About GTZAN Music Genre Dataset

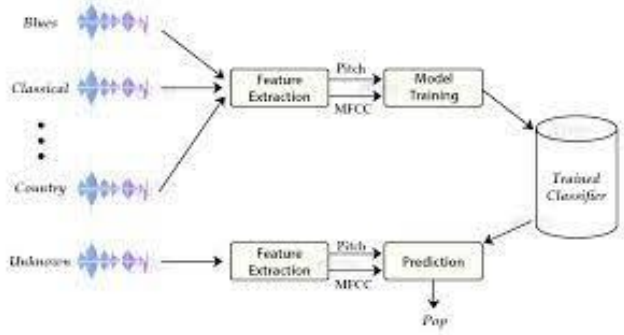
This GTZAN Music Genre Dataset contains 1,000 song samples, each 30 seconds long, belonging to a total of 10 conventional music genres. The samples are classified into blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. The dataset also contains an alternate representation as images of Mel Spectrograms.
Thus, with the GTZAN dataset, we can perform music genre classification using CNN or other computer vision models or use time-series data like MFCCs in training an LSTM or RNN.

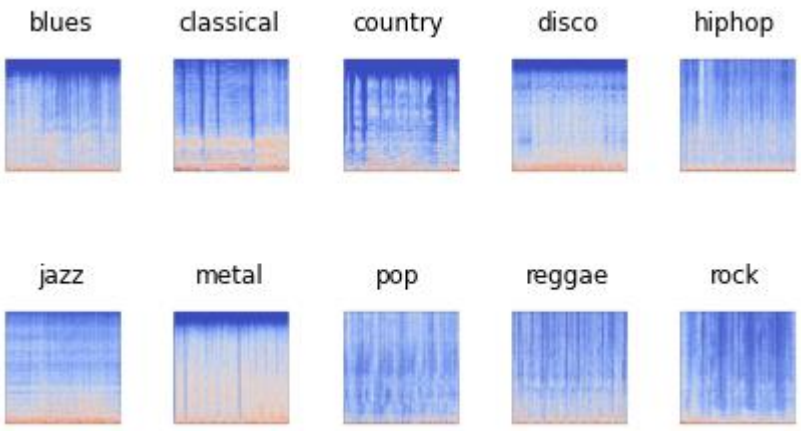
HARDWARE AND SOFTWARE REQUIREMENTS

Minimum Requirements:

a) Hardware: CPU- Intel Core i3 5th gen/ AMD Ryzen 3 1300 or above.
GPU- Nvidia GTX 1050 or above.
RAM- 4 GB or above.
HDD- 2GB free space.

b) Software:
Windows 7/8/8.1/10 (32 bit/64 bit) or Ubuntu Linux or Mac OS.
Python3.1 with Librosa, Keras & Tensor flowlibraries installed.



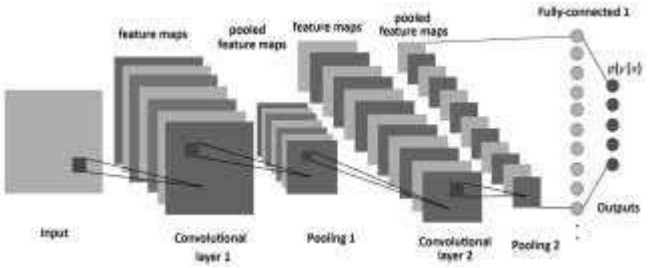


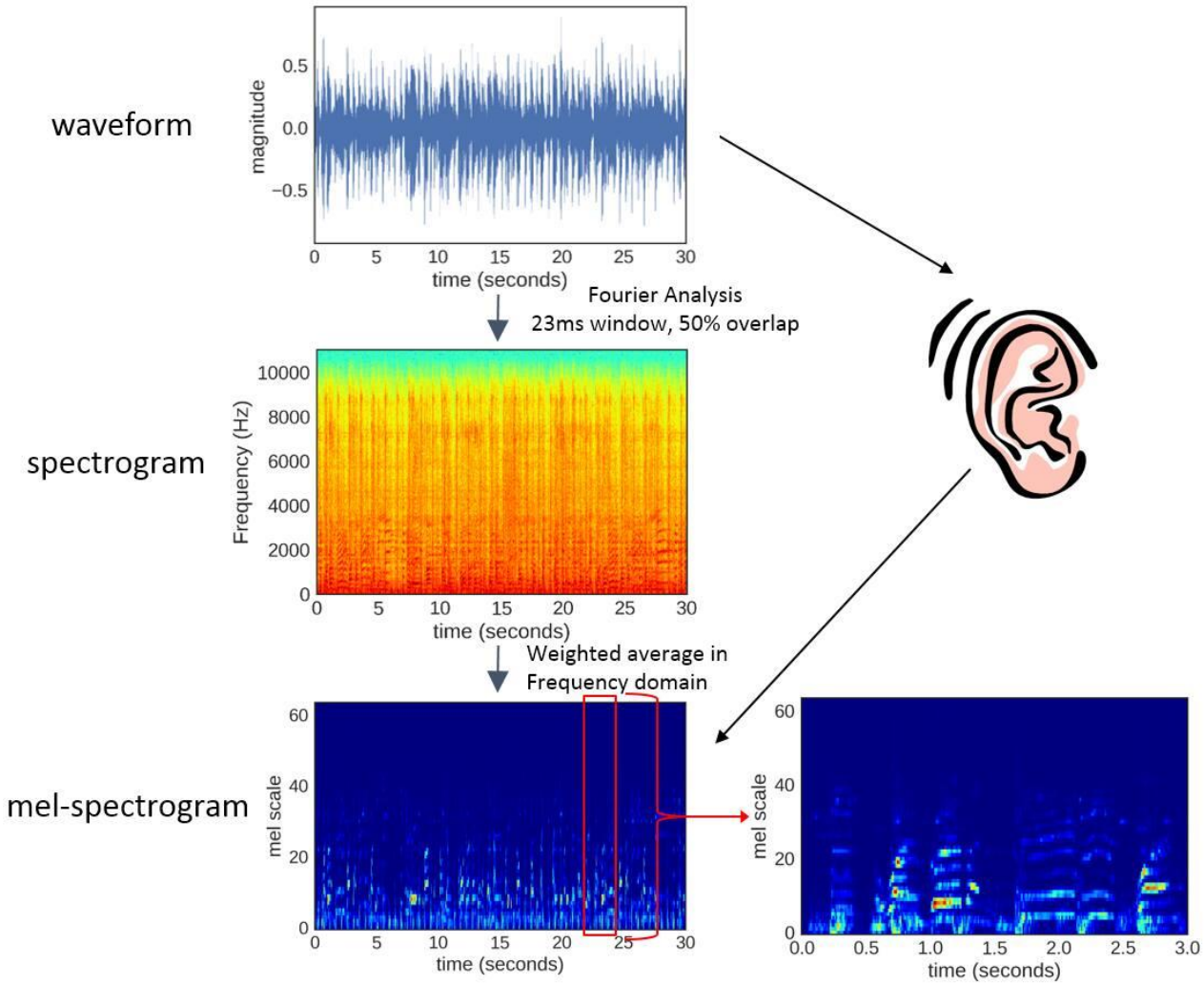
Gtzan music dataset

Librosa
extracts MFCCs

Tensorflow
LSTM (trainning)

Trained LSTM model





Performance analysis:

Model Accuracy			
Model	**Train Accuracy**	**Test Accuracy**	**Remarks**
Simple ANN	92.98%	53.57%	Overfitting
Regularized Simple ANN	68.57%	56.74%	Low Accuracy. Model is not complex enough.
CNN	83.71%	75.43%	Can be improved with more data.
LSTM	90.82%	78.57%	

LSTM MODEL has been selected to be deployed as it has the maximum Test Accuracy of all the models.

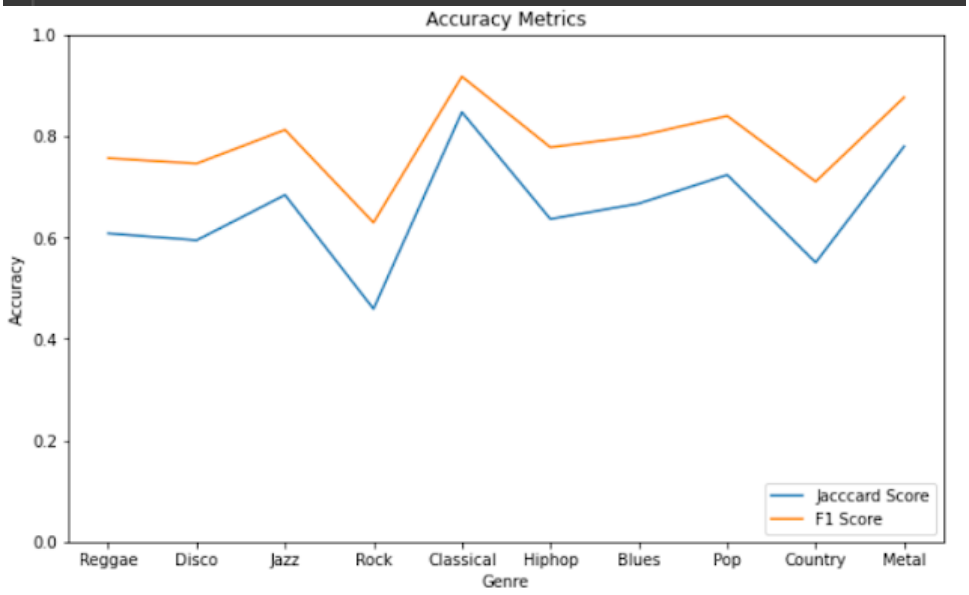
Accuracy Metrics for the LSTM: -

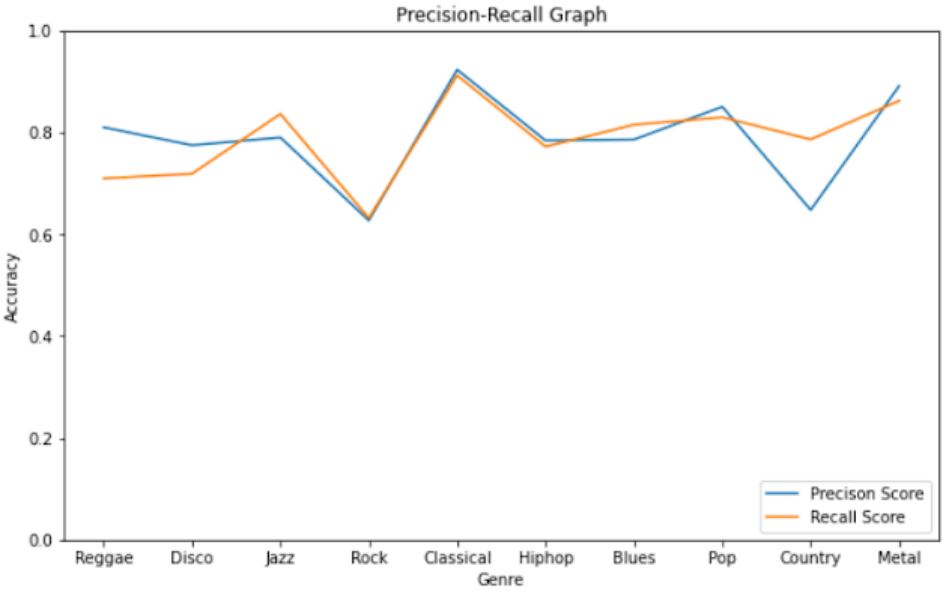
```
score = model.evaluate(X_test, y_test, verbose = 0)

print('Test loss:', score[0])
print('Test accuracy:', score[1])
```

```
Test loss: 0.9284936785697937
Test accuracy: 0.7857428789138794
```

```
[93] print("Model Accuracy Metric: -")

     print(f"Total Accuracy Score              :  {acc_score}" )
     print(f"Weighted Average Jaccard Score    :  {j_score}" )
     print(f"Weighted Average F1 Score         :  {f1}" )
     print(f"Weighted Average Precision Score  :  {precision}" )
     print(f"Weighted Average Recall Score     :  {recall}" )
```

```
Model Accuracy Metric: -
Total Accuracy Score              :  0.7857428914697637
Weighted Average Jaccard Score    :  0.6547766647993974
Weighted Average F1 Score         :  0.786337052332904
Weighted Average Precision Score  :  0.789268106523539
Weighted Average Recall Score     :  0.7857428914697637
```





Model Architecture: -

Model: "sequential_5"

Layer (type)	Output Shape	Param #
lstm_10 (LSTM)	(None, 130, 64)	19968
lstm_11 (LSTM)	(None, 128)	98816
dense_12 (Dense)	(None, 32)	4128
dropout_1 (Dropout)	(None, 32)	0
dense_13 (Dense)	(None, 32)	1056
dropout_2 (Dropout)	(None, 32)	0
dense_14 (Dense)	(None, 12)	396

Total params: 124,364
Trainable params: 124,364
Non-trainable params: 0