An AI based approach to classifying genre and emotion of music using spectrograms

Pranavi Tadigadapa

Abstract

Music classification has become an essential feature of modern technology, from recommendation systems and mood-based playlists to therapeutic interventions. Beyond genre classification, recent work emphasizes the importance of recognizing the emotional tone of music. This study investigates whether spectrogram-based image classification can be applied using simple, accessible AI tools. Specifically, I ask: (1) Can spectrograms serve as reliable features for both genre and emotion classification? and (2) How accurate can a lightweight AI tool be in this task? Using the Emotify dataset, I trained two models—one for genre and one for emotion. Results showed moderate success in genre classification (57% accuracy vs 25% chance levels) but poor performance in emotion classification (21% accuracy vs 11% chance levels). Findings suggest that spectrograms capture some, but not all aspects of genre and emotion related differences, that can be reliably detected using simple AI tools.

1. Introduction

Artificial intelligence (AI) has reshaped how we interact with music. Platforms like Spotify and YouTube use recommendation systems[1] to recommend songs based on individual preferences. Music therapy is increasingly used for improving emotional and mental well-being, and AI is increasingly used to personalize music therapy recommendations based on individual preferences and emotional states. Traditionally, personalization of music preferences often relies on genre metadata, but there has been a shift towards recognizing the emotional characteristics of music.

Here I examine the utility of lightweight AI tools to classify music, and compare their ability to classify music genres and emotions, using statistical properties of the music. Specifically, I utilize spectrograms[2], which are visual representations of the audio frequency spectrum changing over time, and provide a way to represent a song visually[3]. This makes it possible to apply classical image classification models such as convolutional neural networks (CNNs)[4]. This project explores whether simple CNNs can utilize spectrograms effectively for music classification. I focus on two tasks: (1) classifying music by genre, and (2) classifying music by the key perceived emotion.

2. Methods

2.1 Dataset: The dataset used in this project was the Emotify Dataset[5] on induced musical emotion through a game. The dataset consists of 400 song excerpts that are each 1 minute long across 4 different genres: pop, rock, classical, and electronic. Each song excerpt includes metadata including the genre as well as emotions experienced by multiple different participants. This was collected through a game in which the participant could listen to song excerpts and indicate up to 3 emotions they felt most strongly using the GEMS scale (Geneva Emotional

Music Scales)[6]. The songs were classified into the following 9 emotions: amazement, calmness, joyfulness, sadness, nostalgia, power, solemnity, tenderness, tension. For this project, tracks were sorted into their emotions by identifying the most frequently reported emotion for each track.

2.2 Spectrograms: Spectrograms transform audio into a time-frequency visual representation, enabling image-based models like CNNs to process sound data. Because they encode both temporal and frequency information, spectrograms have been widely used in tasks like speech recognition and environmental sound classification. For this project, audio clips were converted to spectrogram images using MATLAB[7]. **Figure 1** shows examples of two spectrograms, the first represents a classical music track, identified by human raters as producing a joyful emotional response, and the second an electronic music track producing a tension-like emotional response.

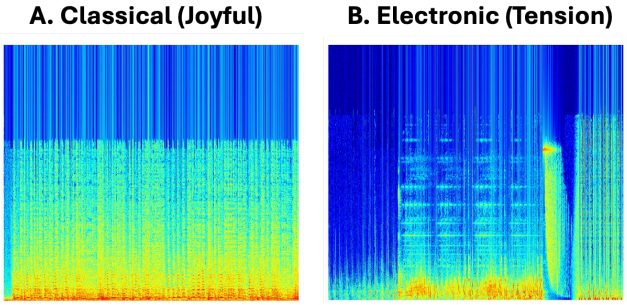

Figure 1: Warmer colors indicate greater amplitudes, while cooler colors indicate lower amplitudes. The x-axis represents time, while the y-axis represents frequencies.

2.3 Model Training and Testing: I used a lightweight online AI tool (Teachable Machine)[8] to specify a CNN model for image classification. I ran two separate models, one for Genre classification (4 categories), and one for Emotion classification (9 categories). Approximately 70% of the spectrograms were used for training and the remaining were used for testing. For genre classification, all genres had equal amounts of data for training. For the emotion classification, some emotions had more data than others (e.g. “Amazement” had the least amount of data for training, while “calmness” had the most).

Model specification: The models were standard image classification models trained using Teachable Machine version 2.4.10. They employed a MobileNet based CNN backbone. 224 x 224 px color images were used as the input and could be exported to TensorFlow, TF Lite, or TF.js formats. Each model size was approximately 5 MB. 4 classes were used to train for genre classification and 9 classes were used to train for emotion classification.

2.4 Evaluation: I examined both the model validation accuracy as well as detailed confusion matrices, which allow us to examine the nature of errors. That is, what type of genres / emotions were misclassified into which categories.

3. Results

Teachable Machine reported its prediction confidence for each track. To evaluate its accuracy, the most confident prediction was compared to the actual track label. For the genre classification, the model had an accuracy of about 57% (compared to chance levels of 25% based on 1 in 4 categories). The machine struggled much more with classifying emotion, with an accuracy of about 21% (compared to chance levels of 11% based on 1 in 9 categories). Thus while both models were significantly better than chance levels, they still produced a significant level of errors. The confusion matrices are reported below.

3.1 Confusion Matrix: Genre Classification:

	Pred. Pop	Pred. Rock	Pred. Electronic	Pred. Classical
Pop (31)	39%	52%	9%	0
Rock (29)	24%	48%	28%	0
Electronic (31)	23%	6%	61%	9%
Classical (29)	14%	0	7%	79%

3.2 Confusion Matrix: Emotion Classification:

	Pred. Amazement	Pred. Calmness	Pred. Joyful	Pred. Nostalgia	Pred. Power	Pred. Sadness	Pred. Solemnity	Pred. Tenderness	Pred. Tension
Amazement (1)	0	0	100%	0	0	0	0	0	0
Calmness (19)	0	32%	11%	5%	0	0	15%	11%	26%
Joyful (18)	0	28%	50%	0	5%	5%	0	0	12%
Nostalgia (11)	0	46%	27%	0	0	9%	9%	0	9%
Power (7)	0	0	43%	0	14%	0	29%	0	14%
Sadness (6)	0	33%	33%	17%	0	0	0	17%	0
Solemnity (6)	0	33%	17%	17%	0	0	0	17%	17%
Tenderness (6)	0	67%	0	17%	0	0	0	17%	0

Tension (9)	0	33%	44%	0	0	0	11%	0	11%

4. Discussion

The model was much better at classifying genre than emotion. For a simple AI tool, it was reasonable at classifying genres, with classical music being the most easily identified and pop being the hardest to identify. Pop music was most often misclassified as rock music.

For emotion classification, basic and universal emotions like *joyful* were the most easily predicted, while emotions like *nostalgia*, which are not just complex, but may also be highly subjective, were most often misclassified. It was an interesting outcome that another basic universal emotion like sadness was highly misclassified. This suggests that some emotions are more strongly tied to personal experience than any elements within the song itself.

A key limitation was the relatively small and skewed sample size for some emotions compared to others. Future work should compare a spectrogram based approach versus a model that can directly process audio data, or use audio features. Future work should also consider more complex model architectures, and explore multi-label classification (simultaneous prediction of genre and emotion).

5. Conclusion

5.1 Educational value: This study demonstrates how complex tasks can be approached with simple, accessible tools, making it a valuable option for students interested in engineering, computer science, and data analysis.

5.2 Engineering perspective: The project highlights how engineering methods, such as transforming audio into spectrograms, allow us to translate one form of information (sound) into another (visual data), to take advantage of existing tools (CNNs).

5.3 Practical applications: Music classification is central to technologies we use daily, from streaming platforms to mental health apps. Understanding the limits and strengths of AI systems is relevant as we design tools that affect real-world usability.

5.4 Broader impact - understanding human emotion: By exploring the challenges of emotion recognition in music, this study underlines the complexity of modeling human experiences and emotions. It provides a window of insights into how different emotions may overlap and be related to each other.

6. References

1. [1] G. Björklund, M. Bohlin, E. Olander, J. Jansson, C. E. Walter, and M. Au-Yong-Oliveira, “An exploratory study on the Spotify recommender system,” in Information Systems and Technologies (WorldCIST 2022), A. Rocha, H. Adeli, G. Dzemyda, and F. Moreira, Eds., Lecture Notes in Networks and Systems, vol. 469. Cham: Springer, 2022, pp. 169–178.
2. [2] S. A. Fulop and K. Fitz, “A spectrogram for the twenty-first century,” Acoustics Today, vol. 2, no. 3, pp. 26–33, 2006.
3. [3] L. Wyse, “Audio spectrogram representations for processing with convolutional neural networks,” arXiv preprint, arXiv:1706.09559, 2017.
4. [4] K. O’Shea and R. Nash, “An introduction to convolutional neural networks,” arXiv preprint, arXiv:1511.08458, 2015.
5. [5] Emotify Dataset. [Online]. Available: https://www2.projects.science.uu.nl/memotion/emotifydata/.
6. [6] A. Lykartsis, A. Pysiewicz, H. von Coler, and S. Lepa, “The emotionality of sonic events: testing the Geneva Emotional Music Scale (GEMS) for popular and electroacoustic music,” in Proc. 3rd Int. Conf. on Music & Emotion (ICME3), 2013.
7. [7] MathWorks, “Spectrograms in MATLAB: Documentation,” [Online]. Available: https://www.mathworks.com/help/signal/ref/spectrogram.html.
8. [8] Google Creative Lab, “Teachable Machine,” [Online]. Available: https://teachablemachine.withgoogle.com/.