Advanced Foundations for Machine Learning Course Project

Two-Factor Audio Verification Using Biometric Voiceprints and GAN-Based Watermarking

Abhinandan Sudharsan, PES1UG23AM013
Abhinav Bhargava, PES1UG23AM014
Anjali H Ramurs, PES1UG23AM053
Anurag Senapati, PES1UG23AM059

Advanced Foundations for Machine Learning

Submission Checklist

No	Feature Description	Drive Shared (Y/N)
1	Code notebook	Y
2	Dataset or dataset source	Y
3	This PPT	Y
4	The 5 mins Video presenting your paper	Y
5	Brief Project Report in IEEE Format	Y

Note: Mention the **email-ids** you shared the drive with (your course TAs and your faculty)

Course TA - namitaachyuthan@gmail.com
Faculty - bhaskarjyotidas@pes.edu

Problem Statement

Develop a robust system to verify the authenticity and integrity of audio recordings by combining biometric speaker recognition with imperceptible audio watermarking to detect manipulation, deepfakes, and tampering.

Uniqueness

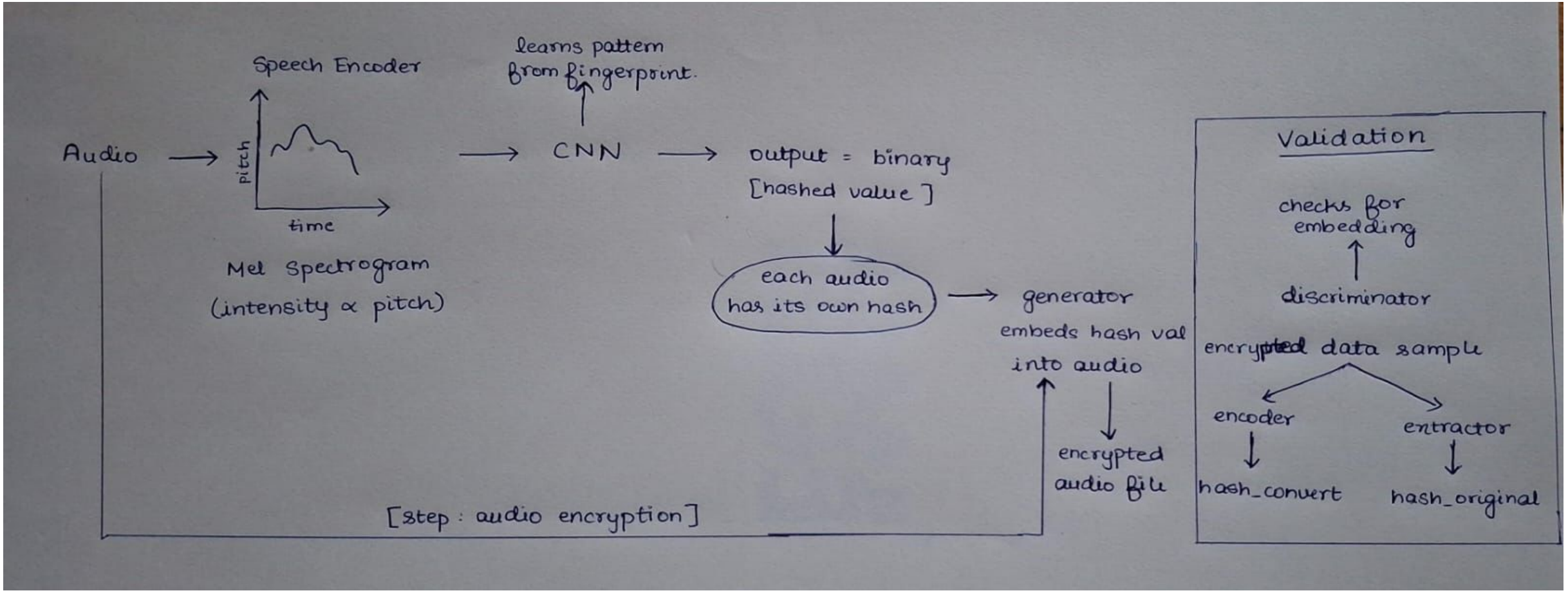
- Combines two separate verification methods—biometric speaker recognition and digital watermarking—into one system.
- Uses deep neural networks for voiceprint extraction and GANs for imperceptible watermarking.
- Achieves authentication with very low error rates, robust to ambient noise and tampering.
- Addresses novel challenges posed by audio deepfakes and highly realistic forged recordings.

Dataset

- Source: LibriSpeech (train-clean-100 and test-clean subsets)
- Size: Thousands of diverse speech samples from multiple speakers
- Attributes: Clean, labeled audio files; augmented with MUSAN noise dataset for robustness

Advanced Foundations for Machine Learning

Overall Design or Approach



Final Results

- Initial bit error rate (BER) reduced from 0.20 to 0.0000 after fine-tuning, showing perfect watermark recovery in tests.
- Similarity score for speaker verification: 0.8303 (authentication threshold=0.7).
- Throughput: ~295 samples/second; average batch time ~108 ms.
- System correctly classified authentic recordings, reliably rejected tampered ones.

Advanced Foundations for Machine Learning

Features : Done vs. Remaining to be done

No	Description	Done or To be Done ?
1	Speaker Encoder model training	Done
2	GAN Watermark embedding and module development	Done
3	System integration and pipeline	Done
4	Noise augmentation experiments	Done
5	Real-world/complex audio testing	To be done

Features : Who did what

No	Feature Description	Contributed By
1	Speaker Encoder	Anjali & Anurag
2	GAN watermarking	Abhinandan & Abhinav
3	Integration and evaluation	Abhinandan & Abhinav
4	Dataset preprocessing	Anjali & Anurag

Advanced Foundations for Machine Learning

Quantity and Quality of Work

No	Code Functionality	% Complete	Runs w/o Issues (Y /N)	State minor issues
1	Speaker Encoder training	100%	Y	None
2	GAN training and watermarking	100%	Y	Needs further noise tuning
3	Verification pipeline	90%	Y	Latency optimization needed
4	Dataset collection/prep	100%	Y	None

Top Few Learnings from this Project

No	Description
1	Combining biometric and watermarking improves audio security.
2	GANs can create imperceptible yet robust watermarks.
3	Data augmentation enhances model robustness to noise.
4	Real-time throughput can be achieved with model optimization.
5	End-to-end integration is essential for reliability.

Top Unresolved Challenges

No	Description
1	Handling severe real-world noise and compression
2	Comprehensive attack simulation and adversarial testing
3	Generalizing to longer audio segments
4	Further speeding up the verification process
5	Extending watermarking to multi-speaker/speech scenarios

References, if any

No	Paper title	Year of publication
1	Deepfake Detection in Call Recordings: A Deep Learning Solution for Voice Authentication	2025
2	Deepfake Audio Detection via MFCC Features Using Machine Learning	2022
3	Robust Audio Watermarking Against Manipulation Attacks Based on Deep Learning.	2024
4		
5		

THANK YOU

November 2025

Department of Computer Science and Engineering in AI &ML