Music Genre Classification

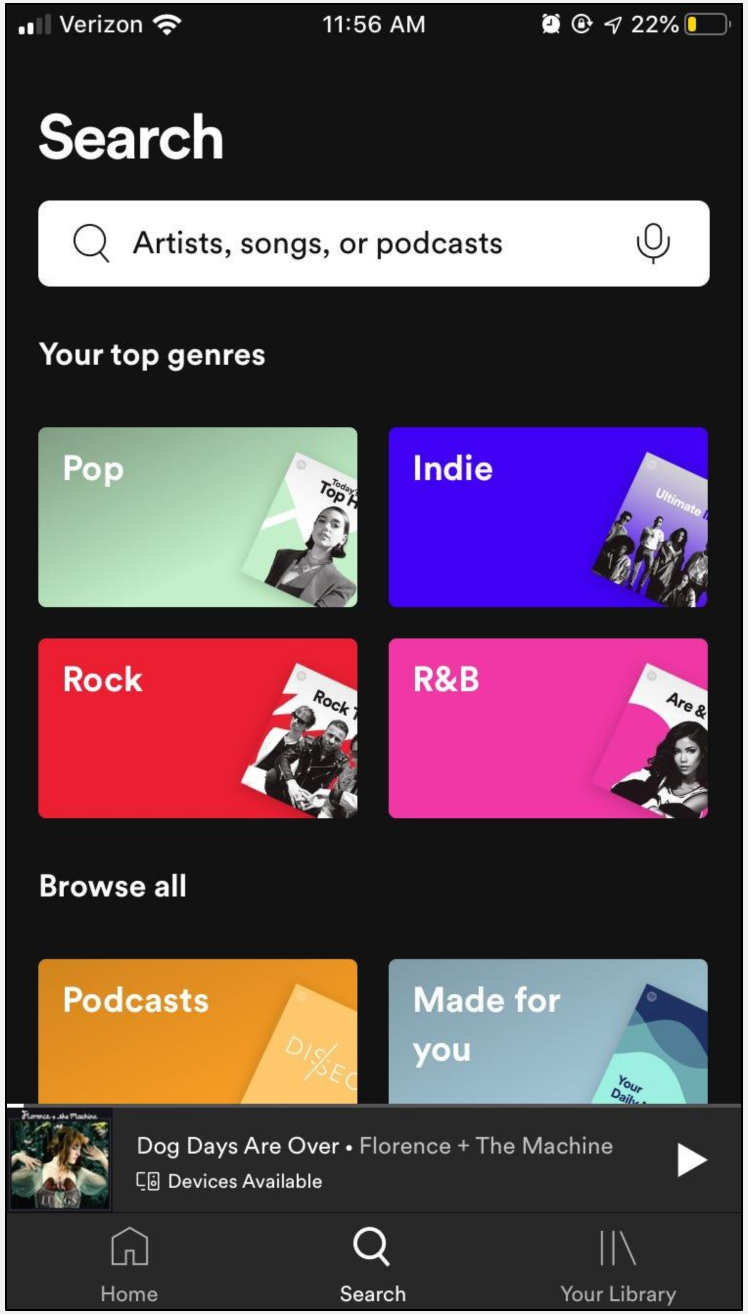
Matthew Caro, Ebube Jack-Davis, Emily Root

Music Genre Classification

- Understanding genres helps to:
  - Personalize **recommendations**
  - **Discover** new music

- Creates better **user experience**, which increases **profit**

- What makes a song **fit into specific a genre**?
  - Instruments used?
  - Artist?
  - Lyrics?
  - Manually label?



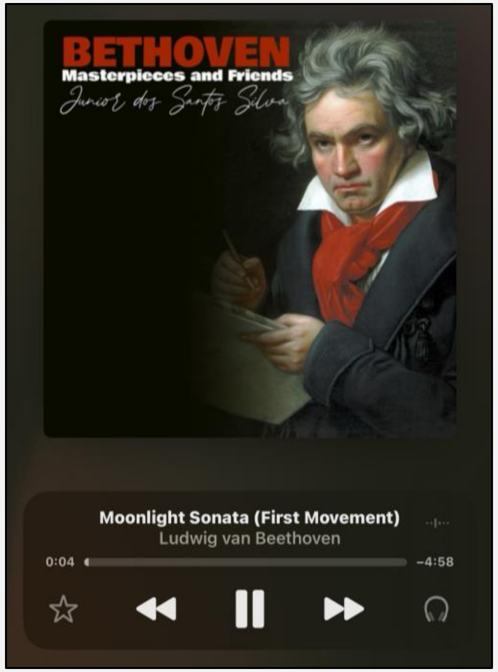
Music Genre Classification

- Understanding genres helps to:
  - Personalize **recommendations**
  - **Discover** new music

- Creates better **user experience**, which increases **profit**

- What makes a song **fit into specific a genre**?
  - Instruments used?
  - Artist?
  - Lyrics?
  - Manually label?

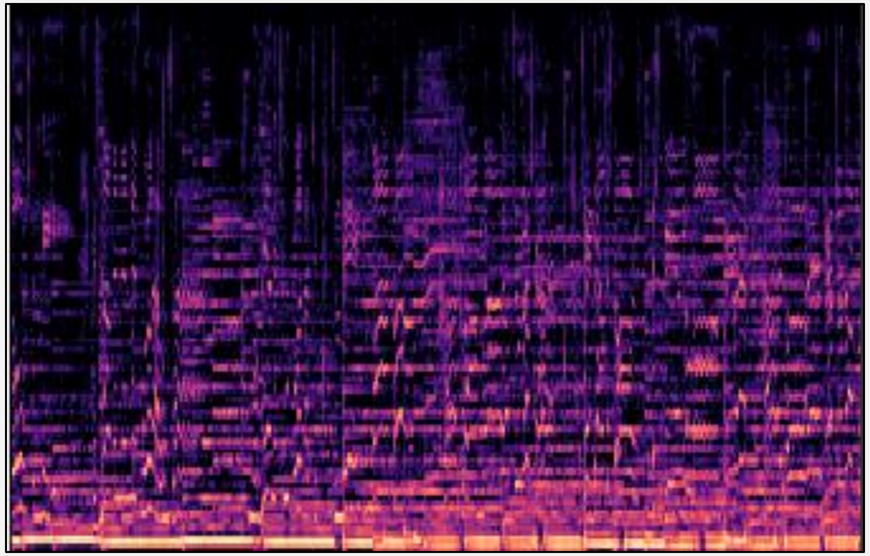
- Spotify looks at **logistics of the song** rather than the **song itself**



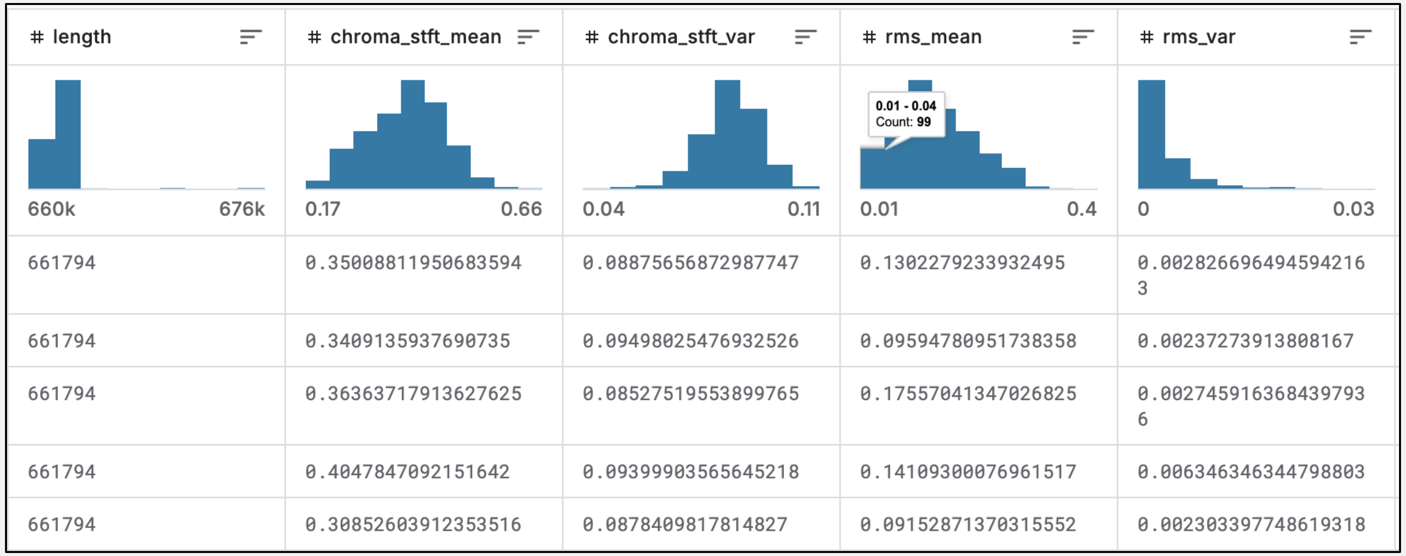
Music Genre Classification – Dataset and Tools

- GTZAN Dataset - collection of **10 genres with 100 audio files each**
  - Blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock

- Looking at data representation of the music
  - **Mel spectrograms** - visualize the frequency of the audio over time
  - **Audio features** - mean, variance, spectral bandwidth, etc.



Spectrogram



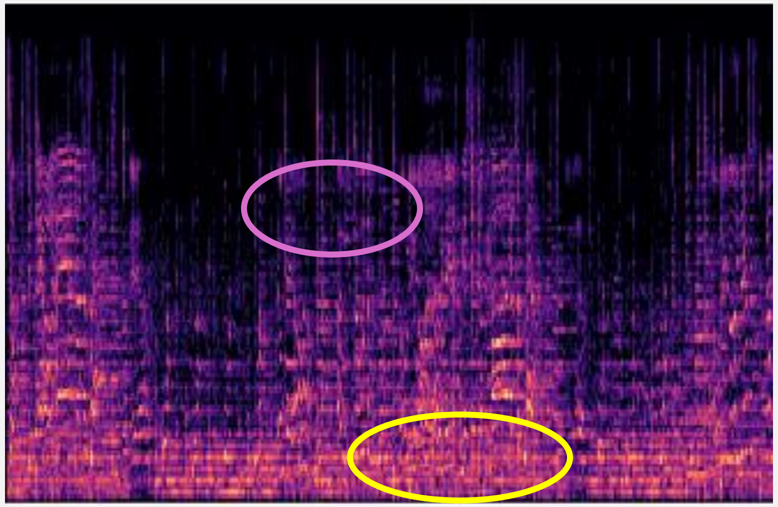
# length	# chroma_stft_mean	# chroma_stft_var	# rms_mean	# rms_var
660k – 676k	0.17 – 0.66	0.04 – 0.11	0.01 – 0.4	0 – 0.03
661794	0.35008811950683594	0.08875656872987747	0.1302279233932495	0.0028266964945942163
661794	0.3409135937690735	0.09498025476932526	0.09594780951738358	0.00237273913808167
661794	0.36363717913627625	0.08527519553899765	0.17557041347026825	0.0027459163684397936
661794	0.4047847092151642	0.09399903565645218	0.14109300076961517	0.006346346344798803
661794	0.30852603912353516	0.0878409817814827	0.09152871370315552	0.002303397748619318

Audio Feature - CSV File

Data Preprocessing

- Resized images to (128, 128, 3)
  - Consistent and small images
  - **3 channels** – RGB
- NumPy array: **Images**, **Features**, and **Labels**
- Normalized the data to ensure **all features carried same weight**
- Split data into **80% training, 20% validation**

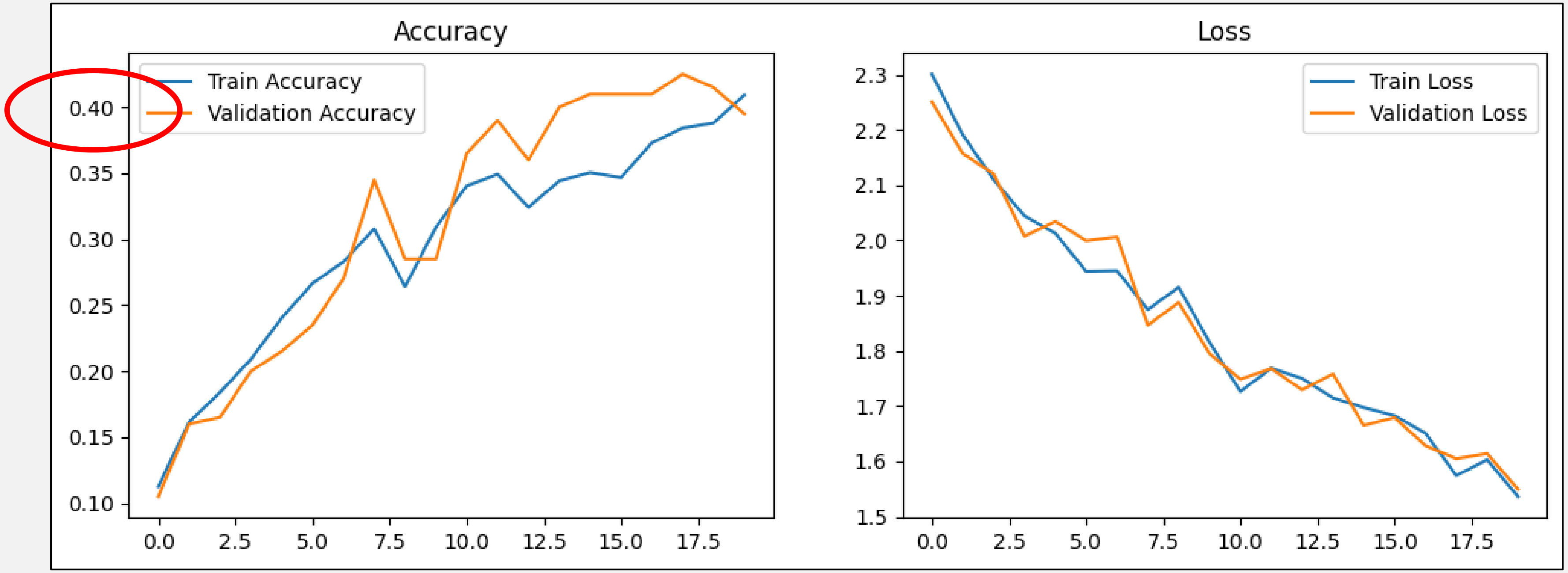
Lower Frequency = Smaller Weight



Higher Frequency = Larger Weight

Classification Algorithms - CNN

- Convolutional Neural Network – **CNN**
  - Parsed through **images** to find patterns
  - Using **only** the CNN was **not able to find a high accuracy (~40%)**



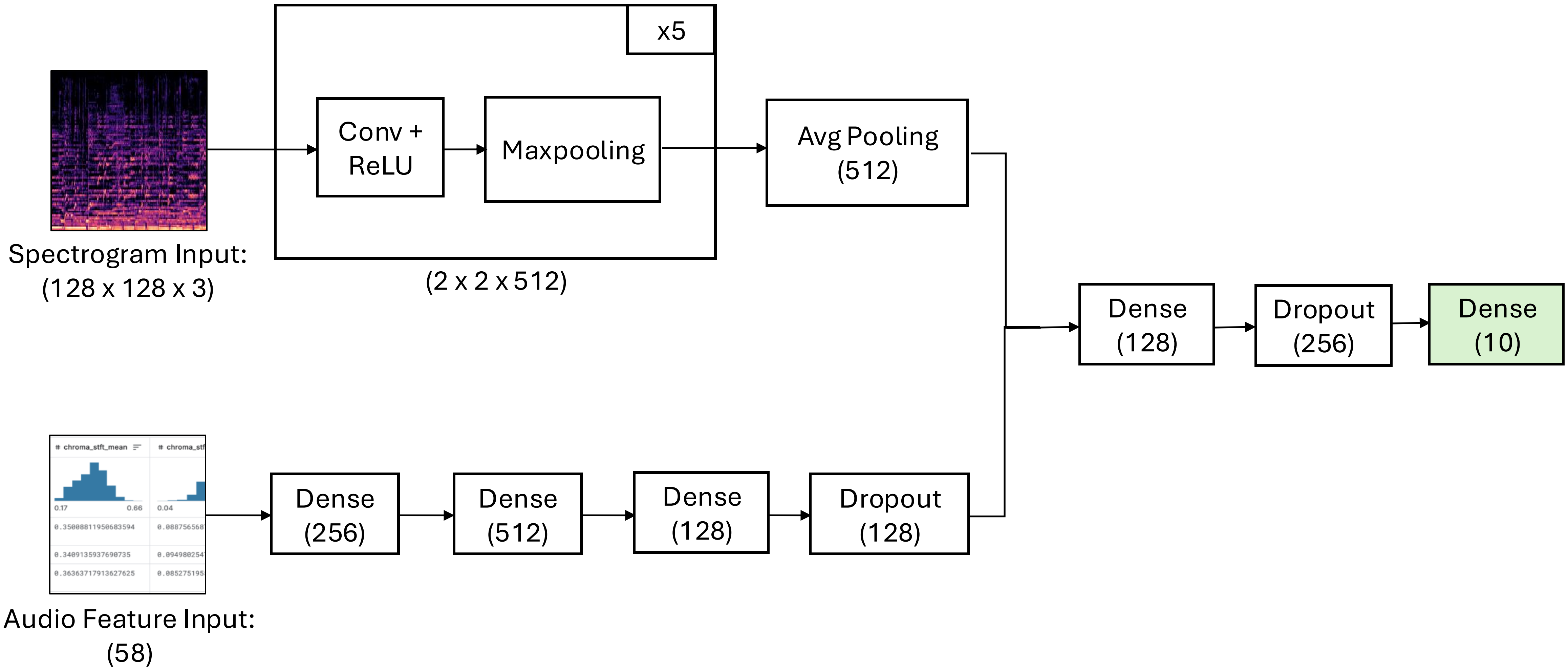
Classification Algorithms – CNN + MLP

- Convolutional Neural Network – **CNN**
  - Parse through the spectrograms to identify **consistent patterns across genres**

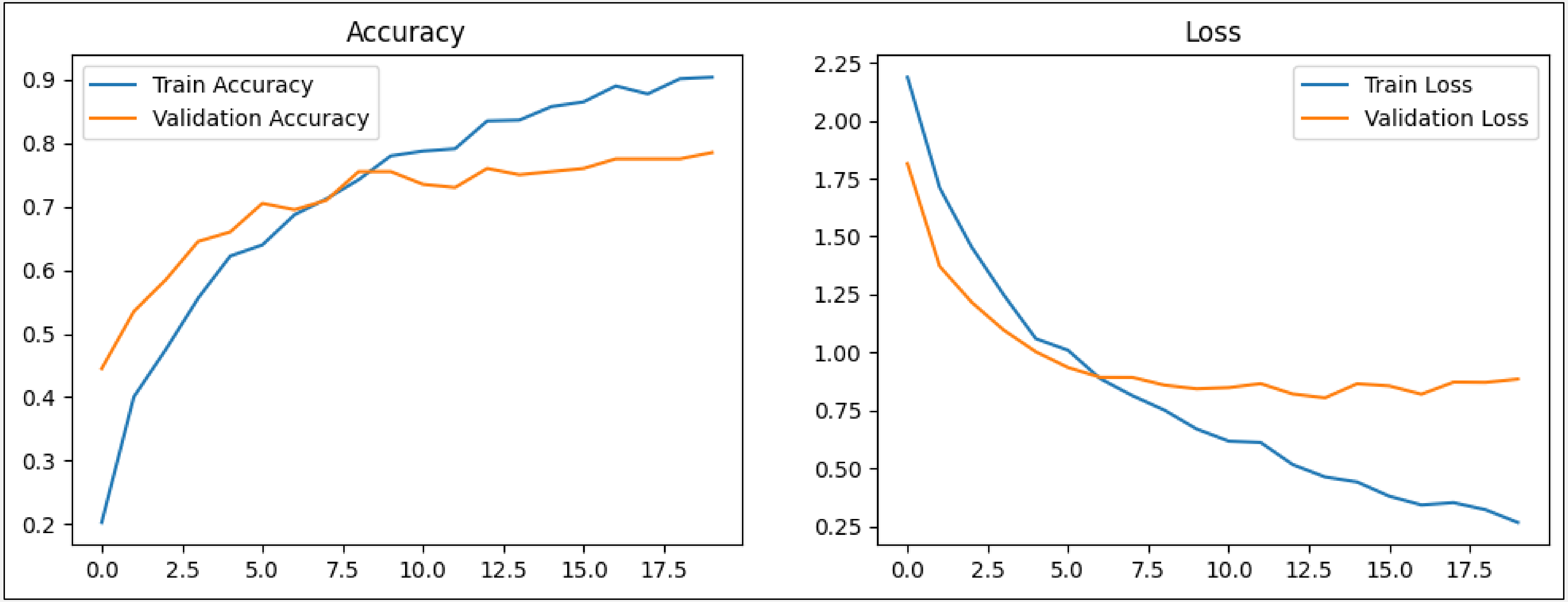
- Multilayer Perceptron (Feed-Forward Network) - **MLP**
  - Look at the audio features that might not appear on the spectrogram (tempo, volume, etc.)

- Together these achieved a **much higher accuracy** than the CNN alone

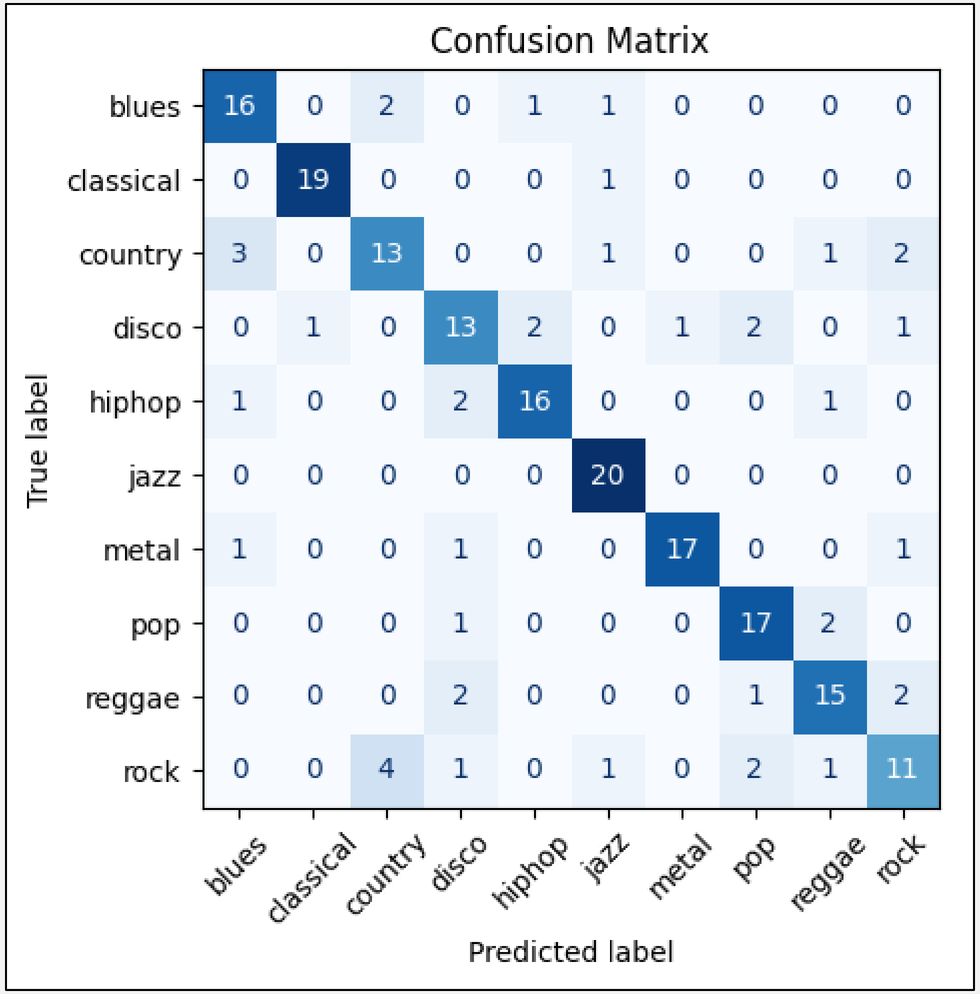
Classification Algorithms



Experiment Results – Accuracy and Loss



Evaluation Metrics



Label	Precision	Recall	F1-Score
Blues	0.75	0.80	0.78
Classical	0.95	0.95	0.95
Country	0.68	0.65	0.67
Disco	0.65	0.65	0.65
Hip-Hop	0.84	0.80	0.82
Jazz	0.83	1.00	0.91
Metal	0.94	0.85	0.89
Pop	0.77	0.85	0.81
Reggae	0.75	0.75	0.75
Rock	0.65	0.55	0.59

Weighted Avg F1 = 0.78

Demo