International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 12 Issue 6, June-2023, Impact Factor: 7.957

A Comprehensive Survey of Music Genre Classification Using Audio Files

Santosh Shirol[1], Kathiresan R S[2]

[1]Student, REVA Academy for Corporate Excellence (RACE), REVA University Bengaluru, Karnataka, India
[2]Student, REVA Academy for Corporate Excellence (RACE), REVA University Bengaluru, Karnataka, India

ABSTRACT

This survey extensively studies music genre classification, a critical task in music information retrieval, to automatically categorize audio recordings into various genres. It provides a comprehensive review of approaches, methodologies, and recent advancements in genre classification from audio data. Scholars and practitioners in the field will find this study to be a valuable resource as it covers various aspects of the discipline, including feature extraction, classification methods, dataset exploration, evaluation metrics, and recent developments. The survey aims to enhance the understanding of music genre classification and foster further research and progress in the field by critically evaluating state-of-the-art techniques discussed in research papers, discussing their strengths and limitations, and providing a comprehensive overview of the field.

Keywords: Music genre classification, feature extraction, classification techniques, Machine Learning, Deep Learning.

INTRODUCTION

In the realm of music information retrieval, a fundamental and critical task is music genre classification, which entails the automated categorization of audio data into diverse genres. It is essential to many applications, including music indexing, content organization, and music recommendation systems. The requirement for precise and effective genre categorization methods has become more critical as digital music collections have become more widely accessible. This survey study provides a thorough analysis of the methodology, strategies, and most recent developments in the subject of music genre categorization. It serves as an invaluable resource for people interested in researching or making contributions to this field of study by aiming to provide scholars and practitioners with a thorough grasp of the subject. The paper covers various aspects of music genre classification, including feature extraction, classification algorithms, dataset exploration, evaluation metrics, and recent developments. Methods for extracting important characteristics from audio signals are critical for the extraction of discriminative features for genre categorization. To evaluate their appropriateness and effectiveness in genre classification tasks, classification algorithms spanning from classic machine learning techniques to deep learning approaches are investigated. Finally, the purpose of this survey is to help scholars and practitioners advance in music genre categorization by providing them with essential knowledge and expertise on the issue. This resource is intended to aid in developing more accurate, efficient, and resilient music genre categorization systems, resulting in improved music recommendation systems, personalized music experiences, and improved music organization in various applications.

DATASET

Several publicly accessible datasets have considerably contributed to breakthroughs in music genre categorization research. These datasets have provided researchers with standardized and diverse collections of music samples, enabling the development and evaluation of machine-learning models and algorithms. Here are some notable datasets commonly used for music genre classification research.

GTZAN datasets

The GTZAN dataset is a widely recognized benchmark for music genre classification. It comprises 1,000 audio clips, each lasting 30 seconds, evenly distributed across ten different genres like blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Each audio clip in the GTZAN dataset is stored in the WAV format and has a sampling rate of 22.05 kHz. The dataset maintains a balanced distribution of samples across genres, ensuring fair representation for each category. This balance is crucial for evaluating the performance of classification algorithms and assessing their generalization capabilities.

Million Song Dataset
As the name suggests, the Million Song Dataset has one million popular music tracks. This monumental collection, developed in collaboration with The Echo Nest and various institutions, has proven to be a highly valuable resource for research in music information retrieval and genre classification.With detailed information on artists, song duration, tempo, pitch, and timbre, this dataset has enabled researchers to delve into complex music analysis and develop sophisticated algorithms. Its vast scale and diverse range of tracks have contributed significantly to the advancement of machine-learning techniques in the realm of music, fostering breakthroughs in automatic genre classification.

Free Music Archive (FMA) dataset
The FMA dataset is a vast collection of over 100,000 freely available music tracks, making it a valuable resource for music genre classification research. The dataset offers a wide range of genres, with tracks categorized into 16 distinct categories and numerous sub-genres. Along with the audio files, the FMA dataset provides extensive metadata, including artist information, album details, track titles, and genre labels. This rich combination of audio samples and accompanying metadata has enabled researchers to explore various approaches in music genre classification and has contributed significantly to advancements in the field.

Music4All dataset
The Music4All dataset comprises a collection of 109,269 songs gathered from the MusicMatch application. The dataset encompasses a wealth of information, including metadata, tags, genre details, 30-second audio clips, lyrics, and more [11].

Table 1: Dataset Summary

Dataset Name	Description	Number of Audio Clips	Number of Genres
GTZAN dataset	Widely used benchmark dataset for music genre classification	1,000	10
Million Song Dataset	Collection of audio features and metadata for a million songs	1,00,000	NA
FMA dataset	Large collection of freely available music tracks	1,00,000+	16
Music4All	Large collection of music clips gathered from the musicMatch application	109,269	Multiple

These datasets have significantly contributed to the advancement of music genre classification research by providing researchers with large-scale, diverse, and labelled collections of music samples. By leveraging these datasets, researchers have been able to develop and evaluate machine learning models, leading to improvements in music analysis, recommendation systems, and other related applications.

FEATURE EXTRACTION TECHNIQUES

Feature extraction techniques play a crucial role in audio signal processing by extracting pertinent information that finds applications in diverse domains such as speech recognition, music analysis, and sound classification. This content emphasizes a range of frequently employed feature extraction techniques.

Time-domain Features
Time-domain features analyze an audio signal's amplitude variations over time, providing insights into its temporal properties. Commonly used features include the zero-crossing rate, which counts sign changes in the signal, revealing frequency content or waveform changes. The root mean square (RMS) energy calculates the overall energy by averaging squared amplitudes, aiding in detecting sound presence and intensity. The temporal centroid represents the energy distribution's centre of mass along the time axis, indicating the signal's temporal balance and aiding in differentiating transient and sustained sounds.

Frequency-domain Features
Frequency-domain features capture the spectral characteristics of an audio signal. They provide information about the distribution of frequencies present in the signal. Notable examples of Frequency-domain features include the spectral centroid (representing the spectrum's centre of mass), spectral roll-off (showing the frequency below which a specific portion of total energy resides), spectral flux (measuring the spectrum's temporal changes), and spectral flatness (evaluating the signal's tonal or noisy characteristics).

Time-Frequency Features
Time-frequency features capture the variations in the frequency content of an audio signal as it evolves over time, incorporating information from both the time and frequency domains. Several illustrative instances of time-frequency features include the short-time Fourier transform (STFT), which examines the signal's frequency content at different time

intervals, the constant-Q transform (CQT), which offers a logarithmically spaced frequency representation, and the wavelet transform, which analyzes the signal at multiple scales.

Timbral and Spectral Features

Timbral and spectral features are important in understanding the perceptual qualities and spectral characteristics of an audio signal. These features help capture information about the overall sound colour or texture of the audio. Descriptors such as brightness, warmth, and spectral contrast are used to quantify and describe these features. By analyzing timbral and spectral features, we can gain insights into the unique tonal and textural aspects of an audio signal, which are crucial for tasks like music classification, audio synthesis, and sound analysis.

Mel-frequency Cepstral Coefficients (MFCCs)

MFCCs are essential features employed in music genre classification and speech recognition. They capture the mel-frequency spectrum of an audio signal, mimicking human hearing perception. By applying cepstral coefficients, MFCCs transform the spectrum into a lower-dimensional representation, effectively reducing computational complexity. This process allows for efficient analysis and comparison of audio signals, enabling accurate classification and recognition tasks in various applications.

Chroma-based Features

Chroma-based features are focused on capturing and representing the chromatic content or tonal quality of a sound signal. Chroma features are obtained by analyzing the spectral content of an audio signal, revealing the distribution of energy across various musical pitch classes. They provide a representation of the tonal characteristics of the audio signal, highlighting the relative prominence of different musical notes or pitch classes. They provide a compact representation of the harmonic and melodic characteristics of a piece of music, making them useful in various applications such as music genre classification, chord recognition, and audio similarity analysis. By focusing on chromatic information, chroma-based features offer insights into the musical structure and tonal aspects of audio signals.

Rhythm and Temporal Features

Rhythm and Temporal features focus on capturing the rhythmic patterns and temporal structures of an audio signal. They provide information about the timing and tempo of the signal. Examples of rhythm and temporal features include the beat histogram (distribution of beat times), tempo (the estimated tempo of the signal), and rhythmic pattern descriptors (representations of rhythmic motifs or patterns).

MUSIC GENRE CLASSIFICATION ALGORITHMS

Music genre classification is a fundamental task in the field of music information retrieval, which aims to automatically assign music tracks to predefined genre categories. With the increasing volume of digital music available, automated genre classification algorithms have become essential for various applications such as music recommendation systems, content organization, and music streaming services. This section explores the key algorithms used in music genre classification and their applications. Fig.1 provides an overview of various algorithms that can be employed in music genre classification tasks.

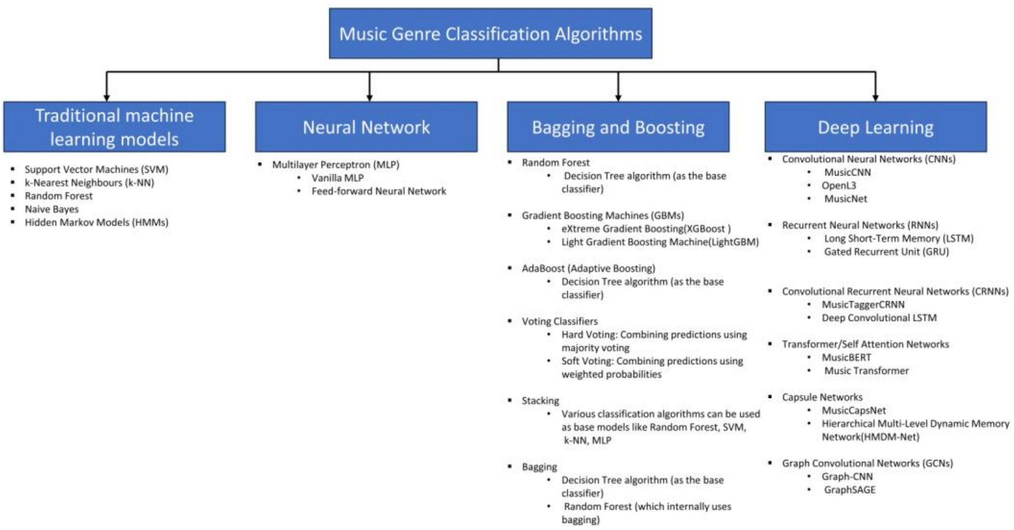

Fig. 1: Music genre classification algorithms

Support Vector Machines (SVM)

SVM, as a widely used supervised learning algorithm for music genre classification, endeavours to map input data into a feature space of higher dimensions and identify an optimal hyperplane that effectively separates distinct genres. SVMs exhibit resilience against noise and can handle intricate decision boundaries. However, their computational complexity can pose challenges when dealing with large datasets.

K-Nearest Neighbors (K-NN)

K-NN is a non-parametric classification algorithm used in music genre classification. It determines the class of an audio file by taking the majority vote of its k nearest neighbours in feature space. This approach assumes that similar audio files are likely to belong to the same genre. The value of k determines the level of local influence on the classification decision.

Decision Trees

Decision tree-based algorithms, such as C4.5 and Random Forests, have been widely used in music genre classification. These algorithms construct a hierarchical tree structure where each node represents a decision based on a specific feature. Decision trees can efficiently handle high-dimensional feature spaces and provide interpretability. Random Forests, a variant of decision trees, combine multiple decision trees to enhance classification accuracy.

Naive Bayes

Naive Bayes is a probabilistic classifier frequently used in music genre classification. It assumes independence between features and models the likelihood of feature occurrences in different genres. By calculating the probabilities of an audio file belonging to each genre based on its features, Naive Bayes determines the most probable genre for the file.

Ensemble Algorithms

Ensemble algorithms combine multiple base classifiers to improve classification performance. Bagging and boosting are two common techniques used in music genre classification. Bagging algorithms, such as Random Forests, create an ensemble of independently trained classifiers and aggregate their predictions. Boosting algorithms, like Ada Boost, iteratively train weak classifiers and assign higher weights to misclassified instances to improve overall accuracy.

Neural Networks

Neural networks, particularly feed-forward neural networks, have been applied to music genre classification. These networks consist of multiple interconnected layers of artificial neurons that learn the complex mappings between input features and output genres. Hidden layers within the neural network capture increasingly abstract representations of the audio data, enabling accurate genre classification.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) have achieved remarkable success in music genre classification. CNNs automatically learn hierarchical representations from raw audio spectrograms or time-frequency representations. By utilizing convolutional layers, these networks can capture local patterns in the audio data, allowing them to discern important features for genre classification.

Recurrent Neural Networks (RNN)

Recurrent Neural Networks (RNNs), particularly the long short-term memory (LSTM) networks, are well-suited for capturing temporal dependencies present in music signals. RNNs process sequential data by retaining a hidden state that preserves information about previous inputs. LSTM networks, which fall under the umbrella of RNNs, excel in capturing long-term dependencies, making them highly effective for music genre classification tasks that demand an understanding of the temporal nature of audio data.

Music genre classification algorithms play a vital role in organizing and recommending music across various applications. Researchers and practitioners have explored a diverse range of techniques, including SVM, KNN, decision trees, random forest methods, and neural networks. The selection of the algorithm depends on specific task requirements, the size and characteristics of the dataset, and the desired level of accuracy. In future research, there is potential to incorporate additional contextual information, such as lyrics and user preferences, to further enhance the accuracy and relevance of music genre classification algorithms.

EVALUATION METRICS

Music genre classification is a challenging task that involves the automatic categorization of music tracks into predefined genres. To evaluate the performance of music genre classification systems, several evaluation metrics are commonly used. Here are some of the key evaluation metrics for music genre classification, along with their explanations:

- Accuracy: Accuracy is a fundamental metric that evaluates the overall correctness of a classification system by calculating the percentage of correctly classified instances out of the total instances. While accuracy provides a general overview of classification performance, it can be misleading when the dataset is imbalanced, meaning some genres have significantly larger representations than others.

- Confusion Matrix: A confusion matrix offers a detailed breakdown of classification results, displaying the counts of true positive, true negative, false positive, and false negative instances for each genre. It aids in identifying specific areas where the classification system may face challenges, such as frequently misclassifying certain genres.

- Precision: Precision measures the proportion of correctly classified positive instances out of all instances predicted as positive. In music genre classification, precision signifies the ability to accurately identify instances of a particular genre.

- Recall: Recall measures the proportion of correctly classified positive instances out of all actual positive instances. It reflects the ability to capture all instances belonging to a specific genre.

- F1-score: The F1-score is the harmonic mean of precision and recall, providing a balanced evaluation metric that considers both precision and recall. It combines these metrics into a single value and is particularly useful when there is an imbalance between classes.

It is crucial to note that the selection of evaluation metrics can differ based on the specific demands of the music genre classification task, the nature of the dataset, and the goals of the system under development. Together, these metrics offer a comprehensive assessment of the performance of the classification system, enabling researchers and developers to analyze and enhance their models effectively.

PAPER SUMMARY

In this paper summary, we provide an overview of 10 papers that have contributed significantly to the advancement of music genre classification.

Paper Title	Dataset	Classification Models	Accuracy	Key Findings
Music genre Classification using Machine Learning[1]	GTZAN 10 genres	ANN, CNN, Decision Tree, MLP, SVM	91%	The proposed CNN model is better than previously proposed models. Some genres were easily identified while others were confused with other styles. The number of audio clips for one class in the dataset is not enough to get good accuracy. Different authors have used different types of datasets and the number of genres considered, making it difficult to compare results. The paper highlights the importance of genre classification in the music industry and how it can be used to recommend songs to users
Music Genre Classification using Deep Learning[2]	GTZAN 10 genres	SVM, Random Forests, XGB (eXtreme Gradient Boosting), and CNN	SVM: 89% CNN: 93.8%	In terms of accuracy and F1-score, CNN demonstrates superior performance compared to other models. Among the extracted features, Mel Spectrogram images were found to reflect audio characteristics more accurately. The custom CNN architecture benefits from the improved performance by increasing the number of filters in each convolutional layer.
PMG-Net: Persian music genre classification	PMG-Data (Persian Music Genre	PMG-Net	86%	There has been limited research conducted on Persian music genre classification, and the absence of a standardized dataset is notable in this field.

using deep neural networks[3]	Data) 5 genres			PMG-Net demonstrates superior performance over traditional algorithms in Persian music genre classification. The classification of music genres plays a significant role in comprehending their creation process and the distinctions between them. People's music preferences can provide reasonably accurate insights into various personalities and cultural aspects specific to a particular region.
Robustness of musical features on deep learning models for the music genre Classification[4]	Carnatic GTZAN Hindustani Homburg 10+ genres	MLP, CNN, and CRNN	Carnatic – 96.38% GTZAN – 98.28% Hindustani – 98.62 Homburg – 81.8	Deep learning (DL)-based models require a substantial amount of data to effectively generalize on new samples. The limited availability of extensive open music datasets highlights the need to analyze the robustness of musical features on DL models. Mel-Scale based features and Swaragram demonstrate significant robustness across datasets and various DL models in the music genre classification (MGC) task. Chromagram and Tonnetz musical features exhibit subpar classification accuracy, primarily due to overfitting concerns. MFCC and Swaragram musical features successfully capture latent patterns underlying musical genres, leading to exceptional performance across datasets and selected models. The design of the MFCC feature, which aims to capture the characteristics of the human auditory system, finds extensive applicability in music genre classification [4].
Holistic Approaches to Music Genre Classification using Efficient Transfer and Deep Learning Techniques[5]	GTZAN ISMIR 2004 MagnaTagATune	Weighted Visibility Graph based Elastic Net Sparse Classifier (WVG-ELNSC) Sequential machine learning analysis with Stacked Denoising Autoencoder (SDA) classifier Riemannian Alliance based Tangent Space Mapping (RA-TSM) transfer learning techniques Transfer Support Vector Machine (TSVM) algorithm Deep learning classifier with Bidirectional Long Short-Term Memory (BiLSTM) cum Attention model with Graphical Convolution Network (GCN) termed as BAG deep learning	GTZAN: 93.51% ISMIR 2004: 92.49% MagnaTagATune: 92.13%	In terms of classification accuracy, the proposed deep learning BAG model demonstrates superior performance compared to other classifiers. The proposed approaches offer robust solutions for music classification, thereby enhancing the user experience when utilizing media players with music files. The TSVM algorithm is selected as the transfer learning target domain classifier due to its simple principle and ease of implementation.

		model		
Evaluating Various Feature Extraction Methods and Classification Algorithms for Music Genres Classification[6]	GTZAN 10 genres	BackpropagationNeural Network (BBNN), Parallel Recurrent Convolutional Neural Network (PRCNN), MINBBNN and PRCNN with an attention mechanism. **MINBBNN** The MINBBNN model is derived from BBNN by making several modifications to the inception block. Specifically, the original 5x5 convolution layer is replaced with two separate 3x3 convolution layers. Additionally, the 1x1 convolution layer and max-pooling layer are eliminated from the inception block. Following the 3x3 convolution, the output is combined with the output of the new version of the 5x5 convolution layer, after undergoing batch normalization. Moreover, a 1x1 convolution layer is applied, and its result is concatenated with the input of the inception block. These adjustments in MINBBNN result in a reduced parameter count from 185,642 to 109,034. **PRCNN with an attention mechanism** ThePRCNN architecture is enhanced by introducing parallel convolution and max-pooling layers alongside the GRU layer, omitting the 32-neuron layer. This modified model contains 177,708	PRCNN self-attention mechanism with Mel spectrogram as input is the best model with an average accuracy of 89.40% and the best accuracy of 94%	The findings of the study highlight the efficacy of various feature extraction methods and classification algorithms in the realm of music genre classification. The researchers compare and evaluate several feature extraction techniques, including statistical features, perceptual features, and spectral features. These methods aim to capture different aspects of music, such as rhythmic patterns, timbral characteristics, and tonal information. The inclusion of two novel feature extraction methods, namely Harmonic Percussive Source Spectrogram (HPSS) and Modulation Spectrogram (MS), did not yield improved performance compared to the Mel spectrogram for music genre classification.

		parameters.		
Improving Music Genre Classification From Multi-Modal Properties OfMusic And Genre Correlations Perspective[7]	Music4All	This paper introduces an innovative multi-modal approach that consists of three essential modules: music representation learning, multi-modal fusion, and genre correlation extraction. In the music representation learning module, features are extracted from both audio and lyrics using a pre-trained BERT model, with a specific emphasis on lyrics encoding. To ensure the effective alignment of these features, an audio-lyrics contrastive loss is applied before inputting them into the multi-modal fusion module. The multi-modal fusion module employs symmetrical mechanisms to seamlessly combine diverse features. Furthermore, the genre correlations extraction module employs a graph convolution network (GCN) to effectively model potential correlations among different genres. Finally, a classifier utilizes the fused feature representation and genre embeddings as input for genre prediction.	incorporating the A-L loss or SCMA (symmetric cross-modal attention) improves accuracy, surpassing the baseline model with scores of 0.396 and 0.408, respectively.	The paper presents an innovative multi-modal approach that integrates audio-lyrics contrastive loss and cross-modal attention to align and merge features extracted from audio and lyrics effectively. Furthermore, a genre correlations extraction module is developed, considering the unique characteristics of multi-label classification. This module captures and models potential genre correlations, allowing for a more comprehensive understanding of genre relationships. The proposed method outperforms previous approaches and establishes new benchmarks in music genre classification. It achieves state-of-the-art performance on the Music4All dataset, demonstrating its superior effectiveness and advancements in the field [7].
A Classification Model of Music Genres Using BP Neural Network Based on Genre Similarity Analysis[8]	COMAP dataset with 90000 songs 17 genres broadly categorized into 5	Backpropagation Neural Network (BP Neural Network)	Decision Tree: 87.5% KNN: 84.2% LDA (Linear Discriminant Analysis): 85.8%	This paper explores the similarity between music genres through two distinct approaches: theoretical analysis and a systematic cluster analysis model. It proposes employing the cluster analysis output as the output of a BP neural network, where the 14 music characteristics are utilized as input features. The study concludes by establishing a BP neural network with two hidden layers as the final model, achieved through iterative training processes [8].

			BP Neural Network: 91.4%	
Classification of Music Genres using Neural Network[9]	GTZAN 10 genres	Recurrent Neural Networks -Long Short-Term Memory (RNN-LSTM)		RNN-based deep learning algorithms possess the ability to anticipate future behaviour based on historical data, enabling the system to comprehend context. However, traditional RNN algorithms have limited memory capacity, making it challenging to handle structures with multiple dependencies. To overcome this limitation, the Long Short-Term Memory (LSTM) architecture is employed, capable of recognizing patterns across longer sequences, spanning up to 100 steps. This study highlights the incorporation of LSTM within RNN-based deep learning algorithms, enhancing the system's contextual understanding, and enabling it to capture long-term dependencies. This makes it a powerful approach for classifying musical genres in audio files. The paper establishes that the most effective method for music genre classification in audio files involves an RNN-based system utilizing LSTM.
Musical Genre Classification Using Melody Features Extracted From Polyphonic Music Signals[10]	GTZAN 10 genres	Support Vector Machines (Sequential Minimal Optimization (SMO) combined with radial basis function (RBF) kernel, also known as the Gaussian kernel), Random Forest (RF), K-Nearest Neighbors (K-NN), and Bayesian Network (B Net)	SVM (SMO combined with RBF): 82%	This study presents an innovative methodology for music genre classification that utilizes high-level melodic features directly extracted from the audio signal.The experimental results indicate a classification accuracy exceeding 90% when applied to a dataset of 500 music excerpts. These findings highlight the effectiveness of the proposed method in accurately categorizing musical genres. Furthermore, the paper delves into a comparative analysis between low-level timbre features and high-level melodic features, exploring their respective contributions to genre classification. Additionally, it investigates the potential benefits of combining these two feature types to enhance classification performance. Overall, the research showcases the viability of utilizing high-level melodic features for music genre classification, achieving impressive accuracy rates. Importantly, the study demonstrated that combining the melodic feature set with the low-level feature set led to improved classification accuracy. This finding indicates that the incorporation of high-level melodic features in conjunction with conventional low-level features shows potential as an effective strategy for genre classification. By employing the SMO classifier and the combined feature set, an accuracy of 82% is attained, outperforming both the melodic

				feature set and the MFCC feature set [10]. While this accuracy may not surpass the highest reported accuracy for the specific dataset, these results serve as a valuable proof of concept. They demonstrate the promising potential of combining low-level features with high-level melodic features to enhance genre classification.

CONCLUSION

The classification of music genres using audio files remains a complex and dynamic field. This extensive survey paper has offered a comprehensive overview of feature extraction techniques, classification algorithms, dataset characteristics, evaluation metrics, and a summary of relevant papers in music genre classification. By reviewing state-of-the-art methodologies and addressing unresolved research inquiries, this survey strives to contribute to the progress of music genre classification and inspire future investigations in this area.

REFERENCES

[1] Anirudh Ghildiyal, Komal Singh, and Sachin Sharma, “Music genre Classification using Machine Learning”, *Proceedings of the 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA 2020) : 5-7, November 2020*.

[2] M. Shah, N. Pujara, K. Mangaroliya, L. Gohil, T. Vyas, and S. Degadwala, “Music Genre Classification using Deep Learning,” in *Proceedings - 6th International Conference on Computing Methodologies and Communication, ICCMC 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 974–978. doi: 10.1109/ICCMC53470.2022.9753953.

[3] N. Farajzadeh, N. Sadeghzadeh, and M. Hashemzadeh, “PMG-Net: Persian music genre classification using deep neural networks,” *Entertain Comput*, vol. 44, Jan. 2023, doi: 10.1016/j.entcom.2022.100518.

[4] Y. Singh and A. Biswas, “Robustness of musical features on deep learning models for music genre classification,” *Expert Syst Appl*, vol. 199, Aug. 2022, doi: 10.1016/j.eswa.2022.116879.

[5] S. K. Prabhakar and S. W. Lee, “Holistic Approaches to Music Genre Classification using Efficient Transfer and Deep Learning Techniques,” *Expert Syst Appl*, vol. 211, Jan. 2023, doi: 10.1016/j.eswa.2022.118636.

[6] M. Bakhtyari, S. Davoudi, and S. Mirzaei, “Evaluating Various Feature Extraction Methods and Classification Algorithms for Music Genres Classification,” in *Proceedings - 2022 27th International Computer Conference, Computer Society of Iran, CSICC 2022*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/CSICC55295.2022.9780487.

[7] G. Ru, X. Zhang, J. Wang, N. Cheng, and J. Xiao, “Improving Music Genre Classification from multi-modal Properties of Music and Genre Correlations Perspective,” in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/ICASSP49357.2023.10097241.

[8] D. Sun, M. Lv, H. Ren, J. Fan, and Q. Liu, “A Classification Model of Music Genres Using BP Neural Network Based on Genre Similarity Analysis,” in *Proceeding - 2021 China Automation Congress, CAC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 6989–6994. doi: 10.1109/CAC53003.2021.9728352.

[9] R. Sharma and Nisha, “Classification of Music Genres using Neural Network,” in *Proceedings of the 2022 11th International Conference on System Modeling and Advancement in Research Trends, SMART 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 142–147. doi: 10.1109/SMART55829.2022.10046811.

[10] Institute of Electrical and Electronics Engineers. and IEEE Signal Processing Society., *2012 IEEE International Conference on Acoustics, Speech, and Signal Processing:proceedings: March 25-30, [2012], Prague Congress Center, Prague, Czech Republic*. IEEE, 2012.

[11] Domingues, Marcos & Pegoraro Santana, Igor & Pinhelli, Fabio & Donini, Juliano & Catharin, Leonardo & Mangolin, Rafael & Costa, Yandre & Feltrim, Valéria Delisandra, *“Music4All: A New Music Database and Its Applications*”, 2020.