Privacy-preserving collection and sharing of unbiased human voice data for automatic assessment of voice disorders

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.

B2AI Voice

INTRO:

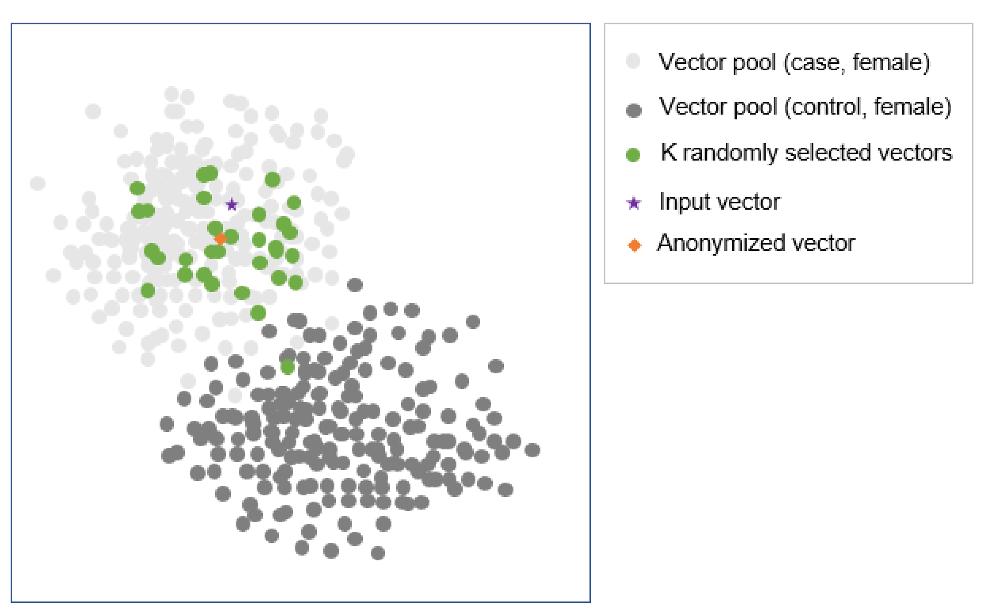
- We aim to explore the potential of utilizing privacy-preserving techniques for safely collecting and sharing human voice data from patients for automatic assessment of voice disorders.

MOTIVATION:

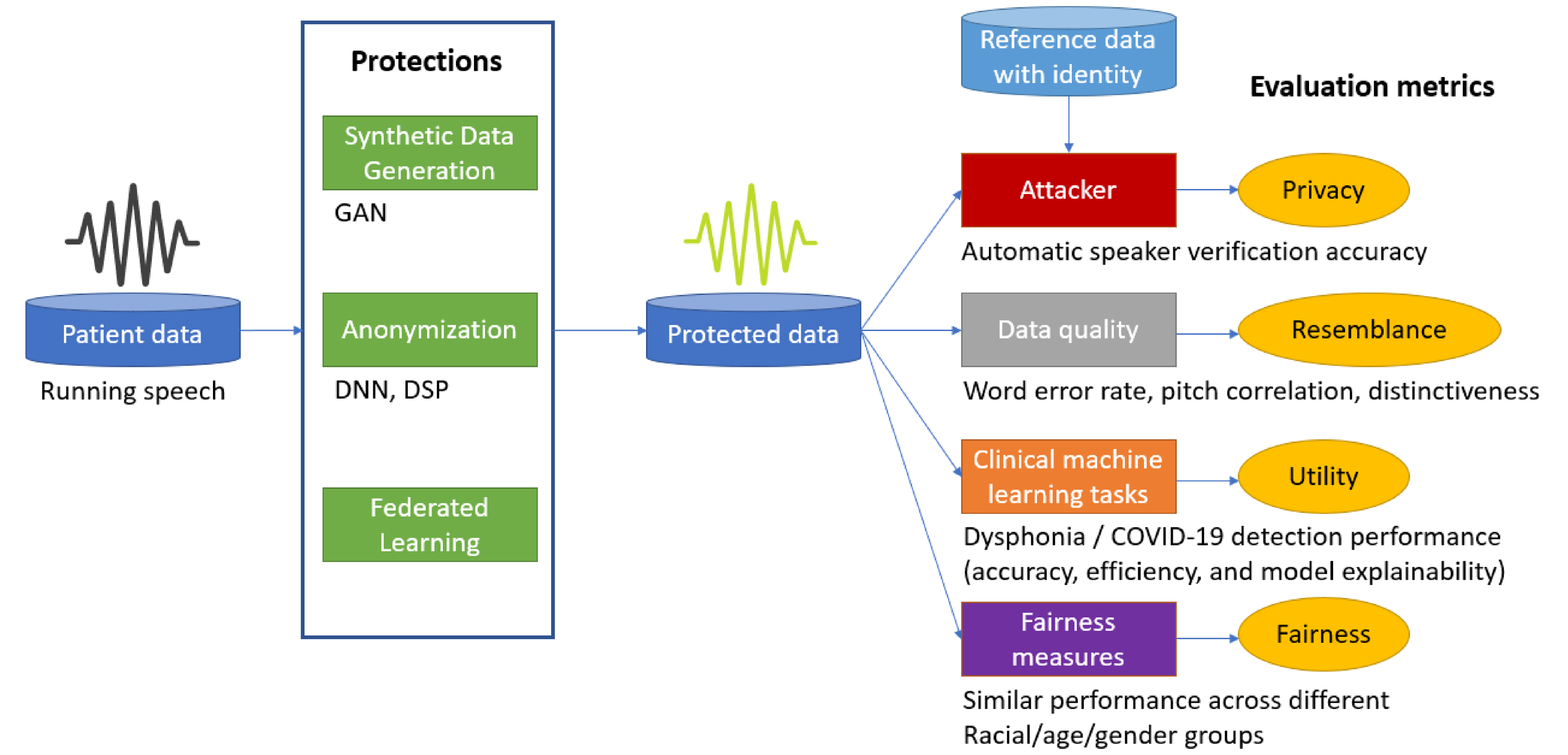
- Sharing voice data from patients with voice disorders/diseases is beneficial.
- Lack of voice data sharing in clinical settings due to privacy concerns.
- Anonymization techniques for human voice data could be used in this case.

METHODS

1. Evaluate the privacy risks of sharing voice data from patients
2. Examine privacy-enhancing techniques for voice data sharing
3. Datasets: (1) 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.
4. **Illustration of the x-vector selection step in the anonymization process**



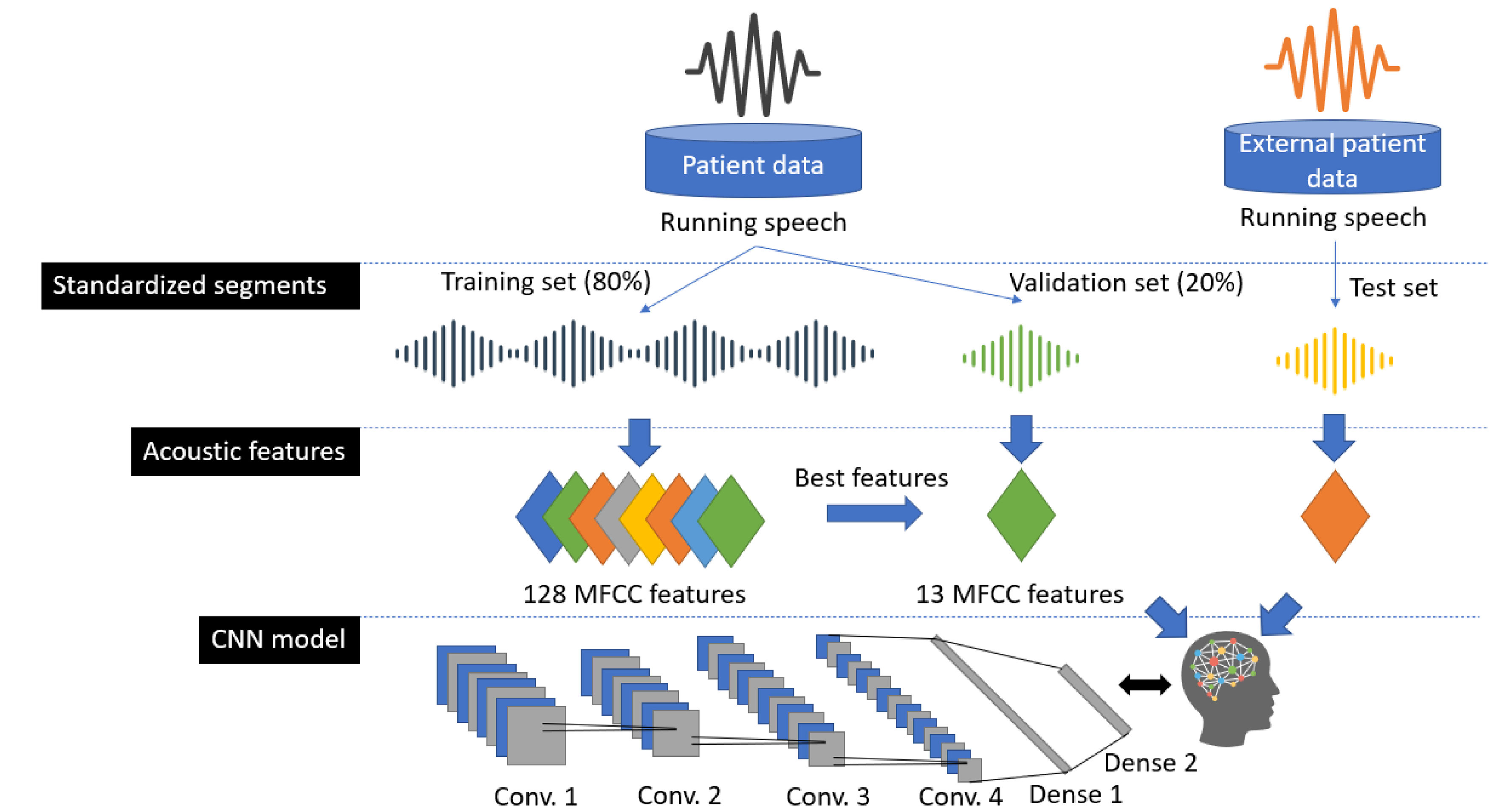
METHODS – System Overview



An overview of the privacy-preserving voice data sharing system

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

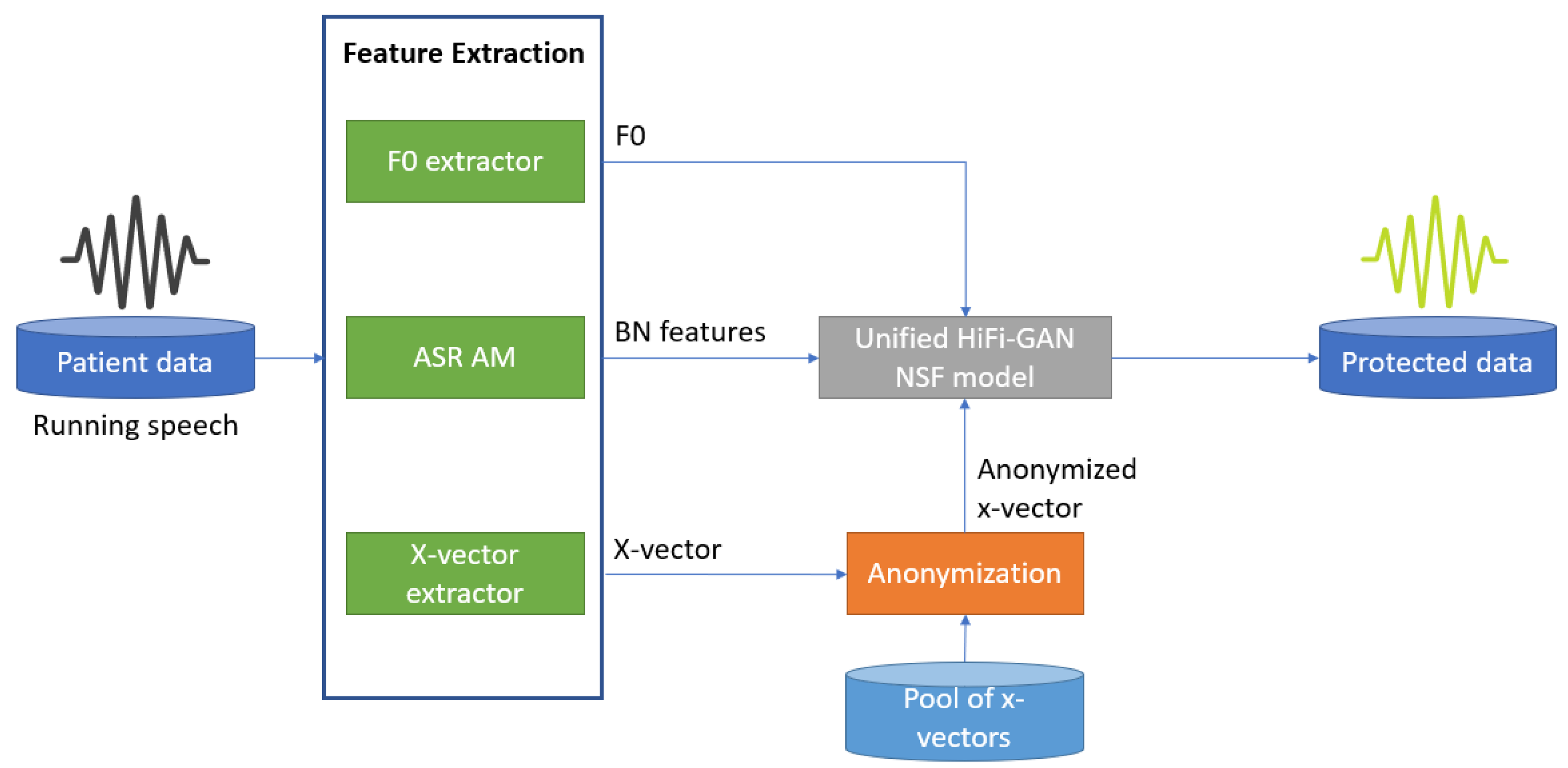
METHODS – Learning Model



A learning model for dysphonia detection

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

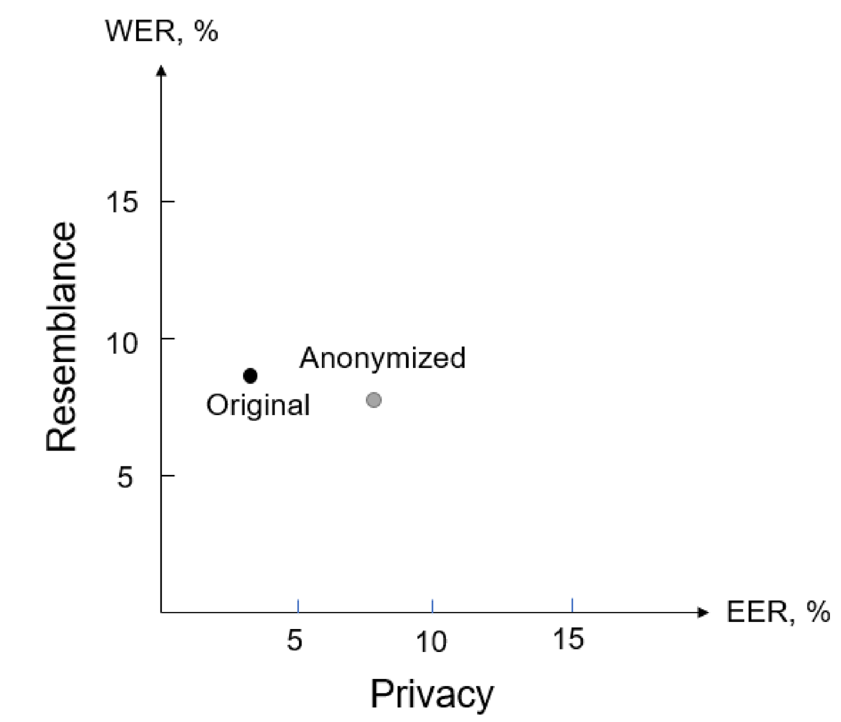
METHODS – Anonymization Model



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.

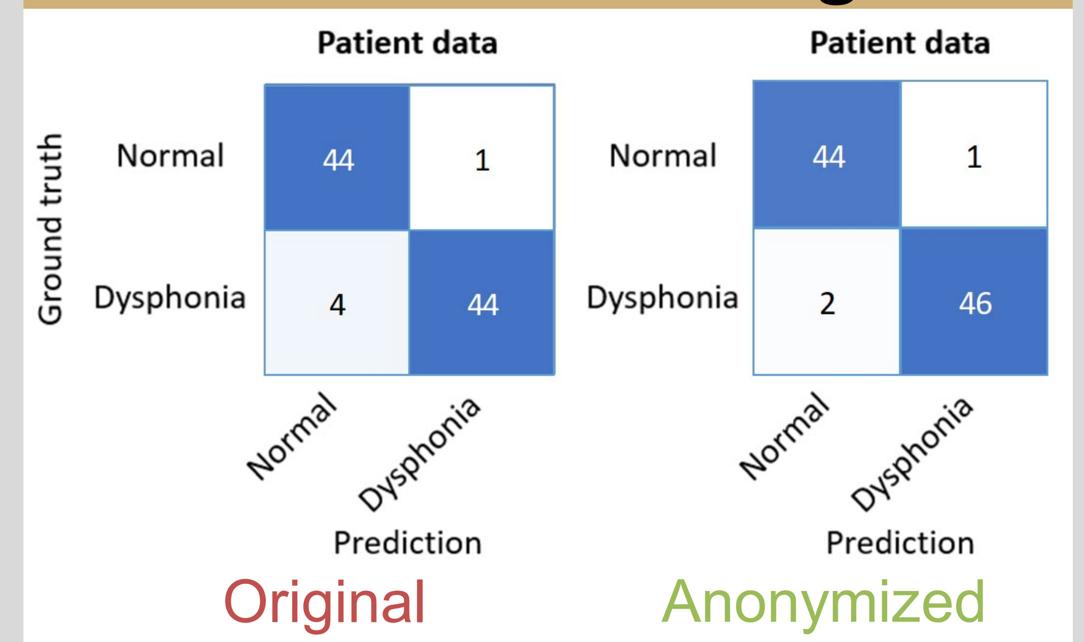
RESULTS – Anonymization



Results for resemblance (WER) and privacy (EER) metrics

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

RESULTS – Learning



Performance evaluation of dysphonia detection

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).

Voice anonymization can be a promising approach to collecting and sharing more voice data while protecting the privacy of patients.

VANDERBILT UNIVERSITY MEDICAL CENTER