A Large-Scale Respiratory Audio Pre-trained Representation Model for Mobile Respiratory Health

Yuwei Zhang
University of Cambridge
yz798@cam.ac.uk

Tong Xia
University of Cambridge
tx229@cam.ac.uk

Cecilia Mascolo
University of Cambridge
cm542@cam.ac.uk

1. INTRODUCTION

Respiratory diseases are a leading cause of morbidity and mortality worldwide. Early diagnosis and screening are crucial for effective treatment and improved patient outcomes. Recently, there has been growing interest in leveraging respiratory sounds, such as breathing and coughing, as non-invasive sources of diagnostic information [1]. They can be easily collected from mobile devices and provide valuable insights into the physiological state of the respiratory system.

Deep learning models have emerged as powerful tools for modelling respiratory sounds. However, their effectiveness is often hindered by the limited availability of labeled data for respiratory condition classification. Recent advances in generative AI demonstrate the potential for large-scale pre-training to enhance performance across various tasks [2], but there is currently a lack of open pre-trained models specifically tailored to respiratory audio data. Existing models, typically trained on large image datasets or general-purpose audio data, may lack the necessary domain-specific knowledge to capture the subtle nuances of respiratory sounds.

By providing a pre-trained model tailored specifically to respiratory sounds, we aim to address the challenge of data scarcity in downstream tasks and offer a valuable resource for the field of respiratory disease diagnosis and screening. Such a model has the potential to empower researchers and developers to build more robust and accurate models for early disease detection.

2. METHOD

The development of a generalizable pre-trained model for respiratory audio-based health monitoring involves several key aspects: leveraging heterogeneous open datasets for pre-training, exploring various pretraining methods, and formulating downstream tasks for evaluation.

Open datasets. A number of open datasets with respiratory audio are currently available, many of which are collected from mobile devices. We believe training with open datasets improves transparency and reproducibility. Moreover, training on multiple datasets makes the model more robust to data heterogeneity and more generalizable for diverse downstream tasks.

Pretraining methods. Several design choices must be considered in modeling audio data for pre-training. These choices include selecting the datasets, defining the training goal (e.g., contrastive and generative objectives), determining the model architecture and size, among others. We aim to explore and benchmark the effects of these design, providing insights for selecting suitable pre-trained models for downstream tasks.

Downstream tasks. We formulate a range of downstream tasks to demonstrate the generalizability of our pre-trained model and provide a benchmark for future research in this area. These tasks include symptom prediction, disease prediction, disease severity classification, and smoker detection.

3. PRELIMINARY WORK & DISCUSSION

In this work, we plan on release an open-source pretrained deep learning model specifically designed for respiratory sound analysis. Firstly, we curate and preprocess several unlabeled open datasets containing respiratory sounds sourced from the literature. These datasets encompass a diverse range of respiratory conditions and capture the variability inherent in real-world mobile health settings. Through self-supervised learning, our model autonomously learns meaningful representations from this rich data corpus, equipping it with a comprehensive understanding of respiratory sound patterns. We then evaluate the performance on various downstream respiratory condition classification tasks. Preliminary results demonstrate our model's potential in outperforming existing off-the-shelf pre-trained models and feature extraction techniques. We believe our work will become a valuable resource to alleviate data scarcity constraints for audio-based mobile respiratory health monitoring.

4. REFERENCES

[1] T. Dang, D. Spathis, A. Ghosh, and C. Mascolo. Human-centred artificial intelligence for mobile health sensing: challenges and opportunities. *Royal Society Open Science*, 10(11):230806, 2023.

[2] R. Krishnan, P. Rajpurkar, and E. J. Topol. Self-supervised learning in medicine and healthcare. *Nature Biomedical Engineering*, 6(12):1346–1352, 2022.