上海科技大学
ShanghaiTech University

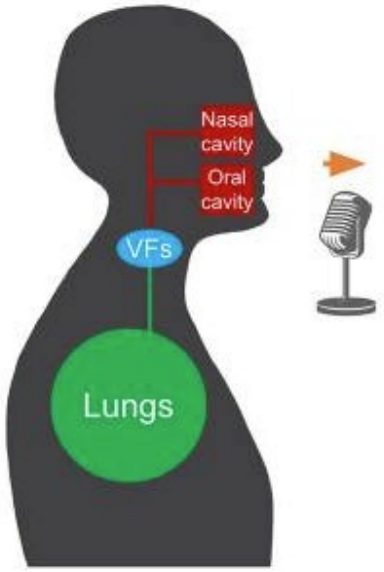
Evaluating Bias and Enhancing Fairness in Acoustic Diagnostic Models

Presenter：Sihan Xie

Date：2025/05/16

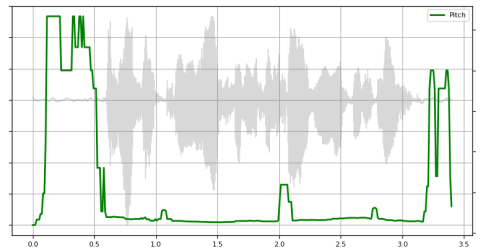
上海科技大学
ShanghaiTech University

Content

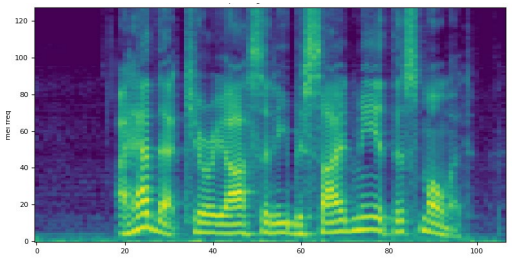


Acoustic features

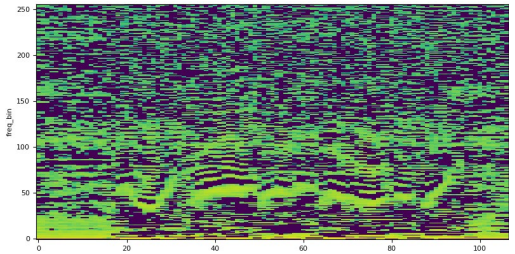
- Vocal folds
- Intensity
- Vocal tract
- Spectral
- Prosodic



pitch

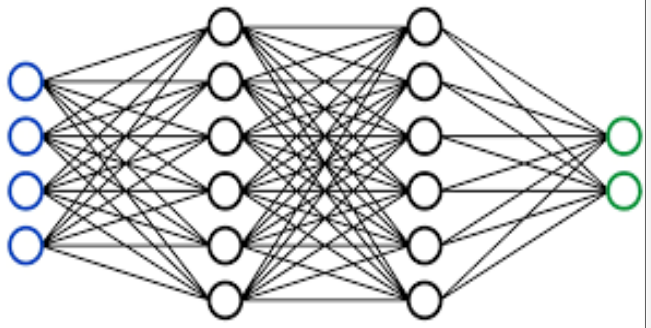


mel



MFCC

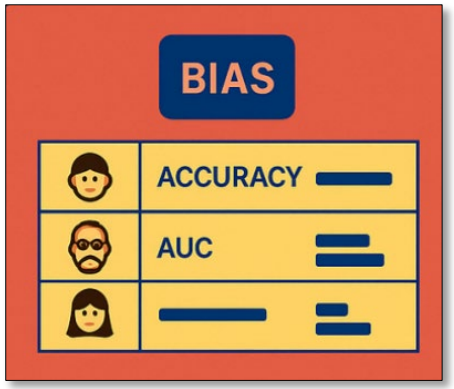
Model



Disease Dignosis

Parkinson, Alzheimer, Dysphonia

.....



Chinese Dataset

Name: EENT

Source: Fudan University

Role: Main Dataset

Disease Type: Dysphonia

Structure

```
├── Normal/
│   ├── AVS51267/                # Participant ID
│   │   ├── chat.wav             # Conversational speech
│   │   ├── vowel-a.wav          # Sustained /a/ phonation
│   │   └── vowel-i.wav          # Sustained /i/ phonation
│   ├── BFG12589/                # Other healthy participants
│   └── ...
├── Functional Dysphonia/
└── Other_Categories/            # Polyps, Sulcus, VFP etc.
```

Statistics

Class	Diagnosis	Number
0	Normal	223
1	UVFP	238
	Nodules	
	.....	

German Dataset

Name: SVD

Source: Saarbruecken Speech Database

Role: External Test Set

Disease Type: Dysphonia

```
├── Healthy/
│   ├── Participant_001/         # Participant ID
│   │   ├── a/                   # Vowel /a/ recordings
│   │   │   ├── 101-a_h.wav      # Normal pitch
│   │   │   └── 101-a_l.wav      # Low pitch
│   │   ├── i/                   # Vowel /i/
│   │   ├── phrase/              # Phrases
│   │   ├── u/                   # Vowel /u/
│   │   └── iau/                 # Diphthong
│   ├── Participant_002/
│   └── ....
└── Other_Categories/            # VFP, Polyps, etc.
```

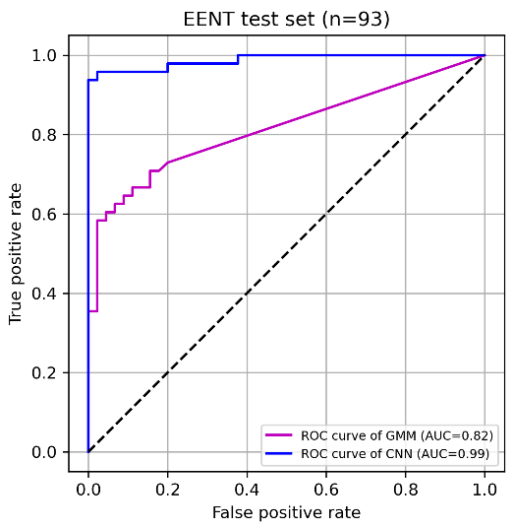
Class	Diagnosis	Number
0	Normal	100
1	Dysphonia	100
	Sulcus	
	.....	

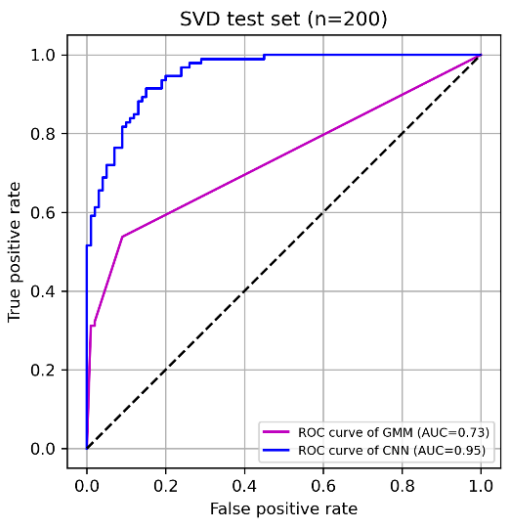
Data Preprocessing

Original data → Segmentation → Standardization → Feature Extraction → pickle files

Model Construction

GMM (pitch) **VS** CNN (mel)





Model Evaluation

	EENT Test Set		SVD Test Set	
Model	accuracy	AUC	accuracy	AUC
GMM	0.72	0.82	0.73	0.73
CNN	0.97	0.98	0.87	0.95

Age

Group		Number	EENT Test	SVD Test
(0, 35)	young	185	0.61	0.56
[35,50)	middle-aged	181	0.60	0.58
[50,+)	old	95	**0.54**	**0.69**

- Chinese Test: The oldest group (50+) performed worst
- German Test: The oldest group (50+) performed best

Gender

Group	Number	EENT Test	SVD Test
female	321	0.57	0.58
male	140	0.63	0.64

Females underperformed males in both datasets

↓

Model shows gender bias

Bias research only on GMM , Not on CNN

Fairness

Not only bias, but also fairness!

- Use statistical metrics
- Comprehensive Evaluation
- Bias Mitigation Techniques

Model

Comprehensive and accurate!

- More Acoustic Features
- Adapt to new dataset.

Potential Risks:
Model Overfitting

Dataset

More Data!

VOICED Database

- 208 samples
- age and gender labels.

- ✓ **Model Architecture:** Add dropout layers
- ✓ **Training Process:** Early stopping mechanism, 5-fold cross-validation
- ✓ **Evaluation:** Independent test set evaluation , multiple prediction strategies

Evaluating Bias and Enhancing Fairness in Acoustic Diagnostic Models

5.19-5.26	✓ Successfully run fairness evaluation code for the CNN. ✓ Use statistical methods to assess the fairness of both GMM and CNN.
5.27-5.31	✓ Investigate sources of bias and explore mitigation strategies. ✓ Expand the dataset for disease diagnosis and evaluate bias.
6.1-6.6	✓ Conduct fairness evaluations through multiple runs. ✓ Organize the results, and prepare for the defense.
6.7-6.12	✓ Organize data and code, and complete the final report.

上海科技大学
ShanghaiTech University

Thanking You