

# Speech Understanding Programming Assignment 3

Bikram Majhi (M23CSA007)

## Links

- **GitHub Repository:** <https://github.com/bikrammajhi/Speech-Assignment-3>
- **Gradio Demo:** <https://huggingface.co/spaces/iiserkbikram/Audio-DeepFake>

## 1 Objective

The objective of this report is to classify audio samples into two categories: Real and Fake. The task involves using a pre-trained Self-Supervised Learning Wav2Vec (SSL W2V) model, finetuning it on a specific dataset, and evaluating its performance on custom and FOR datasets.

## 2 Methods

1. Utilized the pre-trained SSL W2V model, which was previously trained on LA and DF tracks of the ASVSpooof dataset.
2. Downloaded a custom dataset for initial evaluation.
3. Finetuned the SSL W2V model on the FOR dataset.
4. Evaluated the model's performance on both the FOR and custom datasets.

Hyperparameter	Value
Learning Rate	$5 \times 10^{-5}$
Optimizer	Adam
Number of Epochs	10

Table 1: Hyperparameters used for finetuning the model on the FOR dataset.

### 3 Results

Model	AUC	EER
Pretrained Model (on Custom dataset)	0.3576	0.6111
Finetuned Model (on FOR dataset)	0.6947	0.3768
Finetuned Model (on Custom dataset)	0.1651	0.7333

Table 2: Performance of the pre-trained and finetuned SSL W2V models on custom and FOR datasets.

### 4 Analysis

The pre-trained SSL W2V model performed poorly on the custom dataset, with a high EER of 0.6111 and a relatively low AUC of 0.3576. After finetuning the model on the FOR dataset, there was a significant improvement in performance when evaluated on the FOR dataset, with the EER dropping to 0.3768 and the AUC increasing to 0.6947.

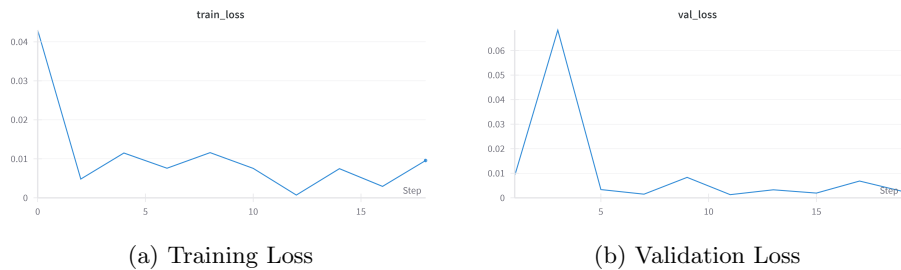


Figure 1: Comparison of Training and Validation Loss

However, when the finetuned model was tested on the custom dataset, the performance dropped significantly, with an AUC of 0.1651 and an EER of 0.7333. This suggests that the finetuned model may have overfit to the FOR dataset, resulting in poor generalization to the custom dataset.

### 5 Conclusion

The pre-trained SSL W2V model did not perform well on the custom dataset, but finetuning it on the FOR dataset led to a significant improvement in performance on the FOR dataset. However, the finetuned model did not generalize well to the custom dataset, indicating potential overfitting to the FOR dataset. Further work could involve finetuning the model on a more diverse dataset or implementing regularization techniques to improve generalization.