Privacy-preserving collection and sharing of unbiased human voice data for automatic assessment of voice disorders and respiratory diseases: A pilot study

Zhiyu Wan, PhD[1], Yike Li, MD, PhD[2], Bradley A. Malin, PhD[1,3,4]

[1]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [2]Department of Otolaryngology-Head and Neck Surgery, Bill Wilkerson Center, Vanderbilt University Medical Center, Nashville, TN, USA. [3]Department of Computer Science, Vanderbilt University, Nashville, TN, USA. [4]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA.

ABSTRACT

This pilot study aims to explore the potential of utilizing privacy-preserving techniques for safely collecting and sharing human voice data from patients for automatic assessment of voice disorders and respiratory diseases.

MOTIVATION

1. Sharing voice data from patients with voice disorders/diseases is beneficial.
2. Lack of voice data sharing in clinical settings due to privacy concerns.
3. Anonymization techniques for human voice data could be used in this case.

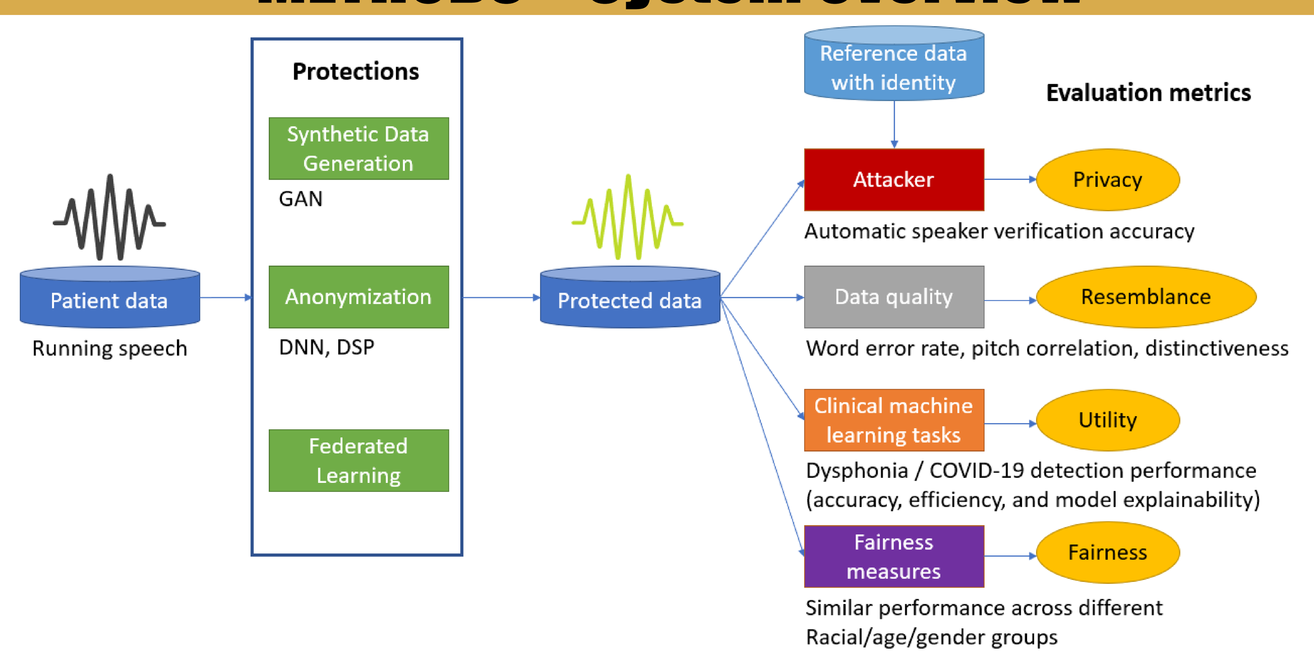
OBJECTIVES

1. Evaluate the privacy risks of sharing voice data from patients
2. Propose and examine privacy-enhancing techniques for voice data sharing
3. Optimize the utility-privacy tradeoff
4. Explore the fairness and the explainability

Datasets

1. A subsect of the LibriSpeech dataset: 363 hours, 921 speakers.
2. Saarbruecken Voice Database: 2000 German-speaking individuals.
3. A dataset from Eye, Ear, Nose and Throat Hospital of Fudan University: 461 people.

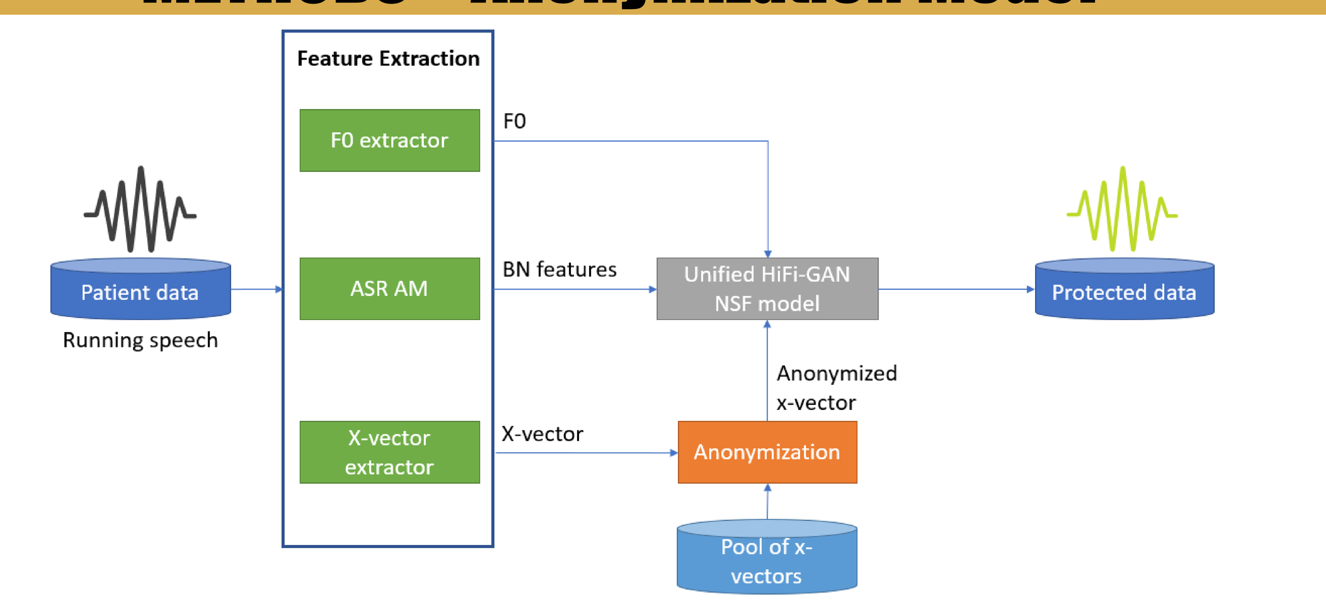
METHODS – System Overview



An overview of the privacy-preserving voice data sharing system

GAN: generative adversarial network. DNN: deep neural network. DSP: digital signal processing.

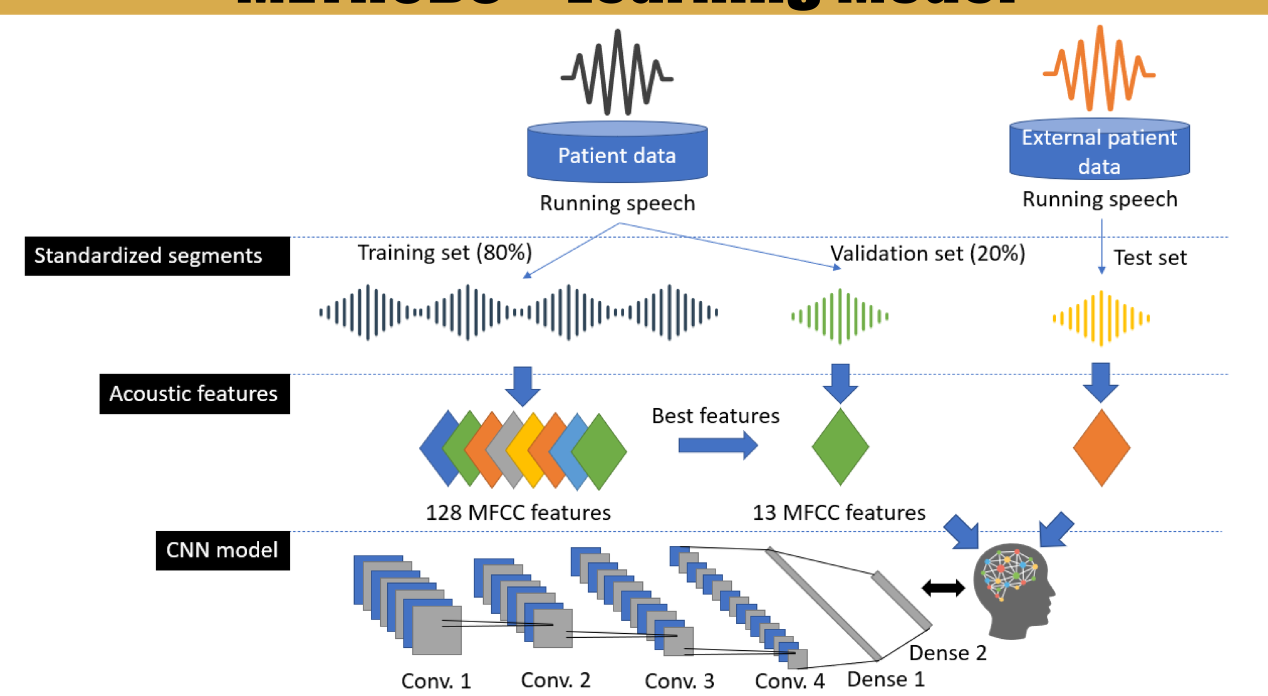
METHODS – Anonymization Model



A DNN-based anonymization model for human voice data

F0: fundamental frequency. ASR: automatic speech recognition. AM: acoustic model. BN: bottleneck. X-vector: DNN embeddings. DNN: deep neural network. GAN: generative adversarial network. NSF: neural source-filter.

METHODS – Learning Model



A learning model for dysphonia detection

MFCC: Mel-frequency cepstral coefficients. CNN: convolutional neural network.

RESULTS – Anonymization Model

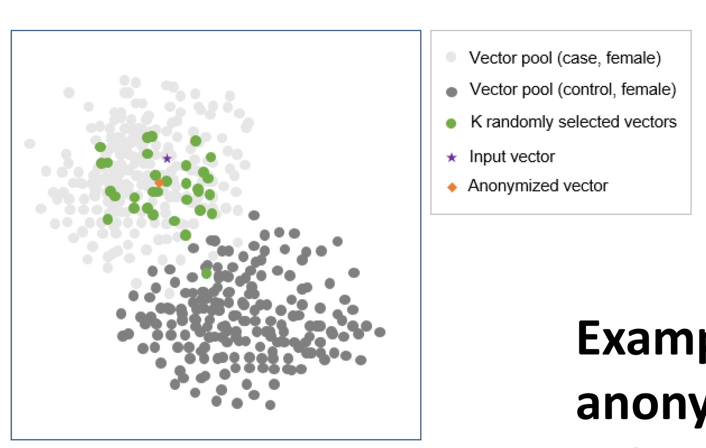
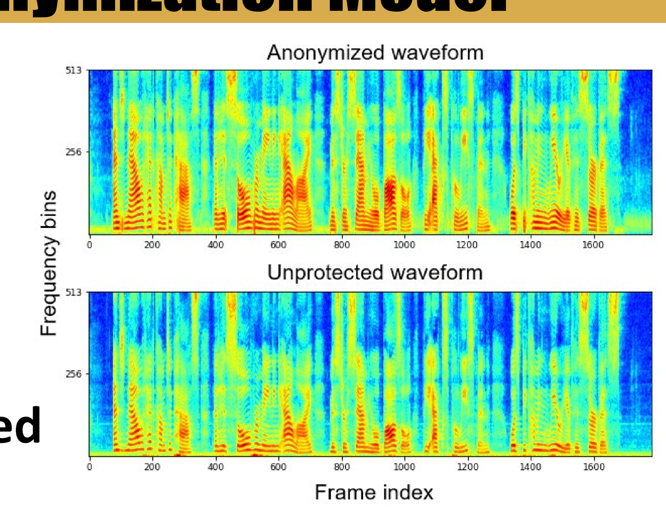
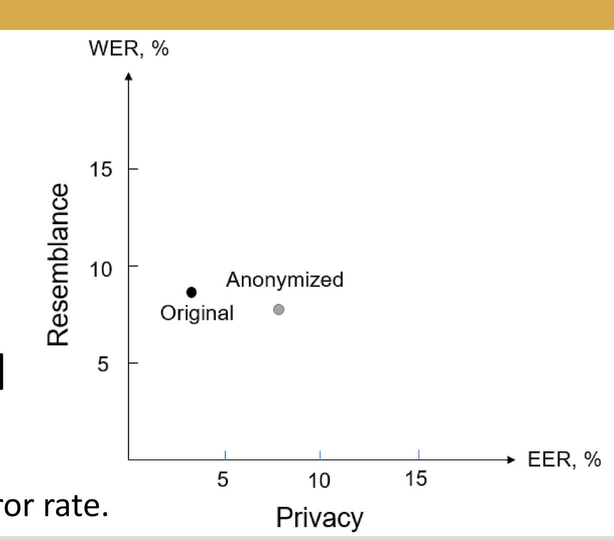


Illustration of the x-vector selection step in the anonymization process



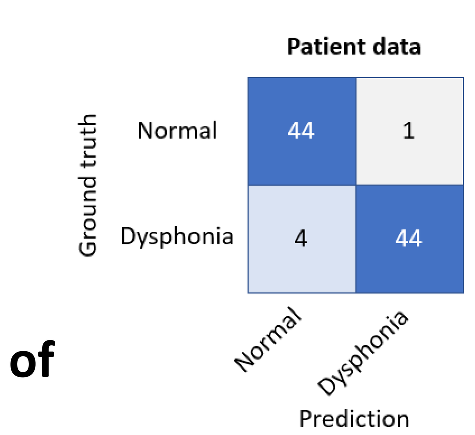
Example comparison for anonymized and unprotected voice data in the waveform



Example results for resemblance (WER) and privacy (EER) metrics

WER, word error rate. EER: equal error rate.

RESULTS – Learning Model



Performance evaluation of dysphonia detection

CONCLUSIONS & DISCUSSION

1. The privacy-preserving human voice data sharing for clinical decision support is an explorable research direction.
2. Anonymization can be a promising approach to sharing more data while protecting the privacy of patients.
3. Sharing voice data from patients with disorders/diseases might be more challenging than sharing voices from other individuals.
4. Finding the perfect balance between utility and privacy is essential while achieving the model's explainability and fairness.

REFERENCES

1. Chen, Z., et al. Deep learning in automatic detection of dysphonia: Comparing acoustic features and developing a generalizable framework. *Int. J. Lang. Commun. Disord.* 1-16 (2022).
2. Tomashenko, N., et al. The VoicePrivacy 2022 Challenge Evaluation Plan. *arXiv* 2203.12468 (2022).

This work was supported by NIH grant RM1HG009034. Please contact the corresponding author at zhiyu.wan.1@vumc.org.

VANDERBILT UNIVERSITY®