Music Genre Classification

NIKITA PATEL

The Problem

- 40,000 songs are added to Spotify everyday, the platdorm hosts over 50 million tracks
- There needs to be an autonomous way to sort through enormous amounts of data
- How are deep learning methods used to solve this problem and which are best

Related Work

- [1] Used audio data and social annotations with a series of machine learning algorithms to identify moods and themes in songs
- K-Nearest Neighbors algorithm achieved 69.5% accuracy with K=5 [2]
- [3] Implemented CRNN for overall music tagging and compared it to other CNN architectures
- There is a lot of literature on audio classification and tagging but none use the smae data or pre-processing methods so there is not way to properly compare the models used

Approach

PREPARE DATA

Find a way to represent audio data in a way a learning architecture could interpret

EXPERIMENT WITH MLP

Varying the number of layers and accommodating to overfitting, find an optimal architecture for classification

EXPERIMENT WITH CNN

Find the optimal configuration of a CNN for classification

EXPERIMENT WITH RNN-LSTM

Find the optimal configuration of a RNN for classification

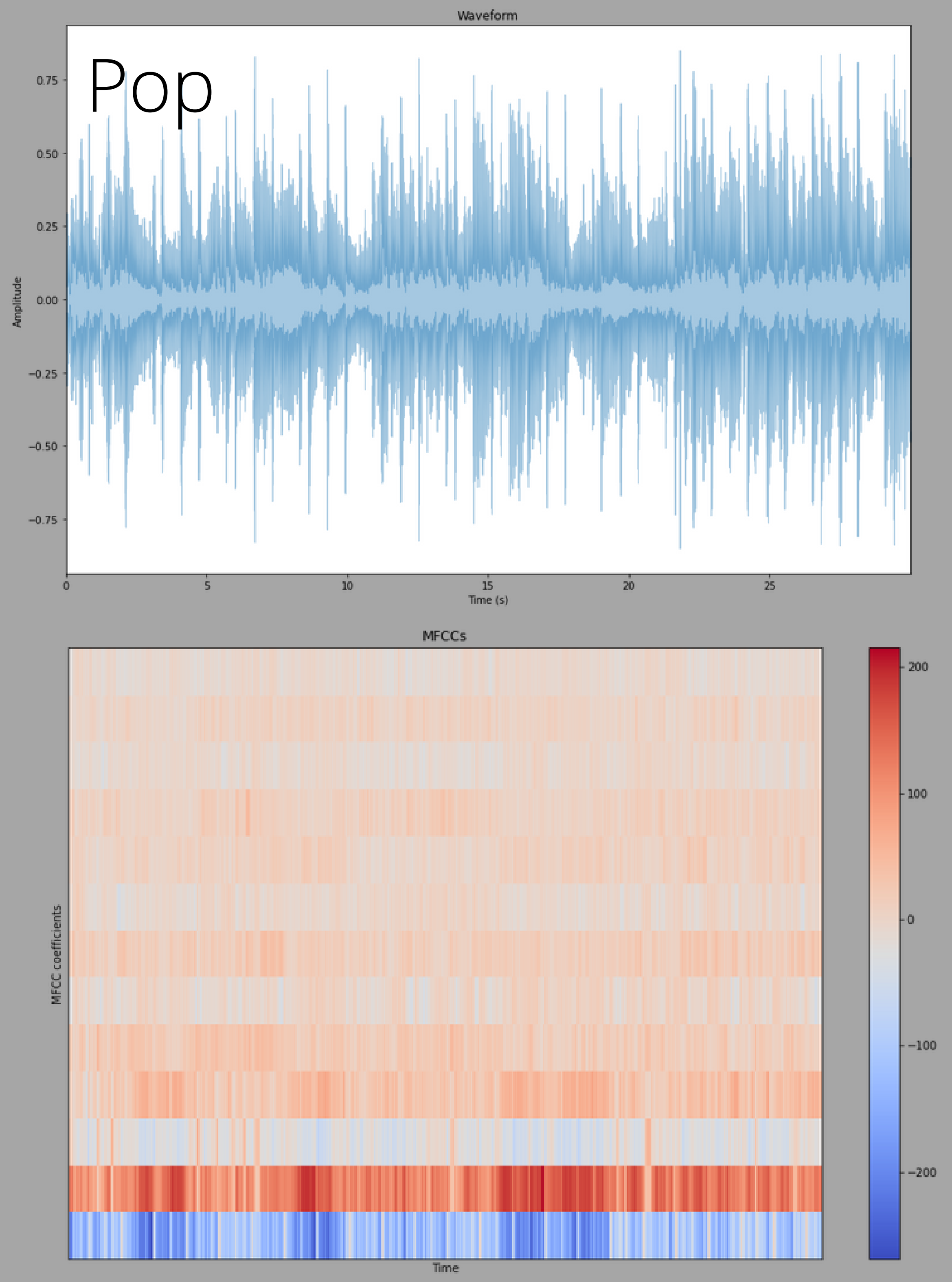
COMPARE AND ANALYZE RESULTS

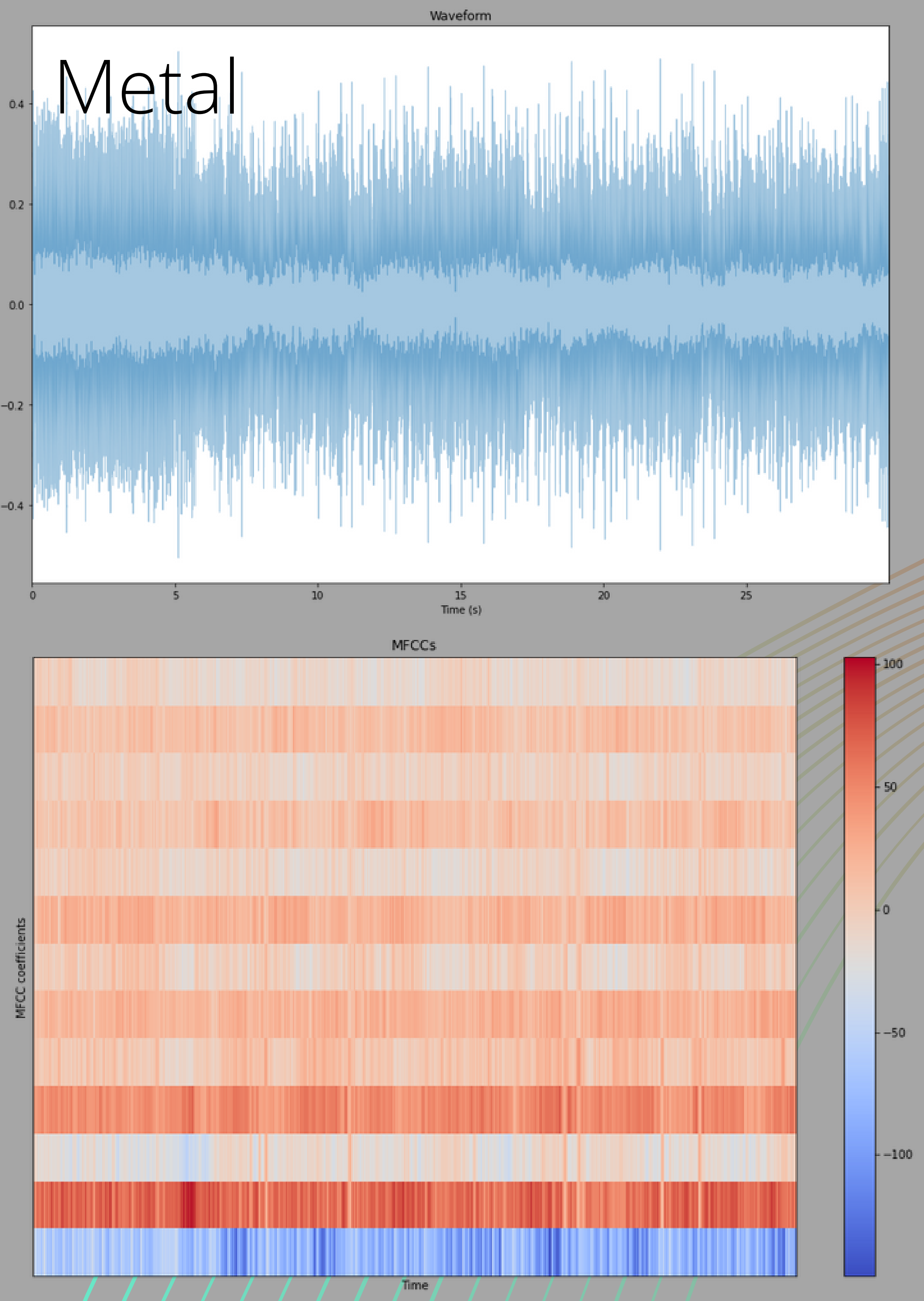
Determine which architectures are superior for audio classification and why

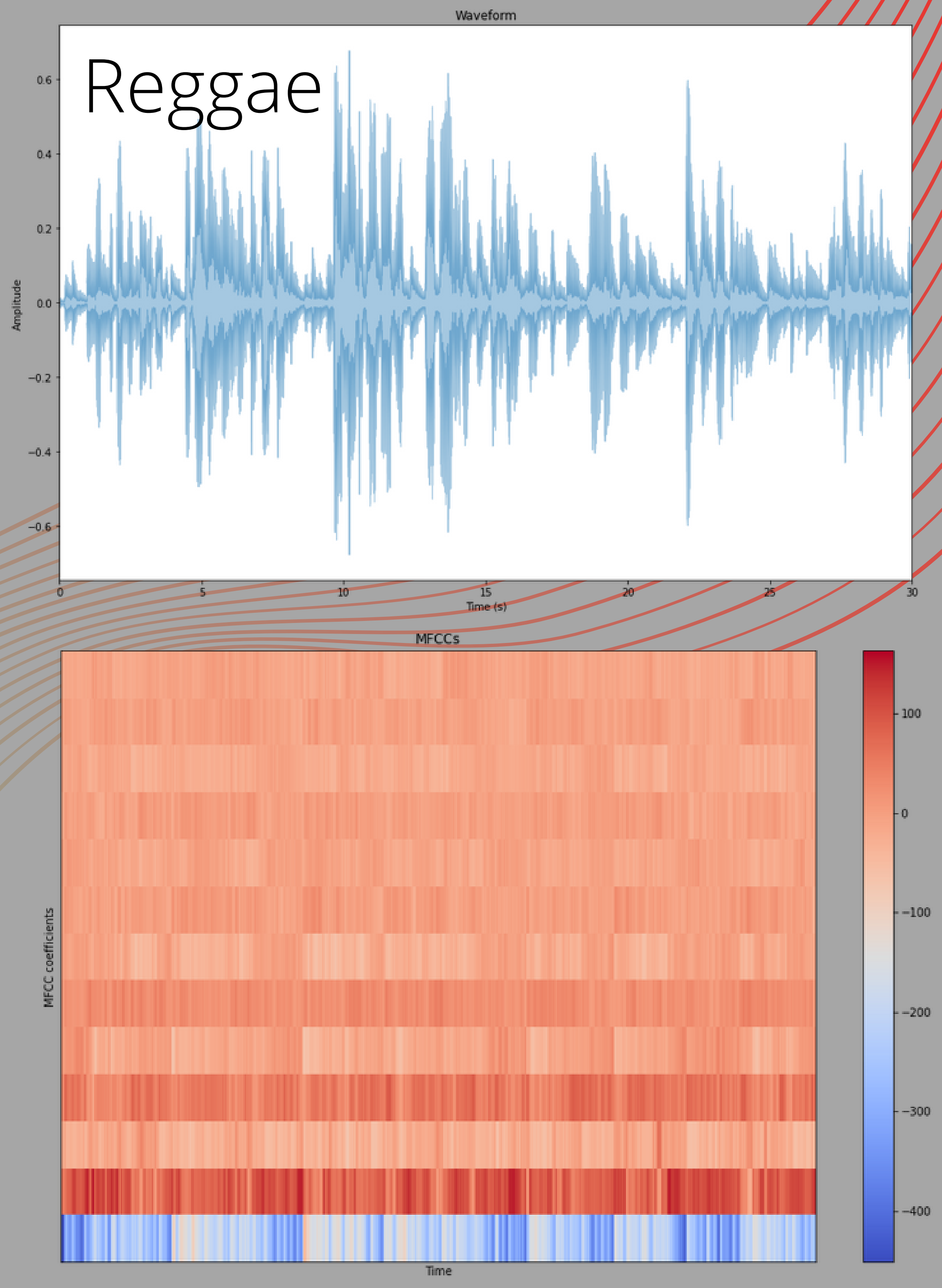
Data

- GTZAN Dataset was used
- 10 genres, 100 audio files per genre, 30s each
- Each audio was broken into 5 segments
- Converted into Mel-frequency cepstrum coefficients (MFCC)
- 30% set aside for tesing

Data Continued







Models

MLP MODEL

Attributes:

- Hidden Layers: ReLu activation function
- Output Layer: Softmax activation function
- Cross-Entropy Objective Function

Experimentation:

- Varied number of hidden layers: 1, 3 and 5
- Overfitting was detected:
  - Regularization Term = 0.0001
  - Dropout added after each hidden layer
    - varied dropout: 0 to 0.5

CNN MODEL

Attributes:

- 3 Convolution Layers: ReLu activation
- Flattening Layer
- 1 Hidden Layer: ReLu activation
- Output Layer: Softmax activation
- Cross-Entropy Objective Function

Experimentation:

- Overfitting was detected:
  - Dropout added after hidden layer
    - varied dropout: 0 to 0.8

RNN MODEL

Attributes:

- 2 LSTM layers
- 1 Hidden Layer: ReLu activation
- Output Layer: Softmax activation function
- Cross-Entropy Objective Function

Experimentation:

- Overfitting was detected:
  - Dropout added after hidden layer
    - varied dropout: 0.1 to 0.5

Results

- MLP objective and accuracy functions need to be re-evaluate
- CNN and RNN had similar accuracy and loss
- Due to training time, CNNs are preferred to RNN model
- Outperformed KNN

Model	Accuracy	Loss
MLP	0.575414	2.691900
CNN	0.760837	0.786095
RNN	0.777377	0.791233

Future Work

- Move away from single genre classification to tagging
- See how transferable this work is to mood identification
- Integrate Natural Language Processing

The End

References

[1] Music Mood and Theme Classification - a Hybrid Approach

[2] Music Genre Classification

[3] Convolutional recurrent neural networks for music classification