Project Overview

With the rapid development of medical AI, disease screening technologies based on voice biomarkers have become important auxiliary diagnostic tools. These technologies analyze acoustic features in voice signals to identify diseases, supporting early screening for conditions like Parkinson's disease, depression, and respiratory disorders. However, there are two key challenges in practical applications: data privacy and diagnostic performance. On one hand, voice data contains sensitive information that can infer age, gender, emotional state, and health characteristics, posing privacy risks. On the other hand, privacy-preserving techniques may impact diagnostic accuracy and cause fairness disparities among different demographic groups.

Research Objectives

This project focuses on the triadic balance of privacy, performance, and fairness in medical AI. It aims to develop a general evaluation framework to systematically assess the impact of privacy-preserving techniques on the performance of voice-based diagnostic models, and reveal their influence on the fairness across demographic groups.

Regarding the impact of privacy protection on fairness, consider the case of the elderly. Their voices typically exhibit features like slower speech rates and diminished high-frequency energy. However, common frequency-domain filtering techniques may unintentionally remove key acoustic markers related to diseases while eliminating identity information, thus increasing the misdiagnosis risk for this group. To access this, the project will first train a diagnostic model using raw voice data to investigate its performance and fairness. Subsequently, comparative experiments will examine group disparities after introducing different privacy-preserving methods, establishing a model that correlates privacy intensity with fairness metrics.

Research Phases

1. Diagnostic model Construction and Validation: Select a pathological voice dataset with age and gender annotations, extract common acoustic features (such as MFCC), and train a disease diagnosis model. Quantify the model's initial fairness across age and gender dimensions.

2. Privacy Intervention Experiments: Two types of privacy-preserving strategies will be implemented—data-level protection and algorithm-level interventions. Systematic evaluations will be conducted to examine the impact of each strategy on overall model performance .

3. Fairness Analysis: Analyze the performance variations of the diagnostic model across different demographic groups under various privacy-preserving methods. Compare the fairness changes before and after applying these methods to determine their specific impact on model fairness.

In summary, this project aims to uncover fairness issues inherent in medical data protection. The insights gained will contribute to enhancing data security, reducing health disparities, and addressing ethical challenges in AI applications.