Term Project Proposal: Respiratory Sound Classification

111061702 常安彥

I. Problem Description

1. Dataset: Respiratory sound with three classes

   Coarse (871), Wheezing (250) and Normal (4615)

2. Motivation

   Respiratory sounds with different features may indicate various diseases. By the model's diagnosis, it is hoped that doctors can effectively screen the potential causes of a patient's condition, or identify patterns that may have been overlooked by human observation.

3. Background

   The dataset is collected in collaboration with NTUH, and includes respiratory sounds from patients. The sounds range from coarse to wheezing and normal healthy breaths, with a highly unbalanced distribution where normal breaths constitute the majority.

   As the collection is still ongoing, the dataset will grow continually, and there will be some difference in number of samples by the end of the semester. Furthermore, this dataset is currently not publicly available, and will be released to the world later this year, focusing on respiratory sounds of Asian individuals.

II. Methods

1. Embedding Features
   a. Construct the spectrogram of each audio wave file from a big public dataset, such as AudioSet[1].
   b. Train an audio spectrogram transformer[2] with these spectrograms.
   c. Use this transformer to get the embedding features of our task.
2. Feature Extraction

   Reduce the dimension of the embedding features with PCA, LDA or Decision tree.

3. Classification

   Classify with SVM, KNN, perceptron or MLP, using 10-holdout folds.

III. Possible Results

1. Evaluation: ICBHI score (mean of specificity and sensitivity)
2. Score may land on 0.6, and specificity will be higher due to imbalance.

IV. References

[1] https://research.google.com/audioset/download.html

[2] (arXiv:2104.01778) AST: Audio Spectrogram Transformer, Yuan Gong, Yu-An Chung, James Glass