Deep Learning for Automated Music Genre Classification Using Spectrograms

Ramya V
PG Student, Department of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Chennai, Tamil Nadu, India.
Email: ramya.v2023@vitstudent.ac.in

Abstract

Music genre classification is a crucial task in music information retrieval, with applications including recommendation systems, content organization, and media categorization. This study aims to develop an automated music genre classification system using deep learning, specifically utilizing the VGGish model, a pre-trained neural network for audio feature extraction. The system is trained and evaluated on the GTZAN dataset, which contains 1,000 audio files uniformly distributed across ten distinct music genres. The methodology involves preprocessing the raw audio files into Mel spectrograms, a time-frequency representation that captures essential audio features. The spectrograms are fed into the VGGish model to extract high-level audio embeddings. A fully connected neural network is then trained on these embeddings to classify each audio track into its appropriate genre. The study improves the feature extraction capabilities by fine-tuning the VGGish model on the GTZAN dataset for improved music genre classification. The model's performance is assessed using metrics including accuracy, precision, recall, and F1-score, with cross-validation used to enhance generalization and robustness. The study aims to provide a highly accurate and efficient system for classifying music genres, with potential applications in music recommendation engines and other audio analysis platforms.

Keywords: Deep Learning; Mel spectrograms; Music genre classification; Transfer learning; VGGish

Objective

The key objectives of this study are:

- Explore the potential of deep learning in solving the complex task of music genre classification by identifying relevant patterns and features in audio data.
- Develop an automated system for music categorization, capable of accurately classifying music into various genres, contributing to more efficient content organization.
- Contribute to the field of music information retrieval, offering insights into how advanced computational techniques can improve audio data processing and analysis.
- Improve existing music classification systems, providing more reliable and accurate methods for genre identification, which can be applied to various audio-based applications.
- Demonstrate the applicability of AI-driven solutions in real-world scenarios, such as music recommendation engines and media categorization, enhancing user experience and system efficiency.

Literature Review

Elbir and Aydin (2020) [1] investigate the use of deep learning for enhancing music genre classification and recommendation systems. They apply Convolutional Neural Networks (CNNs) to analyze spectrograms of audio signals, demonstrating improved accuracy in genre classification compared to traditional methods. Their approach not only refines genre classification but also integrates with collaborative filtering for more effective music recommendations. The study highlights the effectiveness of advanced neural network models in processing complex audio data and providing personalized music suggestions, contributing significantly to the fields of music data analysis and recommendation systems. Yang, Feng, Wang, Yao, and Luo (2020) [2] propose a parallel recurrent convolutional neural network (RRCNN) approach for music genre classification tailored for mobile devices. Their method combines convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to leverage both spatial and temporal features of music data effectively. The RRCNN model is designed to be computationally efficient, making it suitable for deployment on mobile platforms. The authors demonstrate that their approach significantly improves classification accuracy while maintaining low computational requirements, making it feasible for real-time genre classification on mobile devices. This work advances the field by providing a robust and efficient solution for genre classification that is well-suited for resource-constrained environments. Wijaya and Muslikh (2024) [3] present a music genre classification method that combines Bidirectional Long Short-Term Memory (BiLSTM) networks with Mel-Frequency Cepstral Coefficients (MFCCs). Their approach leverages BiLSTM networks to capture both past and future context in sequential audio data, enhancing the model's ability to classify genres accurately. MFCCs are used as input features to represent the audio signal's spectral characteristics. The study shows that the integration of BiLSTM with MFCCs leads to significant improvements in classification performance compared to traditional methods. This work contributes to the field by providing a robust technique for genre classification that effectively utilizes sequential dependencies and spectral features, thereby advancing the accuracy and reliability of music genre classification systems. Zhang (2021) [4] introduces a music style classification algorithm that combines music feature extraction with deep neural networks (DNNs). The study focuses on extracting meaningful features from music data, which are then fed into a deep neural network for classification. Zhang's approach emphasizes the importance of effective feature extraction in enhancing the performance of deep learning models. The algorithm demonstrates improved accuracy in classifying various music styles by leveraging complex patterns identified through the DNN. The research highlights the effectiveness of integrating sophisticated feature extraction techniques with advanced neural network architectures, advancing the field of music style classification by providing a more accurate and robust solution for categorizing music based on its stylistic elements. Zhuang, Chen, and Zheng (2020) [5] explore the application of a transformer-based model for music genre classification. The authors utilize the transformer architecture, known for its success in natural language processing, to capture both local and global dependencies within music data. The model processes sequences of audio features and efficiently classifies music genres by learning long-term dependencies. Their results demonstrate that the transformer classifier outperforms traditional machine learning models and some deep learning architectures in terms of classification accuracy. This work highlights the potential of transformer models for music classification tasks, contributing to the field by providing an innovative approach that leverages attention mechanisms for improved performance in genre classification. Mehta, Gandhi, Thakur, and Kanani (2021) [6] review various music genre classification methods, initially discussing traditional approaches based on handcrafted features like timbre, pitch, and rhythm, which often fail to fully capture the complexity of musical audio. They then focus on deep learning methods, particularly convolutional neural networks (CNNs), which have enhanced classification performance. The authors highlight the importance of mel spectrograms in representing audio data. A key part of the literature review involves transfer learning models, such as VGG16, ResNet50, and InceptionV3, which are pre-trained on large image datasets like ImageNet. These models are fine-tuned to extract deep audio features from log-based mel spectrograms, significantly improving classification accuracy and generalization in music genre tasks, especially with limited labeled data. Transfer learning thus stands out as an efficient solution for audio classification. Prabhakar and Lee (2023) [7] review the advancements in music genre classification by comparing traditional and modern approaches. Early methods relied on handcrafted features such as rhythm, pitch, and timbre, but these techniques struggled with the complexity of audio signals and often resulted in suboptimal performance. With the rise of deep learning, convolutional neural networks (CNNs) have become a standard for improving classification accuracy by automatically learning relevant features. The authors emphasize the role of transfer learning in modern classification systems, discussing models such as EfficientNet and MobileNet, which are pre-trained on large-scale datasets like ImageNet. These models are adapted to audio tasks using mel spectrograms and other feature representations. The literature highlights how transfer learning reduces computational complexity and enhances performance, making it more efficient for real-time applications. The paper also discusses ensemble methods and hybrid deep learning techniques, offering holistic solutions for improving music genre classification outcomes across diverse datasets. Jena, Bhoi, Mohapatra, and Bakshi (2023) [8] review various methods for music genre classification, emphasizing the hybrid deep learning approach that combines wavelet analysis with spectrogram analysis. Traditional methods often rely on handcrafted features and struggle with the complexity of musical data. The paper highlights how recent advancements in deep learning, particularly through hybrid models, enhance classification accuracy. The authors discuss the benefits of integrating wavelet transforms with spectrograms, which helps in capturing both time-frequency representations and hierarchical feature extraction. By combining these techniques with deep learning models, such as CNNs, the paper demonstrates improvements in classification performance and robustness. The review focuses on how this hybrid approach leverages the strengths of both wavelet and spectrogram analyses to handle diverse and complex music datasets effectively. Ashraf, Abid, Din, Rasheed, Yesiltepe, Yeo, and Ersoy (2023) [9] review recent developments in music genre classification, particularly focusing on hybrid models that integrate convolutional neural networks (CNNs) and recurrent neural networks (RNNs). They discuss the limitations of traditional methods, which rely on handcrafted features and often struggle to capture complex patterns in musical data. The paper highlights how CNNs are effective at extracting spatial features from audio spectrograms, while RNNs are adept at modeling temporal dependencies. By combining these two approaches, the authors demonstrate significant improvements in classification performance and robustness. The review explores various hybrid architectures, their performance metrics, and how these models enhance the accuracy of music genre classification by leveraging both spatial and temporal features of the audio data. Sharma, Aggarwal, Bhardwaj, Chakrabarti, Chakrabarti, Abawajy, and Mahdin (2021) [10] review methods for classifying Indian classical music, focusing on time-series matching deep learning approaches. They discuss traditional classification methods that use handcrafted features and often struggle with the intricate patterns of Indian classical music. The paper highlights advancements in deep learning, specifically time-series models, that better capture temporal and sequential characteristics of music. The authors examine how models like Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCNs) are used to analyze time-series data effectively. They explore how these models handle the unique aspects of Indian classical music, such as complex rhythmic patterns and melodic variations, thus improving classification accuracy. The review provides insights into how deep learning techniques tailored for time-series analysis offer significant advantages over traditional methods in classifying Indian classical music. Ceylan, Hardalaç, Kara, and Firat (2021) [11] review automatic music genre classification methods and their connections to music education. They begin by discussing traditional genre classification approaches that rely on handcrafted features and their limitations in accurately categorizing diverse music genres. The paper then focuses on the application of machine learning and deep learning techniques, such as convolutional neural networks (CNNs) and support vector machines (SVMs), to improve classification accuracy. The authors explore how these modern methods handle complex audio features more effectively than traditional approaches. Additionally, they examine the impact of automatic classification systems on music education, highlighting how these technologies can support music educators in analyzing and understanding musical content. The review provides insights into the benefits and challenges of integrating automatic genre classification with educational practices, emphasizing the potential for enhancing both music analysis and pedagogy. Kostrzewa, Kaminski, and Brzeski (2021) [12] review the search for optimal neural network architectures for music genre classification. They discuss the evolution of classification methods from traditional techniques that rely on handcrafted features to modern approaches utilizing deep learning. The paper provides an overview of various neural network models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid architectures that combine different types of networks. The authors evaluate the strengths and limitations of these models in terms of their ability to handle diverse and complex musical features. They also review recent advancements and innovations aimed at improving classification accuracy, such as deeper network architectures and novel training strategies. The review highlights ongoing challenges in finding the "perfect" network architecture that can effectively and efficiently classify a wide range of music genres. Vishnupriya and Meenakshi (2018) [13] review methods for automatic music genre classification, with a focus on convolutional neural networks (CNNs). They discuss traditional classification approaches that rely on handcrafted features, such as spectral and rhythmic characteristics, and their limitations in capturing complex musical patterns. The paper highlights the advantages of using CNNs, which are designed to automatically learn and extract relevant features from raw audio data, particularly from spectrogram representations. The authors examine the effectiveness of CNNs in improving classification accuracy compared to traditional methods. They also discuss various CNN architectures and their performance metrics, demonstrating how these models enhance the ability to classify different music genres by leveraging deep learning techniques. The review emphasizes the significant improvements achieved with CNNs in automatic music genre classification tasks. Yang and Zhang (2019) [14] review advancements in music genre classification, specifically focusing on the use of duplicated convolutional layers in neural networks. They discuss traditional methods that rely on handcrafted features and their challenges in handling the complexity of music data. The paper introduces an innovative approach where neural networks are enhanced by duplicating convolutional layers to improve feature extraction and classification accuracy. By applying this method, the authors aim to capture more nuanced patterns and representations of musical features. The review highlights how duplicated convolutional layers address issues related to depth and feature learning in neural networks, resulting in more effective and precise music genre classification. The authors evaluate the performance of this approach and compare it with existing methods, demonstrating its potential for advancing classification tasks in the field. Ghosal and Kolekar (2018) [15] explore advancements in music genre recognition, emphasizing the role of deep neural networks and transfer learning. They analyze the shortcomings of traditional methods that rely on handcrafted features, which often fail to capture the intricate and diverse characteristics of musical data. The paper highlights the advantages of employing deep neural networks, particularly through transfer learning, where pre-trained models are adapted for genre recognition tasks. This approach leverages features learned from large-scale datasets, improving performance on smaller, music-specific datasets. The authors examine various deep learning architectures and their effectiveness, illustrating how transfer learning enhances accuracy and reduces training time. The study provides a comprehensive look at how integrating deep learning with transfer learning addresses challenges in music genre recognition and pushes the boundaries of current techniques.

References

[1] A. Elbir and N. Aydin, "Music genre classification and music recommendation by using deep learning," *Electronics Letters*, vol. 56, no. 12, pp. 627–629, 2020.

[2] R. Yang, L. Feng, H. Wang, J. Yao, and S. Luo, "Parallel recurrent convolutional neural networks-based music genre classification method for mobile devices," *IEEE*

Access, vol. 8, pp. 19629–19637, 2020.

[3] N. N. Wijaya and A. R. Muslikh, “Music-genre classification using Bidirectional long short-term memory and mel-frequency cepstral coefficients,” *Journal of Computing Theories and Applications*, vol. 1, no. 3, pp. 243–256, 2024.

[4] K. Zhang, “Music style classification algorithm based on music feature extraction and deep neural network,” *Wireless Communications and Mobile Computing*, vol. 2021, no. 1, pp. 9298654, 2021.

[5] Y. Zhuang, Y. Chen, and J. Zheng, “Music genre classification with transformer classifier,” in *Proceedings of the 2020 4th international conference on digital signal processing*, 2020, pp. 155–159.

[6] J. Mehta, D. Gandhi, G. Thakur, and P. Kanani, “Music genre classification using transfer learning on log-based mel spectrogram,” in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1101–1107.

[7] S. K. Prabhakar and S. W. Lee, “Holistic approaches to music genre classification using efficient transfer and deep learning techniques,” *Expert Systems with Applications*, vol. 211, p. 118636, 2023.

[8] K. K. Jena, S. K. Bhoi, S. Mohapatra, and S. Bakshi, “A hybrid deep learning approach for classification of music genres using wavelet and spectrogram analysis,” *Neural Computing and Applications*, vol. 35, no. 15, pp. 11223–11248, 2023.

[9] M. Ashraf, F. Abid, I. U. Din, J. Rasheed, M. Yesiltepe, S. F. Yeo, and M. T. Ersoy, “A hybrid cnn and rnn variant model for music classification,” *Applied Sciences*, vol. 13, no. 3, p. 1476, 2023.

[10] A. K. Sharma, G. Aggarwal, S. Bhardwaj, P. Chakrabarti, T. Chakrabarti, J. H. Abawajy, et al., “Classification of Indian classical music with time-series matching deep learning approach,” *IEEE Access*, vol. 9, pp. 102041–102052, 2021.

[11] H. C. Ceylan, N. Hardalaç, A. C. Kara, and H. Firat, “Automatic music genre classification and its relation with music education,” *World Journal of Education*, vol. 11, no. 2, pp. 36–45, 2021.

[12] D. Kostrzewa, P. Kaminski, and R. Brzeski, “Music genre classification: looking for the perfect network,” in *International Conference on Computational Science*, Cham: Springer International Publishing, 2021, pp. 55–67.

[13] S. Vishnupriya and K. Meenakshi, “Automatic music genre classification using convolution neural network,” in *2018 International Conference on Computer Communication and Informatics (ICCCI)*, 2018, pp. 1–4.

[14] H. Yang and W. Q. Zhang, “Music genre classification using duplicated convolutional layers in neural networks,” in *Interspeech*, 2019, pp. 3382–3386.

[15] D. Ghosal and M. H. Kolekar, “Music genre recognition using deep neural networks and transfer learning,” in *Interspeech*, 2018, pp. 2087–2091.