Speech Understanding Assignment 1

Vikas Kumar Singh
M23AIR545

Code Repo:

https://github.com/vikas-vikkyd/SU_A1

Objective:

Objective of this assignment was to understand different windowing techniques to process audio data and using the same windowing technique get the STFT of signal, which will help to extract different features present in spectrograms like frequencies and their amplitude at different times. These features will help any Machine learning model to classify which type of sound it is.

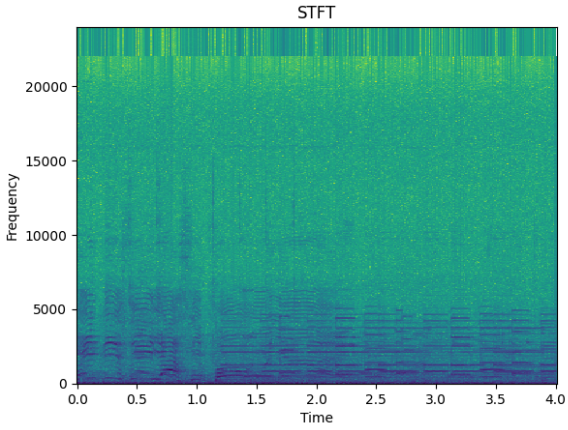
Data:

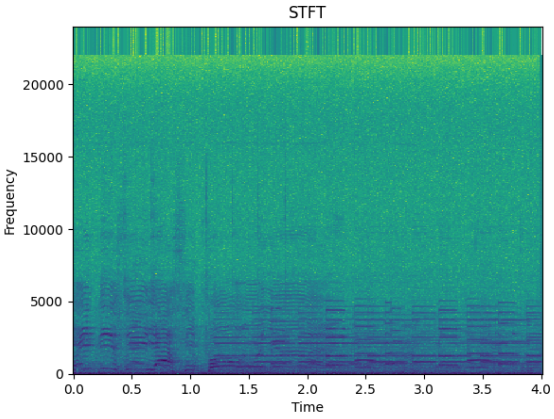
Shared data contains approximately 8000 urban audio files for 10 category like

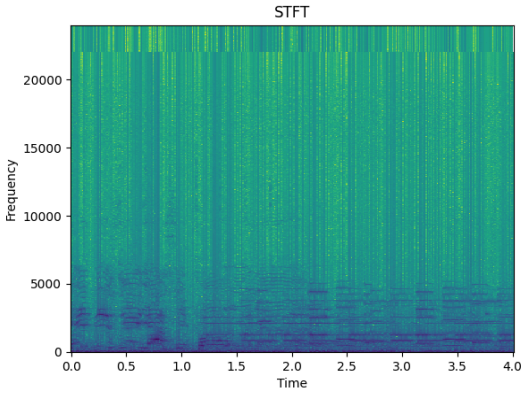
- air_conditioner
- car_horn
- children_playing
- dog_bark
- drilling
- engine_idling
- gun_shot
- jackhammer
- siren
- street_music

Preprocessing:

To extract features which will help to understand the sound to any algorithm we first applied three windowing techniques hamming, hanning and Rectangular.
Spectrogram from different techniques can be found below:

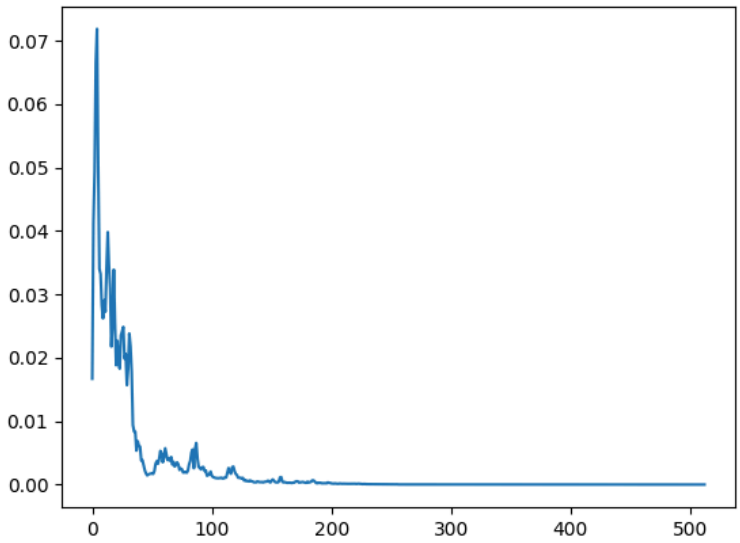






Feature extraction:

First we took the magnitude of data got from STFT and the for all the frequencies we have extracted the maximum amplitude present across entire time so feature will look like below



Model:

Trained a Random Forest model to classify the sound type.

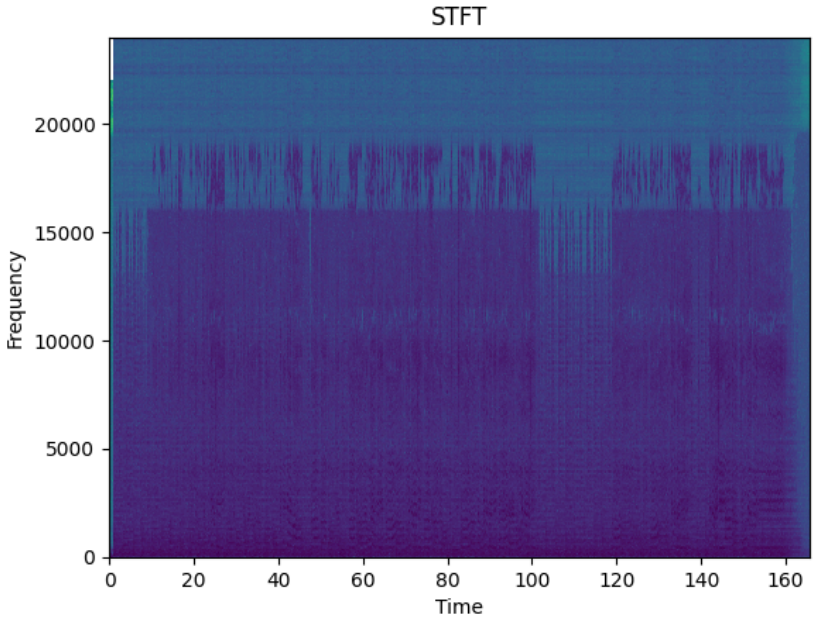
Evaluation:

Model was evaluated on 20 % data from all the classes which were approximately 1700 sound files. Metrics chosen to evaluate the model was accuracy score and F1 score. On Test dataset model accuracy was **75%** and on the train dataset it was 96%.

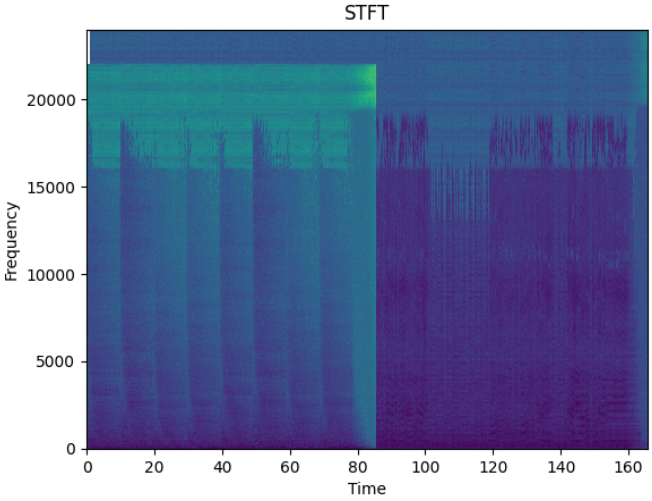
```
###################### Model Evaluation ######################

Train accuracy: 0.9662133142448103
Test accuracy: 0.7492844876931883
Train f1 score: 0.9660906637821117
Test f1 score: 0.7473739230427743
```

Song Analysis:

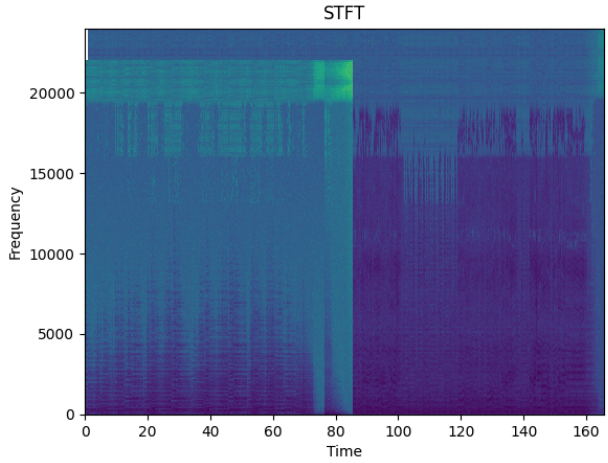
Song eterna-cancao-wav-12569



Song sci-fi-background-258999



Song see-you-later-203103



Song twisterion-b1-221376