Minutes of the last meet

- **Modulation Spectrogram Approach : CV approach**

- Multichannel input considering images with different window sizes.

Window sizes used in Extracting Spectrogram Images from Audio

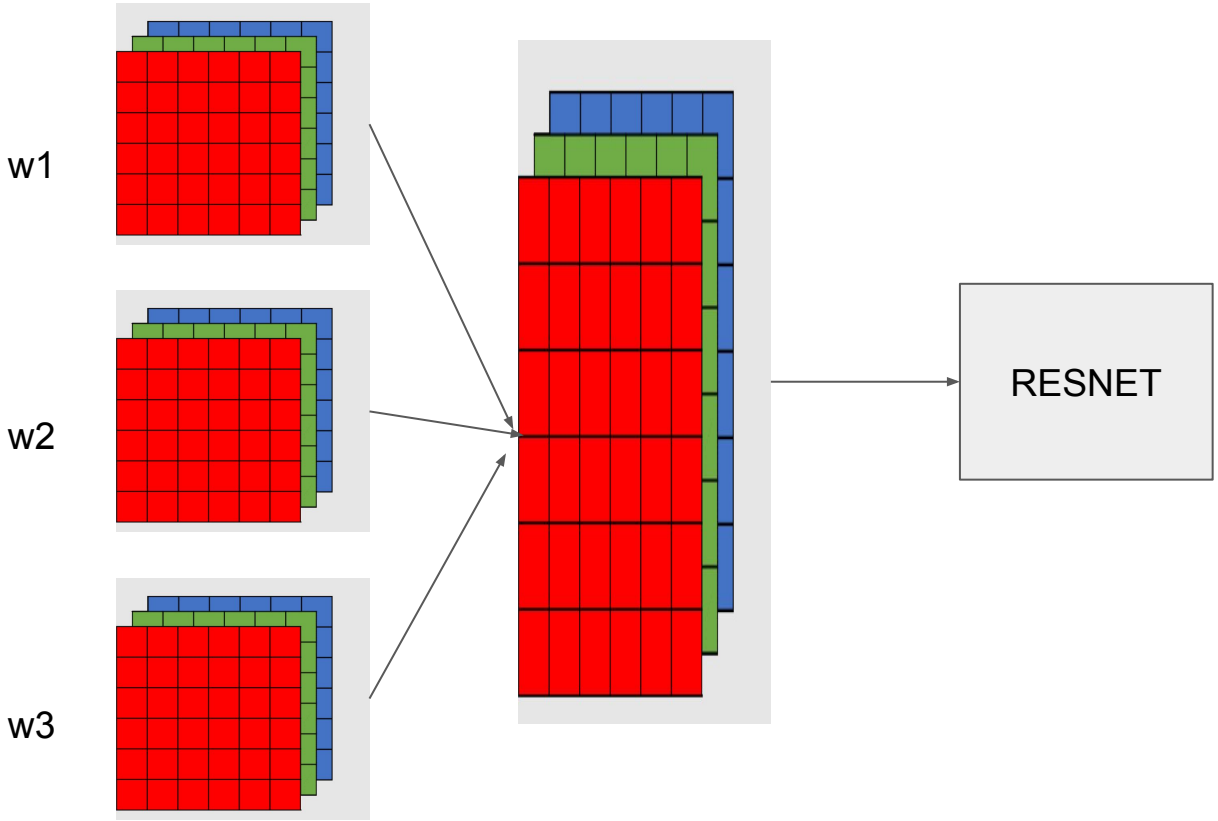
- 0.04 (was default)
- 0.004
- 0.01

```python
# FUNCTIONS FOR MODULATION SPECTROGRAM
def modSpec(x, fs,win_size_sec=0.04):
    # win_size_sec = 0.04  # window length for the STFFT (seconds)
    win_shft_sec = 0.01  # shift between consecutive windows (seconds)

    stft_modulation_spectrogram = ama.strfft_modulation_spectrogram(
        x,
        fs,
        win_size=round(win_size_sec * fs),
        win_shift=round(win_shft_sec * fs))

    return stft_modulation_spectrogram
```

Multichannel approach using 3 Images with Different Window sizes



STACKED OVER AXIS 0

NEW INPUT DIM :

```
(1440, 640, 3)
```

PHQ 8 Dataset

Window sizes used this week
with the new data set

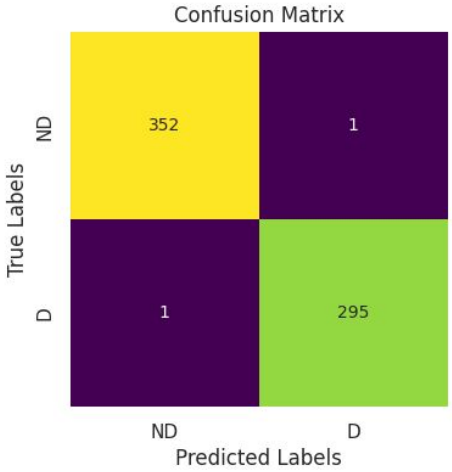
- 25ms
- 400ms
- 800ms

RESNET 50

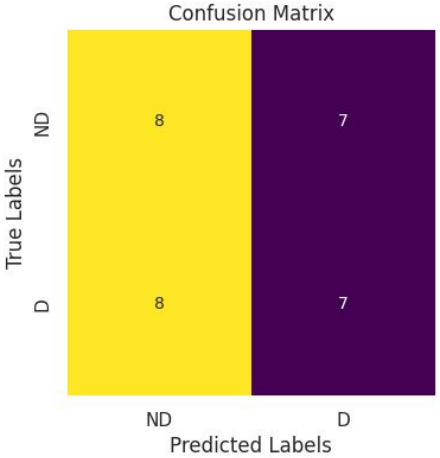
TRAIN ACCURACY : 99.69

TEST ACCURACY : 50

TRAIN



TEST

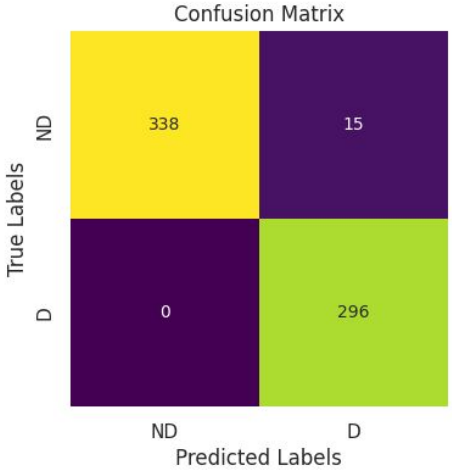


RESNET 18

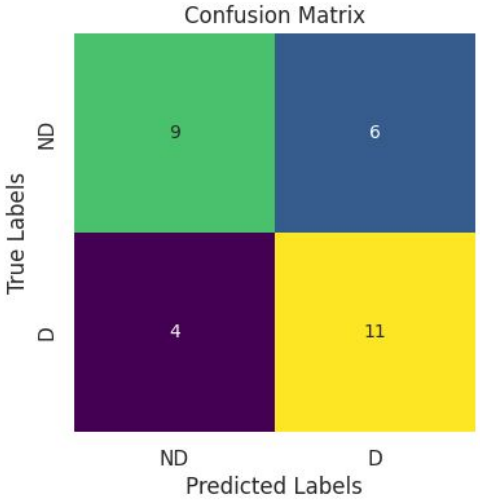
TRAIN ACCURACY : 97.68

TEST ACCURACY : 66.67

TRAIN



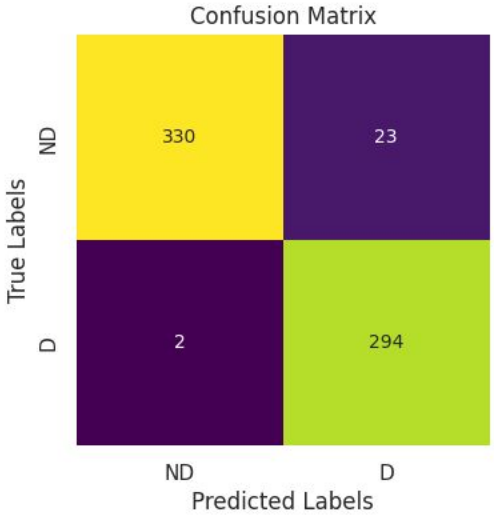
TEST

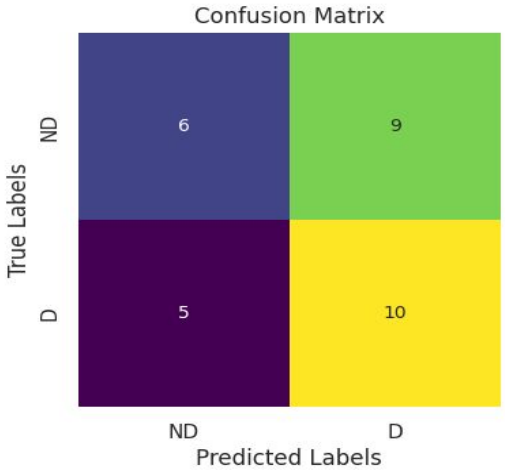


Vision Transformer

TRAIN ACCURACY : 96.1

TEST ACCURACY : 53.37

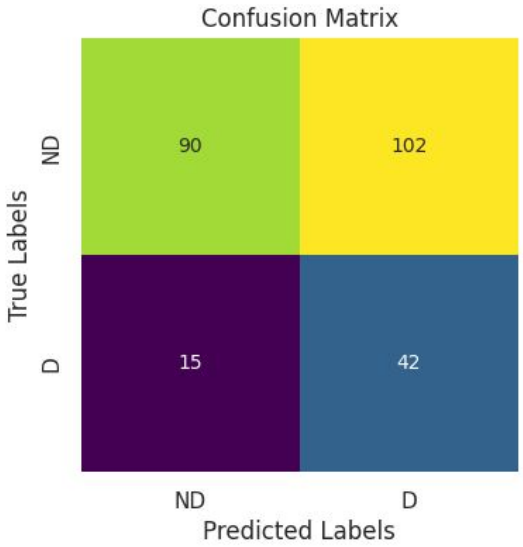


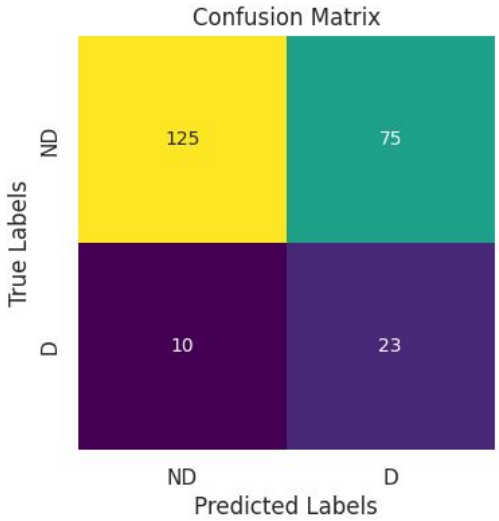


Vision Transformer

EATD Dataset
TRAIN ACCURACY : 63%

TEST ACCURACY : 53.37





Summary and Future work

- Resnet152 performs no better than other 2 models
- Above 3 seconds dataset (19 GB)
- Images are generated , needs to be trained and tested on GPU
- Vision Transformer