Indian Institute of Technology, Mandi

School of Computing and Electrical Engineering (SCEE)

Course: Deep Learning (CS-671)

AITECH Hackathon 2025

Powered by HCLTech

Group - 4

A Report Submitted for

Truth or Trap – Fake Speech Detection Using Deep Learning (PS-8)

May 4, 2025

Group Members

Siddharth Rajesh Tiwari - S24035

Pradyumn Singh Sikharwar - S24047

Himanshi - S24038

Shiv Bhaskar - S24024

Samiran Datta - S24037

Shweta Sharma - S24044

Contents

1 Methodology

1.1 Overview of SSL Anti-Spoofing Architecture

To address the challenge of fake speech detection, we implemented an advanced anti-spoofing model that combines:

- A pre-trained wav2vec 2.0 XLSR model for feature extraction
- A hybrid graph-based deep learning architecture for classification

1.2 Advantages Over Conventional CNNs

1. **Self-Supervised Learning Features:** The wav2vec 2.0 XLSR model leverages large-scale unlabeled speech corpora to extract rich contextual audio representations.
2. **Dual-path Spectral and Temporal Processing:** Separate modeling paths for spectral and temporal domains allow detection of unique inconsistencies in fake speech.
3. **Graph Attention Networks (GAT):** GAT layers learn complex intra-frame relationships in the speech signal.
4. **Heterogeneous Graph Attention (HtrGAT):** Enables information exchange between spectral and temporal frames, improving robustness.
5. **Generalization and Performance:** The architecture achieves state-of-the-art results on ASVspoof 2021 LA and DF benchmarks.

1.3 Why We Chose This Model

Compared to traditional CNNs, our approach provides:

- Better generalization using wav2vec features
- Resilience to spoofing attacks across modalities
- Stronger temporal–spectral fusion via graph attention

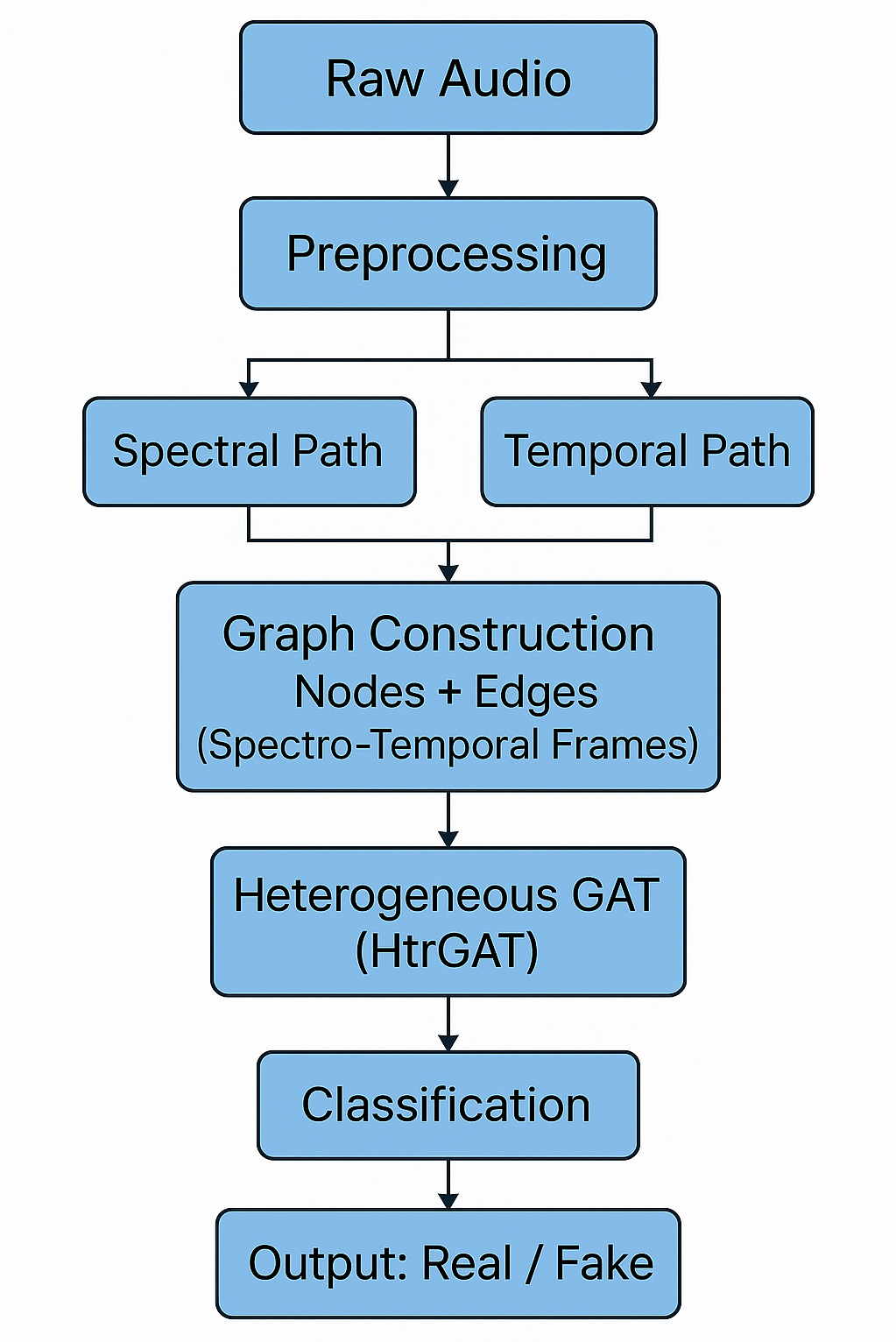

Figure 1: Pictorial Pipeline of Proposed Fake Speech Detection Method

2 Evaluation Metrics and Results

2.1 Confusion Matrix

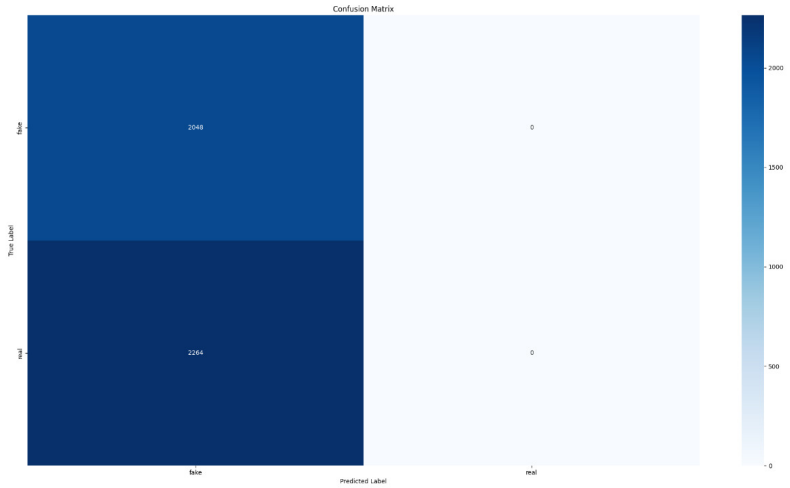

Figure 2: Confusion Matrix – all samples predicted as "Fake"

Inference:

- All 2048 fake samples were correctly identified.
- All 2264 real samples were misclassified as fake.

2.2 Classification Report

Figure 3: Precision, Recall, and F1-Score

Inference:

- **Fake Class**: Precision = 0.47, Recall = 1.00, F1 = 0.64
- **Real Class**: All metrics are zero

2.3 Training and Validation Loss

Class	Precision	Recall	F1-Score	Support
Fake	0.47	1.00	0.64	2048
Real	0.00	0.00	0.00	2264
Accuracy			**0.47**	**4312**
Macro Avg	0.24	0.50	0.32	4312
Weighted Avg	0.23	0.47	0.31	4312

Figure 4: Training and Validation Loss per Epoch

Inference:

- Losses converge rapidly
- Model collapses to single class output

2.4 t-SNE Embedding Visualization

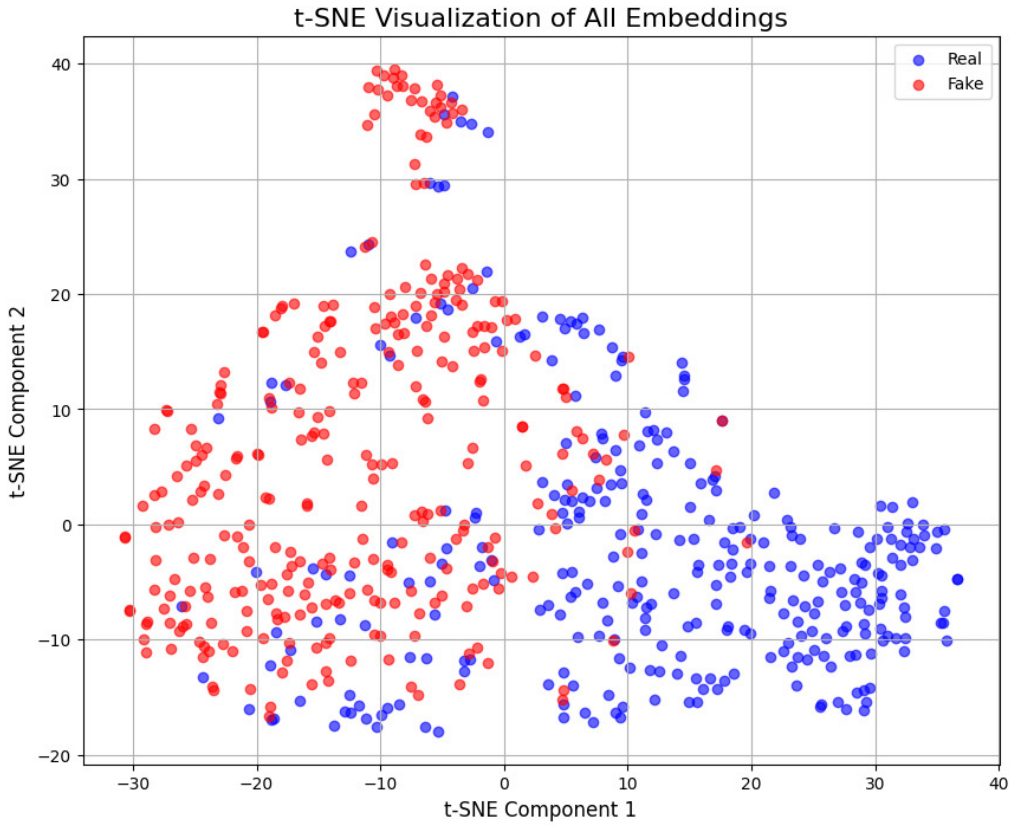

Figure 5: t-SNE Visualization of Real vs Fake Embeddings

Inference:

- Moderate separation in feature space
- Embedding quality not fully exploited by classifier

2.5 Root Cause Analysis

- Class imbalance
- Lack of regularization
- No use of weighted loss functions

3 Conclusion and Observations

- **Training Success, Inference Failure**: Despite low loss, generalization is poor.
- **Bias Observed**: The model predicts everything as Fake.
- **Suggestions for Improvement**:
  - Apply Focal Loss or Class-Balanced Cross-Entropy
  - Add data augmentation
  - Improve sampling strategies
- **Gradio Interface**: A live web-based testing interface has been deployed using Gradio.