Audio Classification System using YAMNet and Wav2Vec2 Models

Front-Era Health Assessment Challenge

Ayush Raina

Indian Institute of Science

February 24, 2025

Dataset

Due to time constraints and limited GPU Compute available we used the total of 3950 datapoints with following distribution:

Class	Number of Datapoints
Cry	457
Not Screaming	2631
Screaming	862

I was not able to download the other datasets whose links were provided in the challenge description, so I proceeded with the above dataset.

Splitting the Dataset

We did the following splits:

- Train: 70%
- Validation: 15%
- Test: 15%

Following are the number of datapoints in each split per class:

Class	Train	Validation	Test
Cry	320	69	68
Not Screaming	1842	394	395
Screaming	603	129	130

YAMNet Model

YAMNet Model

1. Waveform creation using librosa library, sample rate = 16kHz
2. Extracting embeddings which will be used to finetune the model.

Technique Used

I performed transfer learning instead of finetuning the model from scratch. Transfer Learning gave better results. Embeddings are output of the model before the final layer. I used these embeddings to train a simple feedforward neural network to get our final model.

Results

Here are the results of the YAMNet Model:

Model	Train Accuracy	Validation Accuracy	Test Accuracy
YAMNet	97.2%	91.5%	91.2%

Loss Function Used: Sparse Category CrossEntropyLoss
Optimizer Used: Adam
Learning Rate: 0.0005 (default with Adam)
Batch Size: 32

Results

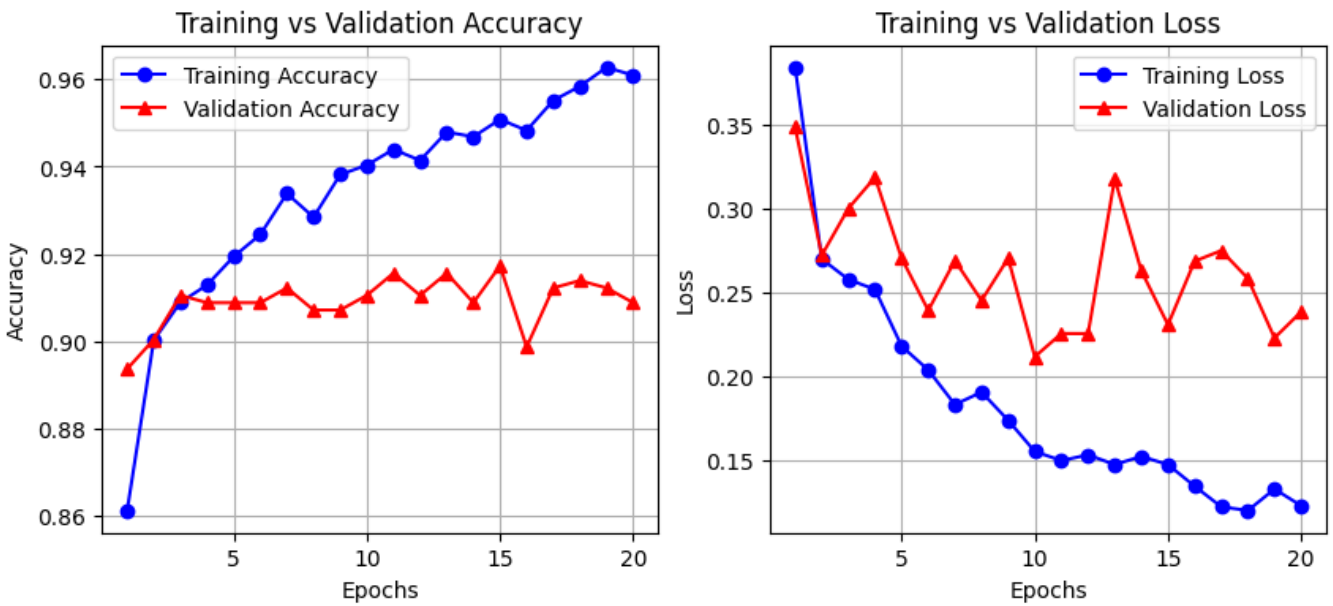

Figure 1: Training Curves

Results

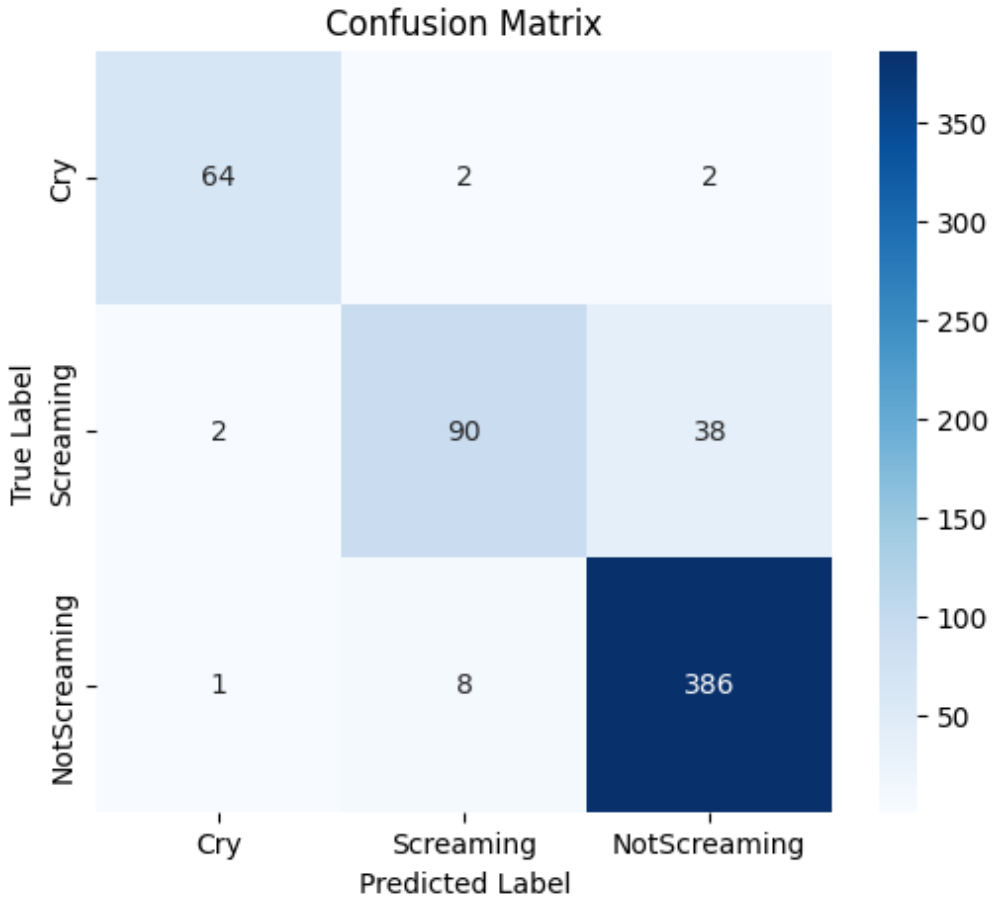

Figure 2: Confusion Matrix of YAMNet Model

Results

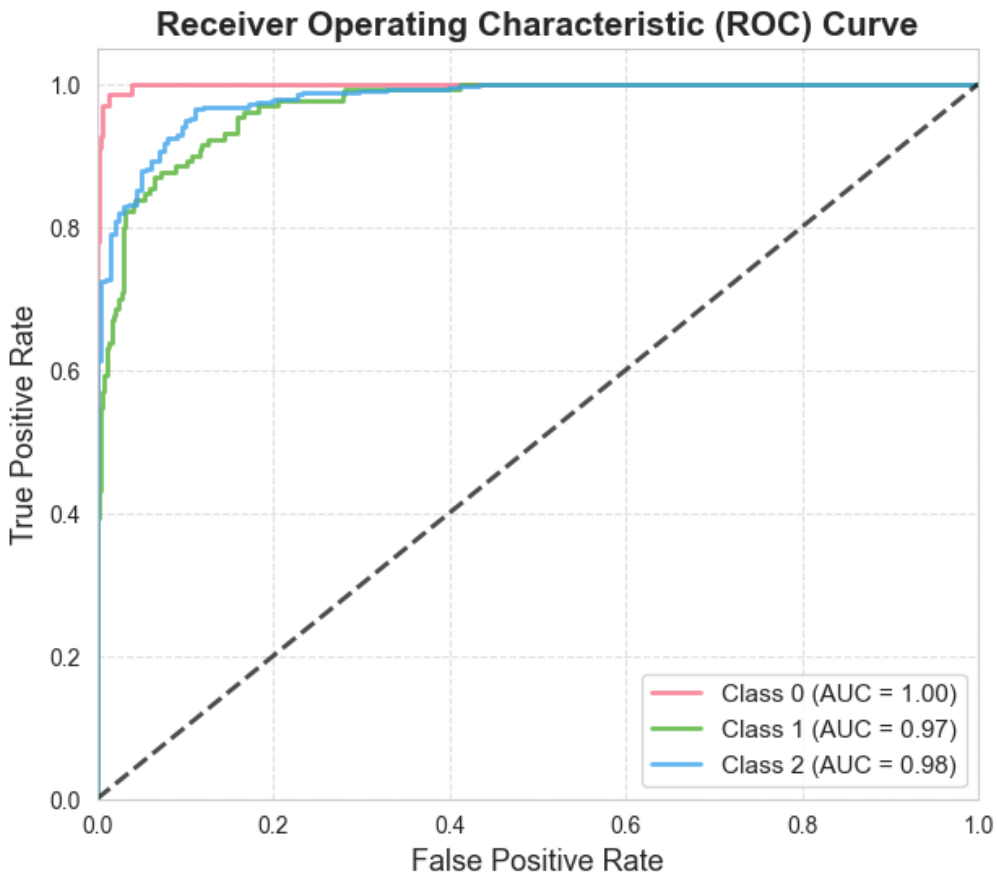

Figure 3: ROC AUC Curve of YAMNet Model

Results

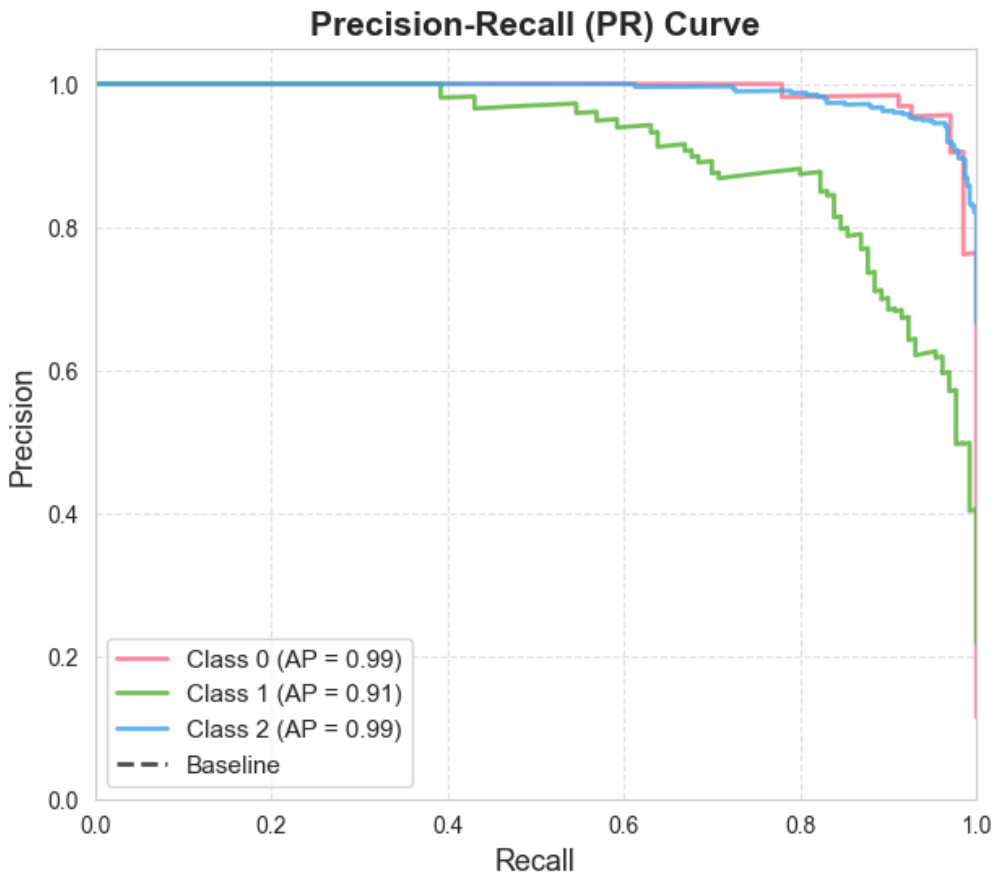

Figure 4: PR Curve of YAMNet Model

Wav2Vec2 Model

Wav2Vec2 Model

1. Waveform creation using torchaudio library, sample rate = 16kHz
2. Here again I used transfer learning by freezing the model and training the multi-layer perceptron on top of it.
3. If 2 channels are present, I combined them to get a single channel by taking the mean of the two channels.

Results

Here are the results of the Wav2Vec2 Model:

Model	Train Accuracy	Validation Accuracy	Test Accuracy
Wav2Vec2	99.17%	90.2%	91.07%

Loss Function Used: CrossEntropyLoss
Optimizer Used: Adam
Learning Rate: 0.0001
Batch Size: 32

Results

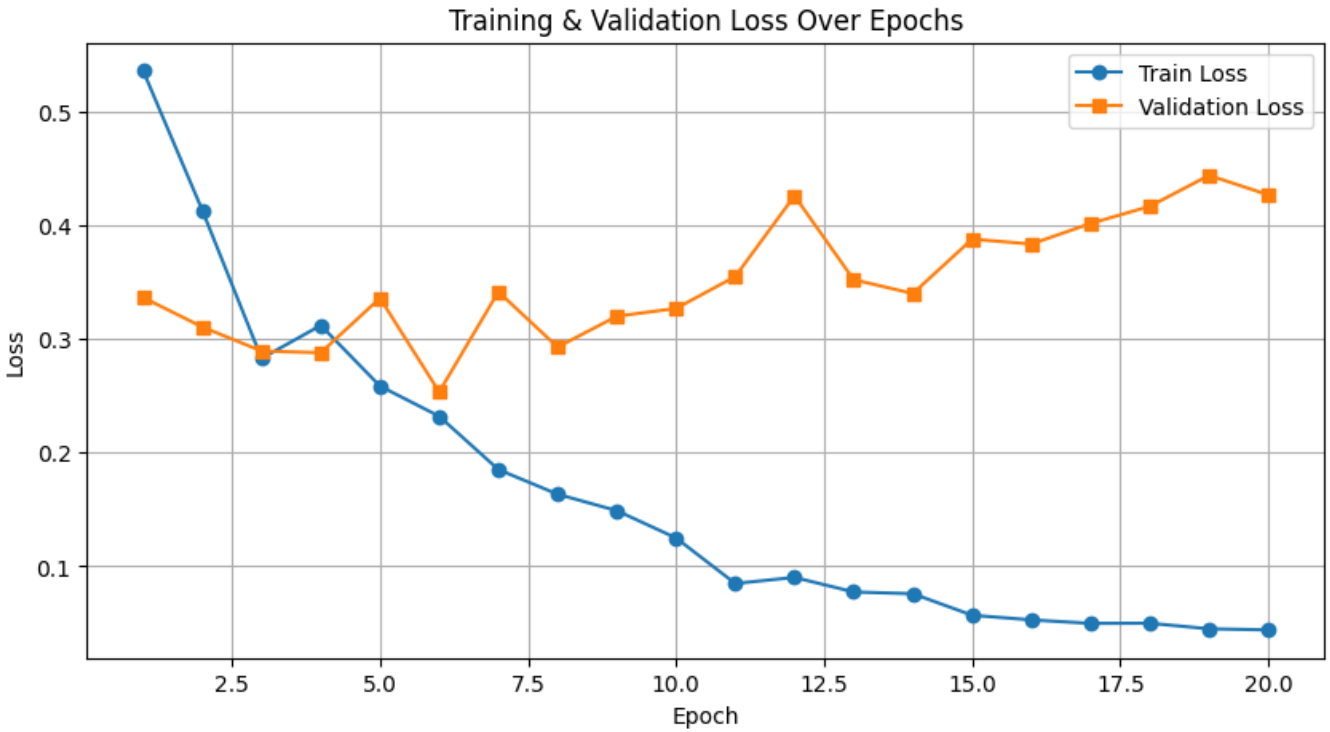

Figure 5: Training Curve 1

Results

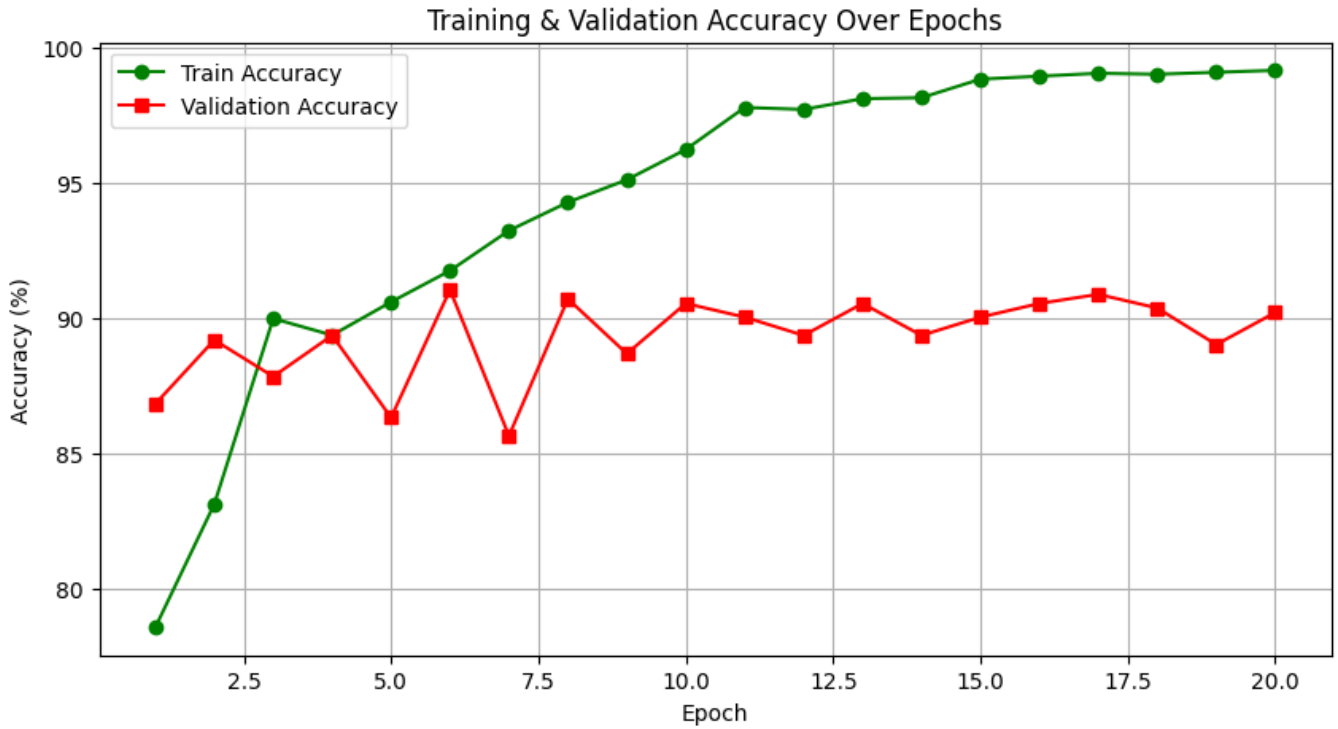

Figure 6: Training Curve 2

Results

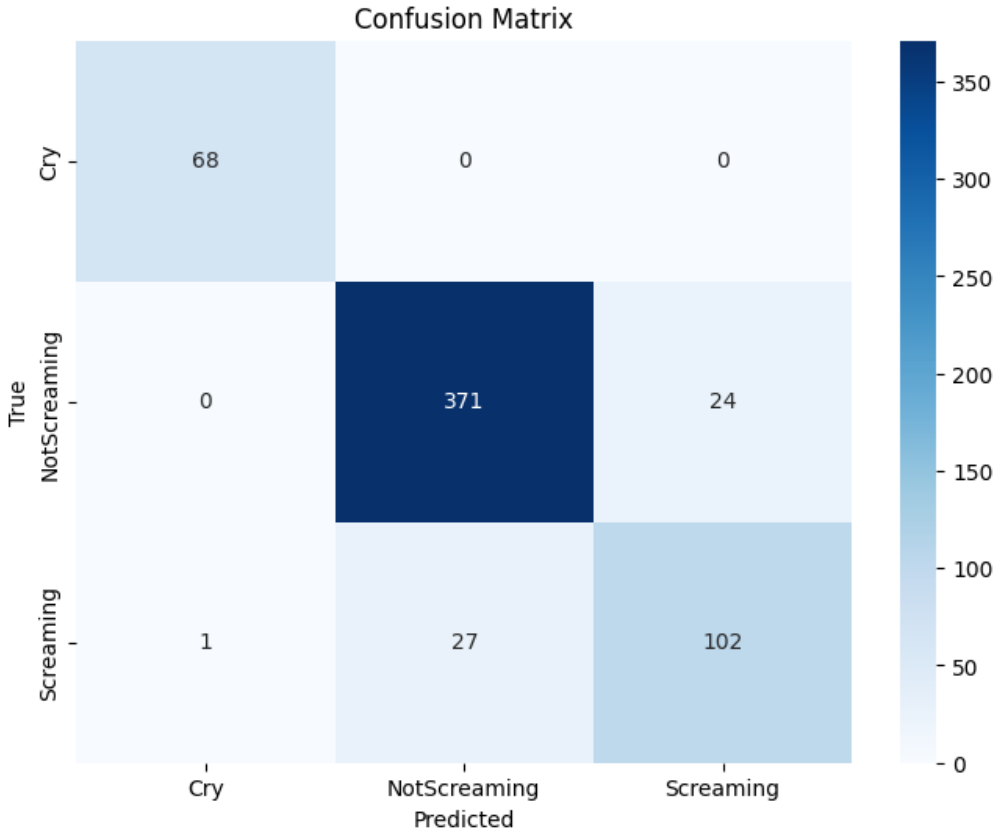

Figure 7: Confusion Matrix of Wav2Vec2 Model

Results

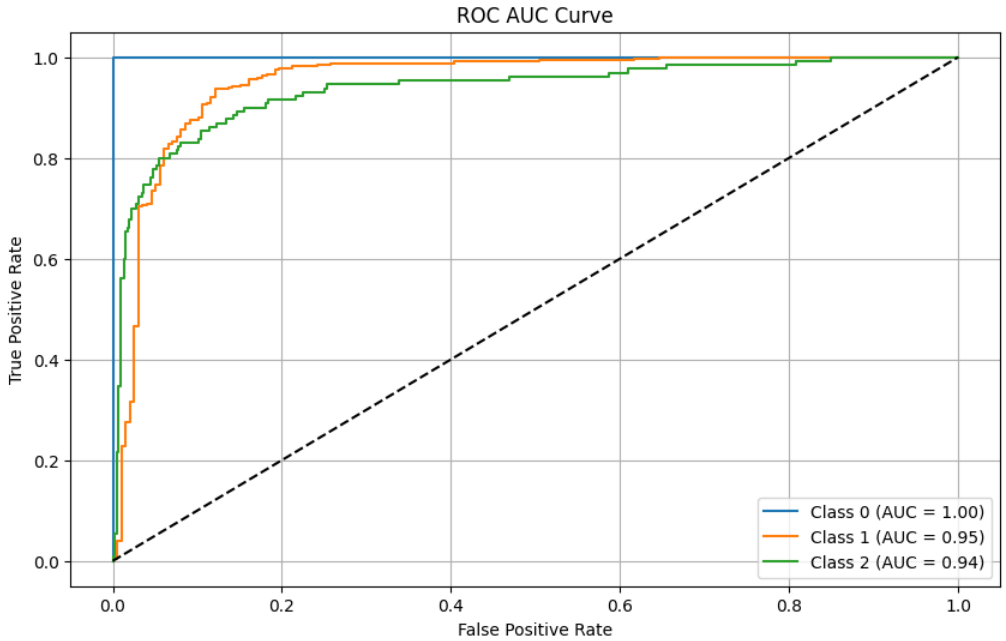

Figure 8: ROC AUC Curve of Wav2Vec2 Model

Results

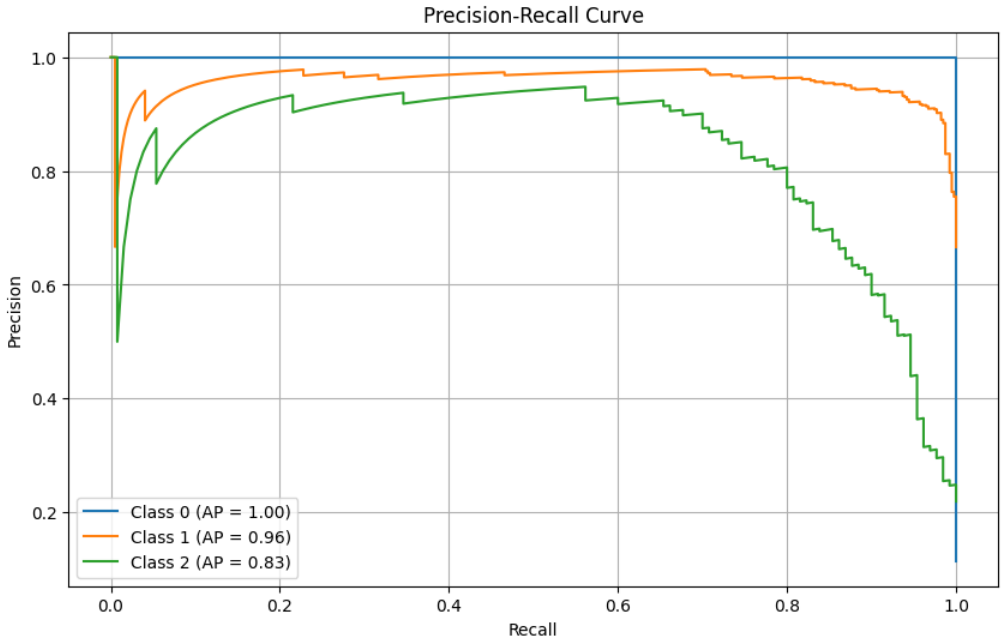

Figure 9: PR Curve of Wav2Vec2 Model

Thank You!

Thank You!