Music Genre Classification

Robert Adragna
Bill Sun

Goal and Motivation

- **Correctly Predict** the genre of any piece of Western music.

- **Auto-tag** playlists, **DJ** events.

- Ill-defined problem = good for Machine Learning

Data

- Used Spotify API to collect **raw audio data** in mp3 format

- 10-second long audio clips

- 4000 songs from 4 genres
  - Hip-hop (rap)
 - Rock
  - Classical
  - Jazz

Initial Experiments

- We first tried training on **raw audio** using a 1D CNN

- Then tried training on **Fourier transformed** audio using a 2D CNN

- Both models did not achieve satisfactory results - almost random

- We looked for better solutions

Time Domain
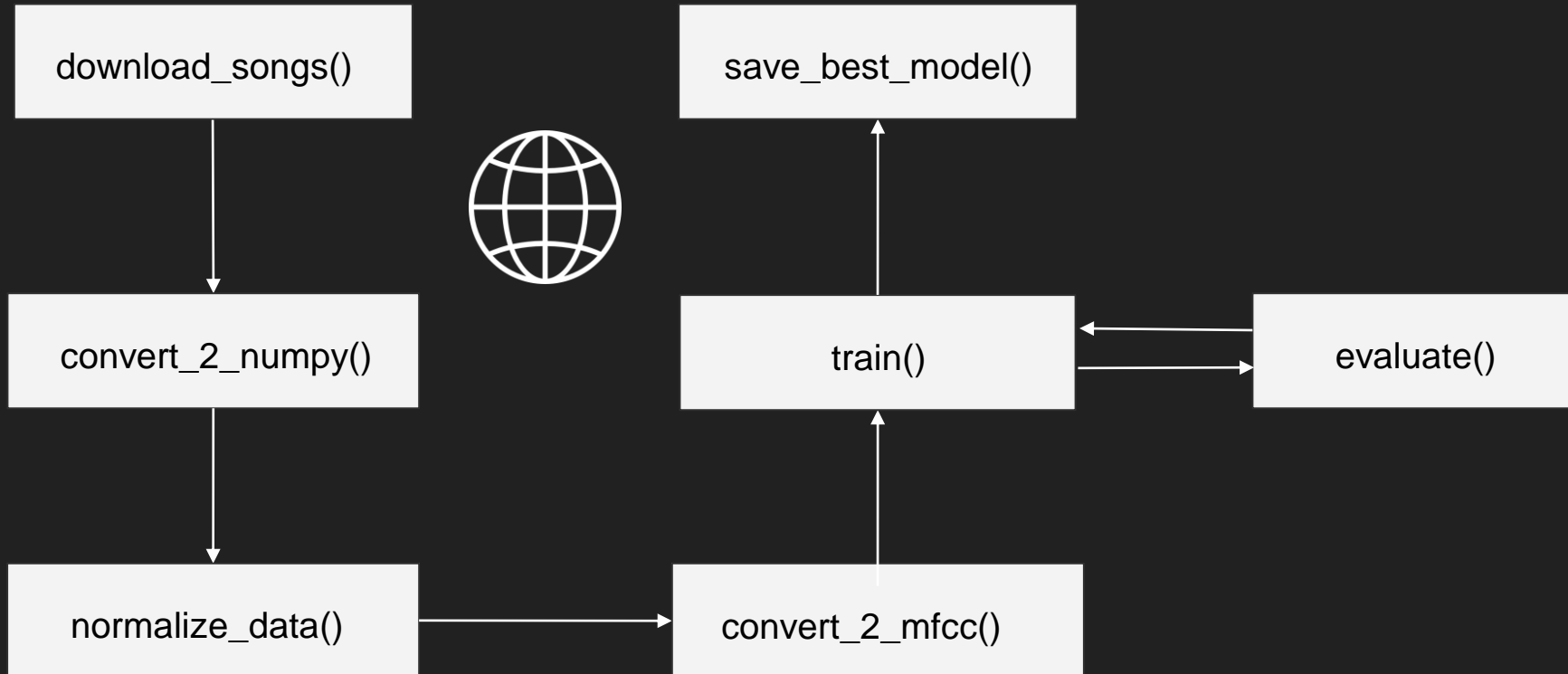s(t)

FT

Frequency Domain
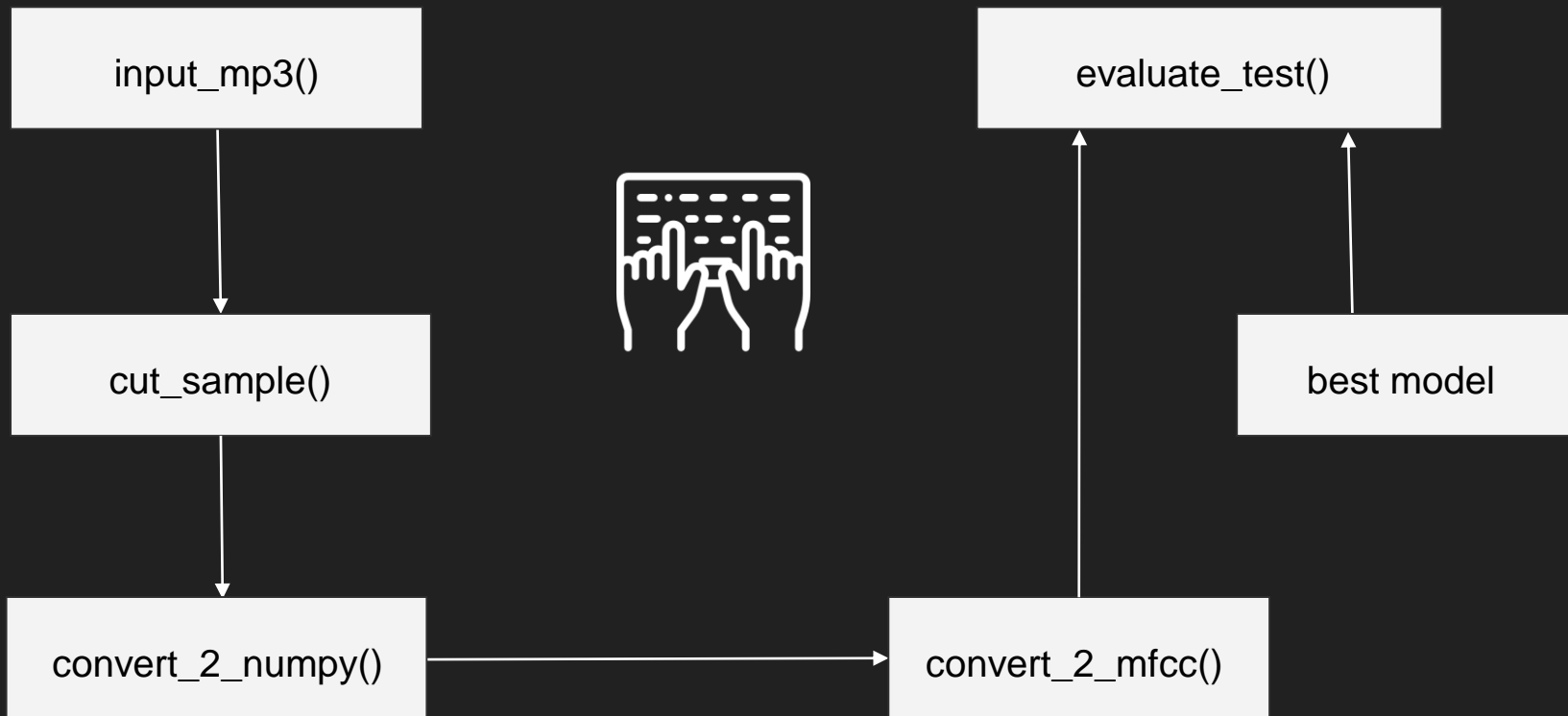S(ω)
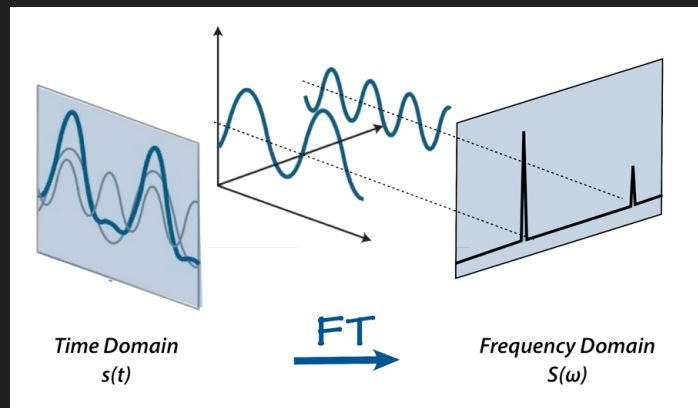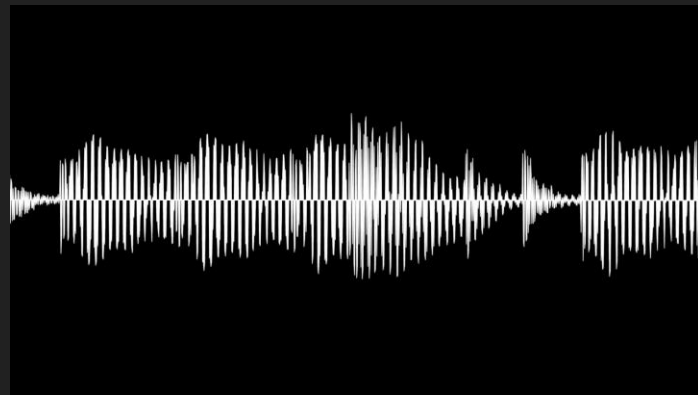
Further Pre-processing

- MFCC: Mel-frequency cepstral coefficients

- A measure of short-term audio power

- We took a time series of MFCC coefficients for every segment of

  0.1 seconds in the audio

- Stack into matrix

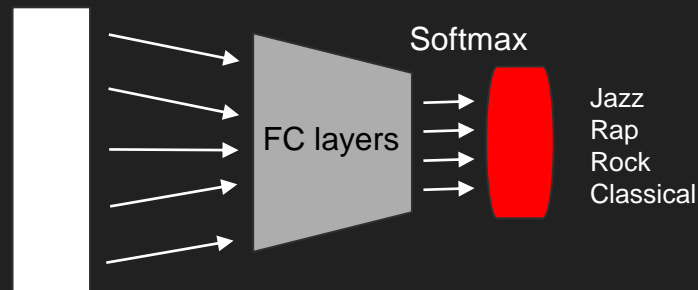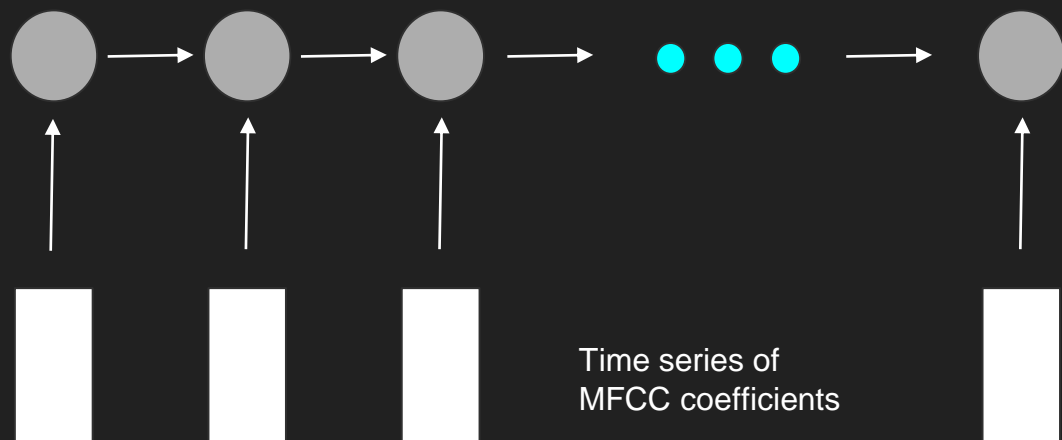MEL POWER OVER TIME
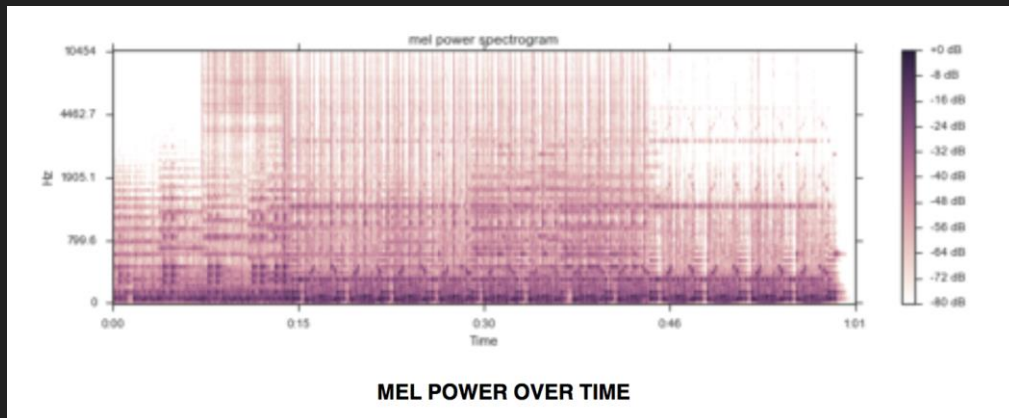
Software Structure Pt.1 - Preprocess & Train

Software Structure Pt.2 - Testing

Model Structure

A multi-layer gated recurrent unit (GRU) RNN followed by fully-connected linear layers

Softmax

FC layers

Jazz
Rap
Rock
Classical

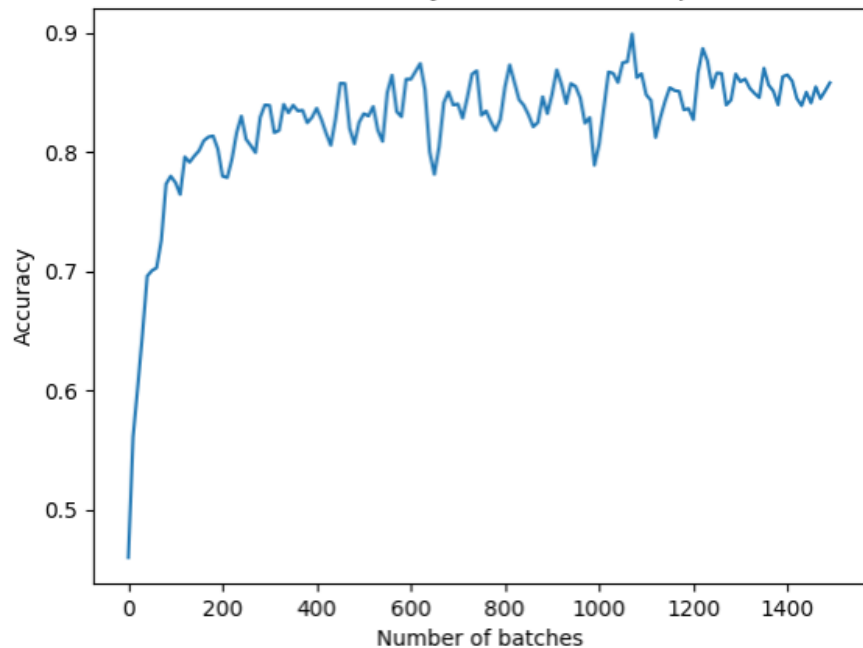Time series of MFCC coefficients
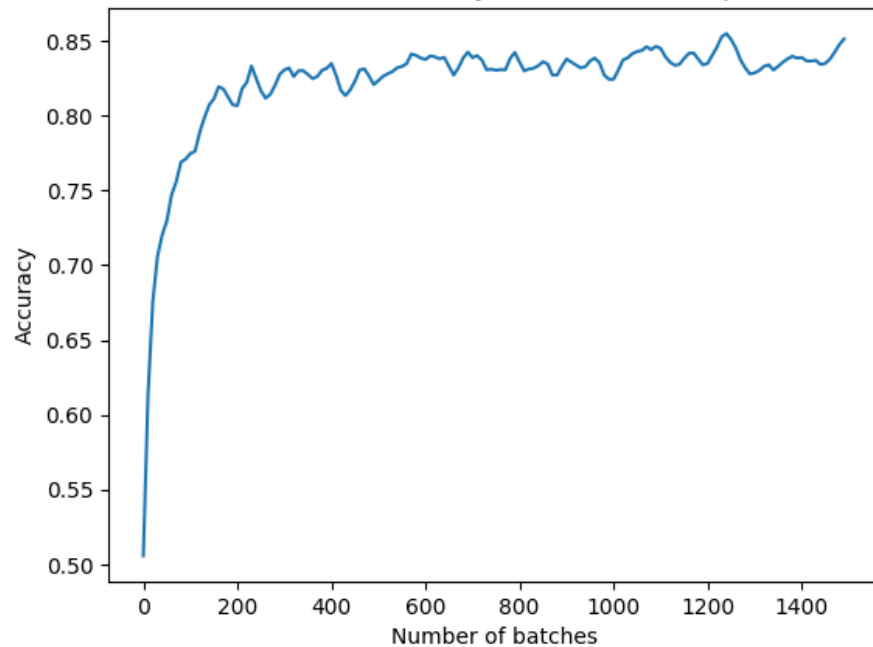
Architecture and Hyper-parameters:

- Batch size = 10
- Learning rate = 0.001
- Num epochs = 10
- Hidden dim = 100
- Embedding dim = 52
- 1 FC layer: size = 100
- Output genres = 4

Training Results

Test Results and Comparisons

- Test set: 640 songs, 160 songs per genre. **72.3% accuracy**

Genre	Classical	Jazz	Rap	Rock	Recall
Classical	107	49	3	1	66.9%
Jazz	33	78	10	39	48.8%
Rap	0	4	150	6	93.8%
Rock	2	7	23	128	80.0%
Precision	75.4%	56.7%	80.6%	73.5%	

Academic standard for
4 genre classification: [1]

63.75%

[1] https://courses.engr.illinois.edu/ece544na/fa2014/Tao_Feng.pdf

Reflections and Takeaways

- Data is **King** (quantity **and** quality)

- ML = trial-and-error + intuition

- Sometimes smaller is better

Demo Time!