



# SPEECH BASED MEDICAL DIAGNOSIS

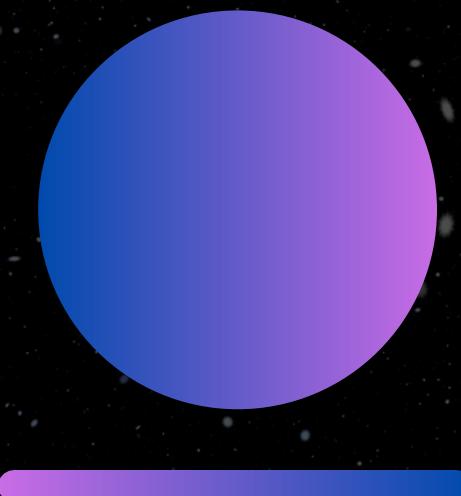
Search 

Soham Niraj Deshmukh (B21EE067)  
Shubh Goyal (B21CS073)



# INTRODUCTION

- Speech understanding used in various fields – speech assistants, automotive industry, education, smart customer service, etc.
- Can also be used as an important tool in medical diagnosis – easier, efficient, reliable and less invasive.
- Identify early symptoms of various medical conditions – Parkinson's and Alzheimer's disease, respiratory, thoracic, and cardio-vascular conditions.
- Our aim – exploring the application and current developments of speech understanding techniques in medical diagnosis.



# Current Works





# Datasets

## COUGHVID

- crowdsourcing dataset
- 25000 cough recordings
- hand labelled

## ADReSS

- spontaneous speech (Alzheimer's Dementia)
- 1955 audios from 78 non-AD subjects
- 2122 audios from 78 AD subjects

## PASCAL HEART SOUND

- 176+656 audio files
- Heart Sound Segmentation
- Normal vs Cardio-vascular diseases

## SAARBRUCKEN VOICE DATABASE

- Two vocal diseases
- Laryngozole
- Vox Senilis

## Parkinson Vocal Database

- 197 datapoints
- 22 features



# **Current Research Works**



# Cardiovascular Diseases



Method -

- feature extraction using MFCC and QCP methods
- 4 Machine Learning models – SVM, Extra Trees, AdaBoost and Feedforward Neural Net.

Metric – Cross Validation accuracy

Strengths –

- Combining MFCC features with glottal features, helps enhance the classification performance
- Feature selection using Gini impurity
- resilient to noise

Limitations –

- Small dataset
- Overfitting
- Deteriorates under low SNR conditions



# Alzheimer Disease

Method -

- GP-Net - feature purification
- Two loss functions - removal of irrelevant features

Metric - Classification accuracy

Strengths -

- Purified features - more discriminative
- Computationally efficient

Limitations -

- Effectiveness depends on purification
- Risk of losing subtle linguistic info
- Not achieved SOTA on some datasets.



# Parkinson Disease

Method -

- ISNDAM – Integrated Smallest Normalized Difference Associative Memory
- 3 phases – Training, Relevance Identification and Testing
- Managing input-output relations using memory matrices

Metric – Classification accuracy (98%)

Strengths –

- Elimination of irrelevant data
- Computationally efficient and noise robust

Limitations –

- Limited generalization
- Black box



# Respiratory Disease

Method -

- RBFNet – 3 components
- CNN-LSTM encoder, a classifier and a bias predictor and remover conditional GAN

Metric – Classification accuracy (5.5 – 8.2% increase)

Strengths –

- Actively removes feature-related bias
- Generalizes well, outperforms traditional CNN-LSTM

Limitations –

- Limited to certain use-cases – no validation over other cases
- High compute requirement
- Careful hyperparameter tuning required

# Open Problems





## AVAILABILITY OF CLINICALLY VERIFIED HIGH QUALITY DIVERSE DATA

- Training models on small, imbalanced datasets often lead to biased predictions or model availability in only one demography.

## GENERALIZATION ACROSS DIVERSE POPULATIONS

- Models often overfit to specific speaker characteristics, and loose effectiveness in real-world applications.

## ROBUSTNESS TO NOISE AND ENVIRONMENTAL VARIATIONS

- Speech data collected in controlled settings v/s noisy, real-world environments
- Distribution shift between training and testing data => wrong inferences.



## COMPUTATIONAL COMPLEXITY OF MODELS

- Transformer based models like GP-Net – difficult to deploy them in resource-constrained environment

## BLACK BOX OF DEEP LEARNING MODELS

- The non-interpretability of the decision made and the lack of reasoning

## STANDARDIZATION OF EVALUATION METRICS

- different studies use varying datasets and methodologies, making it difficult to compare models and assess true performance improvements.

THANK YOU