Leveraging Classic and Modern Learning Techniques for Heart Disease Prediction from Heart Sound Data

Team Cookie Monsters

Md. Asif Haider

Md. Azizur Rahman Anik

CSE-18, BUET

Introduction

- BioMed Datathon by
  - BME-BUET
  - Aurora Consulting

- **Multiple** heart abnormalities prediction from **multiple** heart sound recordings

- Multi label classification task
  - Aortic Stenosis (AS)
  - Aortic Regurgitation (AR)
  - Mitral Regurgitation (MR)
  - Mitral Stenosis (MS)
  - Normal (N)

Motivation

- Heart related conditions have been on an alarming rise globally
- Accurate diagnosis and early intervention is important!
- Data is not so abundant in these cases
- Efficient use of **limited data** is crucial
- Automation is the key!
- Where humans fail, **artificial intelligence** can come to the rescue
- Machine learning
  - Given the input and output, finds the representative function
- Deep learning
  - Complex neural network representation mimicking how human brains learn

Input Dataset

- BUET Multi-Disease Heart Sound Dataset (BMD-HS)

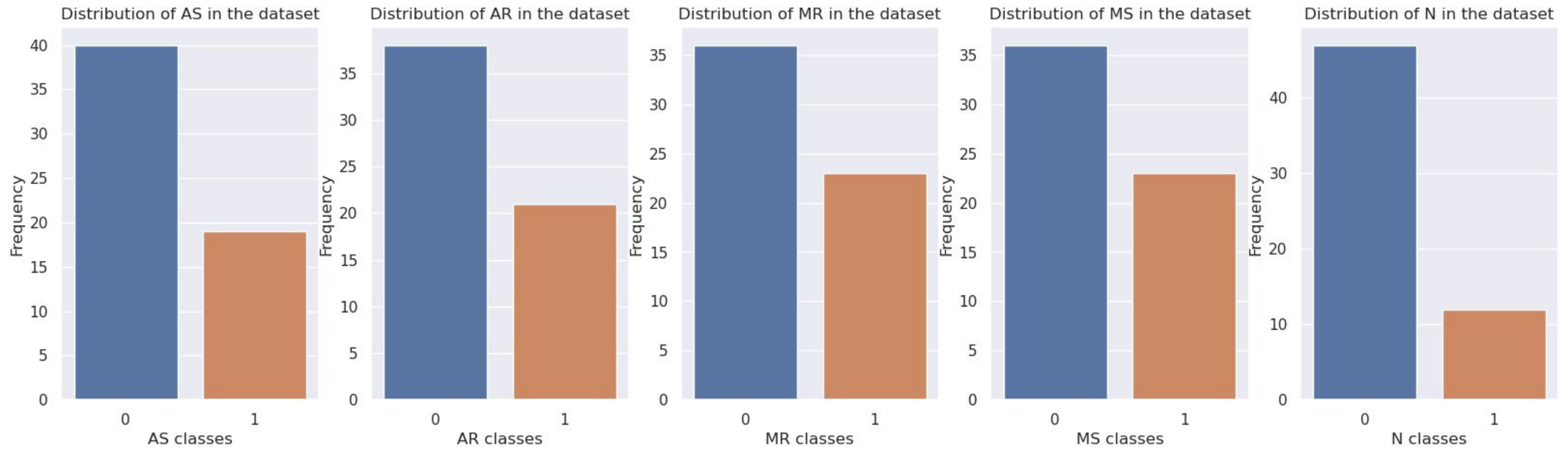
- (20 healthy + 20 x 4 disease) sample patient x (4 locations x 2 postures) heart sound = 800+ audio recordings

- Train set size
  - 59 patients
  - 8 recording each
  - 472 audio files

- Test set size
  - 49 patients
  - 8 recording each
  - 392 audio files

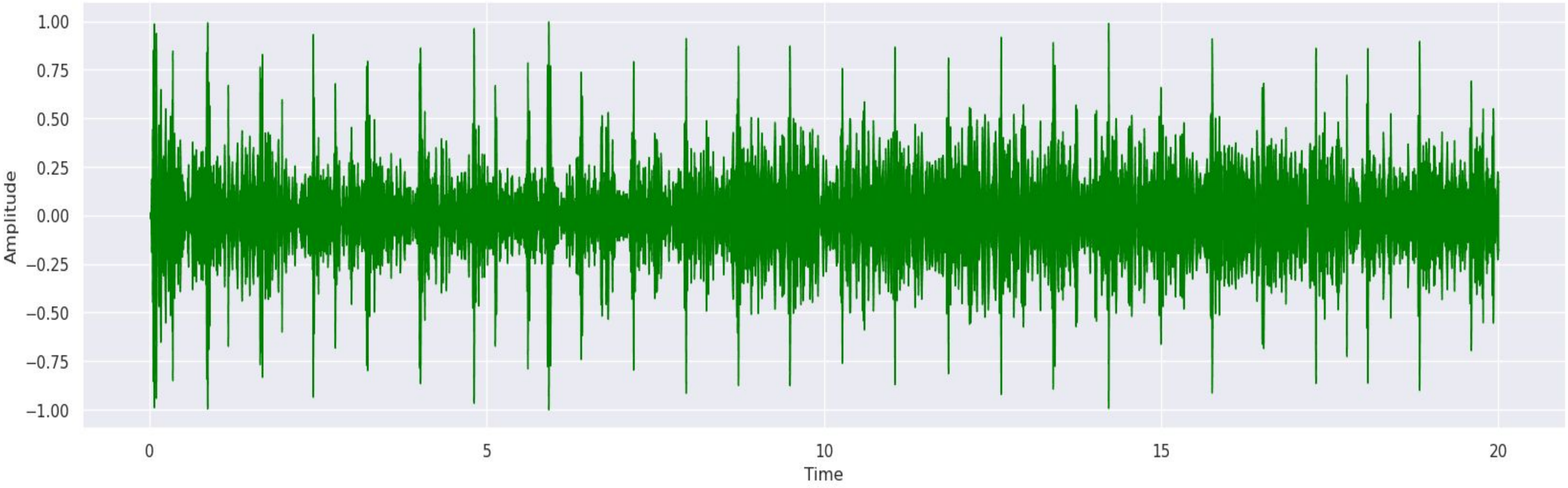
Exploratory Data Analysis

- Python **Numpy, Pandas, Seaborn** package
  - Missing value check
  - Unique value check
  - Label distribution check
  - Metadata check



Loading Audio File

- Python **Librosa** package
- Wave -> Vector representation
  - **Shape** (80000, 1)
  - **Sampling rate** 4KHz
  - **Length** 20 seconds



Machine Learning Approach

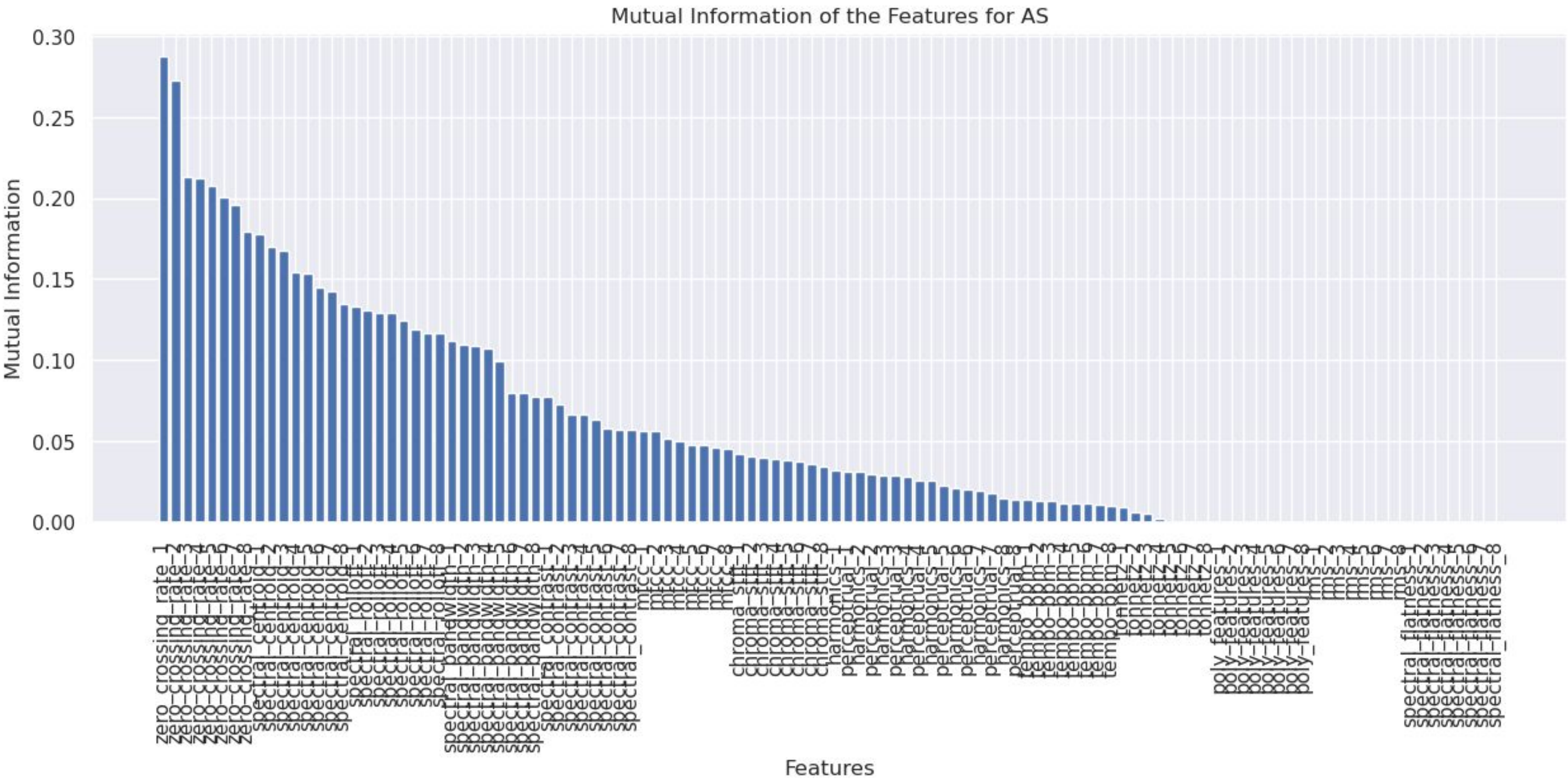
- Decision Tree
- Random Forest
- Extreme Gradient Boost (XGBoost)
- Multi Layer Perceptron

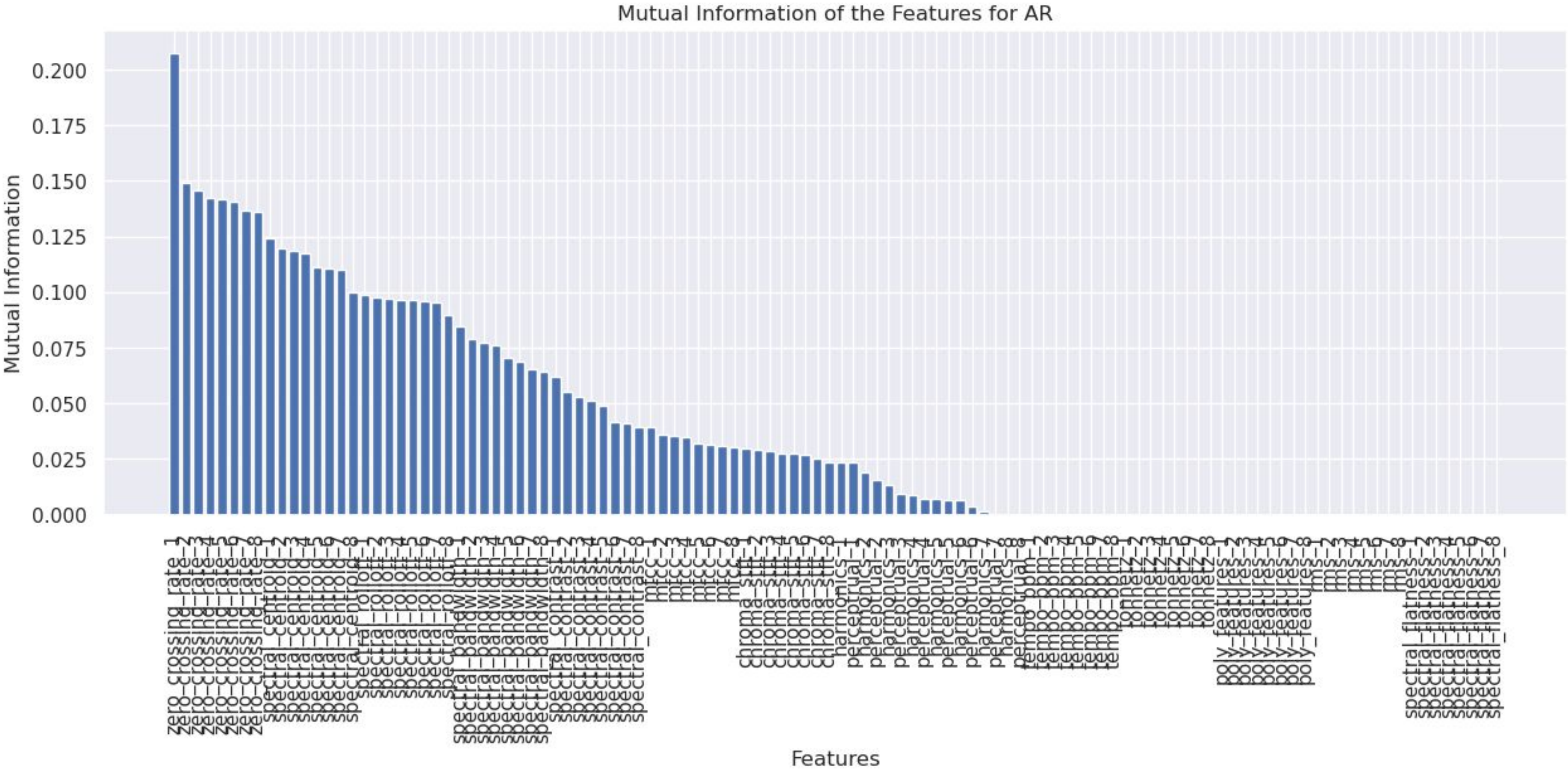
Feature Extraction

- Time Domain Features (basic temporal contents)
    - **Zero Crossing Rate**
        - Sign changing rate
    - **Energy**
    - **Amplitude**
    - Root Mean Square (Average Power)
    - Tempo Beats Per Minute
        - Beat rate
- Frequency Domain Features (rich spectral contents)
    - **Spectral Centroid**
        - Center of mass/brightness
    - **Spectral Roll-off**
        - Energy level
    - Spectral Bandwidth
        - Width of frequency band
    - Spectral Contrast
        - Difference in peak and valley amplitude
    - Spectral Flatness
        - Noisiness
    - **Mel Frequency Cepstral Coefficients (MFCC)**
        - Power level at log mel scale
    - **Chroma***
        - Energy distribution across pitch levels
    - Harmonics and Tonnetz*
        - Overtone series of base frequency
    - Perceptual*
        - Loudness, roughness, sharpness

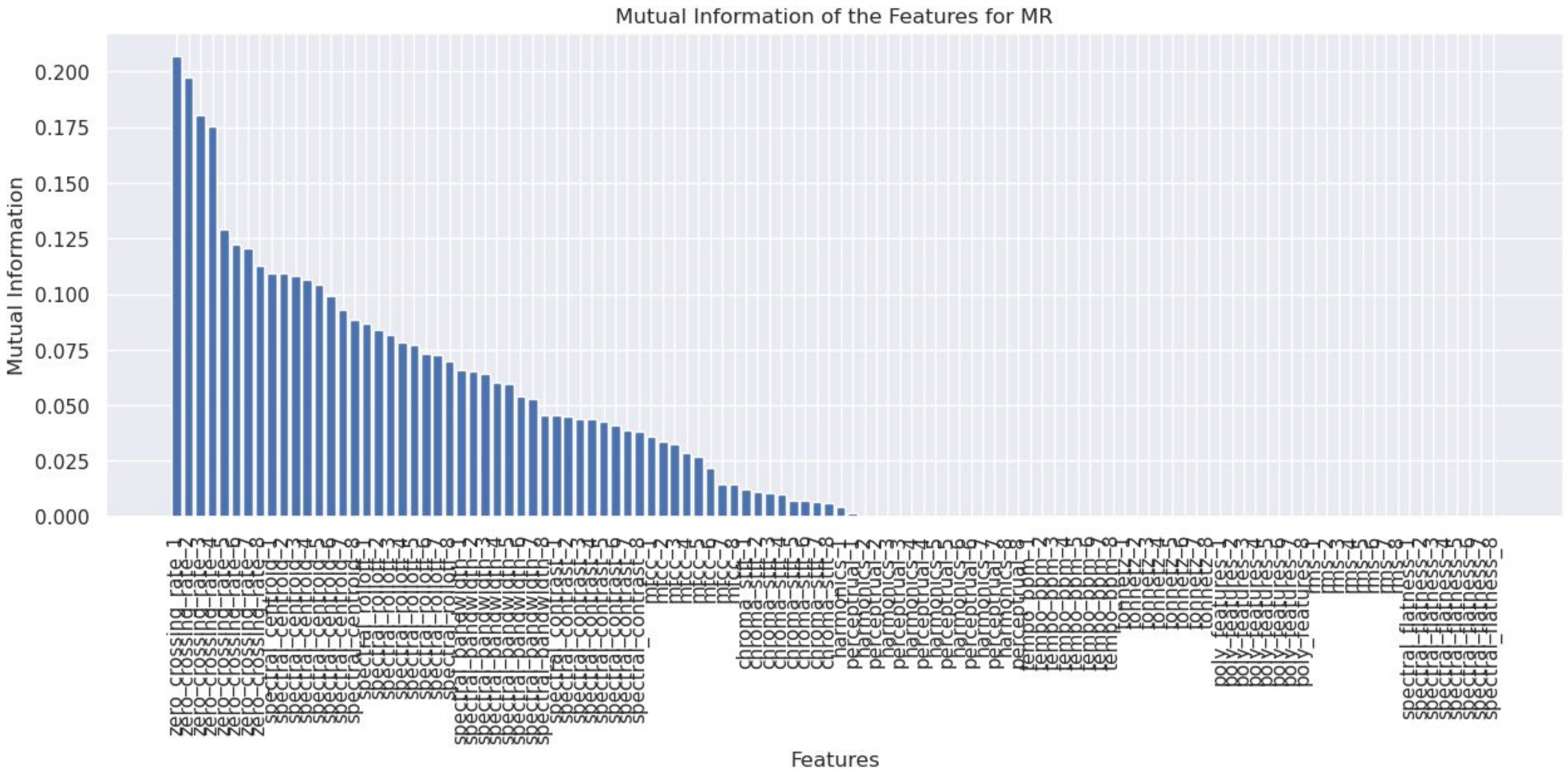
Feature Selection

- Python Scikit-learn package
  - Mutual Information Classifier
  - Selects top features based on Information Gain

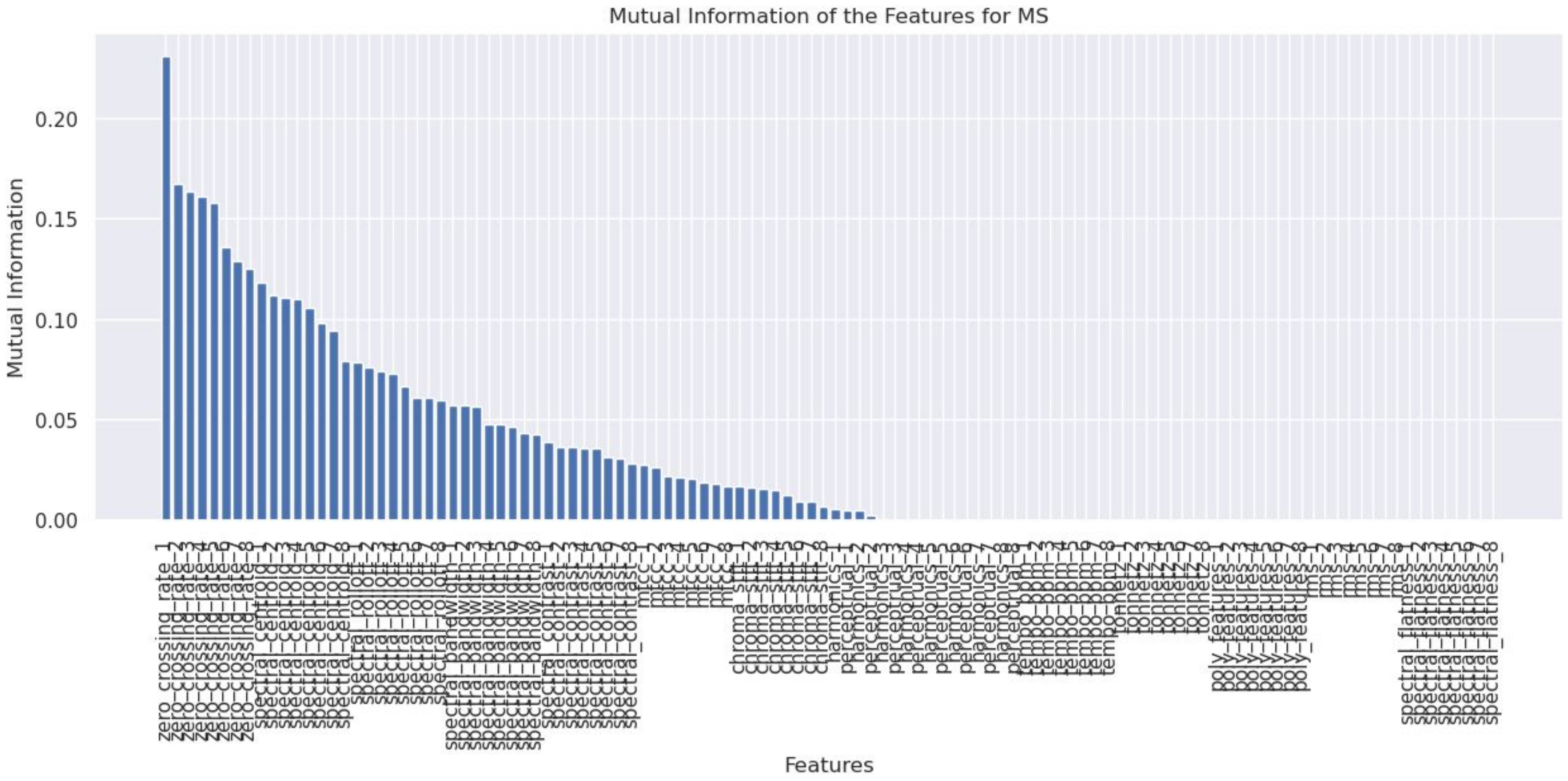

Mutual Information of the Features for AS
Mutual Information
Features
0.30
0.25
0.20
0.15
0.10
0.05
0.00
zero_crossing_rate_1

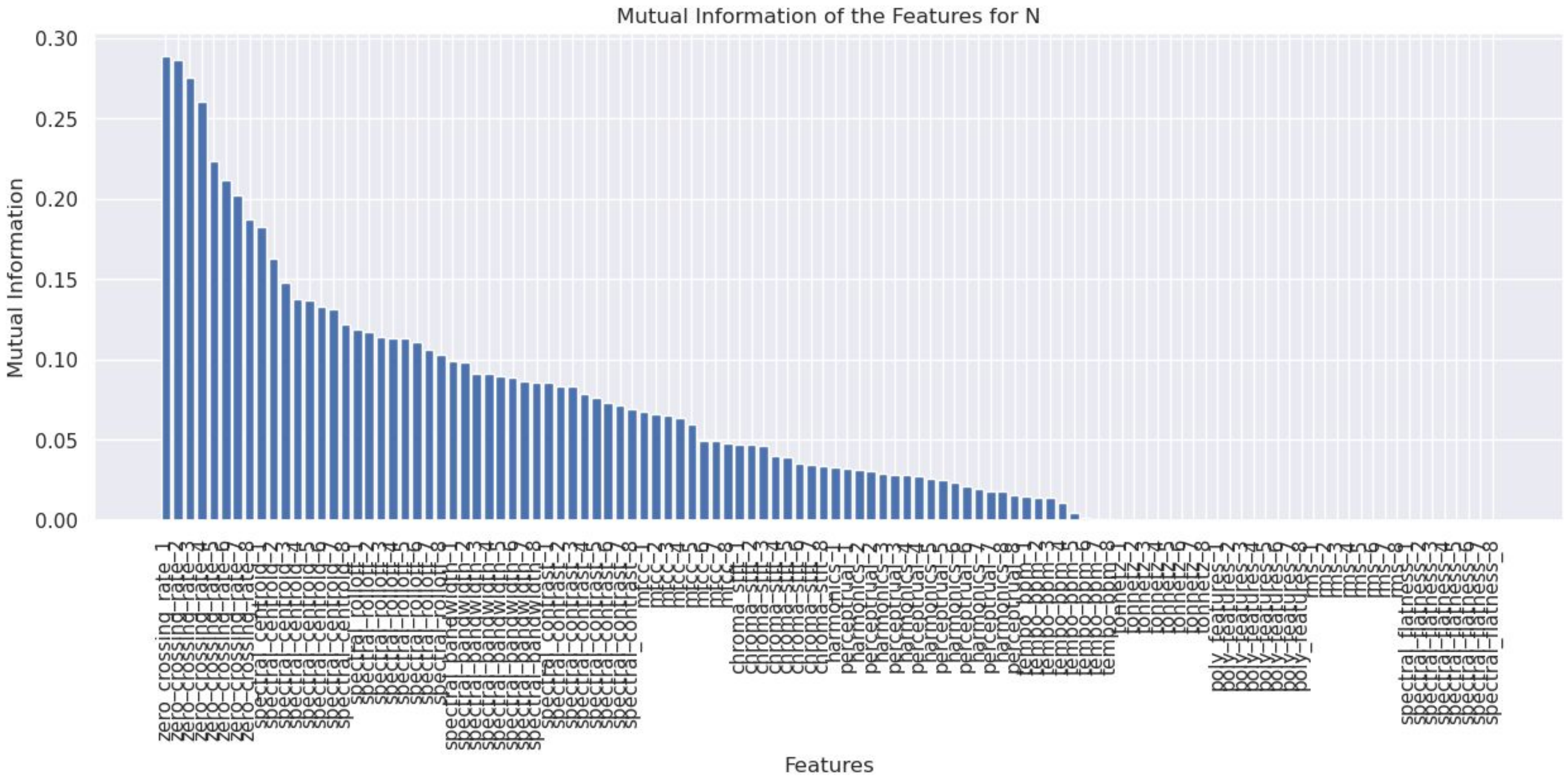


Mutual Information of the Features for AR
Mutual Information
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
Features



Mutual Information of the Features for MR
Mutual Information
Features
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200

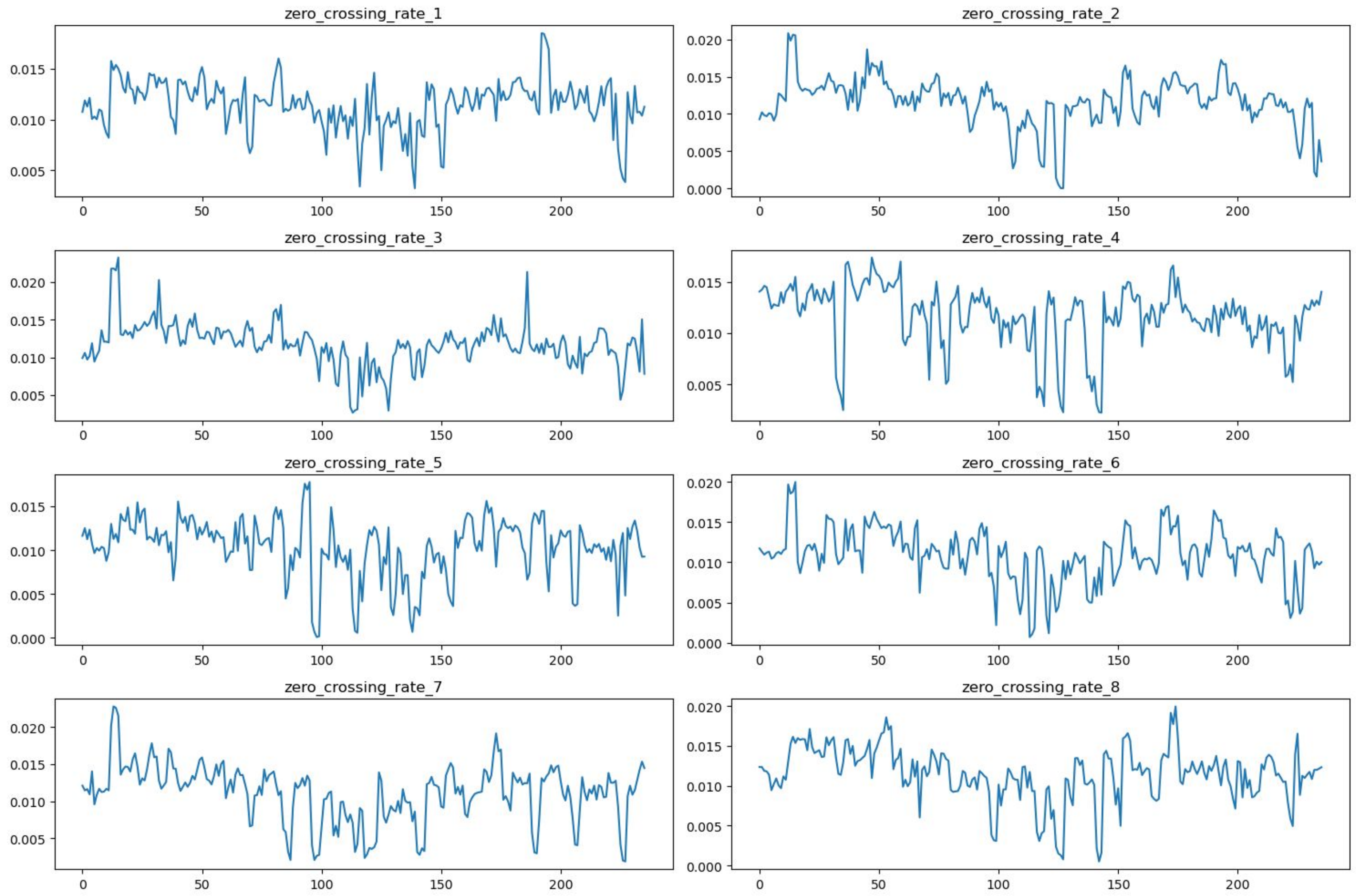

Mutual Information of the Features for MS
Mutual Information
0.20
0.15
0.10
0.05
0.00
Features
zero_crossing_rate_1



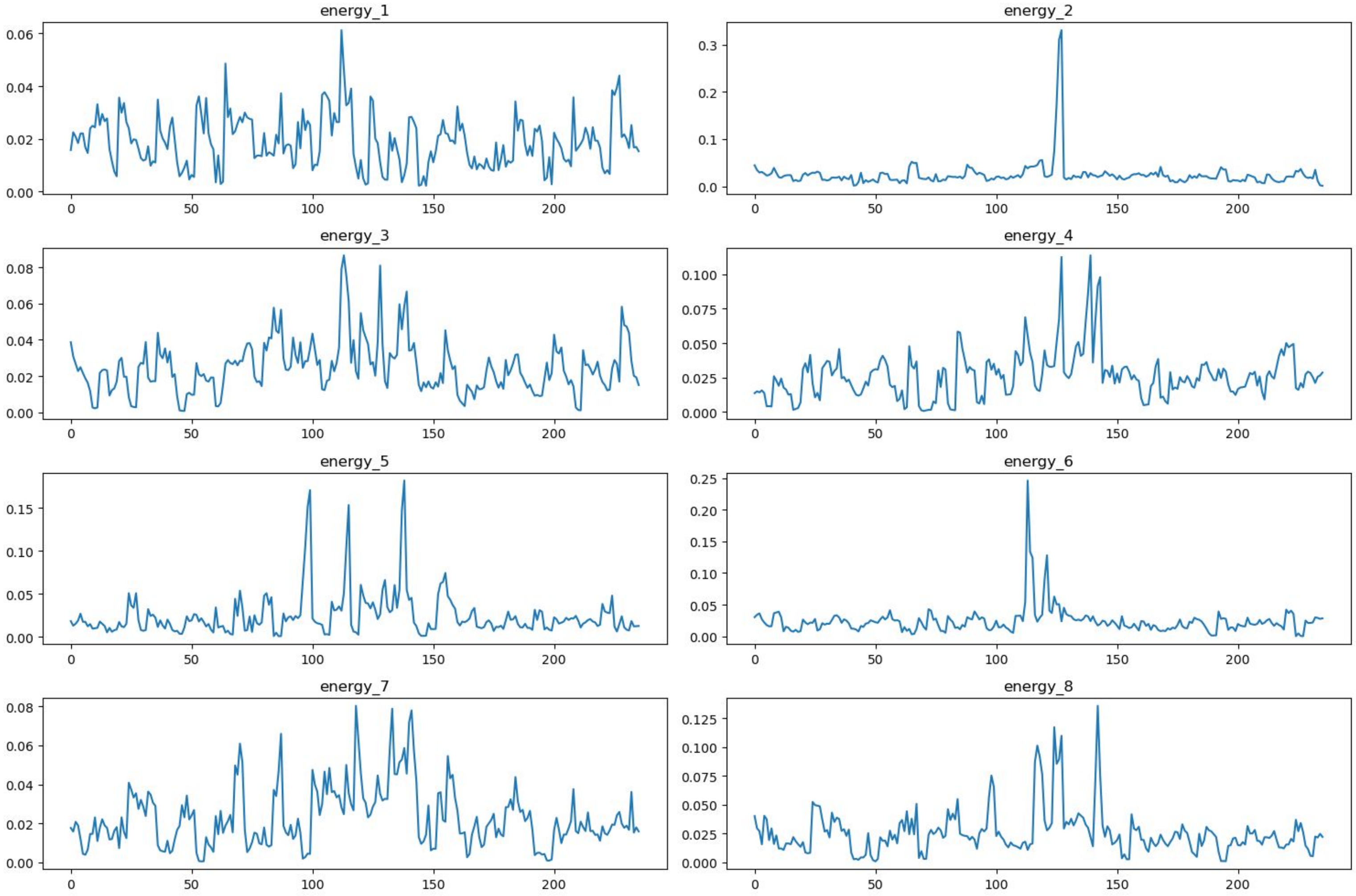
Mutual Information of the Features for N
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Mutual Information
zero_crossing_rate_1
Features

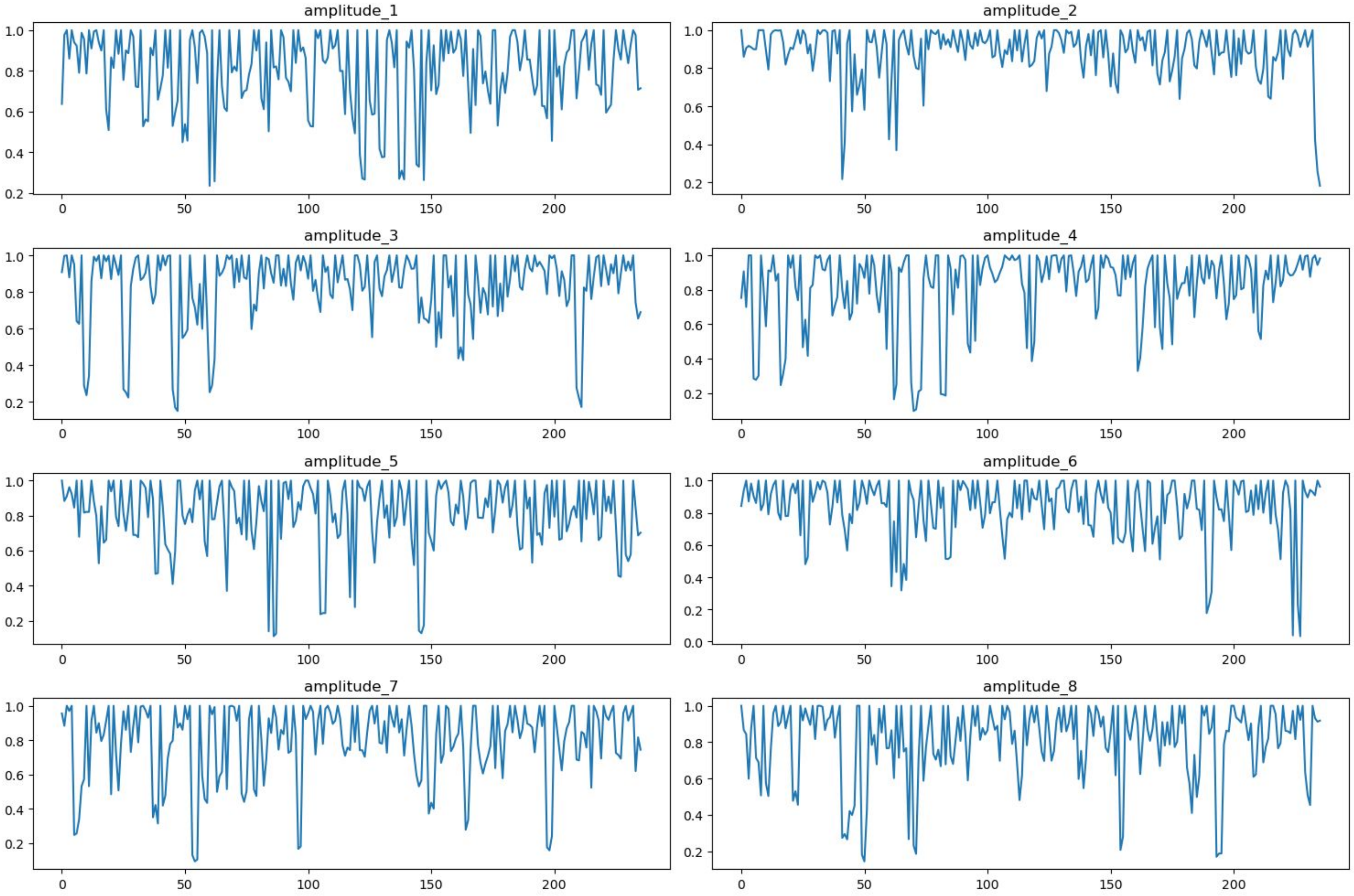
Feature Selection

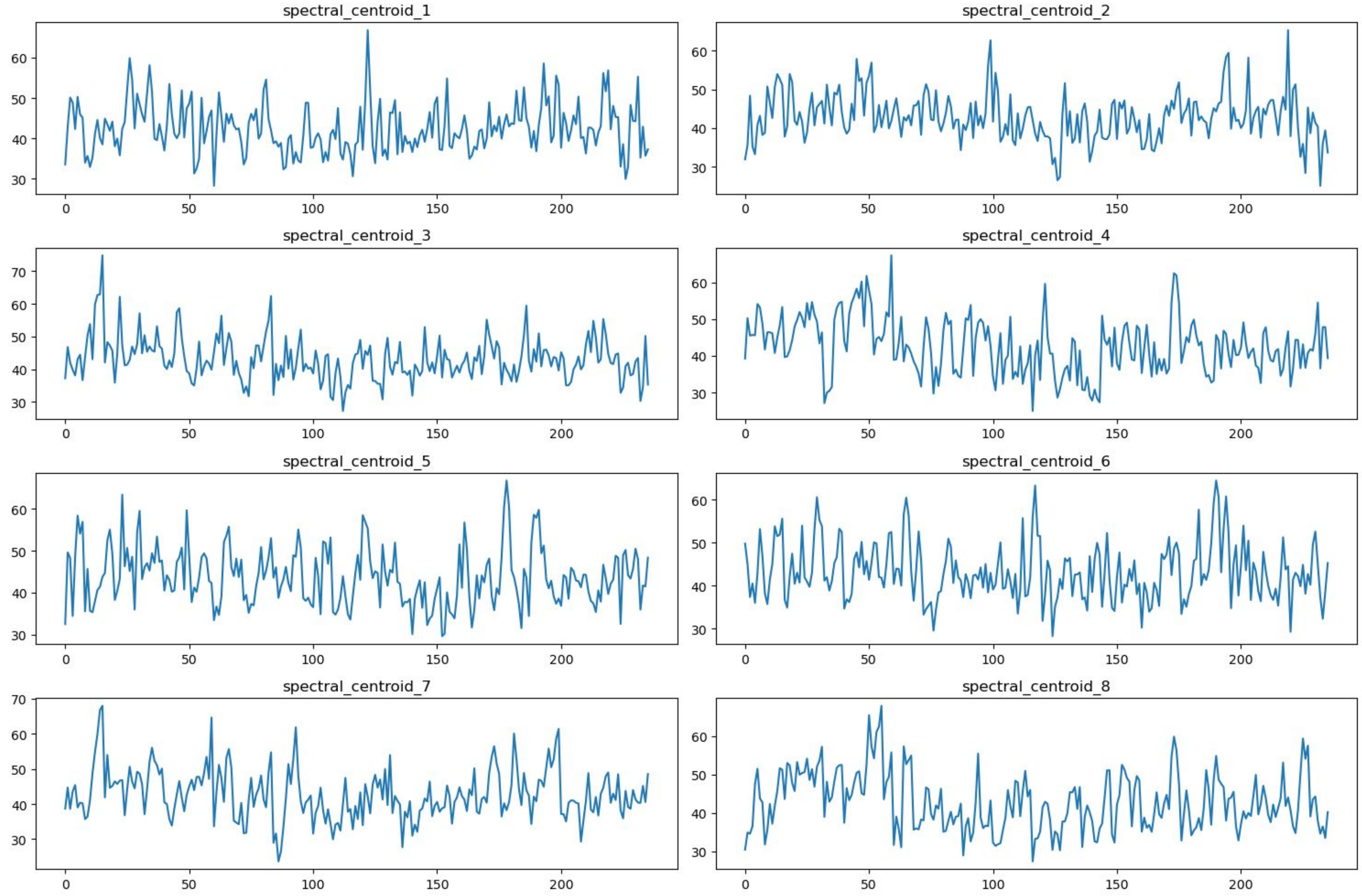
- Python Scikit-learn package
  - Mutual Information Classifier
  - Selects top features based on Information Gain
- 3 Time Domain Features
  - Zero crossing rate
  - Energy
  - Amplitude
- 4 Frequency Domain Features (After STFT, short time fourier transform)
  - Spectral centroid
  - Spectral roll-off
  - Mel frequency cepstral coefficients
  - Chroma



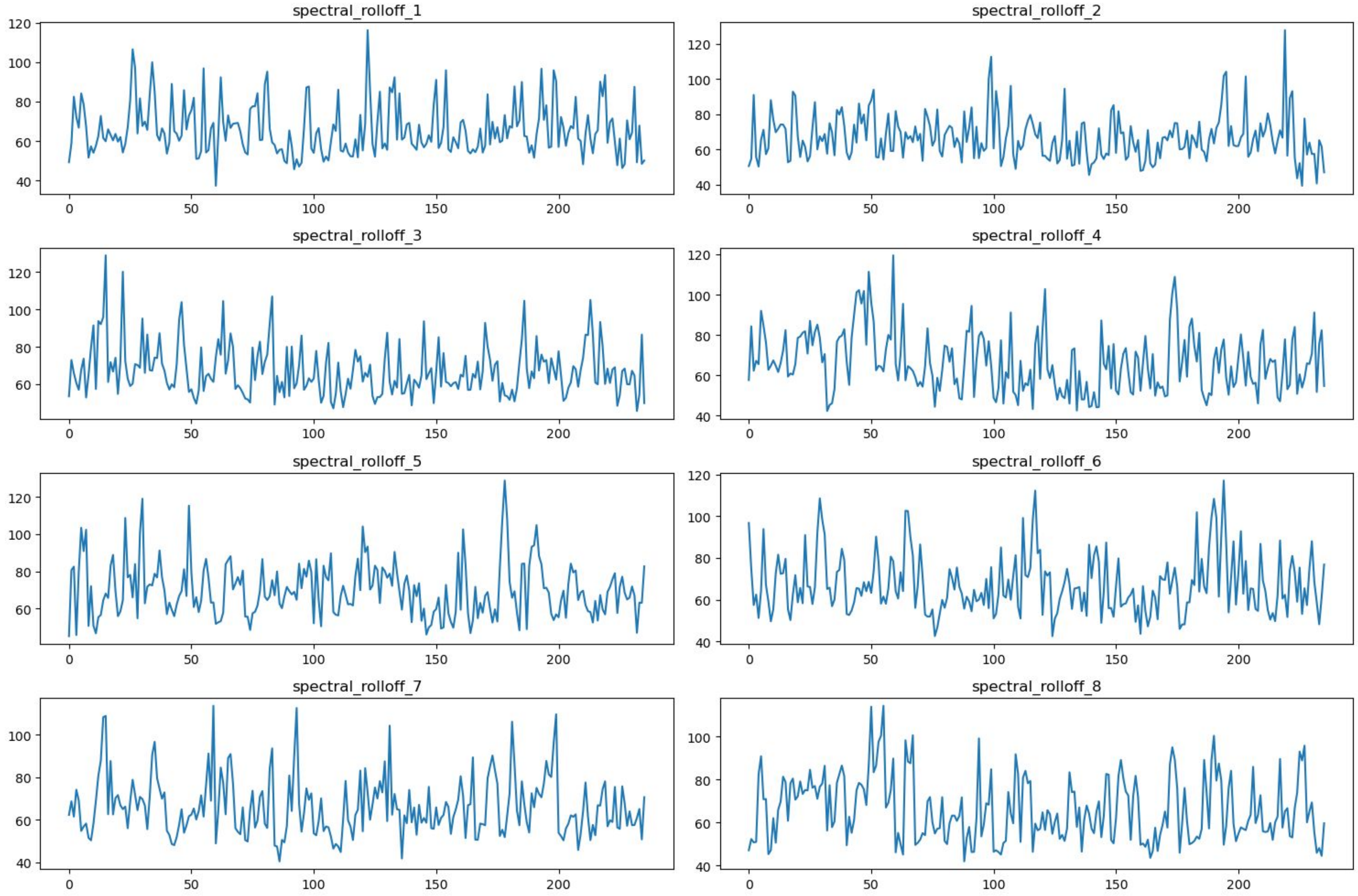
zero_crossing_rate_1
zero_crossing_rate_2
zero_crossing_rate_3
zero_crossing_rate_4
zero_crossing_rate_5
zero_crossing_rate_6
zero_crossing_rate_7
zero_crossing_rate_8

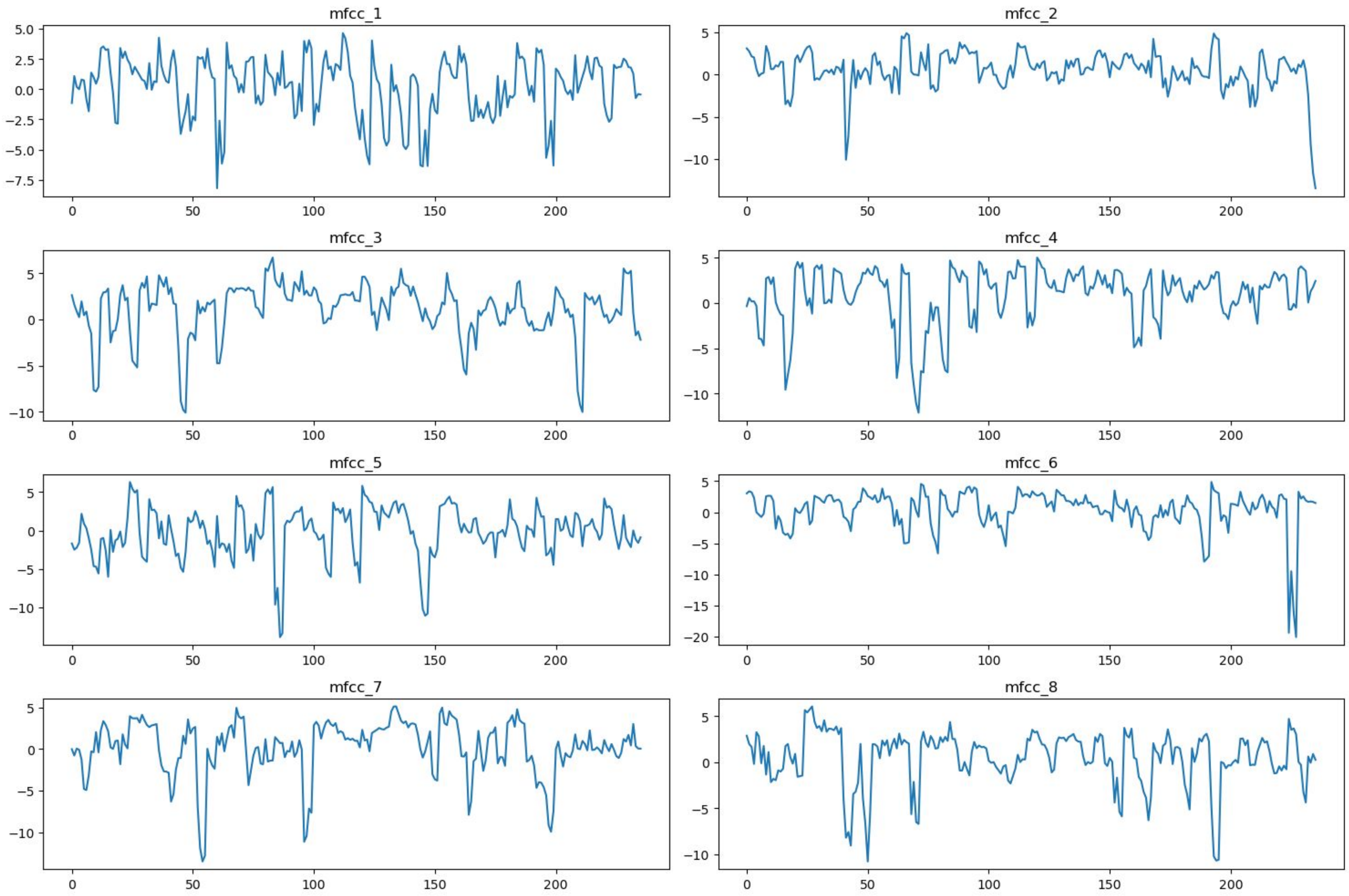

energy_1
energy_2
energy_3
energy_4
energy_5
energy_6
energy_7
energy_8
0
50
100
150
200

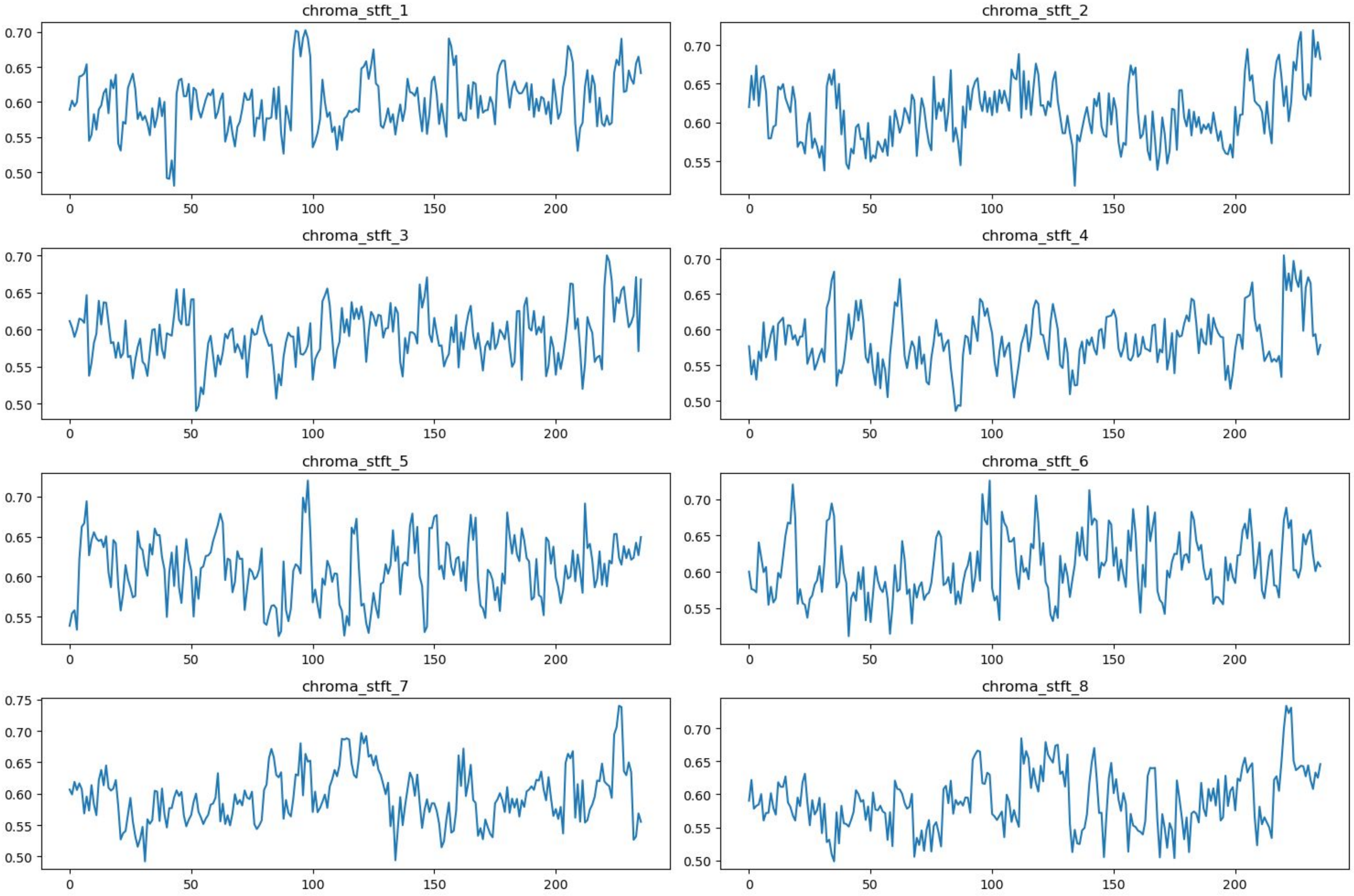

amplitude_1
amplitude_2
amplitude_3
amplitude_4
amplitude_5
amplitude_6
amplitude_7
amplitude_8
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200



spectral_centroid_1
spectral_centroid_2
spectral_centroid_3
spectral_centroid_4
spectral_centroid_5
spectral_centroid_6
spectral_centroid_7
spectral_centroid_8
0
50
100
150
200


spectral_rolloff_1
spectral_rolloff_2
spectral_rolloff_3
spectral_rolloff_4
spectral_rolloff_5
spectral_rolloff_6
spectral_rolloff_7
spectral_rolloff_8


mfcc_1
mfcc_2
mfcc_3
mfcc_4
mfcc_5
mfcc_6
mfcc_7
mfcc_8
5.0
2.5
0.0
−2.5
−5.0
−7.5
5
0
−5
−10
−15
−20
0
50
100
150
200


chroma_stft_1
0.70
0.65
0.60
0.55
0.50
0
50
100
150
200
chroma_stft_2
0.70
0.65
0.60
0.55
0
50
100
150
200
chroma_stft_3
0.70
0.65
0.60
0.55
0.50
0
50
100
150
200
chroma_stft_4
0.70
0.65
0.60
0.55
0.50
0
50
100
150
200
chroma_stft_5
0.70
0.65
0.60
0.55
0
50
100
150
200
chroma_stft_6
0.70
0.65
0.60
0.55
0
50
100
150
200
chroma_stft_7
0.75
0.70
0.65
0.60
0.55
0.50
0
50
100
150
200
chroma_stft_8
0.70
0.65
0.60
0.55
0.50
0
50
100
150
200

Data Preprocessing Pipeline

- **Imbalance Handling**
- **Augmentation**
- **Segmentation**
- Normalization
  - StandardScaler
- Train Validation Split
  - 80-20 split
- Cross Validation
  - Leave one out (LOOCV)
  - K-fold (K=5)
- Machine Learning Model
- Feature Importance Check
- Again Machine Learning Model

Data Imbalance: Label-wise Resampling

- Handling imbalance on original data
  - Resampled to maintain a 90% ratio between label 1 and label 0
  - Augmentation and Segmentation on the resampled data

Data Augmentation

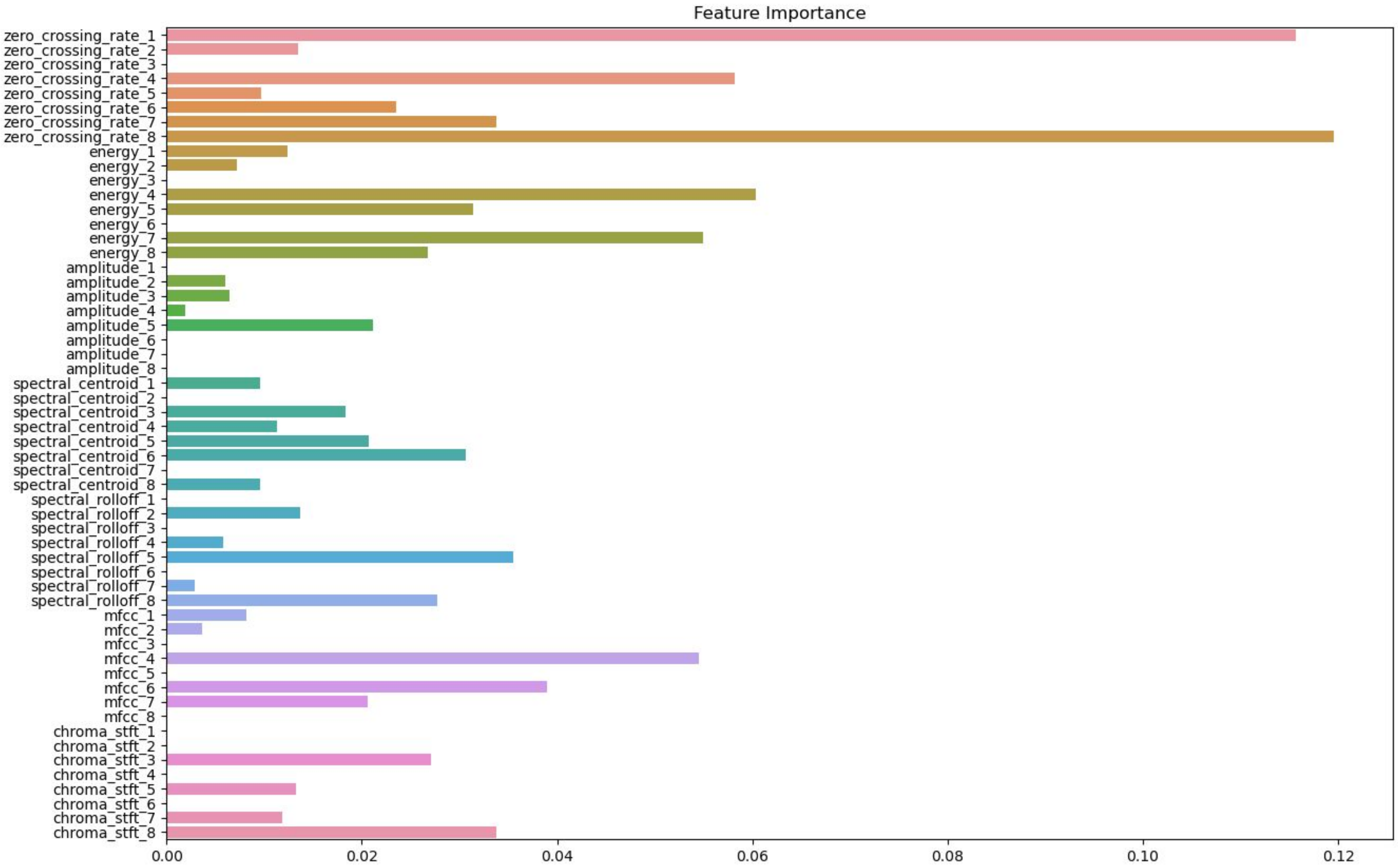
- Pitch Shifting
  - Shifting up and down
- Time Stretching
  - Expansion and Compression
- Adding Noise
  - With factors 0.01 and 0.05
- Increased size after augmenting
  - 413 (7x of the original data)

Data Segmentation

- 4 segments each of 5 second length for every recording entry in each patient sample
- Segmentation on original data
  - Increased size: **236 samples**
- Segmentation on augmented data
  - Increased size : **1652 samples**
- Segmentation on augmented and resampled data to handle imbalance
  - Increased size: **55160 samples**

Data Preprocessing Pipeline

- Imbalance Handling
- Augmentation
- Segmentation
- Normalization
  - StandardScaler
- Train Validation Split
  - 80-20 split
- Cross Validation
  - Leave one out (LOOCV)
  - K-fold (K=5)
- Machine Learning Model
- **Feature Importance Check**
- Again Machine Learning Model


Feature Importance
zero_crossing_rate_1
zero_crossing_rate_2
zero_crossing_rate_3
zero_crossing_rate_4
zero_crossing_rate_5
zero_crossing_rate_6
zero_crossing_rate_7
zero_crossing_rate_8
energy_1
energy_2
energy_3
energy_4
energy_5
energy_6
energy_7
energy_8
amplitude_1
amplitude_2
amplitude_3
amplitude_4
amplitude_5
amplitude_6
amplitude_7
amplitude_8
spectral_centroid_1
spectral_centroid_2
spectral_centroid_3
spectral_centroid_4
spectral_centroid_5
spectral_centroid_6
spectral_centroid_7
spectral_centroid_8
spectral_rolloff_1
spectral_rolloff_2
spectral_rolloff_3
spectral_rolloff_4
spectral_rolloff_5
spectral_rolloff_6
spectral_rolloff_7
spectral_rolloff_8
mfcc_1
mfcc_2
mfcc_3
mfcc_4
mfcc_5
mfcc_6
mfcc_7
mfcc_8
chroma_stft_1
chroma_stft_2
chroma_stft_3
chroma_stft_4
chroma_stft_5
chroma_stft_6
chroma_stft_7
chroma_stft_8
0.00
0.02
0.04
0.06
0.08
0.10
0.12

Primary Results

ML Model	Public macro F1 Score	Private macro F1 Score
MLP	0.3294	0.305
XGBoost	0.4176	0.4425
Random Forest	0.3259	0.4732
Decision Tree	**0.4988**	**0.5138**

Towards Deep Learning

Benchmarks

Add a Result

These leaderboards are used to track progress in Audio Classification

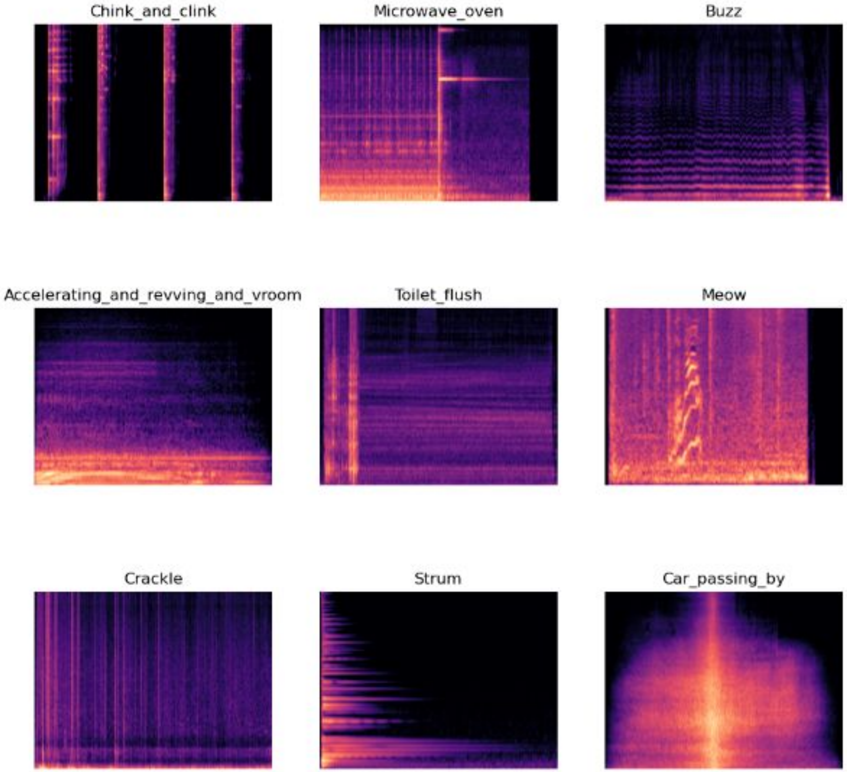
Trend	Dataset	Best Model	Paper	Code	Compare
	AudioSet	OmniVec			See all
	ESC-50	OmniVec			See all
	VGGSound	ONE-PEACE (Audio-Visual)			See all
	ICBHI Respiratory Sound Database	AST (Patch-Mix CL)			See all
	SHD	SNN with Dilated Convolution with Learnable Spacings			See all
	FSD50K	ONE-PEACE			See all

Audio Spectrogram Transformer (AST) vs CNN based models

- Some local developments
  - [MD-CardioNet : BUET EEE and CSE]
    - LSTM (RNN) based cardiovascular disease classifier
  - [SpectNet: BUET BME]
    - Uses CNN based feature extraction beforehand to detect heart anomaly

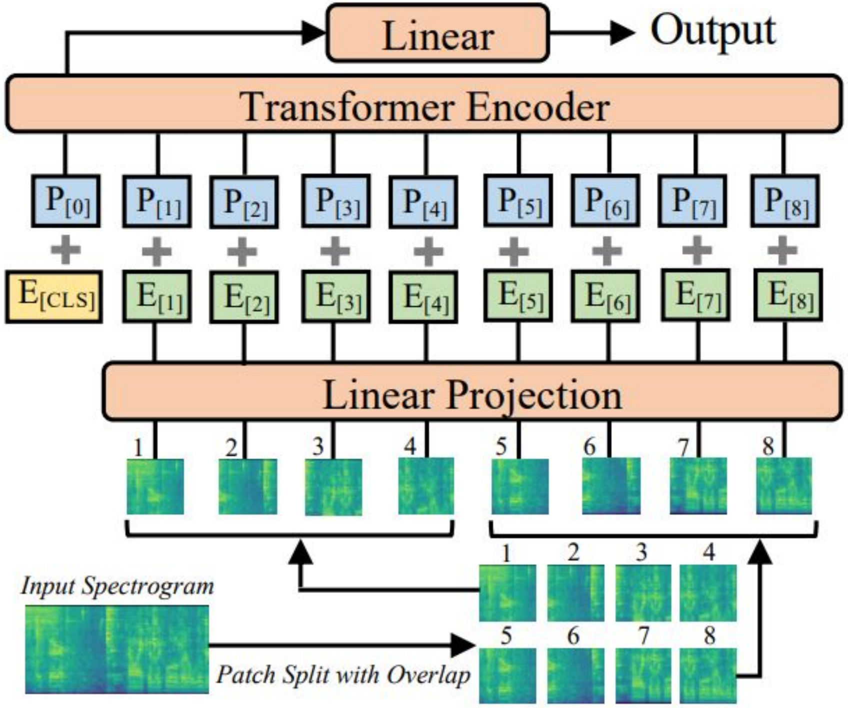
- Audio **Spectrogram Transformer** achieves state-of-the-art audio classification task performance
- In 2021, by MIT CSAILs
- [Audio Spectrogram Transformer Paper]

Spectrogram



- 2D representation of audio signals
- X axis -> Time
- Y axis -> Frequency
- Color intensity -> amplitude/energy for a particular frequency in particular time
- Single image from audio, containing rich temporal and spectral information

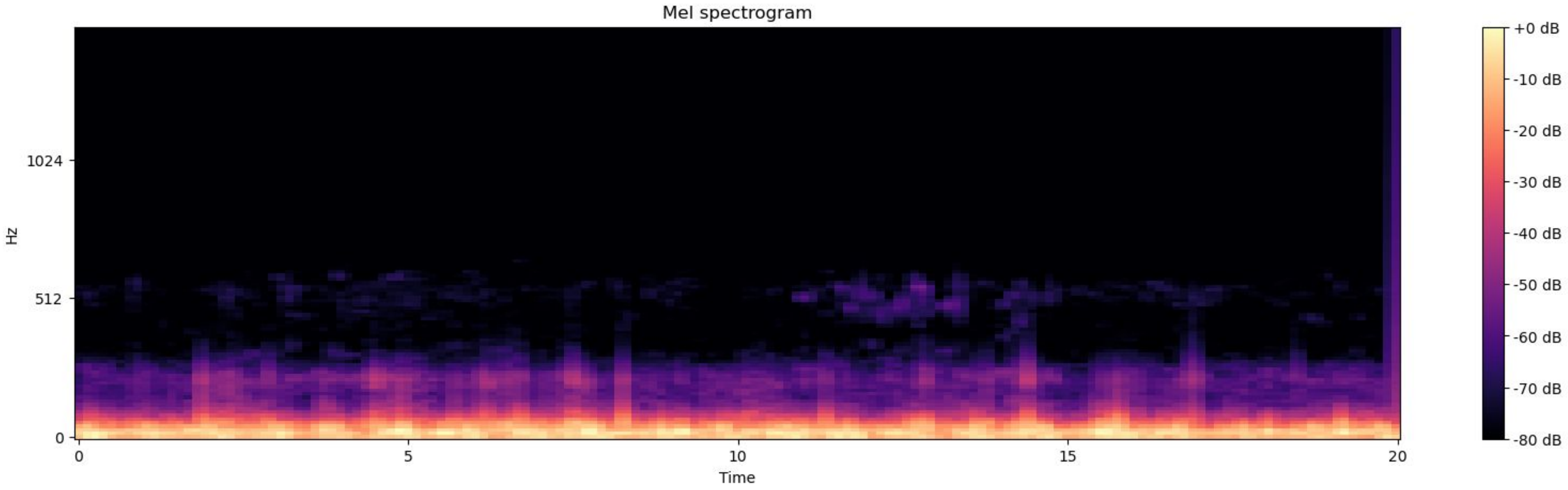
Transformer



- Large scale deep neural network with self attention mechanism
- Started with Natural Language Processing (NLP) domain
- Currently achieves state of the art results in almost every domains including Vision [Vision Transformers]
- The more data, the better
- Audio Spectrogram Transformer (AST) applies a Vision Transformer (ViT) to audio by turning audio into spectrogram image
- Parameter initialization from ImageNet

Huggingface: Everything Here All At Once

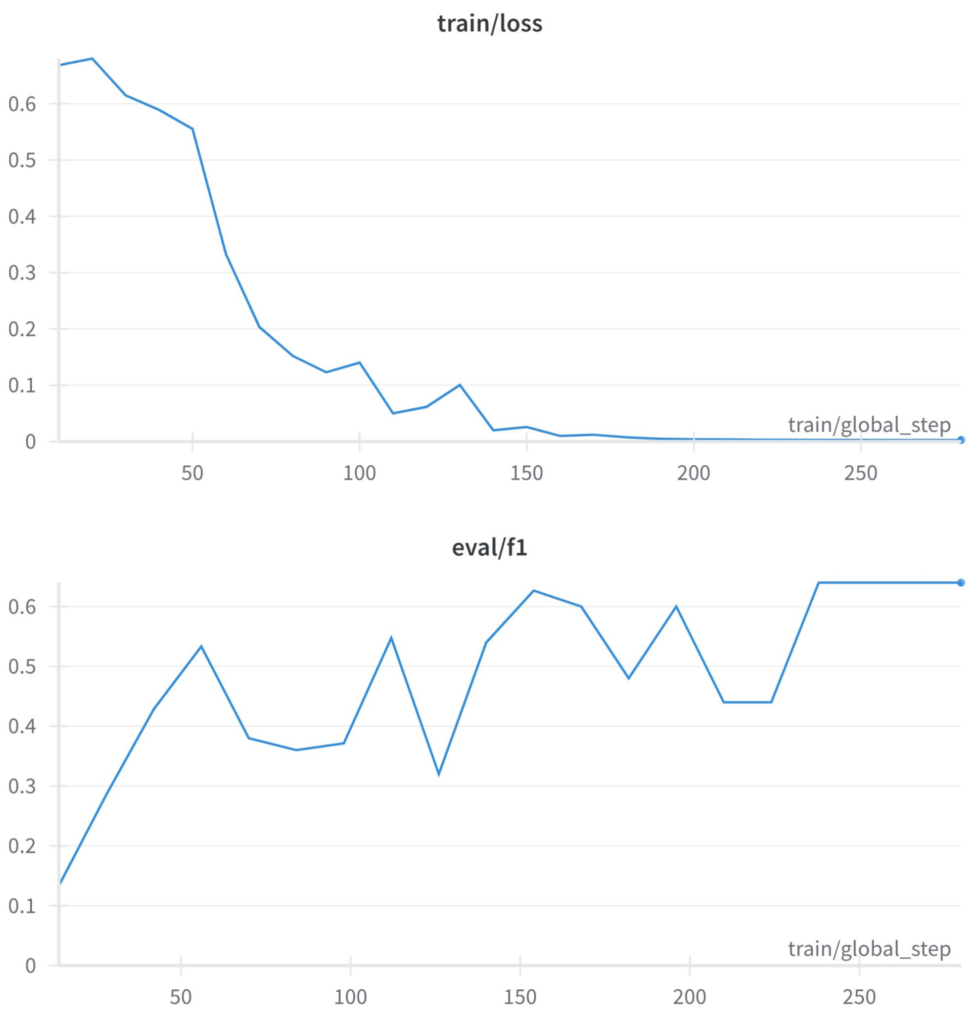
- The global hub of
    - Pretrained and Finetuned model checkpoints
    - Suitable tokenizers
    - Automatic Feature Extractor
        - Extracts mel spectrogram features and converts to corresponding image/ vector
    - Well written Pipeline and Trainer functions



AST Workflow

- Resampled audio to 16 KHz (accepted by AST)
- Feature Extractor, Tokenizer and Model checkpoint loading
- Audio vector -> Image vector transformation
- Loss: Binary cross entropy
- Metric: Macro averaged F1 score
- Hyperparameters:
  - Learning rate: 5e-04
  - Warmup ratio: 0.1
  - Batch size: 4
  - Training epochs: 20
- 8 different trainers for 8 different audio records
- Finally predictions with weighted majority voting based on evaluation macro F1 score

Plots



train/loss
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
250
train/global_step
eval/f1
0.6
0.5
0.4
0.3
0.2
0.1
0
50
100
150
200
250
train/global_step

Final Results

Model	Public F1	Private F1	ICBHI Score
AST	**0.6346**	0.4976	**0.6483**
Decision Tree (7 Features)	0.4988	**0.5138**	0.5945

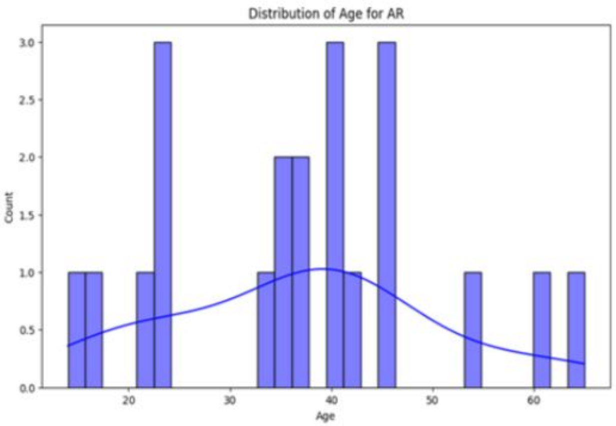
```
# calculate the specificity
specificity = (CAS+CAR+CMS+CMR)/(CAS+CAR+CMS+CMR+NAS+NAR+NMS+NMR)
print('Specificity:', specificity)

# calculate the sensitivity
sensitivity = CN/NN
print('Sensitivity:', sensitivity)

# calculate the icbhi score
icbhi_score = (specificity + sensitivity) / 2
print('ICBHI Score:', icbhi_score)
```

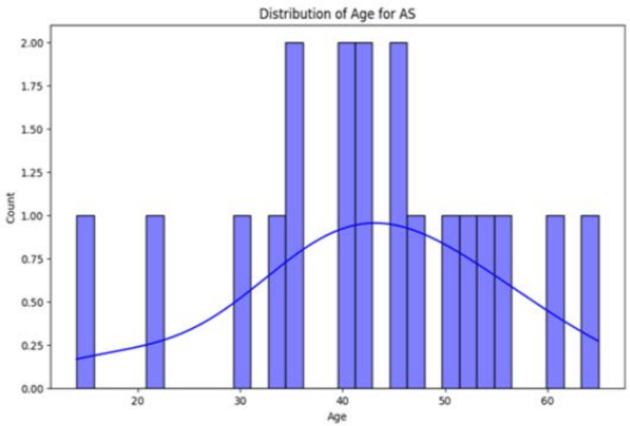
Platforms and Resources

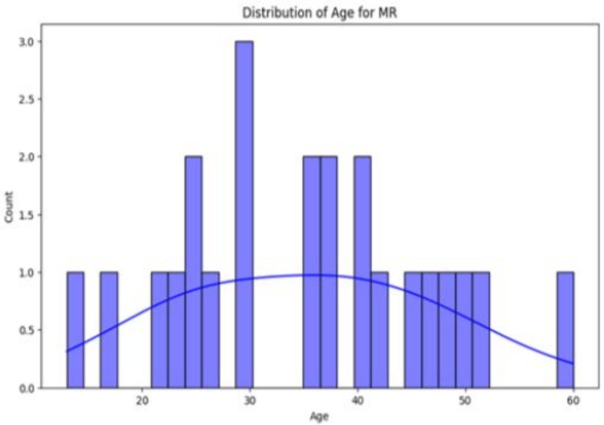
- ML on CPU
  - Core i5 @ 1.00 x 8 GHz
- DL on GPU
  - RTX 5000 (16 x 4 GB VRAM)
- Model, Tutorial and Plot
  - Huggingface
  - Wandb.ai
- Dataset, Testing and Submission
  - Kaggle

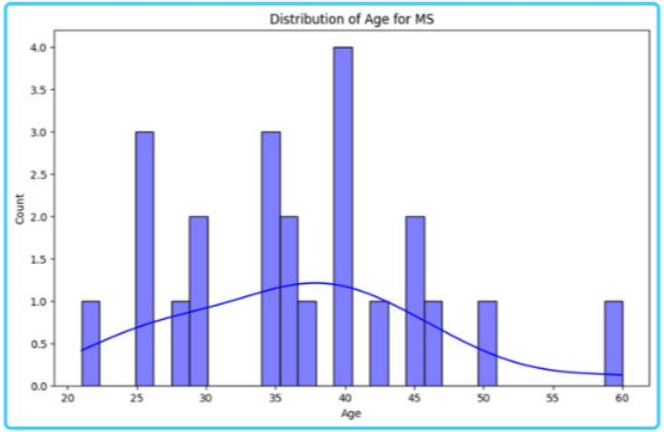
Metadata Insights

- Age
- Gender
  - Male
  - Female
- Smoker
  - Yes
  - No
- Lives in
  - Rural
  - Urban

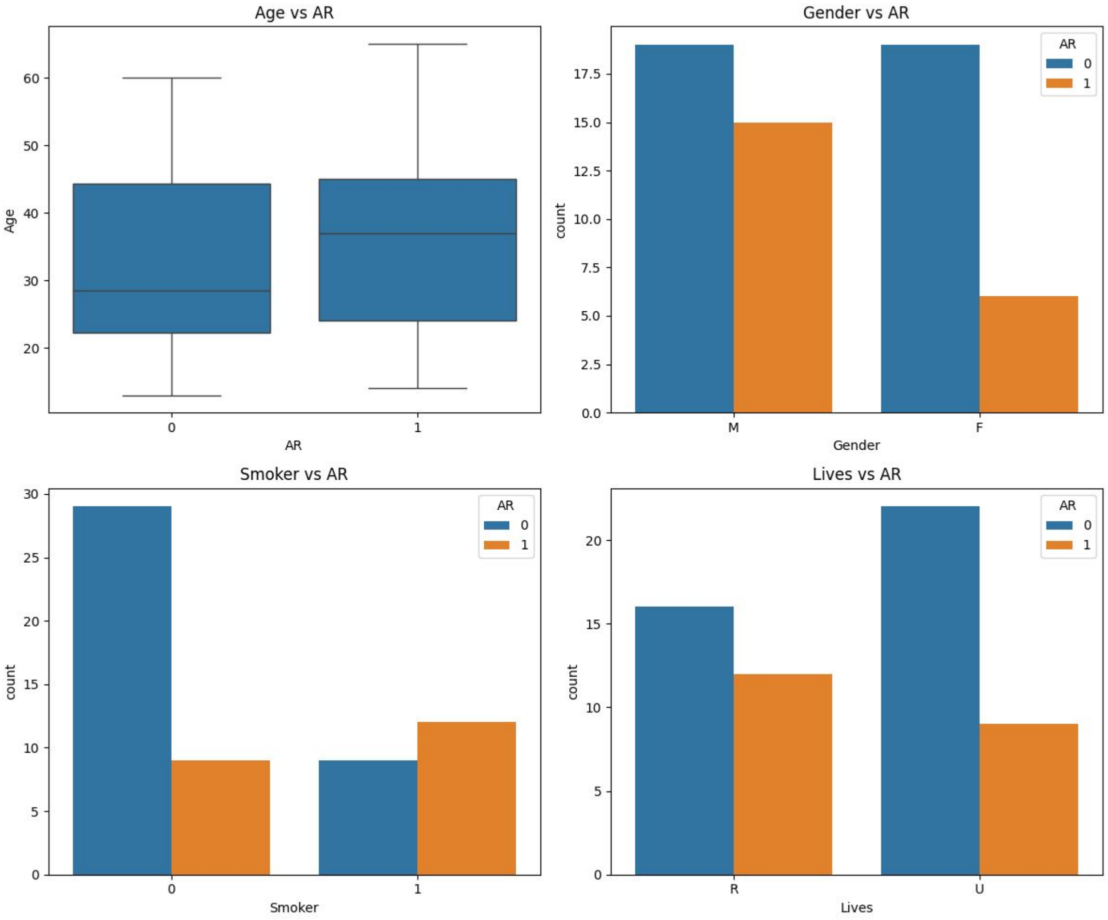




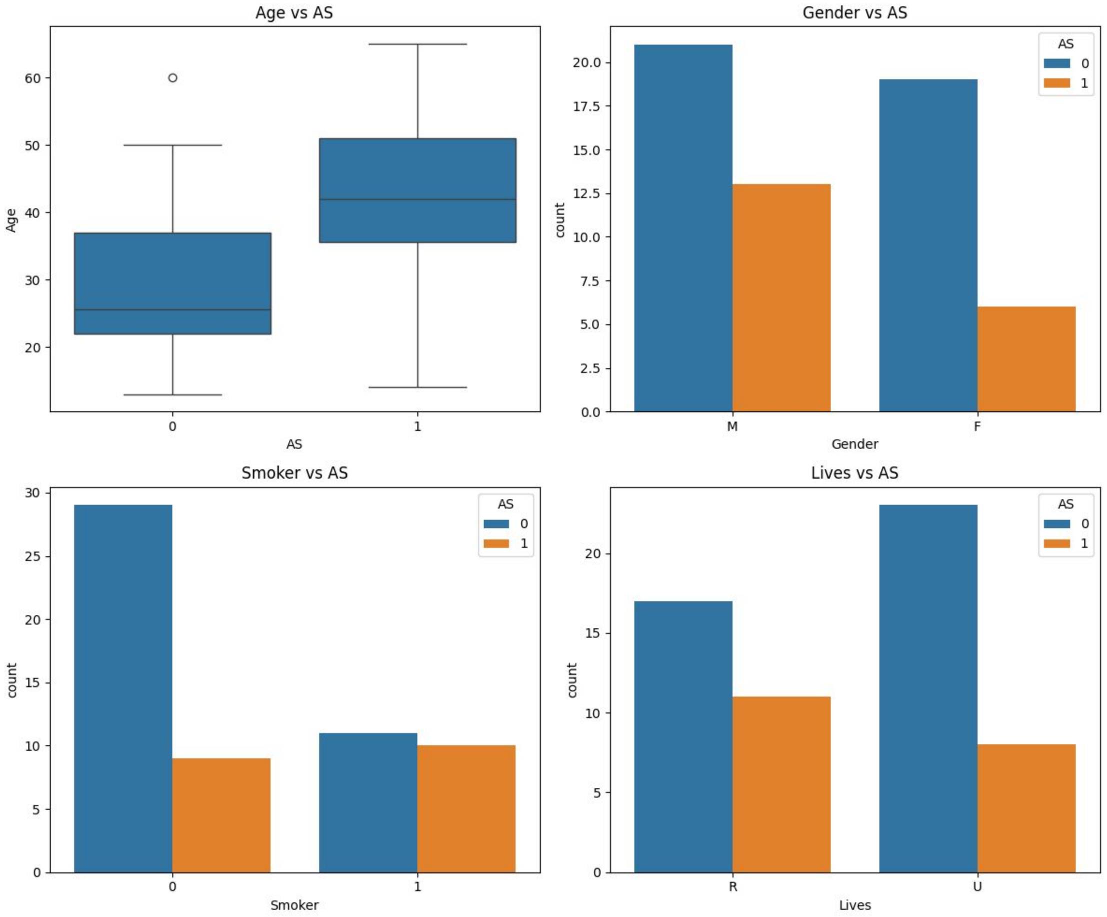




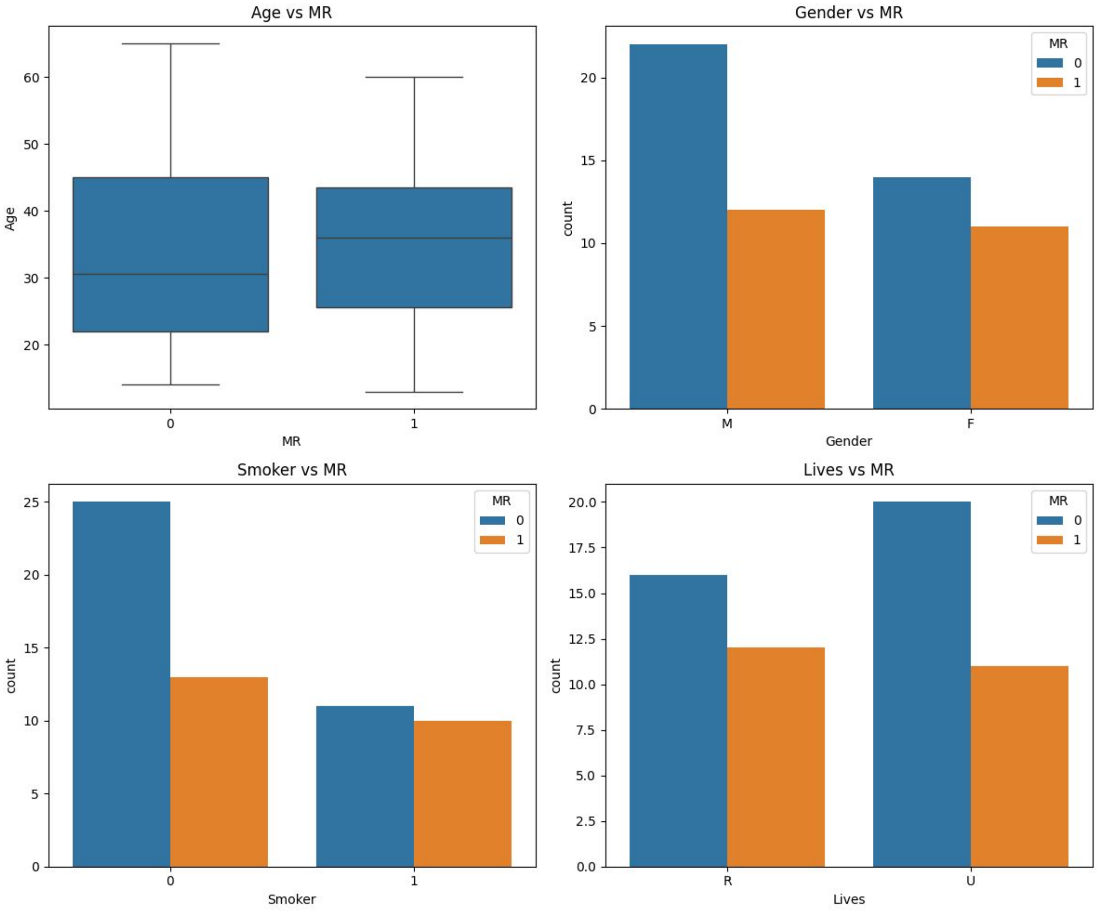
- The distributions of Age over all the disease labels follow approximately a normal distribution



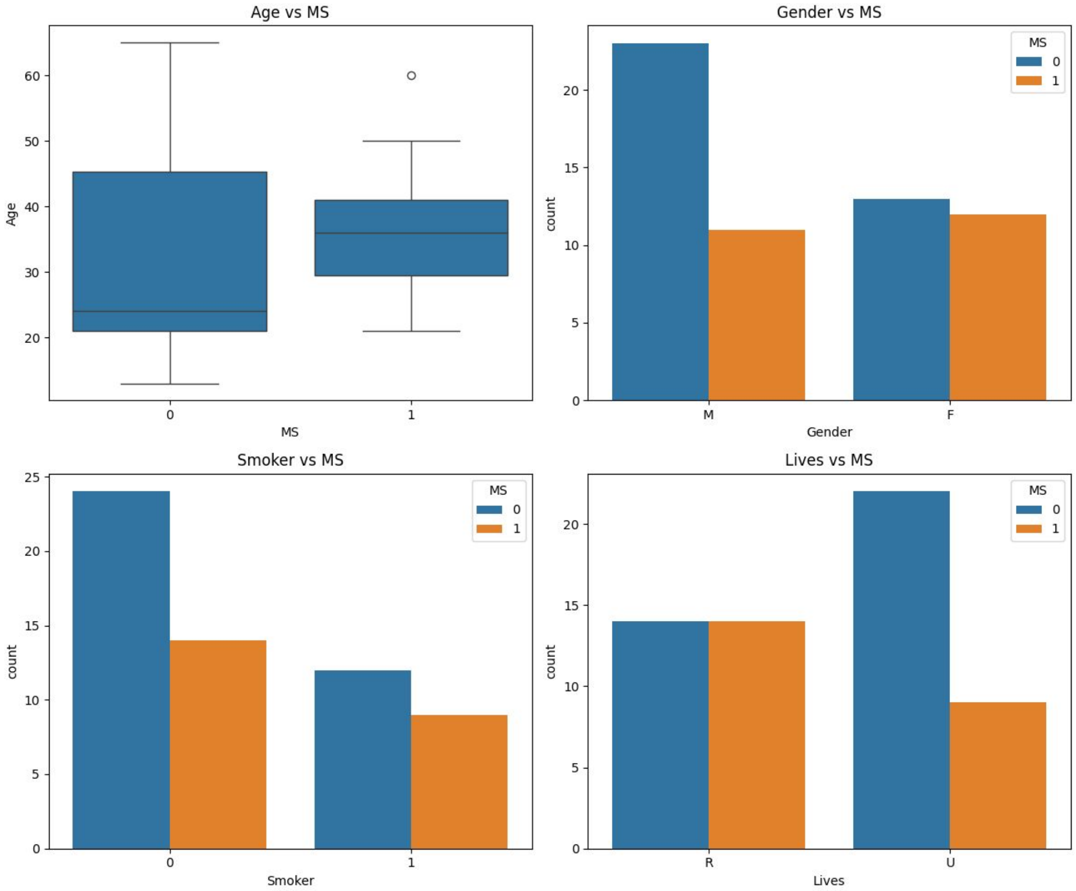
- AR significantly found in males
- Smokers are more prone to be victim



- AS significantly found in males
- Smokers are more prone to be victim



- MR is equally found in Females
- Non smokers also suffer this



- MS is slightly more found in Females
- Non smokers also suffer this
- Significant difference observed for MS disease for living area. R has much higher count than U

Moving Forward

- Possible trial setting
  - Combining all recordings instead of tackling separately
  - Initial weights from more relevant fine tuned tasks like [ICBHI respiratory challenge]
  - Different model hyperparameters

- Larger, more robust dataset

Thank you!

Any questions or suggestions welcomed