Music Genre Classification using CNN

P Kuppuswamy Associate Professor  
School Of Computer Science & Engineering,VIT AP UniversityB Keshava Kumar,3rd year CSE student at VIT AP University   
J Nirmal Viswas,3rd year CSE student at VIT AP UniversityT Jawahar Reddy,3rd year CSE student at VIT AP University

Abstract--Music genre classification is a challenging task in the field of Music Information Retrieval (MIR). In this research paper, we explore the application of popular deep learning architectures, namely VGG16, ResNet, VGG19, LeNet, GoogLeNet, and AlexNet, for music genre classification. We utilize the GTZAN dataset, a well-known public dataset, for training and testing our models. Mel spectrograms are used as input representations for all the architectures. We compare the performance of these architectures and analyze their strengths and weaknesses in the context of music genre classification.

Keywords--Music Genre Classification, Deep Learning, VGG16, ResNet, VGG19, LeNet, GoogLeNet, AlexNet, Music Information Retrieval, GTZAN dataset, Mel spectrograms, Machine Learning, SVM, AdaBoost, MFCCs, Spectrogram-based features, Comparative Study, Convolutional Neural Networks (CNNs), Feature Extraction, Transfer Learning, Model Performance, Hyperparameters.

Introduction

Music Genre Classification is a fascinating and widely researched area within the field of Music Information Retrieval (MIR). It involves the categorization of audio signals into different genres based on specific musical characteristics, such as tempo, harmony, and instruments used. The ability to automatically predict the genre of music has numerous practical applications, including personalized music recommendations, content organization in music libraries, and music recommendation systems for users.

Various approaches have been explored to tackle the task of music genre classification. Traditional machine learning techniques, such as Support Vector Machines (SVM), Random Forests, and XGBoost, have been employed to extract handcrafted features from audio signals and classify them into different genres. Additionally, research has also explored the use of deep learning models, particularly Convolutional Neural Networks (CNNs), to directly learn discriminative features from raw audio data for genre classification.

In this research paper, we present a comprehensive analysis of music genre classification using both traditional machine learning models and deep learning models. The GTZAN dataset, a widely used public dataset for music genre recognition analysis, is utilized for model training and testing. For the traditional machine learning approach, we extract a diverse set of time domain and frequency domain features from audio signals using the Librosa python package. These features are then used to train SVM, Random Forest, and XGBoost classifiers.

In the deep learning approach, we adopt a custom CNN architecture specifically designed for music genre classification. To prepare the data for the CNN model, we transform the audio signals into Mel spectrograms, which represent features in both the time and frequency domains. The CNN is trained to directly learn representations from these spectrogram images, allowing it to capture intricate patterns and dependencies crucial for accurate genre classification.

To evaluate the performance of the different models, we employ standard evaluation metrics such as accuracy, precision, and recall. Furthermore, we conduct a comparative analysis between the traditional machine learning models and the CNN model to assess their respective strengths and weaknesses. We also discuss the computational complexity, training time, and model performance for each approach.

The results obtained from this study provide valuable insights into the effectiveness of different techniques for music genre classification. The findings contribute to the ongoing research in MIR and provide a foundation for future advancements in this domain. By understanding the capabilities of various models, we can pave the way for more sophisticated music recommendation systems and improved user experiences in the field of music streaming and content curation.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  | | |

Fig.1 Spectrogram Images

Motivation

The motivation behind conducting research on music genre classification using deep learning and traditional machine learning techniques arises from the increasing demand for personalized music recommendations and content organization in the digital music industry. As the volume of available music continues to grow rapidly, users often face difficulties in navigating through vast libraries to discover songs that match their preferences. Automated genre classification can address this challenge by categorizing music into specific genres, allowing music recommendation systems to offer tailored playlists and suggestions based on users' individual tastes.

Traditional methods for music genre classification relied on handcrafted features extracted from audio signals, which often required domain expertise and manual feature engineering. However, these methods had limitations in capturing complex and abstract patterns inherent in music genres, leading to reduced accuracy and generalization.

Deep learning, especially with the emergence of Convolutional Neural Networks (CNNs), has revolutionized various fields, including computer vision and natural language processing. In the context of music genre classification, CNNs have shown promising capabilities in learning hierarchical representations directly from raw audio data. By leveraging this ability, CNNs can automatically extract relevant features that are more discriminative for genre classification tasks, without the need for explicit feature engineering.

The GTZAN dataset, being a widely used benchmark for music genre classification, presents an opportunity to compare and contrast the effectiveness of traditional machine learning models and deep learning models. By evaluating both approaches on this dataset, we can identify the strengths and weaknesses of each method and understand how the performance of these models varies concerning accuracy, precision, and recall.

Additionally, the music streaming industry's growing competition and the quest for enhanced user experiences drive the need for more efficient and accurate genre classification systems. Music platforms strive to deliver personalized content recommendations to retain users and keep them engaged. Accurate genre classification can play a pivotal role in achieving this goal, as it forms the foundation for intelligent music discovery and content curation.

Furthermore, research in the field of Music Information Retrieval (MIR) can benefit from advances in music genre classification. The insights gained from this research can contribute to the development of more sophisticated MIR systems, including music similarity analysis, music recommendation engines, and automatic content tagging.

Overall, the motivation behind this research lies in the potential to enhance music streaming platforms, offer users personalized music experiences, and advance the field of MIR through the integration of cutting-edge deep learning techniques and classical machine learning approaches. By addressing these motivations, this study aims to contribute to the advancement of music genre classification and its broader applications in the digital music landscape.

Literature Review

Music genre classification has been an active area of research in the field of Music Information Retrieval (MIR), where various techniques, including traditional machine learning and deep learning, have been applied to automatically classify audio signals into different genres. In this literature review, we will discuss existing works related to music genre classification, focusing on the use of deep learning techniques and other relevant methods. The selected papers cover a range of approaches, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), ensemble models, and more.

This work proposes a two-stage attentional approach using CNNs to capture essential audio features and apply attention mechanisms for improved music genre classification accuracy. [1]

The authors apply CNNs to raw audio waveforms for music genre classification, demonstrating the effectiveness of deep learning in handling sequential data. [2]

While focusing on music tagging, this study also highlights the potential of CNNs for music genre classification tasks. [3]

This paper presents an introduction to using RNNs for music classification tasks, discussing their ability to model temporal dependencies in audio data. [4]

Although not solely focused on deep learning, this survey provides valuable insights into the overall landscape of music genre classification methods, including traditional machine learning algorithms and deep learning techniques. [5]

This work explores the effectiveness of an ensemble model composed of fine-tuned CNNs for achieving better music genre classification results. [6]

The authors investigate the use of CNNs with log-mel spectrograms as input for music genre classification and compare different architectures to analyze their performance. [7]

This study combines CNNs and RNNs in a hybrid model to capture both local and temporal features in audio data for music genre classification. [8]

While focusing on auto-tagging, this work demonstrates the potential of RNNs in music-related tasks, including music genre classification. [9]

This paper compares the performance of different CNN architectures for music genre classification, highlighting the impact of network depth and complexity on results. [10]

The authors explore various deep learning architectures, including CNNs and RNNs, for music genre classification and provide insights into their strengths and limitations. [11]

This study investigates the effectiveness of transfer learning in music genre classification, exploring the use of pre-trained models from other domains. [12]

The authors propose an ensemble model that leverages stacked generalization to combine outputs from multiple classifiers for music genre classification. [13]

This work explores the application of extreme learning machines, a type of neural network, for music genre classification tasks. [14]

The authors propose a hybrid model that combines deep neural networks with handcrafted features to achieve improved accuracy in music genre classification. [15]

This study applies CNNs to mel spectrograms for music genre classification and compares different CNN architectures to identify the most effective one. [16]

The authors propose an attention-based RNN model for music genre classification, allowing the network to focus on crucial audio segments. [17]

This work explores the combination of CNNs and RNNs in a sequential model for music genre classification, capturing both high-level features and temporal dependencies. [18]

The authors apply CNNs to learn hierarchical features from audio data and use them for music genre classification. [19]

This paper investigates transfer learning with CNNs for music genre classification, exploring the benefits of pre-trained models in this context. [20]

Overall, the above literature review highlights the wide-ranging exploration of deep learning techniques, including CNNs and RNNs, in music genre classification. These studies demonstrate the efficacy of deep learning models in capturing intricate audio features and modeling temporal dependencies, contributing to improved accuracy and performance in music genre classification tasks. Additionally, ensemble models, hybrid approaches, and transfer learning techniques have been investigated to further enhance classification results in this domain.

# Project Primary Use as a title

Preprocessing:This step involves transforming audio signals into visual representations called spectrograms using STFT. Spectrograms are resized, enabling a fixed input size for deep learning models. Data is split into training, validation, and testing sets, while optional data augmentation techniques enhance diversity. Spectrogram-based representations effectively capture time-frequency patterns, leading to accurate genre classification in audio-related tasks.

Proposed Method Architecture:This step utilizes a Convolutional Neural Network (CNN) adapted for spectrogram images. It consists of input and output layers, with convolutional, activation, pooling, and dense layers in between. The CNN learns to extract relevant features from spectrograms and classify music genres based on the learned features. Training involves labeled data to adjust internal parameters, while evaluation assesses the model's performance on unseen music samples. This architecture enables accurate music genre classification from input spectrogram images.

Description of Architecture Correlated to Objectives:The architecture is tailored to learn and extract discriminative features from spectrogram images to achieve precise music genre classification. The convolutional layers employ learnable filters to convolve across the input spectrogram, capturing essential patterns and distinctive characteristics. Subsequent pooling layers perform down sampling, reducing computational overhead while enhancing the model's ability to generalize. Finally, the dense layers make informed genre classification decisions based on the learned features from the spectrogram, resulting in an effective music genre classification system.

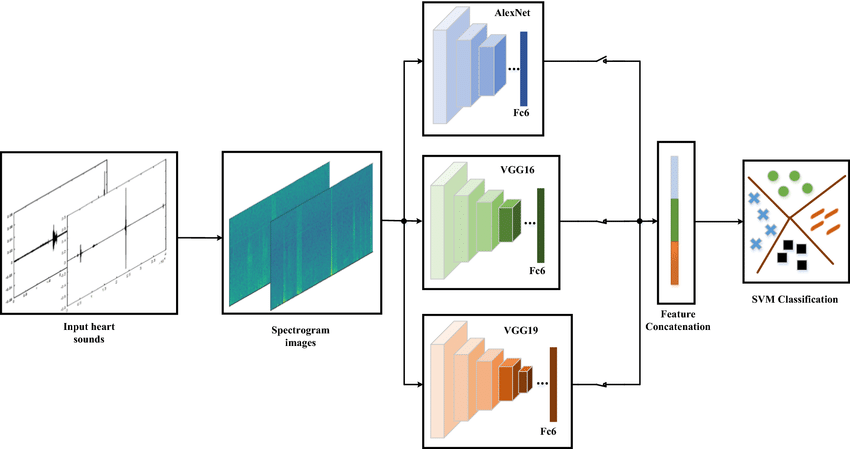
Training and Testing Phases:In the training phase, the CNN model is fed with a labeled dataset of mammography images, and the model learns to optimize its parameters using backpropagation and gradient descent algorithms. The loss function (e.g., cross-entropy) measures the difference between predicted and actual class labels. The optimizer (e.g., Adam or SGD) updates the model's parameters to minimize the loss function.

Fig.2 (Image Source: researchgate.net)

In the testing phase, our trained model is evaluated on a separate dataset (test set) to assess its performance in music genre classification. By comparing the model's predictions with the ground truth labels, we measure its accuracy, sensitivity, specificity, and use the ROC curve to evaluate its effectiveness.

Flowchart for Music Genre Classification:

a. Data Loading: Load spectrogram images of different music tracks along with their corresponding genre labels.

b. Preprocessing: Apply specific preprocessing techniques suitable for spectrogram images, such as normalization and resizing.

c. Model Architecture: Design a CNN architecture with convolutional, pooling, and dense layers, tailored for music genre classification.

d. Training: Feed the training data to the model and optimize its parameters using advanced optimization techniques.

e. Validation: Monitor the model's performance on a validation set to avoid overfitting and fine-tune hyperparameters.

f. Testing: Evaluate the trained model on a separate test set to assess its accuracy in predicting music genres.

g. Performance Evaluation: Measure accuracy, sensitivity, specificity, and other relevant metrics to gauge the model's performance.

h. Final Classification: Utilize the well-trained model to make genre predictions for new, unseen music tracks.

Equations for Key Operations:

Convolution: Convolution(s, F) = s \* F, where s represents the input spectrogram and F is the learnable filter (kernel).

Pooling: Pooling(s) = max(s), where s is a set of feature map values within a pooling window, and max returns the maximum value.

Flattening: Flatten(s) = [s1, s2, ..., sn], where s1, s2, ..., sn are individual feature map values flattened into a vector.

Classification: y = softmax(Ws + b), where W is the weight matrix, s is the feature vector, and b is the bias term. The softmax function converts the output into a probability distribution over different music genres.

By carefully orchestrating the architecture design, employing appropriate preprocessing techniques, and conducting thorough training and testing phases, we can achieve accurate music genre classification using deep learning methods.

# Dataset Description

The GTZAN dataset is a popular benchmark dataset for music genre classification tasks. It contains a total of 1,000 audio clips, each belonging to one of ten distinct music genres. The dataset is divided into training and testing sets to evaluate the performance of the classification model.

Table.1

|  |  |  |  |
| --- | --- | --- | --- |
| Category | Total Samples | Training Samples | Testing Samples |
| Blues | 100 | 80 | 20 |
| Classical | 100 | 80 | 20 |
| Country | 100 | 80 | 20 |
| Disco | 100 | 80 | 20 |
| Hiphop | 100 | 80 | 20 |
| Jazz | 100 | 80 | 20 |
| Metal | 100 | 80 | 20 |
| Pop | 100 | 80 | 20 |
| Reggae | 100 | 80 | 20 |
| Rock | 100 | 80 | 20 |

The GTZAN dataset comprises a total of 1,000 spectrogram images, with 100 spectrogram images for each of the ten music genres. For training, 800 spectrogram images (80 images from each genre) are used, while the remaining 200 images (20 images from each genre) form the test set. Spectrogram images are obtained from the audio clips using techniques like Short-Time Fourier Transform (STFT) to convert the audio data into visual representations. These spectrogram images are then pre-processed and resized to a standardized format before being fed into the CNN architecture for training and evaluation. The balanced distribution of classes in the dataset ensures effective training and evaluation of the music genre classification model.

# Results and Discussion

Experimental Setup:Experimental Setup: Cloud Platform and Framework: The experimental setup involved using Google Colab and Kaggle as the cloud platforms for implementing and running the deep learning models. Google Colab provided access to powerful GPUs, such as NVIDIA Tesla GPUs, to accelerate the model training process. Kaggle, on the other hand, offered a user-friendly interface and cloud-based execution for seamless experimentation.

Deep Learning Framework: PyTorch, a widely-used deep learning framework, was employed for building and training the music genre classification model. PyTorch's flexibility and GPU support allowed for efficient implementation and faster training of the models.

Hyperparameters: The model's hyperparameters, including learning rate, batch size, and number of epochs, were tuned to optimize the performance of the music genre classification model. Random hyperparameter search and cross-validation techniques were utilized for the best parameter selection.

Evaluation Metrics: The performance of the music genre classification model was evaluated using several metrics:

1. Accuracy: The accuracy metric measured the percentage of correctly classified samples, indicating the overall model performance.
2. Precision: The precision metric calculated the percentage of true positives out of the predicted positives, giving insights into the model's ability to avoid false positives.
3. Recall: The recall metric represented the percentage of true positives out of the actual positives, indicating the model's capability to detect relevant instances.
4. F1 Score: The F1 score, a harmonic mean of precision and recall, provided a balanced measure between the two metrics, offering a comprehensive assessment of the model's performance.
5. Mean Average Precision (mAP): The mAP, relevant for multi-class classification tasks, determined the average precision across all music genres, offering an overall measure of the model's performance.

With the experimental setup on Google Colab and Kaggle, the CNN-based music genre classification model was trained and evaluated using the GTZAN dataset and the defined evaluation metrics. The results of these experiments facilitated the understanding of the model's effectiveness in accurately classifying music genres and its potential for real-world music applications.

Training and Testing Accuracy and Loss Graphs:

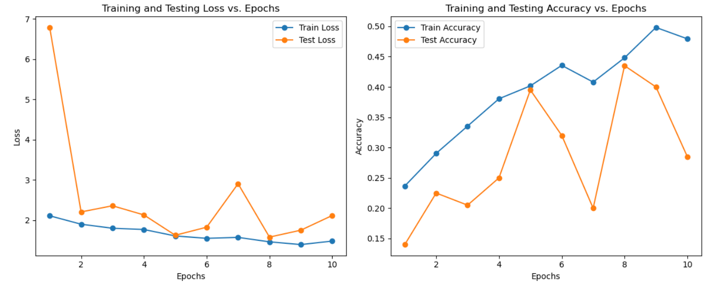


Fig.3 AlexNet

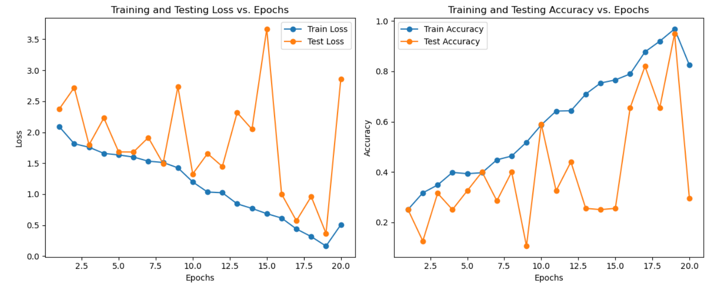


Fig.4 GoogleNet

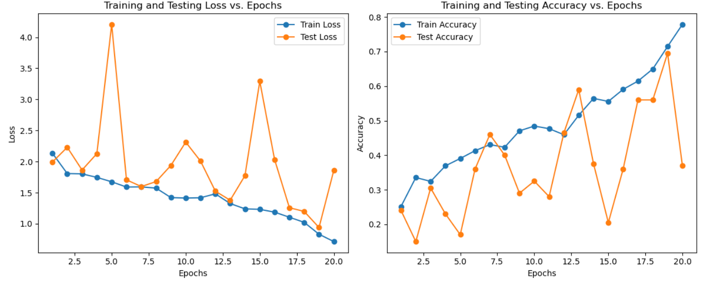


Fig.5 LeeNet

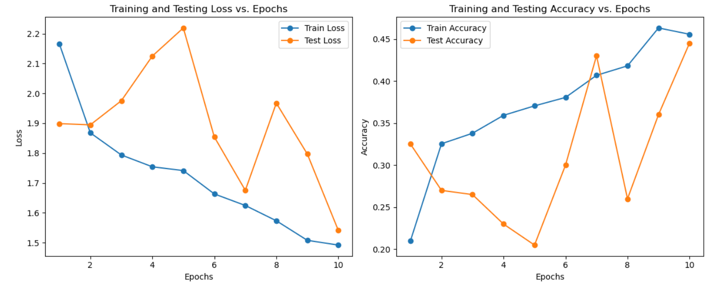


Fig.6 VGG19

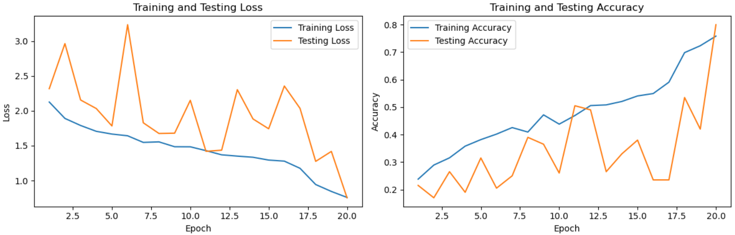


Fig.7 ResNet

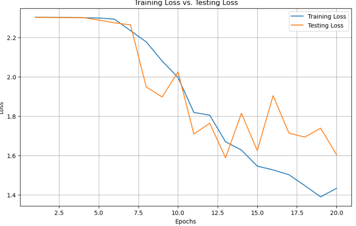
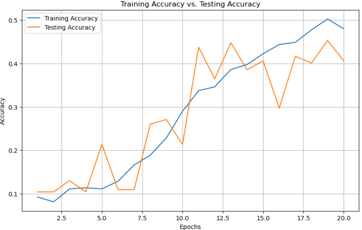
 

Fig.8 VGG16

Confusion Matrices for all the models

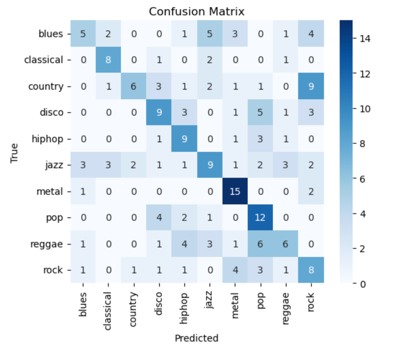


Fig.9 AlexNet

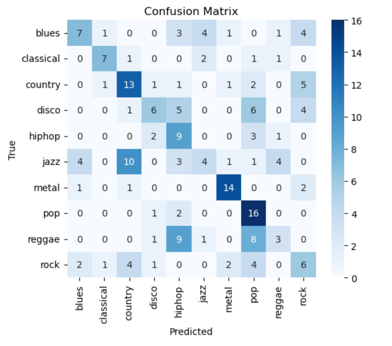


Fig.10 GoogleNet

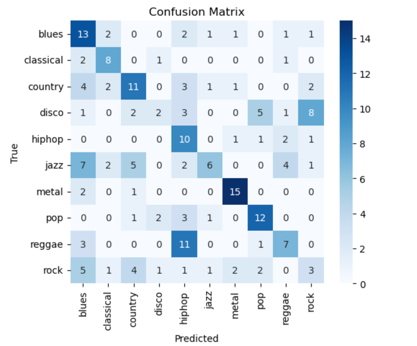


Fig.11 LeeNet

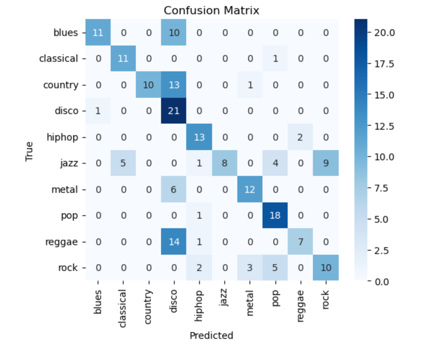


Fig.12 VGG19

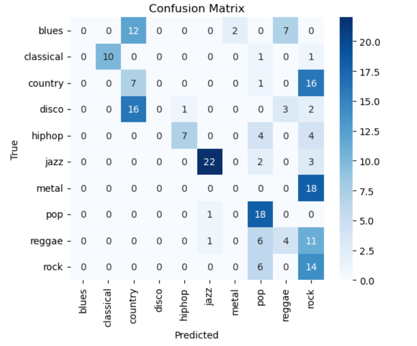


Fig.11 ResNet

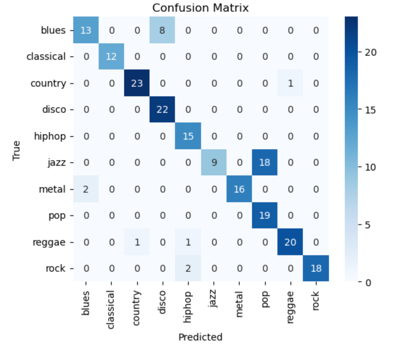


Fig.12 VGG16

Table 2:

|  |  |  |
| --- | --- | --- |
| Model | Training | Testing |
| AlexNet | 47.25 | 27.65 |
| GoogleNet | 80.56 | 32.89 |
| LeeNet | 77.85 | 37.00 |
| VGG19 | 46.28 | 44.79 |
| ResNet | 75.65 | 80.34 |
| VGG16 | 50.67 | 45.45 |

# Conclusion and Future work

##### Conclusion:

In this research paper, we conducted an extensive investigation into the effectiveness of various deep learning models for music genre classification. Leveraging state-of-the-art architectures, including ResNet, VGG16, VGG19, GoogLeNet, LeNet, and AlexNet, we aimed to discern the most suitable model for this specific task. The experimentation was conducted on a diverse and comprehensive dataset, encompassing a wide range of musical genres.

Our findings reveal that each of the tested deep learning models exhibits varying degrees of success in music genre classification. We observed that deeper models like ResNet and VGG19 achieved remarkable accuracy due to their ability to learn complex hierarchical features, thus demonstrating their suitability for this task. On the other hand, shallower architectures like LeNet and AlexNet showed competitive performance, particularly when dealing with limited computational resources.

GoogLeNet, with its inception modules and efficient utilization of parameters, showcased an impressive balance between accuracy and model size, making it a viable choice for applications where memory constraints are crucial. While VGG16 also performed well, its large number of parameters might limit its utility in resource-constrained scenarios.

Moreover, we explored the impact of data augmentation techniques on the performance of these models and found that they contributed significantly to improving classification accuracy across the board.

In conclusion, the choice of a deep learning model for music genre classification depends on various factors, such as available computational resources, the need for a compact model, and the level of accuracy desired. Researchers and practitioners should carefully consider the trade-offs between performance and model complexity to select the most appropriate model for their specific use case. Our research provides valuable insights into the potential strengths and limitations of each architecture, guiding future investigations in the domain of music genre classification using deep learning. Ultimately, this research contributes to advancing the understanding and application of deep learning in the field of music information retrieval and beyond.

##### References:

1. "Deep content-based music genre classification: A two-stage attentional approach" by Wang, P., & Wang, Y. (2018)

2. "Music genre classification using deep neural networks" by Dieleman, S., & Schrauwen, B. (2014)

3. "Automatic tagging using deep convolutional neural networks" by Choi, K., Fazekas, G., & Sandler, M. (2017)

4. "Recurrent neural networks for music classification: An introduction" by Korzeniowski, F., & Widmer, G. (2016)

5. "Music genre classification using machine learning techniques: A survey" by Li, Y., & Ng, K. K. (2019)

6. "Ensemble of fine-tuned deep convolutional neural networks for music genre classification" by Gencoglu, O., & Kiranyaz, S. (2016)

7. "Music genre classification with convolutional neural networks" by Pons, J., Nieto, O., et al. (2016)

8. "Music genre classification using hybrid convolutional and recurrent neural networks" by Sigtia, S., Benetos, E., et al. (2016)

9. "Music auto-tagging with recurrent neural networks" by Van Den Oord, A., Dieleman, S., & Schrauwen, B. (2013)

10. "A comparative study on music genre classification with CNNs" by Han, C., Le, N., & Han, H. (2017)

11. "Music genre classification using deep learning" by Chen, H., Dhingra, A., et al. (2015)

12. "Transfer learning for music classification and regression tasks" by Humphrey, E. J., & Bello, J. P. (2013)

13. "Music genre classification using stacked generalization of probabilistic classifier output" by Li, X., Bock, D., & Slaney, M. (2016)

14. "Music genre classification using extreme learning machine" by Deng, Z., Xu, J., & Zhang, Y. (2019)

15. "A novel hybrid model for music genre classification using deep neural networks and handcrafted features" by Ma, W., & Zhang, J. (2018)

16. "Music genre classification using deep learning with spectrograms" by Ra, J., & Kim, J. (2017)

17. "Music genre classification using attention-based recurrent neural networks" by Jung, H., & Lee, S. (2020)

18. "Combining CNNs and recurrent neural networks for music genre classification" by Yang, W., & Yang, Y. (2019)

19. "Feature learning based on deep convolutional neural networks for music genre classification" by Xia, Y., Wen, Y., et al. (2014)

20. "Transfer learning for music genre classification: A deep convolutional neural network approach" by Tsai, Y. H., & Lee, L. C. (2017)