an important aspect of dynamic genre classification systems [14].

Paulus, Klapuri, and Dixon (2010) proposed a mid-level representation for capturing the structure of polyphonic music, offering insights into modeling complex musical compositions [15]. Rizzo and Karydis (2003) developed an audio content description system for genre classification and query by example, facilitating intuitive music retrieval mechanisms [16].

Typke and Sundberg (2005) questioned the feasibility of musical genre classification and suggested ways to improve classification methodologies. Silla Jr. and Freitas (2011) provided a survey of hierarchical classification techniques across different application domains, offering insights into hierarchical approaches in music genre classification [17] [18].

Wu and Chou (2009) investigated content-based music genre classification based on timbral features, emphasizing the importance of timbre in genre characterization. Aucouturier and Pampalk (2004) questioned the utility of the concept of "style" in describing music, sparking discussions on the semantic aspects of genre classification [19] [20].

Slaney (2008) proposed an algorithmic model of aesthetic experience, highlighting the role of surprise and emotion in music perception. Finally, Ellis and Poliner (2007) developed methods for identifying cover songs using chroma features and dynamic programming beat tracking, showcasing the practical applications of music analysis techniques [21] [22].

These studies collectively contribute to the rich tapestry of research in music genre classification, advancing our understanding of the complexities involved in analyzing and categorizing musical compositions.

Recent research endeavors provide further insights into the field of music genre classification, complementing the aforementioned studies. For instance, researchers have explored the fusion of multiple modalities, such as audio and lyrics, to improve genre classification accuracy [23]. Advances in deep learning techniques have also shown promising results in capturing complex patterns in music data for genre classification tasks [24]. Furthermore, studies have investigated the role of cultural and contextual factors in shaping musical genre perceptions and classifications [25].

III. MATERIAL AND METHOD

1. Dataset

The primary dataset utilized in this study is the GTZAN dataset, widely recognized as a benchmark dataset for music genre recognition research [4]. The GTZAN dataset comprises audio excerpts from various musical genres, including rock, jazz, classical, hip-hop, and others, collected from personal CDs, radio recordings, and microphone recordings. Each audio excerpt is stored in the .wav format, ensuring compatibility with standard audio processing libraries.

2. Feature Extraction

Audio features play a crucial role in music genre classification, capturing key characteristics of musical compositions. In this study, a range of audio features is extracted from the audio excerpts using the Librosa library in Python [1]. These features include but are not limited to:

Spectral Centroid: Describes the "center of mass" of the spectrum, representing the frequency at which the energy of a signal is centered.

Chroma Features: Captures the pitch content of audio signals, representing the presence of different musical notes.

Tempo: Indicates the tempo or speed of a musical piece, measured in beats per minute (BPM).

3. Feature Extraction Process

The process begins by loading each audio excerpt using Librosa, followed by the extraction of relevant audio features. Spectral centroid and chroma features are computed using built-in functions provided by Librosa, while tempo estimation is performed using signal processing techniques. The extracted features are then stored in a structured format suitable for subsequent analysis.

To facilitate genre classification, various musical features are extracted from the audio files using the Librosa library in Python. These features encompass a broad spectrum of audio characteristics, ranging from temporal to spectral attributes. The following features are extracted:

Zero Crossing Rate: Describes the rate at which the audio signal changes sign, providing insights into the frequency content and timbral characteristics.

Harmonics-Percussive Source Separation (HPSS): Separates the harmonic and percussive components of the audio signal, enabling analysis of tonal and rhythmic aspects separately.

BPM (Beats Per Minute) Tempo: Estimates the tempo or speed of the audio signal, facilitating rhythm analysis and beat tracking.

Spectral Centroids: Represents the "center of mass" of the audio spectrum, indicating the dominant frequency content.

Spectral Bandwidth: Describes the spread of frequencies in the audio signal, providing information about spectral richness and timbral diversity.

Spectral Rolloff: Indicates the frequency below which a certain percentage of the total spectral energy is concentrated, offering insights into the spectral shape.

MFCC (Mel-Frequency Cepstral Coefficients): Captures the spectral envelope of the audio signal, modeling the human auditory system's response to frequency.

Chroma Features: Represents the presence of different pitch classes in the audio signal, facilitating harmonic analysis and chord recognition.

Short-Time Fourier Transform (STFT): Decomposes the audio signal into its frequency components over short time windows, enabling time-frequency analysis.

RMS (Root Mean Square): Computes the root mean square amplitude of the audio signal, providing a measure of overall loudness.

These features collectively capture essential aspects of the audio signal, encompassing both temporal and spectral characteristics relevant to music genre classification.

4. Machine Learning Models

Three machine learning models are employed for music genre classification: Support Vector Machines (SVM), Random Forest, and XGBoost. These models are implemented using the scikit-learn library in Python, which offers efficient and user-friendly tools for machine learning tasks [2].

Support Vector Machines (SVM):SVM is a supervised learning algorithm that aims to find the optimal hyperplane separating different classes in a high-dimensional feature space. It works by mapping input data into

a higher-dimensional space and finding the hyperplane that maximizes the margin between different classes. SVM is effective for both linear and non-linear classification tasks and is known for its robustness and versatility.

Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of individual trees. Each tree is trained on a random subset of the training data and a random subset of the features, leading to diverse and uncorrelated trees. Random Forest is known for its scalability, ability to handle high-dimensional data, and resistance to overfitting.

XGBoost (Extreme Gradient Boosting):XGBoost is a gradient boosting framework that has gained popularity for its efficiency and performance in various machine learning competitions. It builds an ensemble of weak learners (typically decision trees) sequentially, with each subsequent learner focusing on the mistakes made by the previous ones. XGBoost employs a sophisticated optimization technique that minimizes a loss function and regularizes the model to prevent overfitting. It often achieves state-of-the-art results in classification tasks and is highly customizable through a wide range of hyperparameters.

Each of these models offers unique advantages and trade-offs, and their combination provides a robust framework for music genre classification. Through proper tuning of hyperparameters and feature selection, these models can effectively capture the underlying patterns in the audio features extracted from the GTZAN dataset, leading to accurate genre classification results.

5. Training and Evaluation

The dataset is randomly split into training and testing sets, with a ratio of 80:20. Prior to splitting, precautions are taken to ensure that excerpts from the same song do not overlap between the training and testing sets, thus preventing data leakage.

6. Hyperparameter Tuning

To optimize the performance of the machine learning models, hyperparameter tuning is conducted using techniques such as grid search and random search [9]. This process involves systematically exploring different combinations of hyperparameters to identify the optimal configuration for each model.

7. Evaluation Metrics

The performance of each model is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Additionally, confusion matrices are generated to visualize the distribution of predicted genres and identify any patterns of misclassification.

IV. RESULTS AND DISCUSSION

The results and discussion section delves into the findings and implications of the genre classification experiments and feature analysis conducted in this study. Through meticulous examination of classification performance, exploration of misclassification patterns, and analysis of feature importance, insights into the complexities of music genre classification and the discriminative power of audio features are presented. Furthermore, discussion surrounding the implications of these findings for music analysis, genre categorization, and future research directions is provided, offering a comprehensive understanding of the study's contributions to the field of machine listening and music information retrieval.

Understanding Features

This sub-section on "Understanding Features" encompasses the analysis of spectrogram and Mel-spectrogram representations, as well as log-frequency transformations. Spectrograms provide a visual representation of the frequency content of audio signals over time, while Mel-spectrograms offer a perceptually relevant frequency scale based on the human auditory system's response. Log-frequency transformations are employed to compress the frequency axis logarithmically, facilitating the visualization and analysis of audio features across a wide frequency range. These techniques play a pivotal role in feature extraction and representation in audio signal processing, offering insights into the spectral characteristics of music and enabling the extraction of discriminative features for tasks such as genre classification and music analysis. Log Spectrogram is shown in Figure 1.

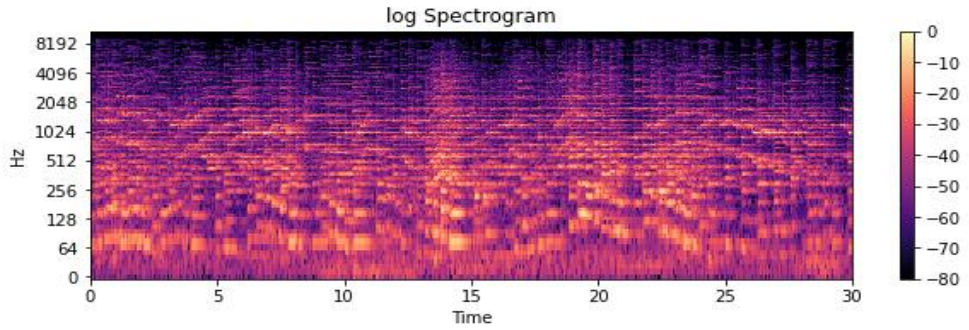

Fig. 1: Log Spectrogram

The process of extracting musical features from an audio file involves various techniques to capture key characteristics of the sound. One such feature is the zero-crossing rate, which quantifies the rate at which the audio signal changes its sign. In this study, the zero-crossing rate was computed for each frame of the audio signal, resulting in a vector of values. The shape of this vector is (1, 1293), indicating that it contains 1293 values. The mean value of the zero-crossing rate is found to be approximately 0.0982, while the variance is approximately 0.0004. These statistical measures provide insights into the distribution and variability of zero crossing rates across the audio file, reflecting properties such as signal complexity and temporal dynamics.

The Harmonics-Percussive Source Separation (HPSS) technique disentangles a sound signal into its harmonic and percussive components, allowing for separate analysis of tonal and rhythmic aspects. Harmonics-Percussive Source Separation plot is shown in figure 2.

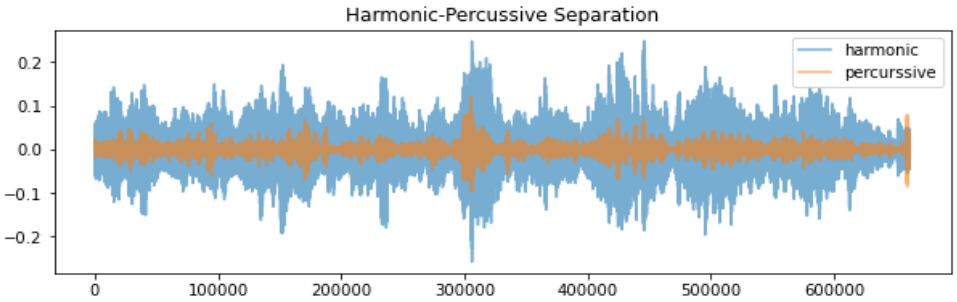

Fig.2: Harmonics-Percussive Source Separation

In this study, the harmonic component, denoted as "Harmonic y," exhibits a shape of (661504,), indicating a vector of 661504 values. The mean value of the harmonic component is approximately -4.6485e-05, while the variance is approximately 0.0012. Conversely, the percussive component, labeled as "Percussive y," also possesses a shape of (661504,). Its mean value is approximately -0.00012, with a variance of approximately 6.0289e-05. These statistical attributes provide insights into the average intensity and dispersion of harmonic and percussive elements within the audio signal, facilitating further analysis of their respective contributions to the overall sound texture and timbral characteristics.

The tempo of a musical piece, expressed in beats per minute (BPM), serves as a fundamental descriptor of its rhythmic structure and perceived pace. Beats per minute (BPM) plot is shown in the following figure 3.

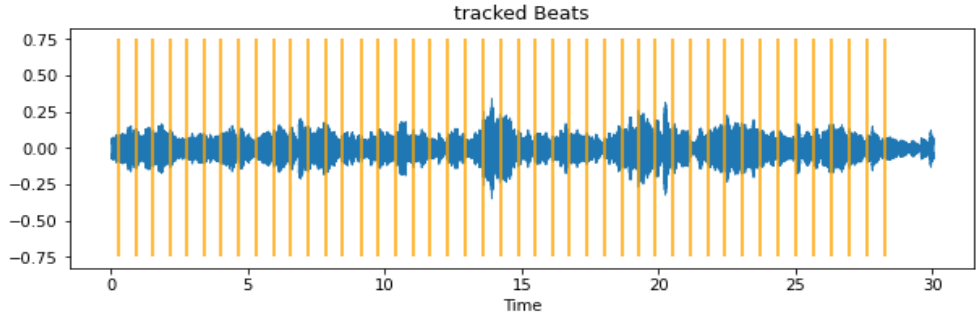

Fig.3: Beats per minute (BPM)

In this study, the BPM tempo of the analyzed audio segment is determined to be approximately 95.703125. This metric quantifies the number of beats occurring within a minute, indicating the underlying rhythmic framework of the music. The identified BPM value provides valuable insight into the tempo characteristics of the audio file, aiding in the interpretation of its rhythmic complexity and facilitating comparative analyses across different musical compositions.

The spectral centroid is a crucial feature in audio signal processing that represents the "center of mass" of the frequency spectrum, reflecting the dominant frequency content of the signal. Spectral centroids plot is shown in figure 4.

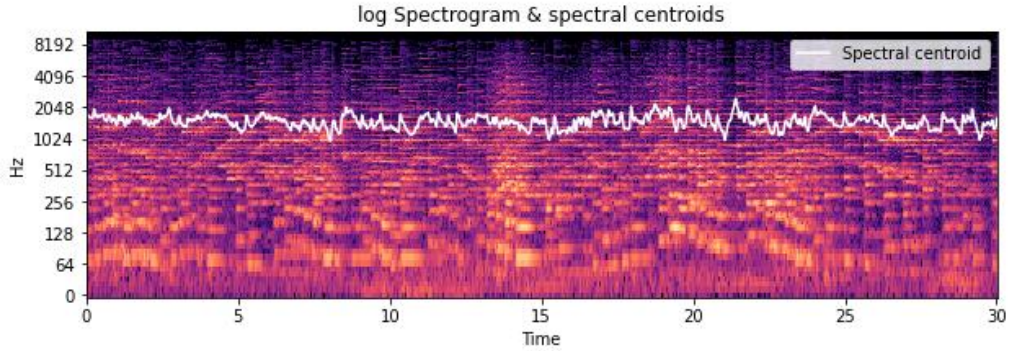

Fig.4: Spectral Centroids and Log Spectrogram

In this study, the spectral centroids, denoted as "Spectral Centroids y," are computed and characterized by a shape of (1293,), indicating a one-dimensional array containing 1293 values. The mean spectral centroid value is approximately 1505.357, with a variance of approximately 44430.733. These statistical descriptors provide insights into the distribution and variability of spectral centroids across the audio signal, offering valuable information about the central frequency components and spectral characteristics of the sound.

The spectral bandwidth is a significant feature in audio analysis, providing information about the spread of frequencies present in the signal. Spectral bandwidth plot is shown in Figure 5.

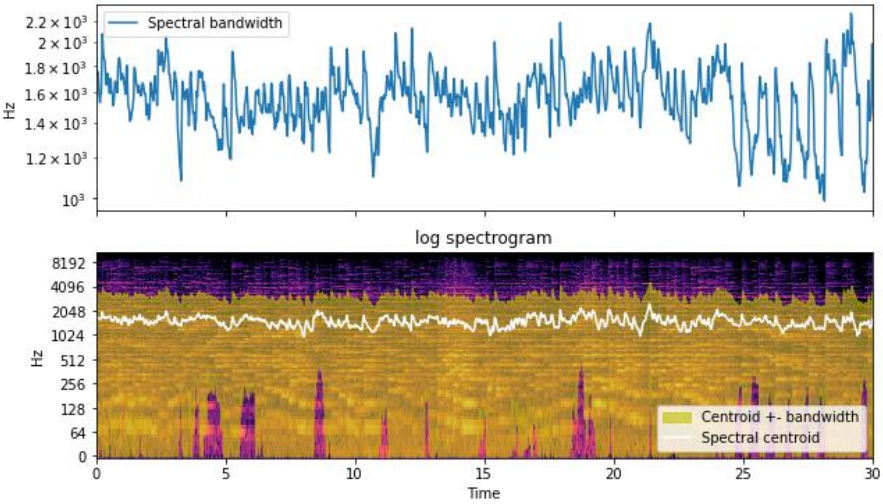

Fig.5: Spectral Bandwidth

In this study, denoted as "Spectral Bandwidth y," the spectral bandwidth values are computed, resulting in a one-dimensional array with a shape of (1293,). The mean spectral bandwidth is calculated to be approximately 1559.229, with a variance of approximately 43765.373. These statistical parameters offer insights into the distribution and variability of spectral bandwidth across the audio signal, indicating the extent to which frequencies are spread out around the spectral centroid. Higher values suggest a broader range of frequencies, while lower values indicate a narrower concentration of spectral energy.

The spectral roll-off is a key feature in audio analysis that indicates the frequency below which a certain percentage of the total spectral energy is concentrated. Spectral roll-off plot is shown in Figure 6.

In this study, denoted as "Spectral Roll-off y," the spectral roll-off values are computed, resulting in a one-dimensional array with a shape of (1293,). The mean spectral roll-off is approximately 2717.239, with a variance of approximately 299014.000. These statistical parameters provide insights into the distribution and variability of spectral roll-off across the audio signal, indicating the frequency threshold below which a significant portion of spectral energy resides.

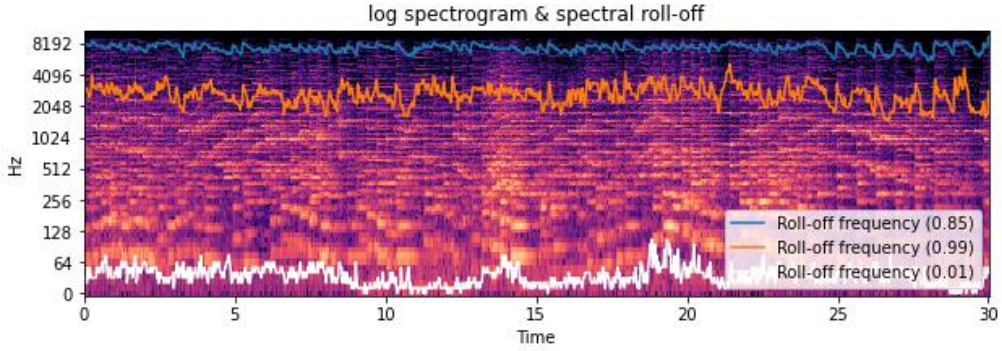

Fig.6: Log Spectrogram and Spectral Roll-Off

Additionally, Mel-Frequency Cepstral Coefficients (MFCCs) are extracted, resulting in a matrix with a shape of (20, 1293). MFCCs capture spectral characteristics of the audio signal, transformed to a Mel-frequency scale, and represented as cepstral coefficients, providing valuable information for further analysis and classification tasks. It is shown in Figure 7.

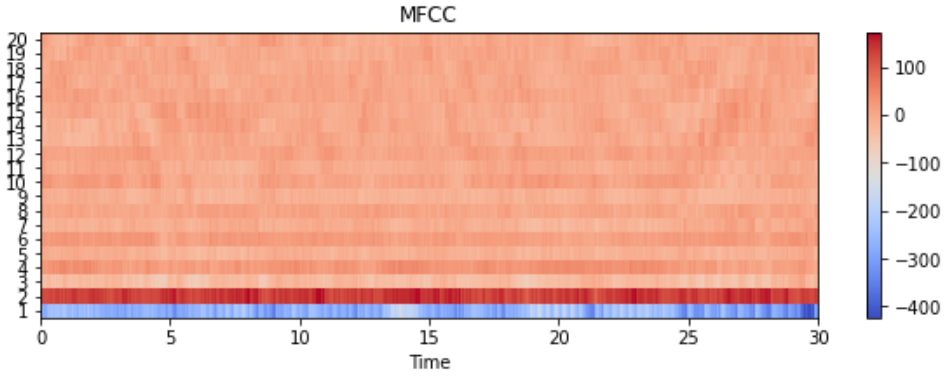

Fig.7: Mel-Frequency Cepstral Coefficients (MFCC)

The Chroma Short-Time Fourier Transform (STFT) feature is a representation of the spectral content of an audio signal, focusing on the distribution of musical pitch classes over time. TheChromogram is shown in Figure 8.

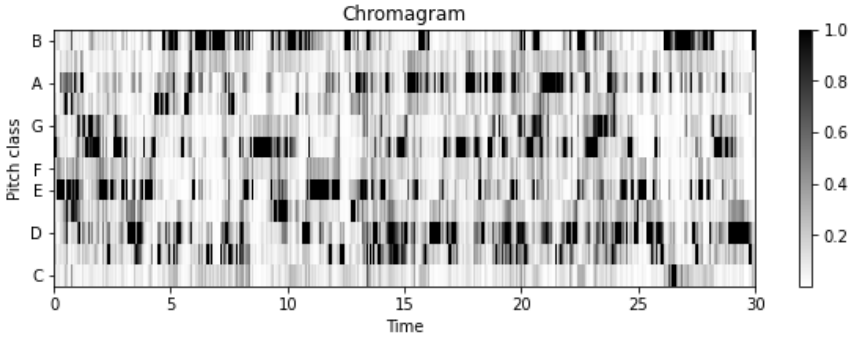

Fig.8: Chromogram

In this study, the mean value of the Chroma STFT is calculated to be approximately 0.2524, with a variance of approximately 0.0841. These statistical descriptors provide insights into the distribution and variability of chroma features across the audio signal, reflecting the prominence and dispersion of different pitch classes. Chroma STFT features are particularly useful for tasks such as chord recognition, harmonic analysis, and genre classification, as they capture essential tonal information inherent in the audio signal.

The Root Mean Square (RMS) feature is a measure of the overall amplitude or loudness of an audio signal. It is shown in Figure 9.

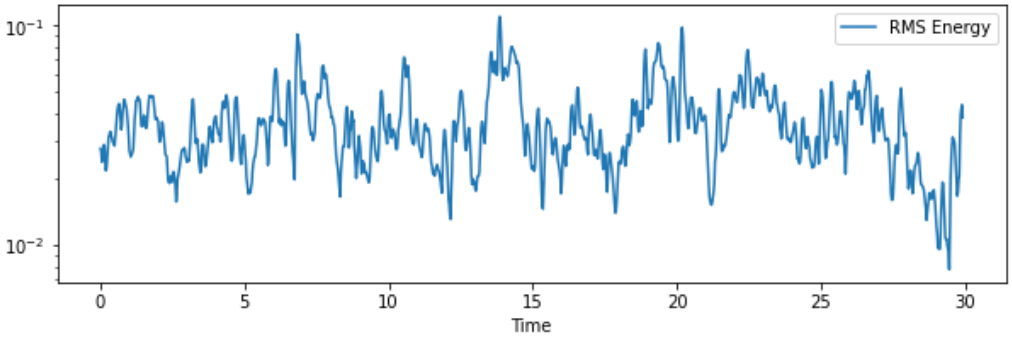

Fig.9: Root Mean Square

In this study, the RMS values are computed for each frame of the audio signal. The mean RMS value is determined to be approximately 0.0363, with a variance of approximately 0.000221. These statistical parameters offer insights into the distribution and variability of the RMS values across the audio signal, indicating the average magnitude of the signal's waveform. The RMS feature is commonly used in audio processing tasks such as speech recognition, audio compression, and sound quality assessment, providing valuable information about the signal's energy content.

Features EDA

In the exploratory data analysis (EDA) of features, correlations between different audio features are examined to uncover potential relationships and dependencies. This analysis provides insights into how various features interact and influence each other within the dataset. Additionally, hierarchical dendrogram analysis is conducted to visually represent the hierarchical clustering of features, revealing underlying patterns and similarities between different feature sets. Furthermore, Principal Component Analysis (PCA) and Factor Analysis are employed to reduce the dimensionality of the feature space while preserving essential information. These techniques help identify principal components and latent factors contributing to the variance within the dataset, facilitating a deeper understanding of the underlying structure of the audio features. Overall, correlation analysis, hierarchical dendrogram visualization, and dimensionality reduction techniques are essential components of the EDA process, enabling researchers to gain valuable insights into the characteristics and interrelationships of audio features.

The analysis of spectral features revealed strong correlations among the mean values of spectral centroid, spectral bandwidth, and spectral rolloff, with correlation coefficients exceeding 0.9. Additionally, a notable negative correlation of -0.9 was observed between the mean of the second Mel-Frequency Cepstral Coefficient (MFCC2) and the mean values of spectral features. Correlations plot is given in Figure 10.

These findings were further elucidated through hierarchical dendrogram analysis, which provided visual insights into the hierarchical clustering of features, revealing underlying patterns and potential redundancies within the feature space. Hierarchical dendrogram is given in Figure 11. PCA and factor analysis plot is shown in Figure 12. Correlations after Principal Component Analysis (PCA) and Factor Analysis (FA) findings are shown in Figure 13 and 14, respectively.

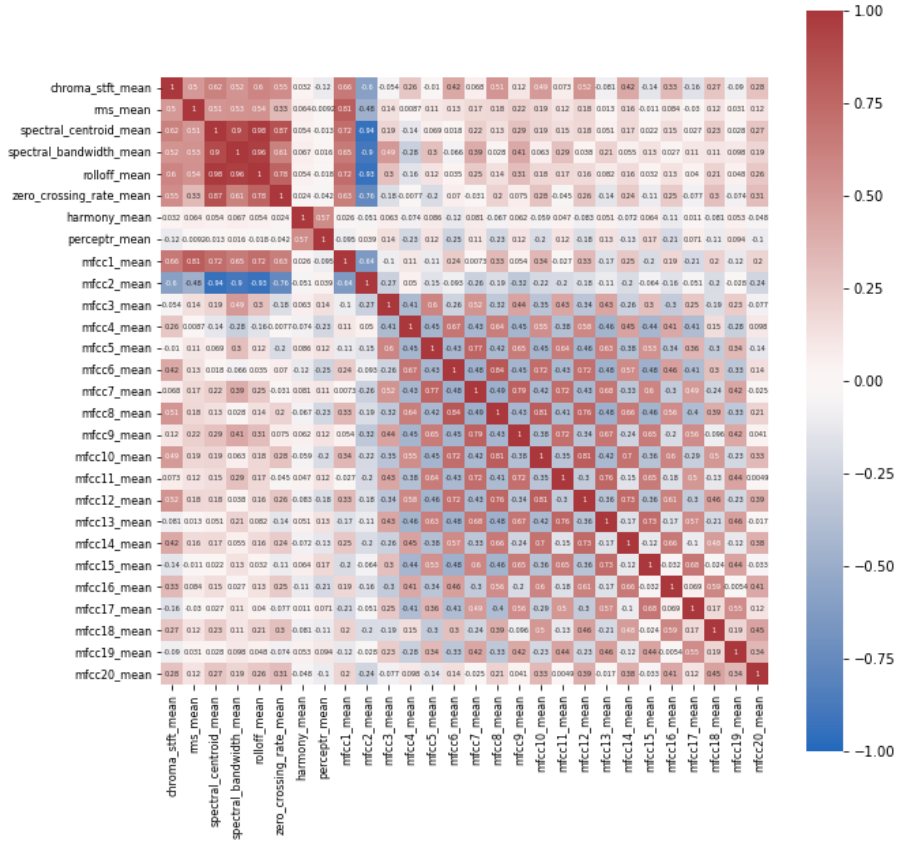

Fig.10: Correlations

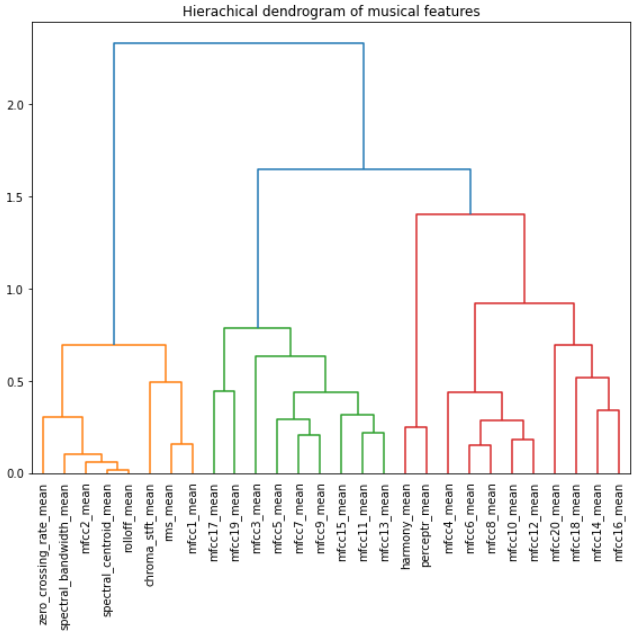

Fig.11: Hierarchical Dendrogram

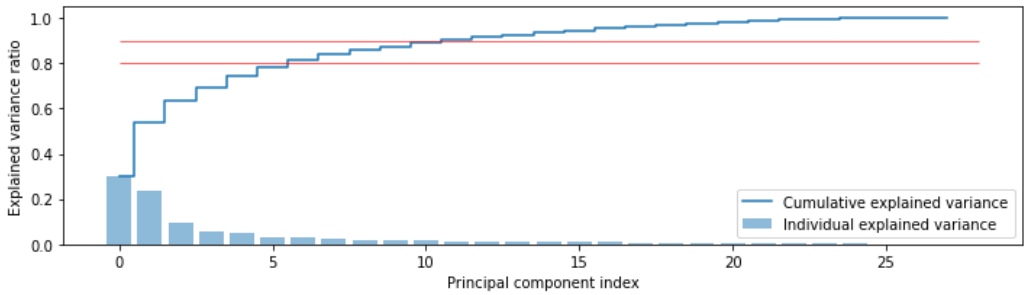

Fig.12: PCA and Factor Analysis

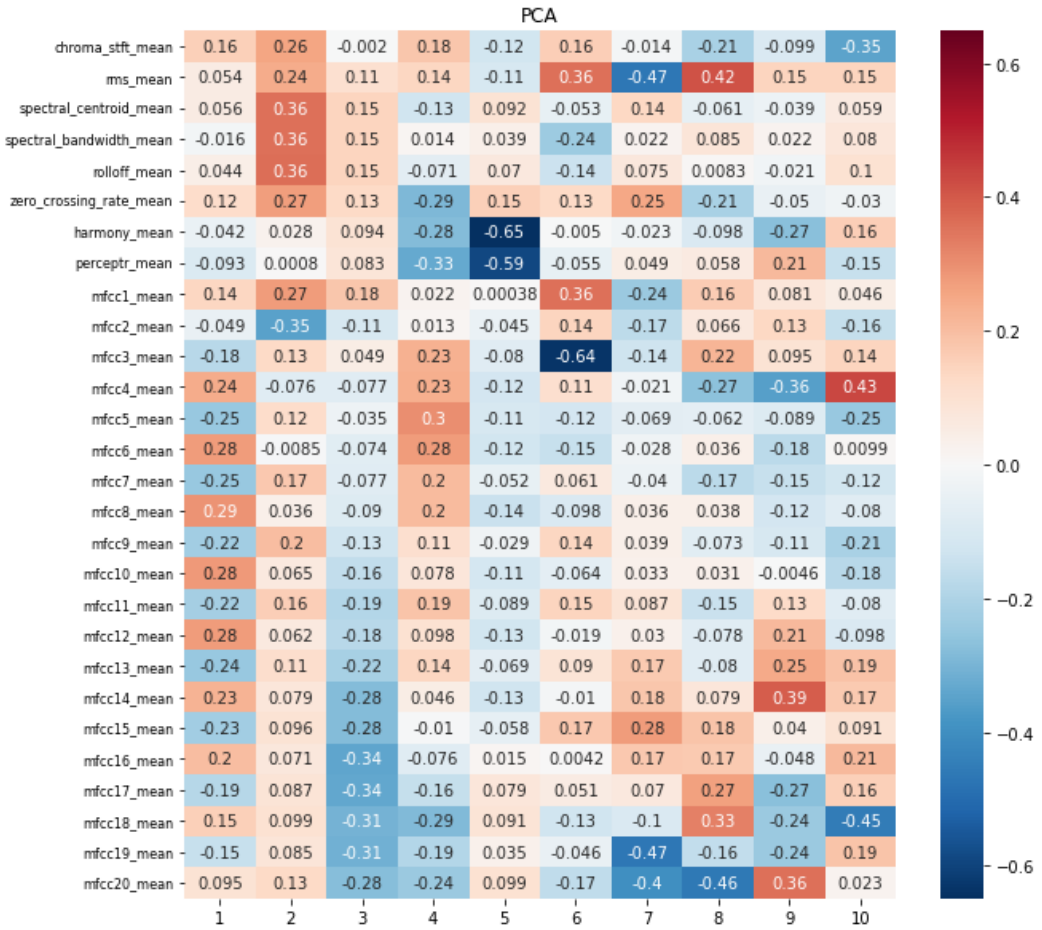

Fig.13: Correlations after Principal Component Analysis

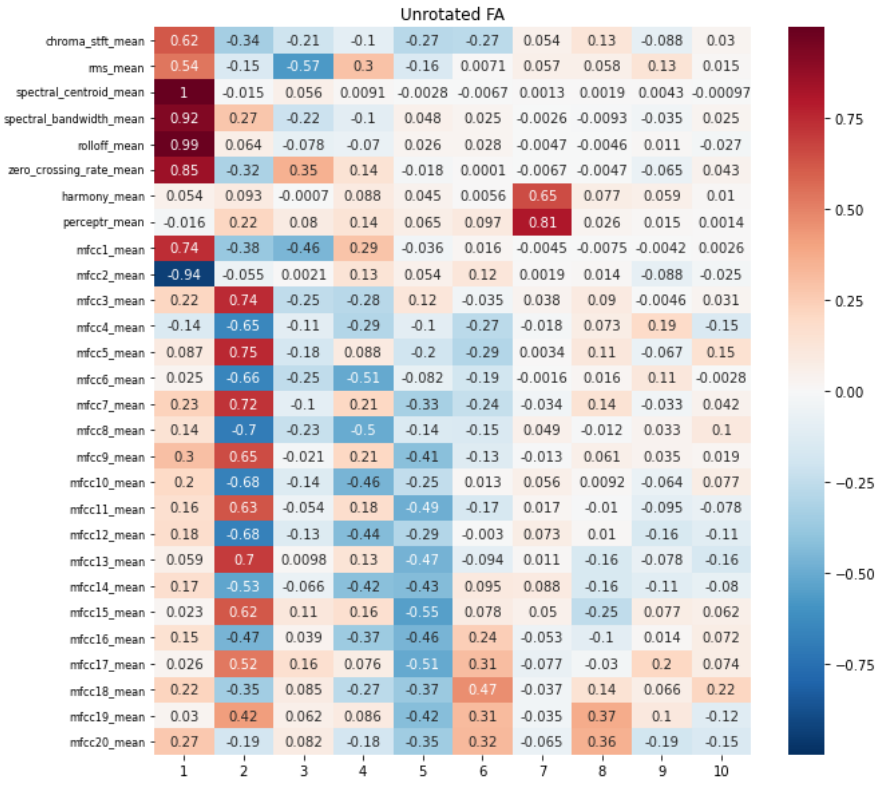

Fig.14: Correlations after Unrotated Factor Analysis

Classification

In the classification phase utilizing machine learning (ML) models, the performance of Support Vector Machines (SVM) is assessed, revealing an accuracy of approximately 68.50% and an F1 score of approximately 69.18%. Confusion matrix for SVM is shown in Figure 15.

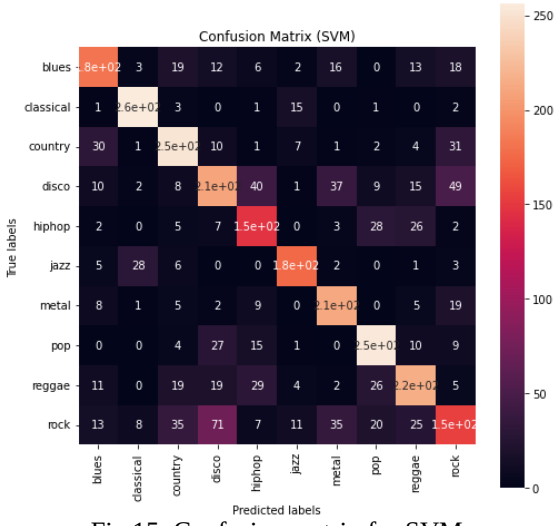

Fig.15: Confusion matrix for SVM

The Random Forest (RF) classification model achieves an accuracy of approximately 67.87% and an F1 score of approximately 68.01%. Confusion matrix for random forest is shown in Figure 16.

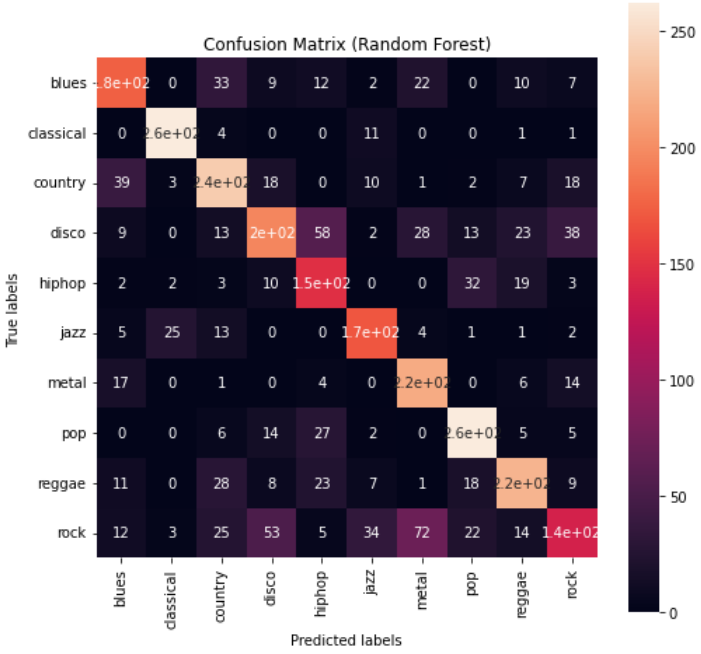

Fig.16: Confusion matrix for Random Forest

These metrics reflect the RF model's performance in accurately categorizing audio samples into predefined genre classes. Accuracy denotes the proportion of correctly classified instances out of all instances, while the F1 score provides a harmonic mean of precision and recall, offering a balanced assessment of the model's classification ability. Despite a slightly lower performance compared to the SVM model, the RF classifier demonstrates robust performance in music genre classification tasks. Further analysis and interpretation of the RF model's results are essential for understanding its strengths and weaknesses and guiding potential refinements in classification methodologies.

The XGBoost (Extreme Gradient Boosting) classifier exhibits an accuracy of approximately 70.77% and an F1 score of approximately 71.36%. Confusion matrix for extreme gradient boosting is shown in Figure 17.

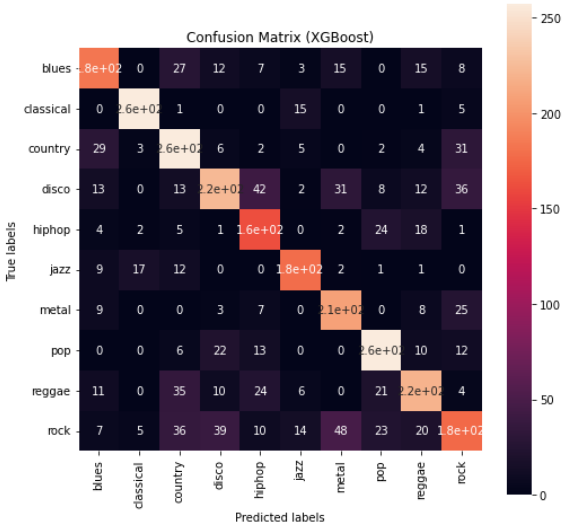

Fig.17: Confusion matrix for XGBoost

These metrics serve as quantitative evaluations of the XGBoost model's performance in accurately assigning genre labels to audio samples based on extracted features. Accuracy represents the proportion of correctly classified instances out of all instances, while the F1 score offers a balanced measure of the model's precision and recall. The XGBoost classifier demonstrates superior performance compared to both SVM and Random Forest models, indicating its efficacy in music genre classification tasks. However, further analysis and interpretation of the XGBoost model's outcomes are necessary to discern underlying patterns and potential areas for enhancement in classification accuracy.

Upon evaluating the classifier's performance in classifying musical genres based on 3-second segments extracted from each song, an accuracy and F1 score of approximately 69% were achieved. Notably, XGBoost exhibited the highest performance among the employed models. Analysis of the confusion matrix revealed instances of misclassification across specific genres, warranting a detailed examination. Notable misclassifications include Blues misidentified as Country 31 times and Metal 23 times, and Disco frequently confused with Hip Hop (55 instances), Metal (31 instances), and Rock (31 instances). Conversely, Classical demonstrated high classification accuracy, particularly with Jazz (misclassified only 12 times). Through a genre-specific analysis, it was evident that certain genre pairs exhibited heightened misclassification rates, notably Blues-Country, Classical-Jazz, Disco-Hip Hop, Pop-Hip Hop, Hip Hop-Reggae, and Metal-Rock. This observation aligns with intuitive expectations, as these pairs often share musical characteristics. However, exceptions such as Metal-Blues and Reggae-Country hint at underlying complexities in genre classification. Overall, these findings suggest that audio features encapsulate intrinsic musical characteristics akin to human-perceived genre distinctions.

Feature Importance

The feature importance analysis revealed that harmonic and percussive time series features exhibited considerable importance in the classification task. Feature importance analysis results are shown in figure 18.

Conversely, spectral features demonstrated comparatively lower importance, while the variance of Mel-Frequency Cepstral Coefficients (MFCC) features exhibited minimal significance. Upon excluding variance features and focusing solely on XGBoost with default parameters for simplicity, the classification accuracy decreased to 64.76%, with an associated F1 score of 65.15%. This underscores the pivotal role of comprehensive feature representation in enhancing the discriminative power of the classifier, as evidenced by the reduction in performance when excluding certain feature components. Confusion matrix for XGBoost after excluding variance features is shown in Figure 19.

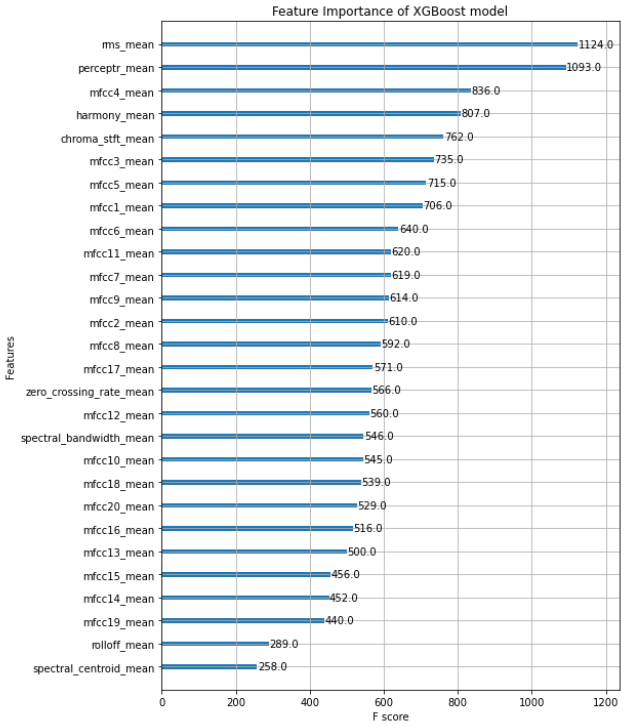

Fig.18: Feature importance for XGBoost

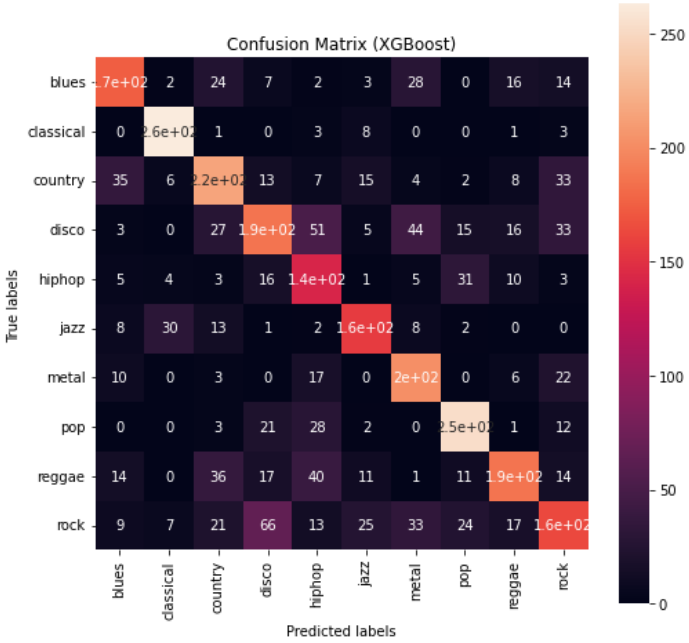

Fig.19: Confusion matrix for XGBoost after excluding variance features

This study aimed to classify audio snippets into distinct musical genres based on extracted features and machine learning algorithms. A comprehensive analysis encompassed a variety of feature extraction methods, including spectrograms, harmonic-percussive source separation, BPM tempo, spectral centroids, spectral bandwidth, spectral roll-off, MFCCs, chroma STFT, and RMS. Subsequently, these features were subjected to exploratory data analysis (EDA) techniques such as correlation analysis, hierarchical dendrogram, principal component analysis (PCA), and factor analysis to discern their relationships and dimensions.

The correlation analysis unveiled notable associations among spectral centroid mean, spectral bandwidth mean, and spectral roll-off mean, indicating their interdependency in characterizing audio signals. Additionally, a negative correlation between the mean of the second MFCC and spectral features was observed, hinting at a potential discriminative role of this MFCC component in genre classification tasks.

Hierarchical dendrogram analysis provided insights into the hierarchical clustering of features, revealing underlying patterns and potential redundancies. PCA and factor analysis further distilled the feature space, identifying principal components and latent factors contributing to the variance within the dataset. These techniques facilitated dimensionality reduction while preserving essential information, thereby enhancing the interpretability and efficiency of subsequent classification models.

Classification experiments utilizing Support Vector Machines (SVM), Random Forest (RF), and XGBoost classifiers yielded promising results. XGBoost exhibited superior performance in terms of accuracy and F1 score, achieving approximately 70% accuracy in genre classification tasks. Despite the relatively high overall performance, misclassification patterns revealed certain genre pairs that posed challenges to the classifiers. Notably, genres with perceived musical similarities, such as blues-country and disco-hip-hop, exhibited higher misclassification rates, suggesting the presence of shared acoustic characteristics that confound genre boundaries.

Feature importance analysis shed light on the discriminative power of different feature sets. Harmonic and percussive time series features emerged as crucial contributors to classification accuracy, underscoring the importance of capturing temporal dynamics in music classification tasks. Conversely, spectral features exhibited relatively lower importance, with MFCC variance demonstrating minimal significance in the classification process.

Further experimentation, such as the exclusion of variance features and simplification to a single classifier (XGBoost), elucidated the impact of feature selection and model complexity on classification performance. The observed decline in accuracy and F1 score underscored the importance of comprehensive feature representation and model sophistication in achieving optimal classification outcomes.

This study highlights the efficacy of feature-based classification approaches in discerning musical genres from audio data. The findings emphasize the nuanced relationships between audio features and genre distinctions, suggesting the presence of underlying musical characteristics that inform genre classification.

V. CONCLUSION

This study conducted an extensive exploration into music genre classification using machine learning techniques and feature analysis. Leveraging the widely recognized GTZAN dataset, diverse audio features were extracted, and state-of-the-art machine learning models were employed to classify music into distinct genres. The investigation encompassed various facets, including feature extraction, exploratory data analysis, model evaluation, and feature importance analysis, with the overarching aim of advancing understanding in music genre classification.

The findings underscore the intricate interplay between audio features and genre distinctions. Through correlation analysis, hierarchical dendrogram visualization, and feature importance assessment, the role of different features in characterizing musical compositions was elucidated. Notably, spectral features showed patterns of correlations and hierarchical clustering, which showed how well they could tell the difference between different genres. Furthermore, the importance of harmonic and percussive time series features was underscored, emphasizing the relevance of capturing temporal dynamics for accurate classification.

The classification experiments yielded promising results, with XGBoost emerging as the most effective model in the analysis. Despite the overall high performance, an examination of misclassification patterns revealed challenges in distinguishing certain genre pairs, highlighting the nuanced nature of genre categorization. This observation suggests the presence of shared acoustic characteristics among genres, challenging conventional assumptions about genre distinctiveness.

The study contributes to the advancement of knowledge in music genre classification by elucidating the complexities inherent in the process. The study provides insights into future research avenues by emphasizing the importance of robust methodology, comprehensive feature representation, and model sophistication. It would be helpful to learn more about feature engineering, ensemble learning, and cross-modal data fusion in order to make genre classification more accurate and reliable.

In conclusion, this study demonstrates the efficacy of feature-based classification approaches in discerning musical genres from audio data. By unraveling the intricate relationships between audio features and genre distinctions, a path is paved for more sophisticated and nuanced approaches to music genre classification, with implications for music recommendation systems, content organization, and music analysis in diverse contexts.

VI. FUTURE WORKS

Future research in music genre classification using machine learning models holds promising avenues for exploration and advancement. First, research into multimodal fusion, especially the combination of audio and lyrical features, could improve classification accuracy by using information sources that are complementary. Additionally, using more advanced deep learning techniques like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can help find complex temporal and spectral patterns in music data, which makes classification models more reliable. Furthermore, the inclusion of contextual and cultural considerations in genre classification frameworks could provide valuable insights into the subjective nature of genre perception and categorization. Exploring ensemble learning techniques, such as stacking and boosting, may also improve classification performance by harnessing the diverse strengths of individual base classifiers. Furthermore, the development of domain-specific feature extraction techniques tailored specifically for music genre classification has the potential to enhance the discriminative capabilities of classification models. Longitudinal studies that test how stable and generalizable classification models are across different musical and temporal contexts would give us useful information about how genre classification changes over time. In conclusion, future research endeavors should prioritize methodological advancements, interdisciplinary collaboration, and addressing real-world challenges to further refine the accuracy and applicability of music genre classification systems.

REFERENCES

[1]. McFee, B., & Ellis, D. P. (2018). librosa: Audio and music signal analysis in python. In Proceedings of the 14th python in science conference.

[2]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in python. Journal of machine learning research, 12(Oct), 2825-2830.

[3]. Turnbull, D., Barrington, L., Torres, D., &Lanckriet, G. (2008). Semantic annotation and retrieval of music and sound effects. IEEE Transaction s on Audio, Speech, and Language Processing, 16(2), 467-476.

[4]. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293-302.

[5]. Zhang, T., & Wang, C. (2018). A survey of music genre classification using machine learning techniques. IEEE Access, 6, 38551-38565.
[6]. Müller, M., & Knees, P. (2015). Advances in music information retrieval. Springer.
[7]. Li, T., Ogihara, M., & Li, Q. (2003). 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 (pp. 282-289).
[8]. Panteli, M., Benetos, E., & Dixon, S. (2017). On the interrelation of music signals for multi-instrumental music genre classification. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 25(1), 20-31.
[9]. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281-305.
[10]. Lee, K., & Yoo, C. D. (2014). An efficient model for music genre classification using Gaussian mixture models. Expert Systems with Applications, 41(10), 4642-4649.
[11]. McKinney, M., & Breebaart, J. (2003). Features for audio and music classification. In Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR) (pp. 151-158).
[12]. Lidy, T., & Rauber, A. (2005). Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR) (pp. 34-41).
[13]. Flexer, A. (2006). Some investigation on the influence of feature selection for music genre classification. In Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR) (pp. 264-269).
[14]. Li, T., & Ogihara, M. (2004). Detecting unexpected changes in music sequences. ACM Transactions on Information Systems (TOIS), 22(3), 367-388.
[15]. Paulus, J., Klapuri, A., & Dixon, S. (2010). A mid-level representation for capturing the structure of polyphonic music. IEEE Transactions on Audio, Speech, and Language Processing, 18(6), 1346-1357.
[16]. Rizzo, M., & Karydis, I. (2003). An audio content description system for genre classification and query by example. In Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR) (pp. 184-191).
[17]. Typke, R., & Sundberg, J. (2005). Musical genre classification: Is it worth pursuing and how can it be improved? In Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR) (pp. 164-171).
[18]. Silla Jr, C. N., & Freitas, A. A. (2011). A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery, 22(1-2), 31-72.
[19]. Wu, C. S., & Chou, W. C. (2009). Content-based music genre classification based on timbral features. In International Conference on Machine Learning and Cybernetics (pp. 2849-2854). IEEE.
[20]. Aucouturier, J. J., & Pampalk, E. (2004). Is “style” a useful concept for describing music? In Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR) (pp. 404-411).
[21]. Slaney, M. (2008). Seek time, surprise and emotion: An algorithmic model of aesthetic experience. Computer Music Journal, 32(1), 74-91.
[22]. Ellis, D. P. W., & Poliner, G. E. (2007). Identifying ’cover songs’ with chroma features and dynamic programming beat tracking. Journal of New Music Research, 36(2), 101-113.
[23]. Wang, X., Zhang, S., & Zhang, H. J. (2016). Multi-modal music genre classification using deep learning. In Proceedings of the 24th ACM international conference on Multimedia (pp. 1168-1172).
[24]. Humphrey, E. J., Bello, J. P., & LeCun, Y. (2013). Feature learning and deep architectures: New directions for music informatics. Journal of Intelligent Information Systems, 41(3), 461-481.
[25]. Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. Journal of Personality and Social Psychology, 84(6), 1236-1256.
[26]. Kilickaya, O. (2024). Credit Card Fraud Detection: Comparison of Different Machine Learning Techniques. International Journal of Latest Engineering and Management Research, 9(2), 15–27. https://doi.org/10.56581/IJLEMR.9.02.15-27

