Optimizing Musical Genre Classification Using Genetic Algorithms

Caio Grasso[1][a], Thiago Carvalho[1,2][b], José Franco Amaral[1][c], Pedro Coelho[1][d], Robert Oliveira[1][e] and Giomar Olivera[1][f]

[1]*FEN/UERJ, Rio de Janeiro State University, Rio de Janeiro, Brazil*

[2]*Electrical Engineering Department, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil*

Keywords: Genetic Algorithm, Music Classification, Signal Processing, Deep Learning.

Abstract: Classifying music into genres is a challenging yet fascinating task in audio analysis. By leveraging deep learning techniques, we can automatically categorize music based on its acoustic characteristics, opening up new possibilities for organizing and understanding large music collections.The main objective of this study is to develop and evaluate deep learning models for the classification of different musical styles. To optimize the models, we utilized Genetic Algorithms (GA) to automatically determine the optimal hyperparameters and model architecture selection, including Convolutional Neural Networks and Transformers. The results demonstrated the effectiveness of GAs in exploring the hyperparameter space, leading to improved performance across multiple architectures, with EfficientNet models standing out for their consistent and robust results. This work highlights the potential of automated optimization techniques in enhancing audio analysis tasks and emphasizes the importance of integrating deep learning and evolutionary algorithms for tackling complex music classification problems.

1 INTRODUCTION

The classification of musical genres is essential for indexing and recommending songs, directly impacting user experience on streaming platforms and the organization of the digital music industry. With distinct traits like rhythm, instrumentation, and structure, musical genres classify works with common elements, enabling the creation of personalized recommendation systems.

Given the growth and diversity of musical genres, advanced classification methods are necessary to overcome challenges related to musical heterogeneity and provide more refined recommendations. This study aims to contribute to the advancement of automated classification solutions, enabling the development of more efficient music curation and retrieval systems.

This study focuses on the classification of musical genres using Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), which analyze audio signals represented visually through spectrograms. To enhance the performance of these models, Genetic Algorithms (GAs) are employed for hyperparameter optimization, automating the search for optimal configurations. The research compares different CNN architectures, such as EfficientNet and ResNet, to identify the most effective models for genre classification. Additionally, the study evaluates the advantages of hyperparameter optimization in improving classification outcomes and verifies the efficiency of transformer-based models for signal classification. The work contributes to more efficient music curation and retrieval systems, offering insights into the integration of DL and evolutionary algorithms for improved accuracy and enhanced user experiences in music streaming platforms.

The primary objective of this study is to optimize both the model selection and hyperparameter tuning with a singular focus: maximizing classification accuracy. The use of Genetic Algorithms is centered on identifying the best-performing configurations that lead to the most accurate genre predictions. By systematically refining architectural choices and training parameters, this approach ensures that improvements are solely driven by their impact on classification accuracy, reinforcing the effectiveness of hyperparameter optimization in deep learning-based music genre classification.

[a] https://orcid.org/0009-0008-3316-1026
[b] https://orcid.org/0000-0001-8689-1438
[c] https://orcid.org/0000-0003-4951-8532
[d] https://orcid.org/0000-0003-3623-1313
[e] https://orcid.org/0000-0003-0000-3001
[f] https://orcid.org/0000-0002-7172-6525

Grasso, C., Carvalho, T., Amaral, J. F., Coelho, P., Oliveira, R. and Olivera, G.
Optimizing Musical Genre Classification Using Genetic Algorithms.
DOI: 10.5220/0013418200003929

In *Proceedings of the 27th International Conference on Enterprise Information Systems (ICEIS 2025) - Volume 1*, pages 881-887
ISBN: 978-989-758-749-8; ISSN: 2184-4992

In the following section, a brief review of the literature on the use of genetic algorithms for model optimization in different DL models and DL models for the music genre classification task will be presented. Section 3 will detail the proposed approach, including the data preprocessing steps, model selection, and the experimental protocol adopted. In Section 4, we discuss the experiments conducted, outlining the configurations and methodology employed. In Section 5, we present the discussion and results, emphasizing how the genetic algorithm influenced hyperparameter selection and model performance. Finally, Section 6 concludes the study and highlights potential directions for future research on applying genetic algorithms to optimize hyperparameters for automatic music genre classification.

2 LITERATURE REVIEW

The study of CNNs for audio processing and audio signal classification has gained attention due to its importance in material retrieval and music recommendation tasks on digital platforms. Early studies focused on manual feature extraction of acoustic properties, such as Mel-Frequency Cepstral Coefficients (MFCCs), timbre, and rhythm, combined with classical machine learning algorithms, including Support Vector Machines (SVM) and K-Nearest Neighbors (KNN) (Tzanetakis and Cook, 2002). While these methods showed some success, their effectiveness was limited by the challenge of capturing the more complex nuances of musical genres.

The representation of audio signals as images has become a widely adopted approach in audio analysis, particularly in the context of musical genre classification (Müller, 2015). One of the most common ways to achieve this representation is through spectrograms, which are visualizations that display the variation of sound frequencies over time. This type of representation transforms the audio signal into a two-dimensional format, enabling a more intuitive analysis of acoustic characteristics (Müller, 2015).

With advancements in Deep Learning techniques, CNNs have emerged as powerful tools for audio analysis, utilizing visual representations like Mel spectrograms to identify patterns that distinguish genres (Choi et al., 2017a). For instance, Choi et al. (2017) demonstrated the effectiveness of CNNs in genre classification, outperforming traditional methods by leveraging the networks' ability to learn hierarchical features directly from raw data. Furthermore, Dieleman and Schrauwen (Dieleman et al., 2011) explored end-to-end approaches, enabling CNNs to operate directly on audio representations without requiring manual feature extraction.

Another critical aspect of musical genre classification is model optimization, where hyperparameter selection, such as network architecture, learning rate, and the number of convolutional filters, plays a pivotal role. Traditional tuning methods, such as grid search or random search, often prove inefficient for deep networks due to high computational costs (Bergstra and Bengio, 2012). In this context, GAs have been explored as a promising alternative, enabling automated selection of optimal hyperparameter configurations. For example, Young et al. (Young et al., 2015) demonstrated the application of GAs for neural network architecture optimization, while Real et al. (Real et al., 2019) showcased advancements in using these techniques to achieve competitive performance on deep learning benchmarks.

These studies highlight the evolution of the field, from manual feature-based methods to the application of deep learning and evolutionary algorithms. However, significant gaps remain in effectively integrating these technologies to capture the diversity and complexity of musical genres, motivating the proposal of this study.

3 PROPOSED APPROACH

This section outlines the methodology adopted for the task of musical genre classification. A step-by-step flowchart is presented in Figure 1, detailing the key processes, starting from the visual representation of audio signals to model optimization. First, audio data is converted into spectrograms to enable visual analysis. In this case, CNNs are leveraged for feature extraction and classification. Finally, a GA is applied to optimize the model's architecture and hyperparameters, ensuring the best configuration for the task.

3.1 Visual Representation of Audio Signals

By representing audio as an image, details such as rhythmic patterns, timbres, and harmonic transitions, often associated with different musical genres, can be captured (McFee et al., 2015). Additionally, many genres share similar characteristics, which can be directly observed in their visual representations. For example, genres like rock and blues may exhibit comparable frequency patterns due to the use of similar instruments, whereas electronic genres may stand out through frequency peaks generated by synthesizers (Pons et al., 2017). In Figures 2, 3, 4 and 5, we can

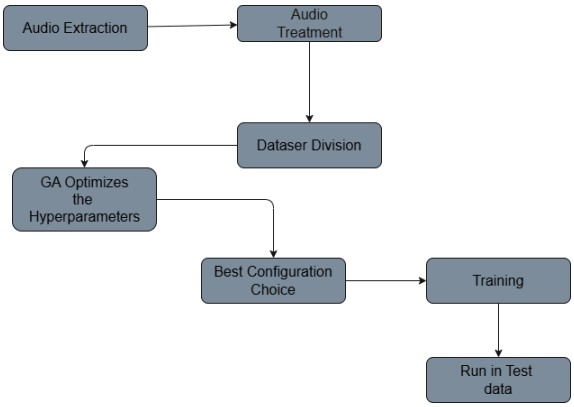

Figure 1: General development flowchart.

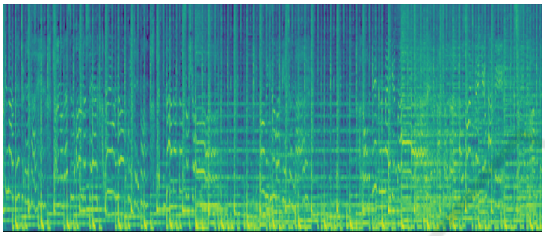

Figure 2: Example of a spectrogram from a Pop sample (1).

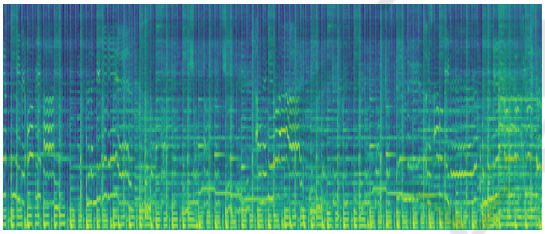

Figure 3: Example of a spectrogram from a Pop sample (2).

observe the visual similarity between musical genres representations.

Another relevant aspect of this approach is the flexibility it provides in data manipulation. By adjusting the frequency scale, such as using the Mel scale, the visual representation can be aligned with human auditory perception, facilitating the identification of relevant patterns (McFee et al., 2015). Additionally, techniques like data augmentation, when applied directly to spectrograms, create variations in visual representations, increasing data diversity and enhancing the robustness of subsequent analyses. Representing audio as images not only offers an efficient analysis method but also provides a unique perspective on the similarities and differences between musical genres, enabling a deeper understanding of the characteristics that define each style (McFee et al., 2017).

Furthermore, the use of Mel scales and other spectrogram-based representations has emerged in the literature as an efficient way to capture relevant sound

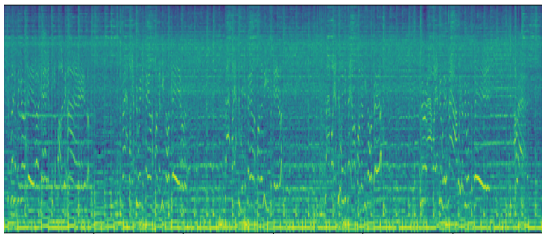

Figure 4: Example of a spectrogram from a Metal sample (1).

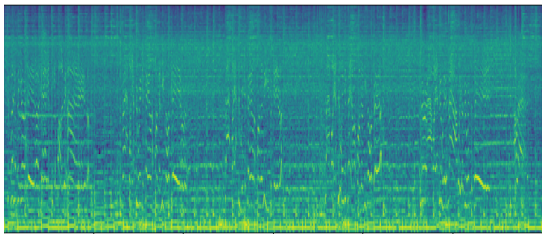

Figure 5: Example of a spectrogram from a Metal sample (2).

signal features for musical genre classification (Choi et al., 2017a). These representations are widely employed because they approximate the audio signal to human perception, making the classification task more intuitive for deep learning models. Recent studies have explored the combination of multiple computer vision approaches to improve representation quality. For example, techniques such as data augmentation and hybrid neural networks combining CNNs with RNNs (Recurrent Neural Networks) have been investigated to explore the temporal and spatial relationships in spectrograms, thereby enhancing model accuracy (Choi et al., 2017b).

3.2 CNN Models

In this study, we used the visual representations of the audio to classify music genres using CNNs. The adopted methodology relied on converting audio signals into spectrograms, which were subsequently used as inputs for image classification models. This approach has gained prominence in the literature due to the ability of CNNs to extract complex and hierarchical features, significantly improving classification performance (Hershey et al., 2017).

Traditional model architectures, such as ResNet (He et al., 2016) and EfficientNet (Tan and Le, 2019), played a crucial role in this study. These models demonstrated a strong ability to generalize to data represented as images, even when derived from non-visual sources such as audio (Choi et al., 2017a). The application of CNNs for spectrogram classification proved particularly effective, capturing temporal and

frequency patterns in sound signals that are vital for identifying musical genres. Additionally, the transfer learning technique might enable the reuse of pretrained models on image datasets such as ImageNet, reducing computational costs and increasing accuracy on the specific datasets used in this study.

3.3 Optimizing Model Architecture Using Genetic Algorithm

Unlike the grid search method, which performs an exhaustive and predefined search across all hyperparameter combinations within a set range, GAs employ a stochastic approach inspired by natural evolution (Darwin, 2023). These algorithms operate based on a population of individuals, each representing a potential solution to the problem. Through operators such as selection, crossover, and mutation, GAs efficiently explore the search space, dynamically adapting to find near-optimal solutions even in high-dimensional spaces, such as hyperparameter tuning for a CNN (Aszemi and Dominic, 2019). This approach allows for the discovery of hyperparameter configurations that may not be easily identifiable through traditional methods.

GAs rely on a binary or numerical representation of individuals—each individual being a sequence of encoded parameters—which simplifies the implementation of mutation and crossover operations that generate potential solutions. The selection process favors individuals with higher fitness values, ensuring that well-performing configurations are propagated through generations. Figure 6 illustrates a flowchart of how a GA is implemented. The genes chosen for optimization involved essential parameters for configuring the model. The following subsections detail the four primary genes optimized: the model architecture, learning rate, weight decay, and training batch size.

The first gene to be optimized refers to the selection of the neural network model to be used. For the task of music genre classification from spectrograms, three different CNN models were chosen: ResNet50, EfficientNetB0, and another baseline model. The optimization of this gene aims to determine which of these architectures is most suitable for the specific task, based on their performance during training and validation. The choice of model directly impacts the network's ability to capture relevant features from the spectrograms and, therefore, is crucial for the effectiveness of the classification.

The second gene to be optimized is the learning rate, a critical hyperparameter that governs the speed of convergence during training. Proper adjustments to the learning rate can significantly enhance the model's

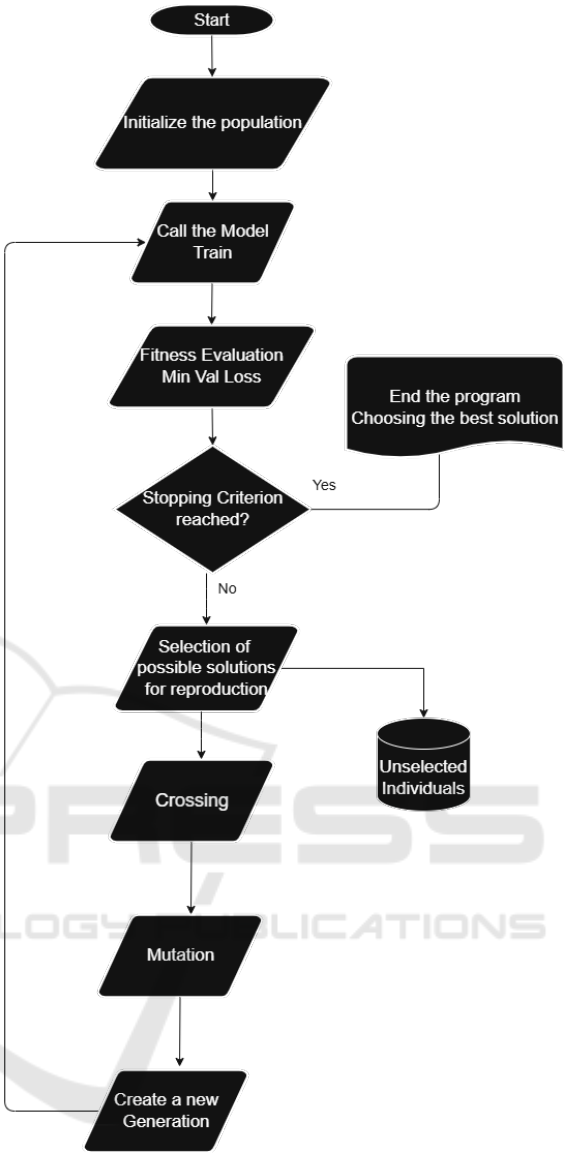

Figure 6: Flowchart of the Genetic Algorithm Used.

convergence, preventing both premature convergence, which may result in suboptimal solutions, and excessive error oscillation, which hinders stable training.

The third gene to be optimized is the weight decay, a regularization technique that prevents overfitting by penalizing large weights in the network. By finding an optimal balance, weight decay helps maintain a model that generalizes well to unseen data while avoiding excessive complexity that could lead to poor performance on new inputs.

The fourth gene to be optimized is the training batch size, a parameter that defines the number of samples processed by the model before updating its weights. Smaller batch sizes can improve generaliza-

tion but may lead to noisier updates, whereas larger batch sizes offer more stable updates but require careful tuning to avoid memory limitations and training inefficiencies. By optimizing this parameter, the genetic algorithm seeks to find a balance between computational efficiency and model performance.

4 EXPERIMENTS

In this section, the experiments conducted for the task of music genre classification are described, including the dataset used, the division of the data sets, and the execution environment. To ensure transparency and facilitate further research, all code, scripts, and configurations used in this study have been made publicly available. [1]

4.1 Dataset

The dataset used in the experiments was GTZAN, widely recognized in the literature as a standard dataset for music genre classification. The dataset contains 1,000 audio files, evenly distributed across 10 music genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Each file is 30 seconds long and is in WAV format, with a sampling rate of 22,050 Hz. The choice of GTZAN is due to its relevance and the diversity of the music genres, which allows for a comprehensive evaluation of the model across different styles. The dataset was split into 70% for training, 20% for validation, and 10% for testing.

4.2 Experiment Environment

The entire experimental procedure was conducted in the Google Colab environment, which provides free GPU support, facilitating the training of deep learning models. Google Colab was chosen for its accessibility, integration with popular machine learning libraries such as PyTorch and TensorFlow, and support for cloud collaboration.

The experiments were segmented to optimize the use of available computational resources, including the pre-processing of audio spectrograms, configuring the GA with the PyGAD library, and training the CNN. Additionally, intermediate results were periodically saved to Google Drive to ensure data integrity and facilitate later analysis.

[1] https://github.com/ciaograsso06/Music-Genre-Classification.git

Table 1: Number of training epochs used in each experiment during the optimization of CNN models.

Experiment	Epochs
Experiment 1	7
Experiment 2	10
Experiment 3	12
Experiment 4	15
Experiment 5	20

To evaluate the performance of the music genre classification model and the effectiveness of the GA in hyperparameter optimization, four distinct experiments were conducted. Each experiment varied the number of training epochs (as shown in Table 1) and produced specific configurations of optimized hyperparameters, such as the model, learning rate, and weight decay. Therefore, the training epochs were not considered as a gene for the optimization step.

For each experiment, the GA aim to optimize the defined hyperparameters, and the model was evaluated based on the validation loss.

4.3 Experimental Protocol

In this study's methodology, the PyGAD library (Gad, 2023) was utilized to implement the GA, enabling the efficient optimization of hyperparameters for the CNN model. This library provides a robust and flexible interface for defining genes, genetic operators, and evaluation criteria, facilitating experimentation with various configurations.

The genetic operators used in PyGAD were the default configurations, requiring no additional setup. These include tournament selection with a size of 2, single-point crossover with a crossover probability of 0.8, and random mutation with a mutation probability of 0.1. The GA was executed for a total of 5 generations, ensuring sufficient exploration of the hyperparameter search space while maintaining computational efficiency.

5 DISCUSSION AND RESULTS

The detailed results of each experiment are presented in this section, including tables and comparative analyses. These tables contain metrics such as accuracy and validation loss for the optimized solution, enabling a detailed analysis of the impact of each experiment on model performance. The Table 2 presents the backbone chosen as the model to be finetuned. Besides, the best fitness of each experiment can be observed in Table 3 and analyzed in more detail in Figure 7.

Table 2: Chosen models in the optimization process.

Experiment	Model
1	google/efficientnet-b0
2	google/efficientnet-b1
3	google/efficientnet-b0
4	microsoft/resnet-50
5	google/efficientnet-b0

Table 3: Parameters Selected by the GA per Experiment and the Corresponding Fitness.

Exp.	LR	W. Decay	Batch Size	Fitness
1	0.0033	0.0072	16	0.8096
2	0.0014	1e-06	22	0.8098
3	0.0012	1e-06	28	0.8159
4	0.0020	1e-06	25	0.7407
5	0.0003	0.0003	16	0.7370

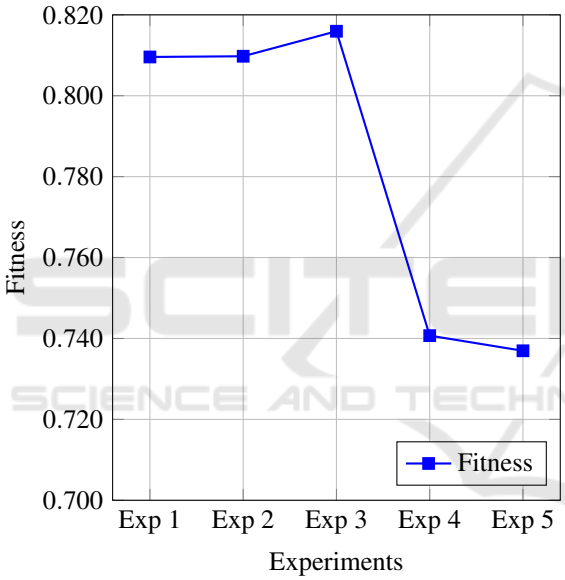

Figure 7: Fitness x Experiments graphic.

Once the GA finished the optimization process, training was carried out with these parameters, and the model was subsequently tested on the test set, with the results shown in Table 4.

Table 4: Test Accuracy for each experiment.

Exp.	Run Time	Test Accuracy
1	32 min 25 sec	66.0000%
2	39 min 14 sec	67.8201%
3	42 min	70.0000%
4	50 min 45 sec	72.0000%
5	1h 10 min	79.4126%

Although the importance of proper training with an adequate number of epochs is evident, it is worth noting that the use of the GA for selecting the best solutions had a significant positive impact. Even with a reduced number of training epochs, the GA was able to identify model configurations that achieved acceptable accuracy results, demonstrating the algorithm's effectiveness in optimizing the training process and obtaining robust models in less time.

By exploring a wider range of hyperparameter configurations, the GA was able to identify high-performing models that would have otherwise been overlooked. This not only reduced the overall training time but also ensured that the models achieved competitive accuracy with fewer iterations, validating the GA's potential as a powerful tool for optimizing DL workflows.

6 CONCLUSION

In this work, we explored the use of GA for hyperparameter optimization of convolutional neural networks in the music classification task. The experiments conducted demonstrated the efficiency of GA in identifying hyperparameter configurations that minimize the validation loss function across different models, such as the EfficientNet and ResNet architectures.

The results indicated that GA successfully identified combinations of learning rate, weight decay, and batch size that significantly improved model performance, as evidenced by reduced validation loss values. The EfficientNet-b0 model exhibited consistent performance across multiple experiments, highlighting its robustness for tasks with different configurations. More complex architectures, such as ResNet-50, also benefited from optimization, underscoring the applicability of GA to various models. This study reaffirms that GA is a powerful tool for hyperparameter optimization strategy. The approach is particularly useful in scenarios where traditional methods, such as grid search and random search, would be computationally expensive or less effective.

For future work, this approach will be evaluated on additional datasets, comparing the performance of the GA-based hyperparameter optimization with other methods presented in the literature, including simpler techniques. This will provide further insights into the generalizability and robustness of the method across different music genre classification tasks. Additionally, Neural Architecture Search (NAS) will be explored to optimize the model architecture within a multi-objective framework, considering not only classification accuracy but also hardware constraints such as model size and inference time. Evaluating these techniques in various domains will further clar-

ify their comparative advantages in terms of computational efficiency and model performance.

ACKNOWLEDGEMENTS

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento e Pesquisa (CNPq) under Grant 140254/2021-8, and Fundação de Amparo à Pesquisa do Rio de Janeiro (FAPERJ)

REFERENCES

Aszemi, N. M. and Dominic, P. (2019). Hyperparameter optimization in convolutional neural network using genetic algorithms. *International Journal of Advanced Computer Science and Applications*, 10(6).

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of machine learning research*, 13(2).

Choi, K., Fazekas, G., Sandler, M., and Cho, K. (2017a). Convolutional recurrent neural networks for music classification. In *2017 IEEE International conference on acoustics, speech and signal processing (ICASSP)*, pages 2392–2396. IEEE.

Choi, K., Fazekas, G., Sandler, M., and McFee, M. B. (2017b). A tutorial on deep learning for music information retrieval. *ACM Computing Surveys*, 51(1):1–34.

Darwin, C. (2023). Origin of the species. In *British Politics and the environment in the long nineteenth century*, pages 47–55. Routledge.

Dieleman, S., Brakel, P., and Schrauwen, B. (2011). Audio-based music classification with a pretrained convolutional network. In *12th International Society for Music Information Retrieval Conference (ISMIR-2011)*, pages 669–674. University of Miami.

Gad, A. F. (2023). Pygad: An intuitive genetic algorithm python library. *Multimedia Tools and Applications*, pages 1–14.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 770–778.

Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., and Slaney, M. (2017). Cnn architectures for large-scale audio classification. *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 131–135.

McFee, B., Jülke, L., Salamon, J., and Ellis, D. P. (2017). Learning multi-scale temporal features for music information retrieval. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, pages 252–258.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. (2015). librosa: Audio and music signal analysis in python. *Proceedings of the 14th python in science conference*, 8(1):18–25.

Müller, M. (2015). *Fundamentals of music processing: Audio, analysis, algorithms, applications*, volume 5. Springer.

Pons, J., Slizovskaia, O., Gong, R., Gómez, E., and Serra, X. (2017). Timbre analysis of music audio signals with convolutional neural networks. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 2744–2748. IEEE.

Real, E., Aggarwal, A., Huang, Y., and Le, Q. V. (2019). Regularized evolution for image classifier architecture search. In *Proceedings of the aaai conference on artificial intelligence*, volume 33, pages 4780–4789.

Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302.

Young, S. R., Rose, D. C., Karnowski, T. P., Lim, S.-H., and Patton, R. M. (2015). Optimizing deep learning hyper-parameters through an evolutionary algorithm. In *Proceedings of the workshop on machine learning in high-performance computing environments*, pages 1–5.