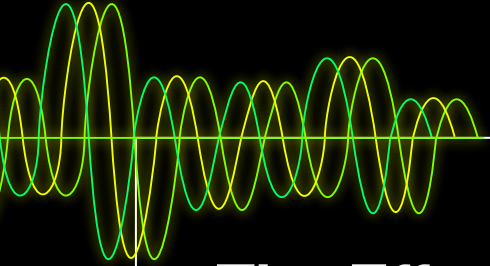


The Effect of Autoencoder Audio Feature Reduction on Deepfake Audio Detection with Convolutional Neural Network

Rabia Gondur & Tony Wen
December 18th, 2023

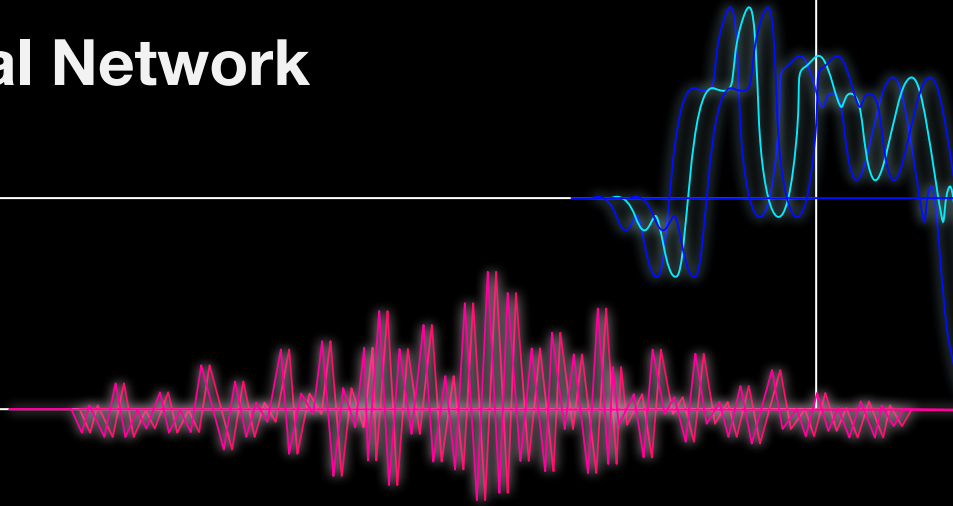
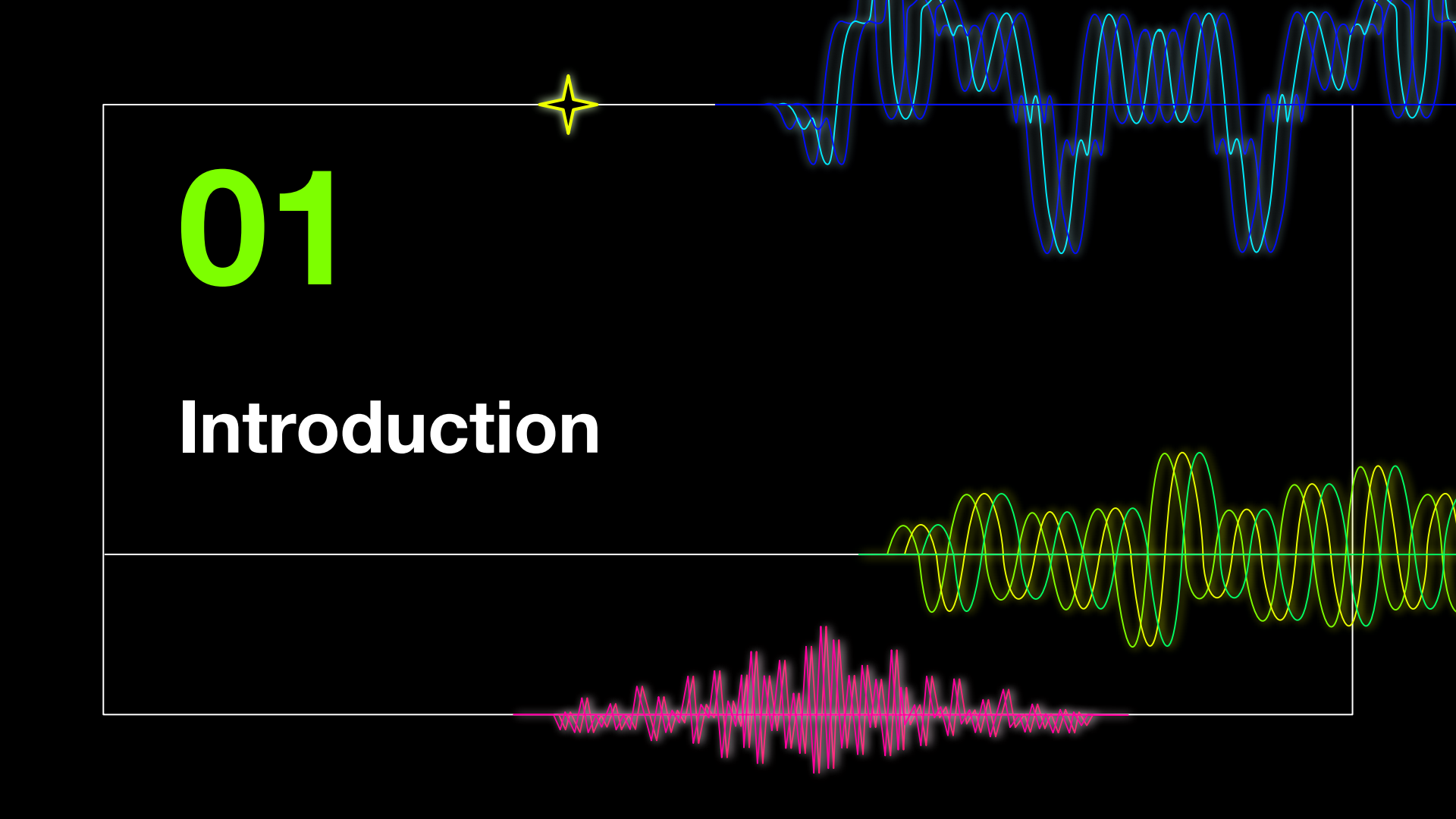


TABLE OF CONTENTS


01
Introduction

Previous Work & Limitations

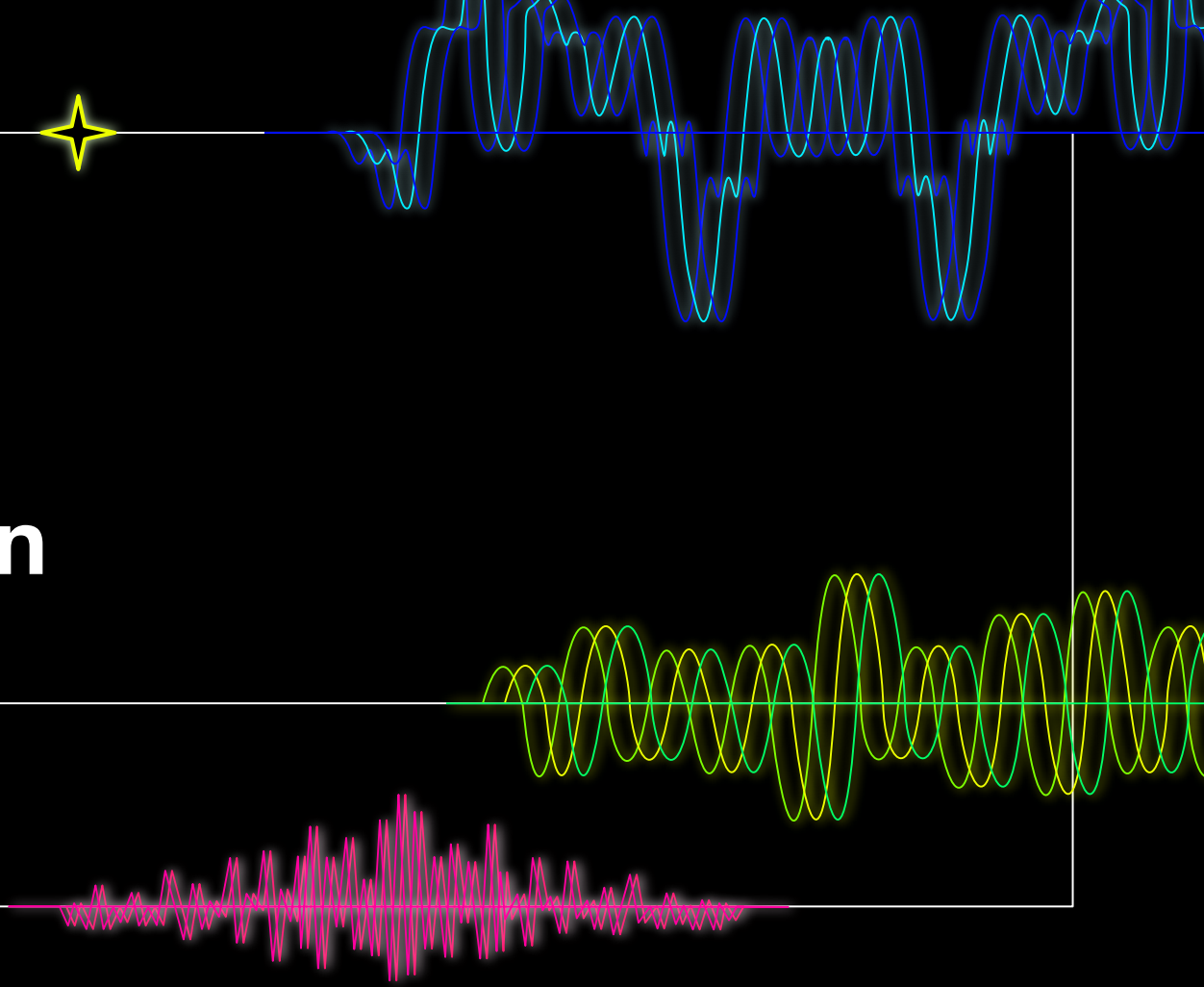
Deepfake audios pose a large threat to public security. Therefore, there is a lot of research on efficient deepfake audio classification

- E.g. : Hamza et al. (2022) investigated deepfake audio detection using CNNs. In their research, they specifically focused on one of the components of audio data called Mel-frequency cepstral coefficients (MFCC features)

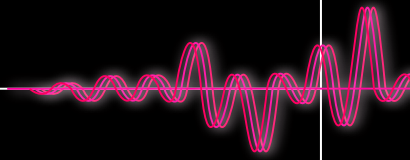
Although these models are efficient and have high accuracy of detecting deepfakes, many of them deploy the same architecture and use popular audio features

02

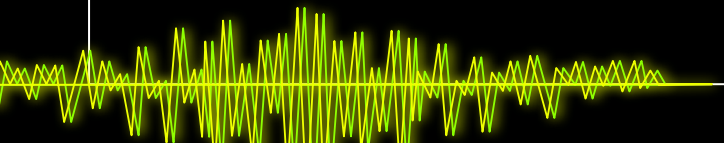
Motivation



Motivation

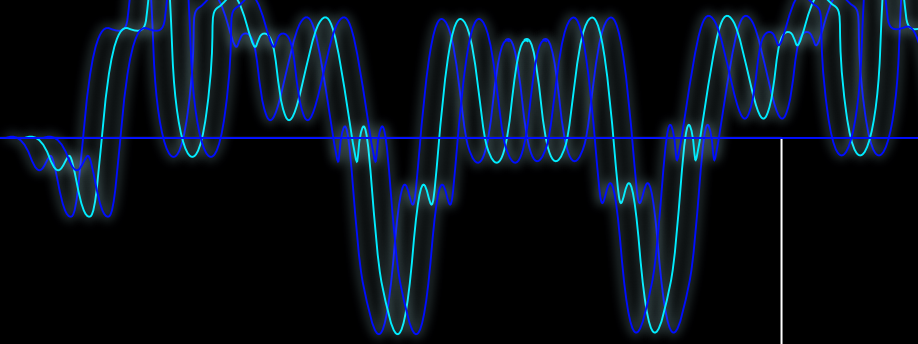


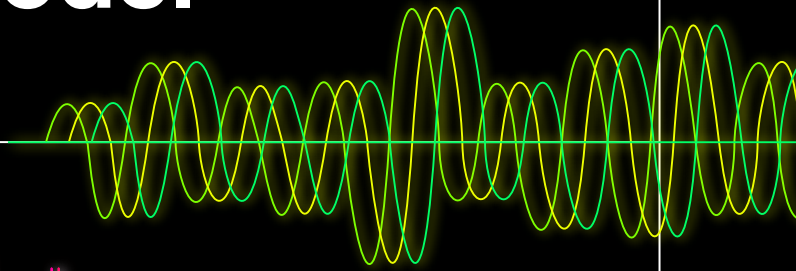
Create a hybrid model composed of an autoencoder and a CNN for more robust classification of deepfake audio, as well as compare two popular audio features (MFCC & LFCC) with two less used features (CQCC & LPCC).

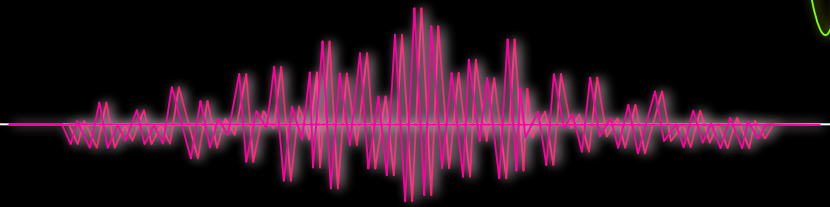


03

Data & Proposed Model







Data Source

- Biennial challenge since 2015
- Logical Access + Physical Access + **DeepFake**
- 611829 .flac audio files (16kHz, 16-bit)
  - 9 codec formats, 3 vocoder types
  - 22600 spoofed + 22600 real for analysis
  - Samples w. metadata:

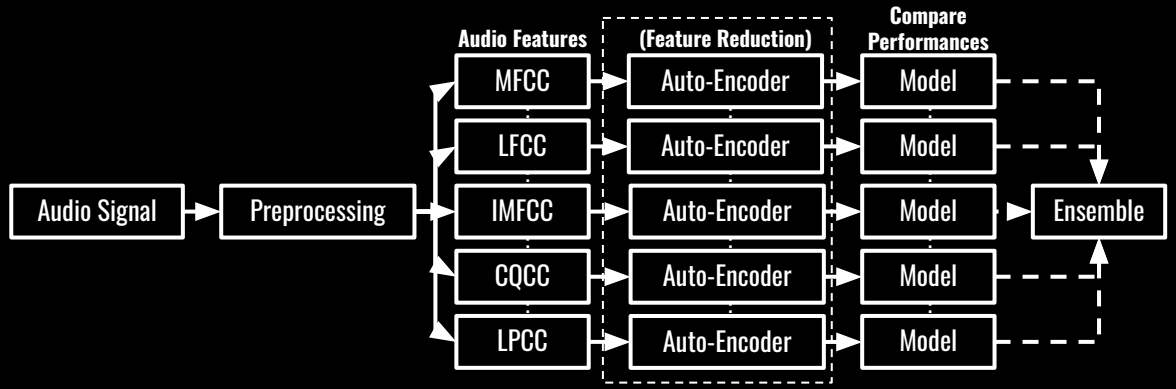
TEM2 DF_E_2000079 high_m4a vcc2020 - **bonafide** notrim eval bonafide - - - -

VCC2TF2 DF_E_2000476 high_ogg vcc2018 SPO-N04 **spoof** notrim eval **traditional_vocoder** SPO N04 FF -

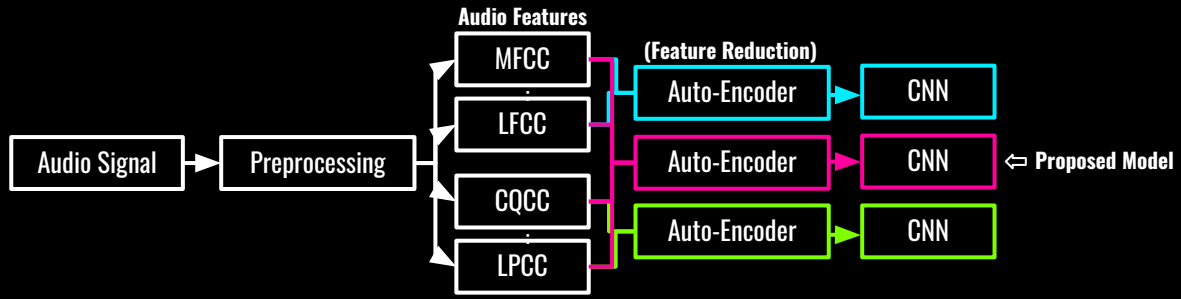
LA_0032 DF_E_2000563 high_ogg asvspoof A15 **spoof** notrim eval **neural_vocoder_autoregressive** - - - -

TEM1 DF_E_2000346 high_mp3 vcc2020 Task1-team27 **spoof** notrim eval **neural_vocoder_nonautoregressive** Task1 team27 FM E

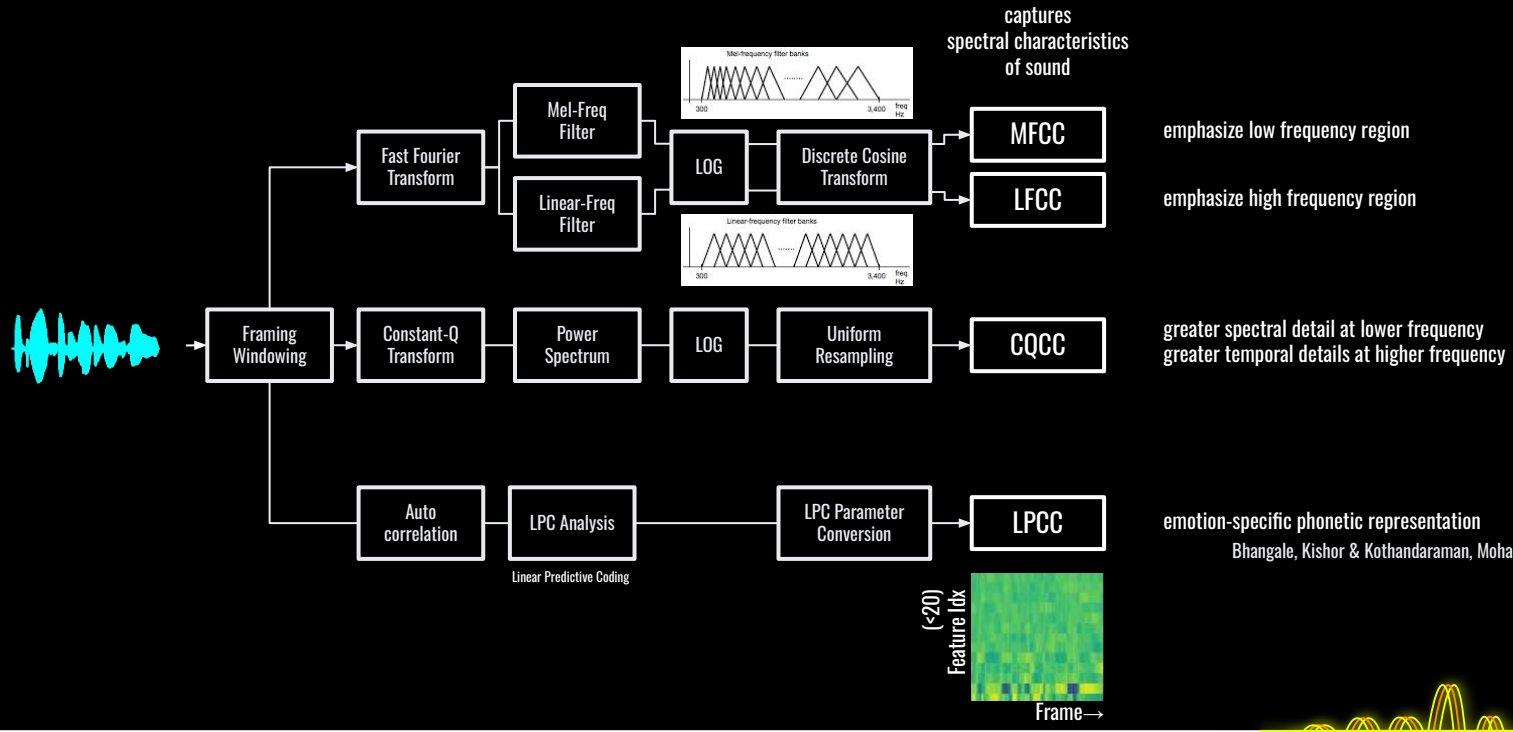
Motivation & Proposed Model



B. T. Balamurali, K. E. Lin, S. Lui, J. -M. Chen and D. Herremans, (2019)



Audio Feature Background



Data Processing

Safe

Pre-Processing

Windowing
MVNorm
Band-Pass Filter
Pre-Emphasis

mfcc-lfcc
(T,13,2)

cqcc-lpcc
(66,13,2)

mfcc-lfcc
(43,13,2)

mfcc-lfcc
cqcc-lpcc
(66,13,4)

Classifiers

Classifiers

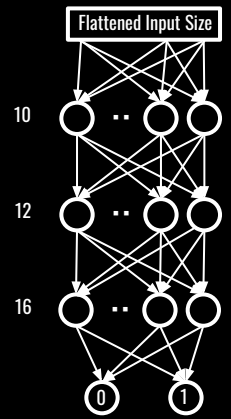
Classifiers

Baselines & Proposed Model

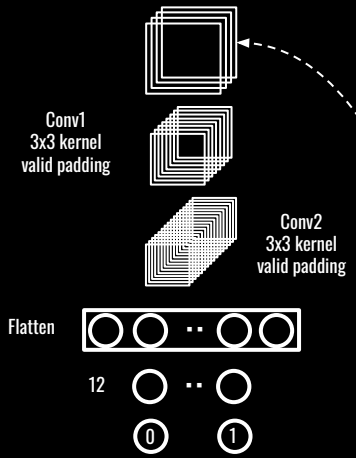
Models

Adam Optimizer
lr = 0.0005, weight decay = 0.001
CE Loss, MSE Loss
batch = 64
50 epochs

Feed-Forward (baseline)

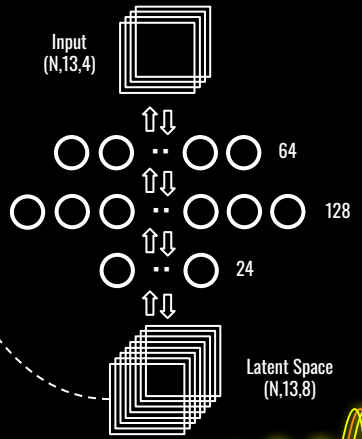


CNN (baseline)



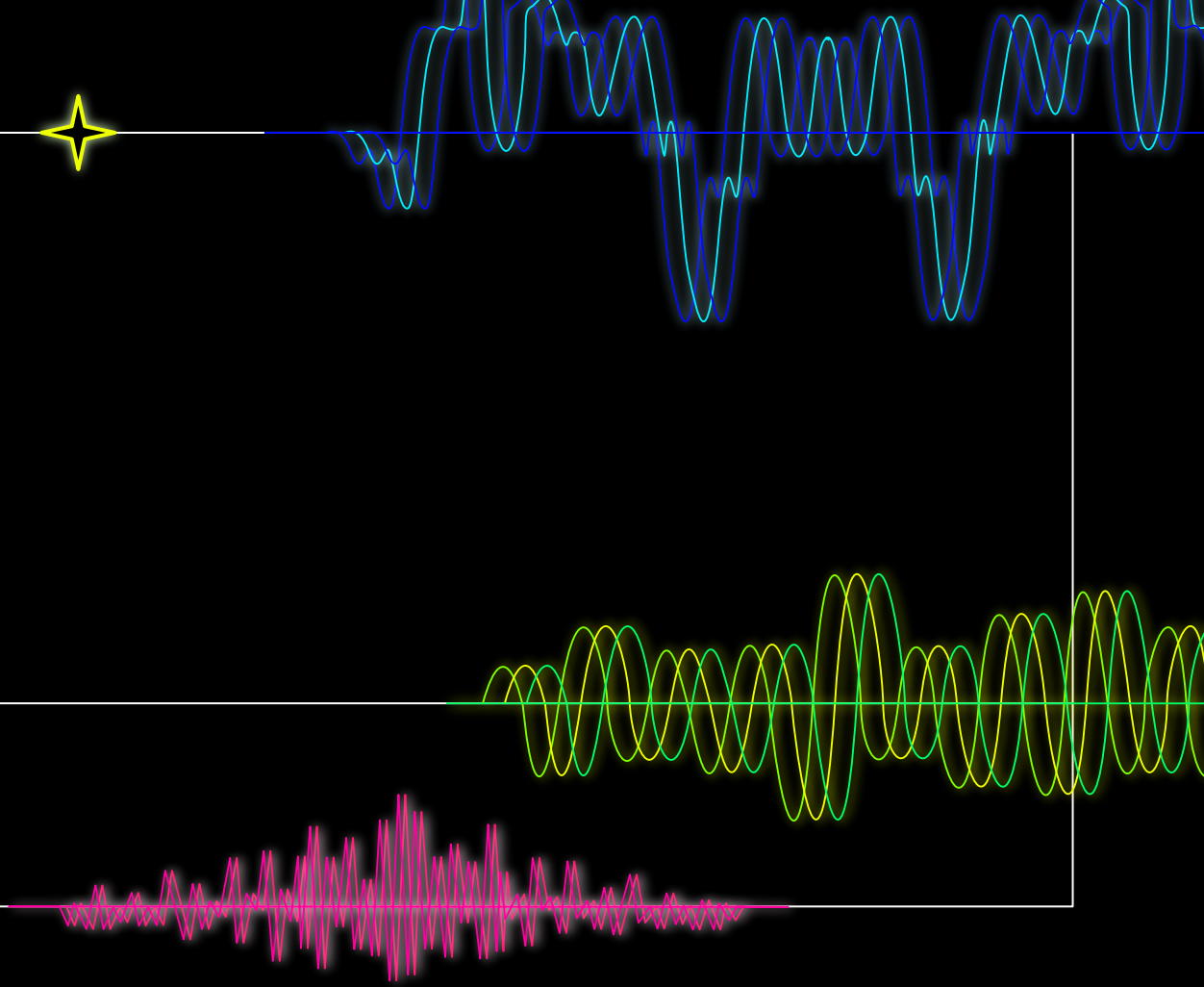
Autoencoder + CNN (proposed model)

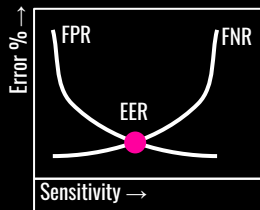
Estimated Size: 3.01MB
+
Estimated Size: 0.25MB
=
Estimated Size: 3.26MB



04

Results





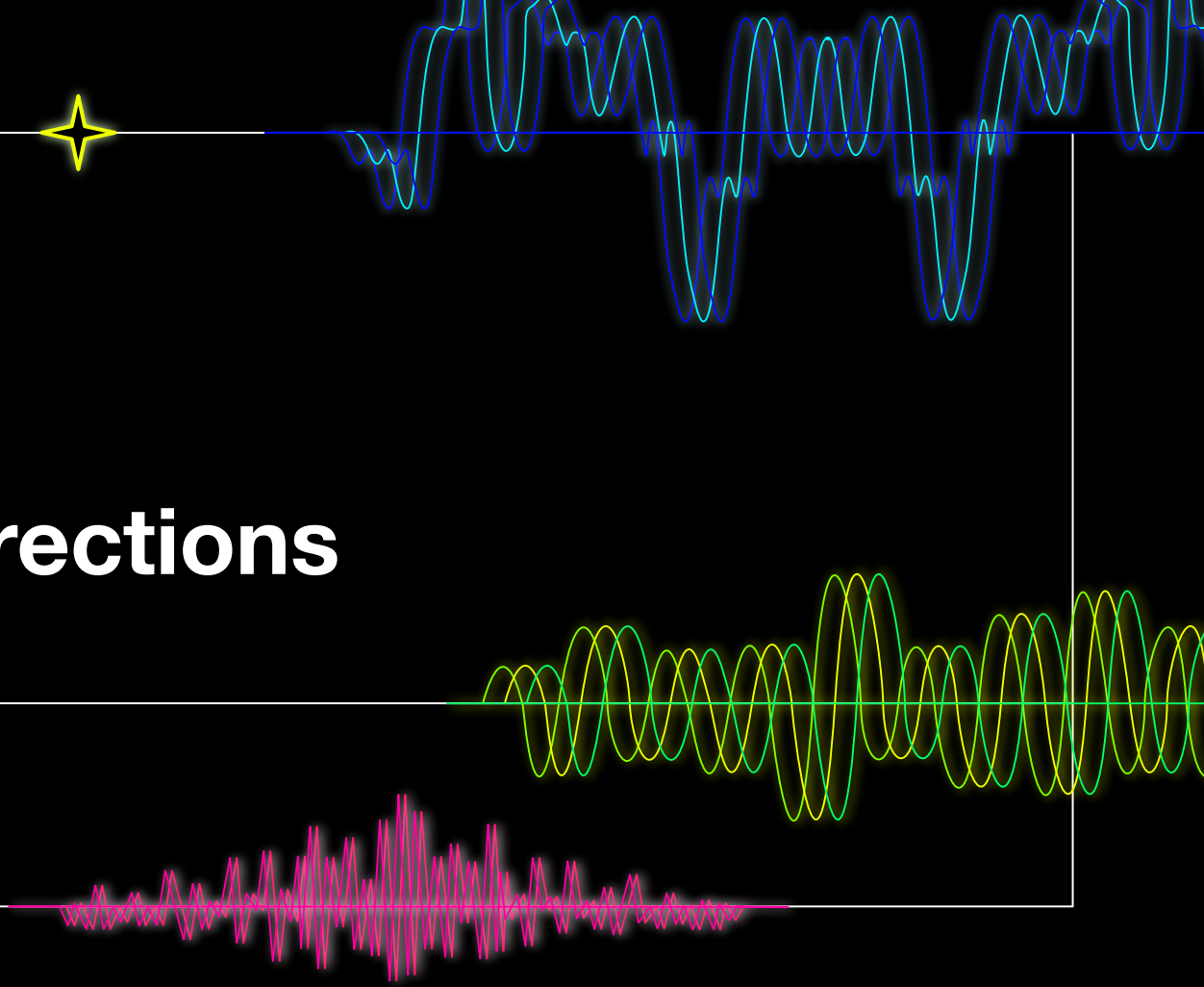
Model Comparisons

		Test Accuracy	Equal Error Rate	F1-Score	Precision
FFN	mfcc-lfcc cqcc-lpcc ALL	0.8308 0.7987 0.8676	18.21% 22.87% 13.61%	0.8308 0.7982 0.8676	0.8314 0.7987 0.8676
CNN	mfcc-lfcc cqcc-lpcc ALL	0.8145 0.8393 0.8691	22.90% 18.36% 14.02%	0.8133 0.8391 0.8691	0.8228 0.8393 0.8694
AE + CNN	mfcc-lfcc cqcc-lpcc ALL	0.8816 0.8682 0.9156	13.19% 15.55% 9.57%	0.8816 0.8680 0.9156	0.8822 0.8682 0.9159
Multi-Feature GMM + Fusion		—	10.8%	—	—

B. T. Balamurali, K. E. Lin, S. Lui, J. -M. Chen and D. Herremans, (2019)

05

Future Directions



Future Directions

- Increasing the data size + increasing the training data
    - When we increased the training data from 50% to 80% we saw that the accuracy went from **88.16% to 92.32%**
- Increasing the epochs + more fine tuning
- Including more features
- Experiment with AE latent vector size
- Manual Audio Feature Extraction wo. Spafe

References

Flitter, E., & Cowley, S. (2023, August 30). *Voice deepfakes are coming for your bank balance*. The New York Times.

Hughes, Alex. (2023). *AI: Why the next call from your family could be a deepfake scammer.* BBC Science Focus.

Hamza, A., Javed, A. R. R., Iqbal, F., Kryvinska, N., Almadhor, A. S., Jalil, Z., & Borghol, R. (2022). Deepfake audio detection via MFCC features using machine learning. *IEEE Access*, *10*, 134018-134028.

Jones, V. A. (2020). *Artificial intelligence enabled deepfake technology: The emergence of a new threat* (Doctoral dissertation, Utica College).

Liu, X., Wang, X., Sahidullah, M., Patino, J., Delgado, H., Kinnunen, T., Todisco, M., Yamagishi, J., Evans, N., Nautsch A., & Lee, K. A. (2023). ASVSPOOF 2021: Towards spoofed and Deepfake Speech Detection in the wild. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 31,* 2507-2522. https://doi.org/10.1109/taslp.2023.3285283

Balamurali, B. T., Lin, K. E., Lui, S., Chen, J. M., & Herremans, D. (2019). Toward robust audio spoofing detection: A detailed comparison of traditional and learned features. *IEEE Access, 7, 84229-84241.* https://doi.org/10.1109/access.2019.2923806

Mcuba, M., Singh, A., Ikuesan, R. A., & Venterm, H. (2023). The effect of deep learning methods on Deepfake audio detection for digital investigation. *Procedia Computer Science, 219*. 211-219, https://doi.org/10.1016/j.procs.2023.01.283

Adiban, M., Shehnepoor, S., Sameti, H. (2020). Replay spoofing countermeasure using autoencoder and siamese networks on ASVSPOOF 2019 Challenge. *Computer Speech & Language, 2019.* https://doi.org/10.1016/j.csl.2020.101105

Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology Innovation Management Review, 9 (11), 39–52.

Yamagishi, Junichi, et al. "ASVspoof 2021: accelerating progress in spoofed and deepfake speech detection." *arXiv preprint arXiv:2109.00537* (2021).