

Music Genre Classification with Deep Learning

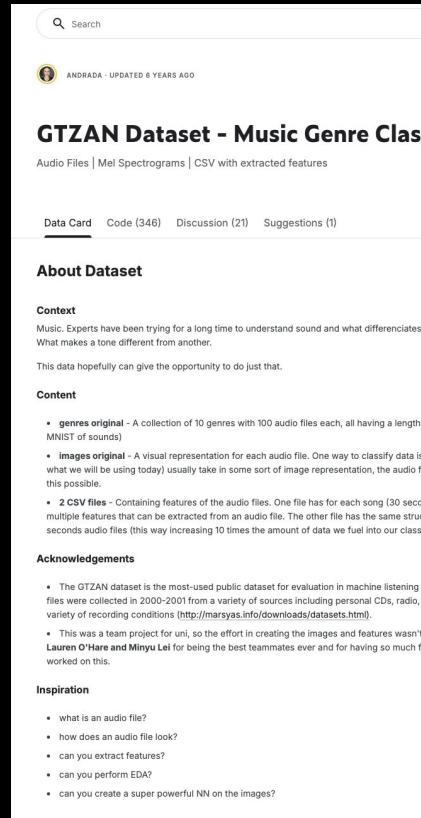
**I'm a Director of User Experience at Google,
and an aspiring student of Artificial
Intelligence at University of Colorado
Boulder, but I'm also a musician with 25
years of guitar playing experience.**



*I play anything
from jazz to metal*

I found on Kaggle the GTZAN Dataset for Music Genre Classification. It's fascinating!

The dataset consists of: 1,000 wav files (100 per 10 genres), generated spectrogram images, CSV file with numeric features for each sample.



A screenshot of a Kaggle dataset page for the GTZAN Music Genre Classification dataset. The page includes a search bar, user profile information, and tabs for Data Card, Code, Discussion, and Suggestions. The main content area shows sections for About Dataset, Context, Content, Acknowledgements, and Inspiration, each containing bullet-pointed lists of details about the dataset.

GTZAN Dataset - Music Genre Class

Audio Files | Mel Spectrograms | CSV with extracted features

Data Card Code (346) Discussion (21) Suggestions (1)

About Dataset

Context
Music. Experts have been trying for a long time to understand sound and what differentiates what makes a tone different from another.

This data hopefully can give the opportunity to do just that.

Content

- **genres original** - A collection of 10 genres with 100 audio files each, all having a length MNIST of sounds
- **Images original** - A visual representation for each audio file. One way to classify data is what we will be using today usually take in some sort of image representation, the audio this possible.
- **2 CSV files** - Containing features of the audio files. One file has for each song (30 seconds audio files (this way increasing 10 times the amount of data we fuel into our classes

Acknowledgements

- The GTZAN dataset is the most-used public dataset for evaluation in machine listening files were recorded in 2000-2001 from a variety of sources including personal CDs, radio, variety of recording conditions (<http://marsyas.info/downloads/datasets.html>).
- This was a team project for uni, so the effort in creating the images and features wasn't Lauren O'Hare and Minyu Lei for being the best teammates ever and for having so much worked on this.

Inspiration

- what is an audio file?
- how does an audio file look?
- can you extract features?
- can you perform EDA?
- can you create a super powerful NN on the images?

The problem

While it's often easy to recognize a genre of music by ear, I was curious if a machine can do it too. This can be useful for automatic classification for services such as YouTube Music or Spotify, but also can be a foundation for all sorts of advanced music tooling (e.g. genre specific track separation).

I decided to use deep learning to solve this issue and build a music genre classifying neural network.

Goal

Build a deep learning model effectively classifying music to genres by analysing spectrograms.

Research Plan

1. **EAD.** After loading the data set I'm going to thoroughly evaluate it. I'll pay a lot of attention to how balanced the data set is (number of observations per type) and visual inspection of spectrograms
2. **Data processing.** I assume that I'll have to process the data to clean it up and prepare it for training.
3. **Multilayer perceptron training.** As a baseline of performance I'm going to train a simple neural net with the numerical features data. I'll use it as a comparison point for my CNN.
4. **CNN training.** I'll define a custom CNN and train it on spectrograms.
5. **CNN optimization.** I'll keep optimizing the CNN until I effectively beat the result of the Multilayer perceptron