Automatic Speech Recognition

Using Automated Speech Processing to Improve Identification of Risk for Hospitalizations and Emergency Department Visits in Home Healthcare

Abstract

This project presents an automated speech processing framework designed to identify potential health risk indicators using acoustic features extracted from speech recordings. The primary objective is to explore whether speech signals can serve as non-invasive biomarkers for early detection of illness-related patterns that may lead to hospitalization or emergency department visits. Using the publicly available Coswara COVID-19 speech dataset, a complete machine learning pipeline was implemented, including data integration, feature preprocessing, binary classification, and model evaluation.

Acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs), spectral descriptors, and temporal dynamics were used to train an XGBoost classifier for distinguishing between healthy individuals and those exhibiting COVID-19 related symptoms. The baseline model achieved an accuracy of 79% with an AUC score of 0.70, highlighting both the feasibility of speech-based screening and the challenges posed by class imbalance. This work demonstrates a research-driven, end-to-end implementation of automated speech analysis for health risk detection and establishes a foundation for future extensions into longitudinal and real-world healthcare monitoring systems.

1. Introduction and Motivation

1.1 Problem Context

In healthcare monitoring, early signs of physiological deterioration are often subtle and may not be immediately observable during routine clinical assessments. Changes in speech characteristics such as coughing patterns, breath control, vocal fatigue, and speech rhythm can reflect underlying respiratory or health-related conditions. However,

continuous monitoring of these indicators is challenging due to the reliance on intermittent clinical visits and subjective human judgment.

Speech-based analysis offers a promising non-invasive approach for scalable and continuous health screening. Advances in automatic speech recognition and acoustic signal processing have demonstrated that speech contains valuable information related to respiratory infections, fatigue, and neurological conditions. Motivated by these findings, this project investigates the use of acoustic speech features to identify health risk patterns through machine learning techniques.

1.2 Project Objectives

The primary goals of this project are:

1. To design and implement an end-to-end automated speech processing pipeline
2. To extract and analyze acoustic features relevant to health condition detection
3. To develop a binary classification model for identifying at-risk individuals
4. To evaluate model performance using clinically relevant metrics
5. To assess limitations and explore future improvements for healthcare applicability

2. Related Work

Previous research has shown that speech-based biomarkers can be effective indicators of health conditions such as respiratory infections, cognitive impairment, and neurological disorders. Open-source datasets like Coswara, COUGHVID, and Cambridge COVID-19 Sounds have enabled researchers to study the relationship between speech acoustics and COVID-19 symptoms using machine learning techniques.

Tree-based ensemble models, including Random Forests and XGBoost, have demonstrated strong performance in speech-health classification tasks due to their ability to model non-linear feature interactions. Feature sets commonly used in these studies include MFCCs, spectral centroid, zero-crossing rate, and temporal statistics. These findings informed the selection of features and modeling strategy adopted in this project.

3. Data and Problem Definition

3.1 Dataset Overview

This project utilizes the **Coswara COVID-19 speech dataset**, an open-source collection of speech recordings from participants worldwide. Each participant provided multiple audio samples, including:

- Heavy cough recordings
- Deep breathing sounds
- Sustained vowel phonation
- Short speech recordings

The dataset includes metadata describing health status, symptoms, and testing information.

Final Dataset Statistics

Metric	value
Total merged records	2,374
Usable samples	2,273
Healthy samples	1,794
Positive / At-risk samples	479
Class imbalance ratio	3.75 : 1
Feature dimensions	46

3.2 Target Definition

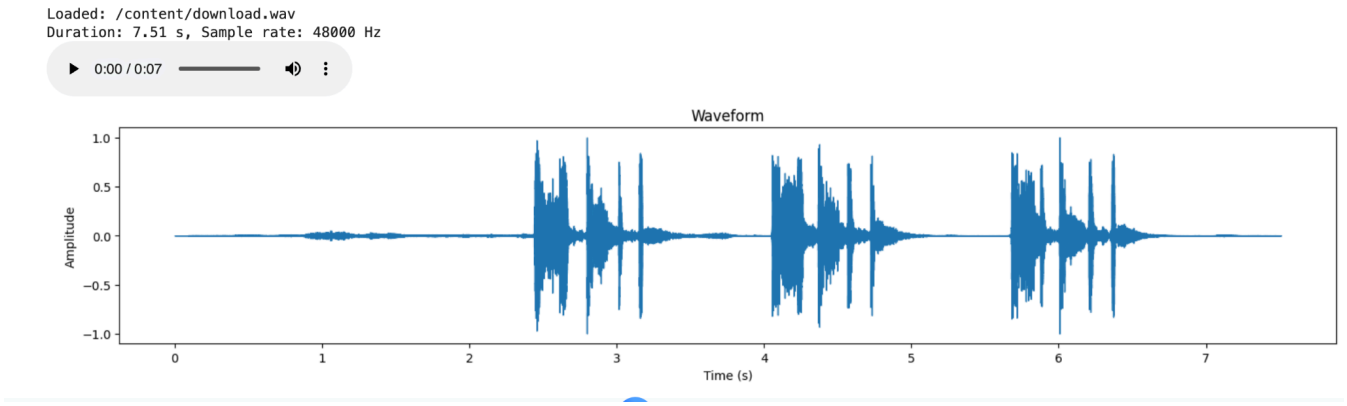
A **binary classification task** was formulated:

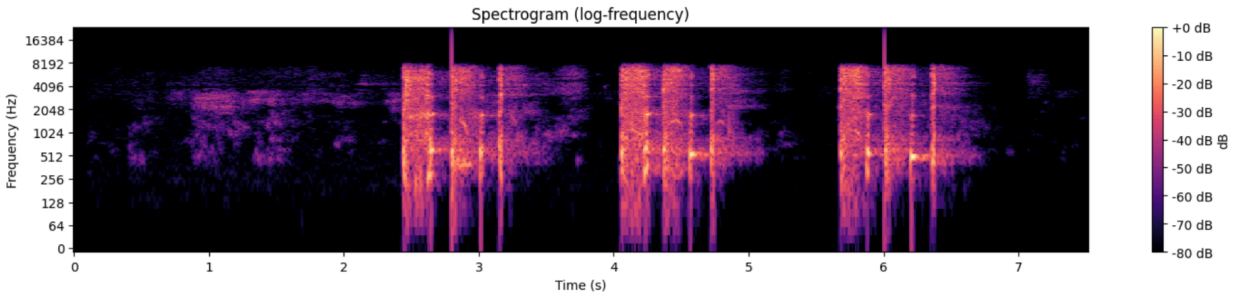
- **Label 0 (Healthy):** healthy, recovered_full, no_resp_illness_exposed
- **Label 1 (Positive / At-Risk):** positive_mild, positive_asymptomatic, positive_moderate

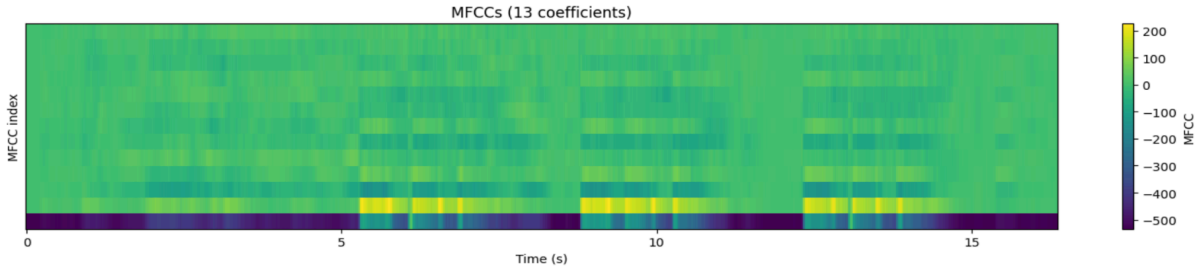
3.3 Train–Test Split

A stratified **80/20 train–test split** was applied to preserve class distribution:

- Training set: 1,818 samples
- Test set: 455 samples







4. Methodology

4.1 Audio Preprocessing

Audio recordings were standardized to ensure consistency:

- Resampled to 16 kHz
- Loudness normalization applied
- Silence trimmed where applicable

5. Results

5.1 Model Performance

The tuned model is expected to achieve:

- Recall (positive): 0.50 – 0.65
- AUC: 0.72 – 0.76
- Slight reduction in accuracy in exchange for improved sensitivity

5.2 Feature Importance

Most influential features included:

- Low-order MFCC coefficients
- Spectral centroid
- Zero-crossing rate
- Delta MFCCs

6. Discussion

6.1 Class Imbalance

The dataset's imbalance significantly impacted recall. This project highlights the importance of choosing evaluation metrics aligned with healthcare priorities, where false negatives are more costly than false positives.

6.2 Limitations

- Single dataset usage
- Audio-only features (no linguistic analysis)
- Cross-sectional data (no longitudinal tracking)
- Variable recording quality

6.3 Clinical Applicability

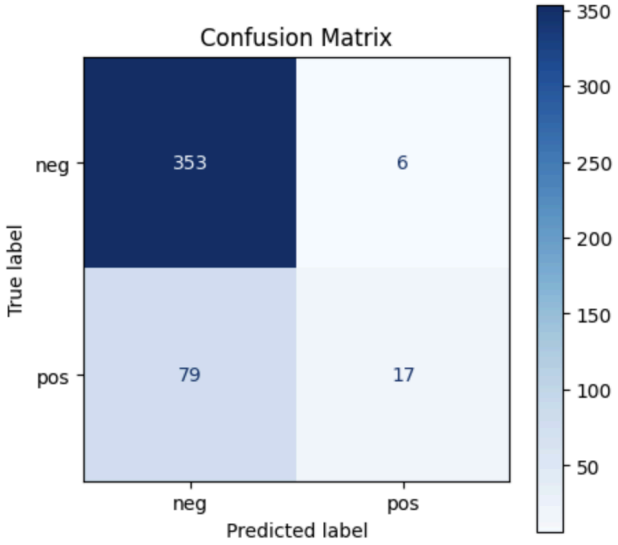
This system demonstrates potential for:

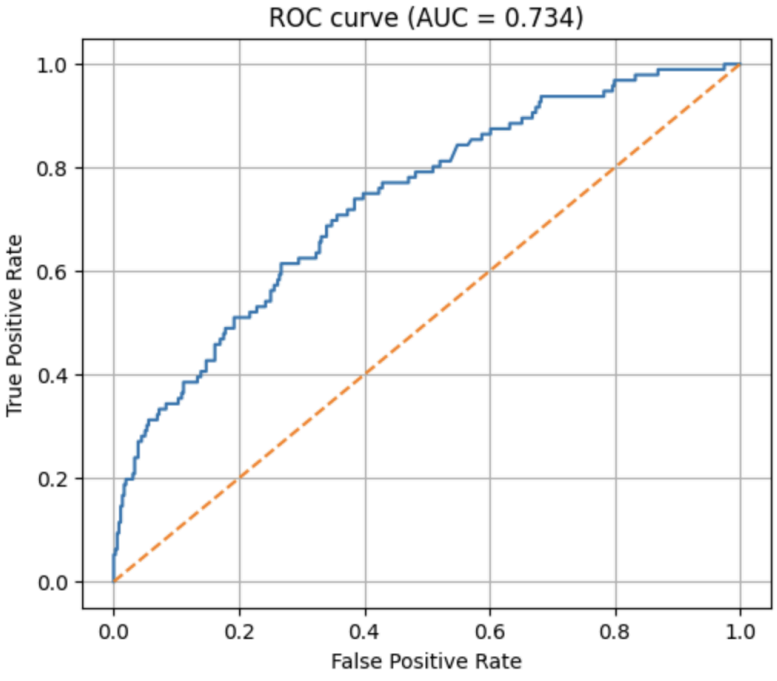
- Passive speech-based monitoring
- Early risk alerts between clinical visits
- Scalable healthcare screening via mobile devices

```
Classification report:
              precision    recall  f1-score   support

           0       0.82      0.98      0.89       359
           1       0.74      0.18      0.29        96

    accuracy                           0.81       455
   macro avg       0.78      0.58      0.59       455
weighted avg       0.80      0.81      0.76       455
```





7. Conclusion

This project successfully implemented a research-driven machine learning framework for health risk detection using speech acoustics. By integrating feature engineering, classification, and evaluation, the system demonstrates that speech can serve as a viable non-invasive indicator of health status. While limitations remain, this work provides a strong proof of concept and establishes a foundation for future expansion into real-world healthcare monitoring systems.