

# Differential-based DEEPPFAKE SPEECH detection



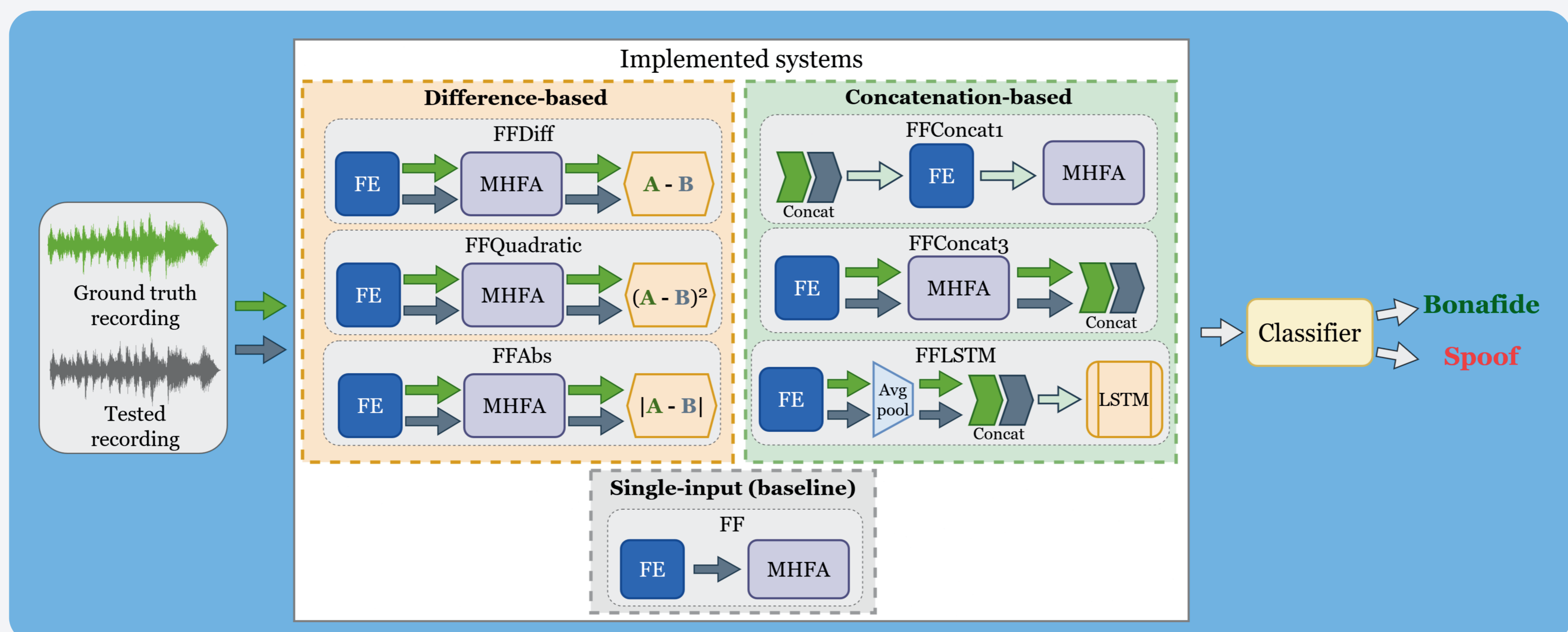
## MOTIVATION

- High-quality realistic deepfakes
- Brand-new approach
- Pairs of recordings
  - Tested and ground-truth
- Prove authenticity
- Expose fraud

## TECH & DATA

- PyTorch
- Self-Supervised Learning (SSL)
- Multi-head factorized attentive pooling (MHFA)
- ASVspooof challenge datasets (2019 + 2021)
- In-the-Wild dataset of circulating deepfakes

## IMPLEMENTATION



## RESULTS

Equal Error Rate (EER)	Best to date	Implemented	Equal Error Rate (EER)	Voice Conversion	Text to Speech
ASVspooof2019 [1]	5.74% (single T04) 0.22% (ens. T05)	0.18% (FF) 0.22% (FFConcat3)	Single input	6.99% (FF)	1.22% (FF)
ASVspooof2021 [2]	15.64% (T23) 2.85% (W2V2 + GAT)	5.51% (FFConcat1) 3.1% (ensemble)	Difference based	6.15% (FFDiff)	1.38% (FFQuadratic)
In-the-Wild [3]	33.94% (RawNet2)	12.36% (FFConcat3) 9.76% (ensemble)	Concat based	4.98% (FFConcat1)	1.29% (FFConcat3)

- State-of-the-art performance
- Pair input is more robust to overfitting
- Pair input significantly better for detecting VC

[1] ASVspooof 2019: Future Horizons in Spoofed and Fake Audio Detection, *Todisco et al., 2019, Interspeech 2019*

[2] ASVspooof 2021: Towards Spoofed and Deepfake Speech Detection in the Wild, *Xuechen Liu et al., 2023, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 31(1):2507-2522*

[3] Does audio deepfake detection generalize?, *Nicolas M. Müller et al., 2022, Interspeech 2022*