# RespiraCheck: Using Audio Analysis as a COVID-19 Testing Tool

**Jennifer Chiou**, **Gabriel McFadyen**, Joseph Yu, Houman Ebrahimi, Krish Chhajer, Armagan Gul, Tasfia Ara, Akshata Kulkarni, Leilia Ho

*University of Toronto*

***Abstract*—To address barriers preventing timely COVID-19 diagnosis, we propose RespiraCheck, a convolutional neural network (CNN) designed to classify COVID-19 based on cough audio. Our approach utilized Mel spectrogram representations of labeled cough recordings to fine-tune the last convolutional and fully connected layers of a pretrained ResNet-18 model, leveraging transfer learning for efficient and accurate classification. Using the crowdsourced Coswara and COUGHVID datasets, we trained on a balanced set of COVID-19 positive and negative samples. To ensure real-world applicability, we also developed a web interface that allows individuals to record or upload cough samples and receive an instant diagnostic assessment. By bridging the gap between clinical research and practical deployment, RespiraCheck aims to provide an accessible, non-invasive, and scalable tool for COVID-19 screening.**

## I. Introduction

### A. Motivation

The recent COVID-19 pandemic has demonstrated the impact of respiratory illnesses on a global scale. As of October 2020, more than 1 million COVID-19 related deaths have been documented worldwide [1], with 8,749 deaths out of 106,804 COVID-19 reported cases in Canada [2]. While the COVID-19 related mortality rates have decreased, the disease remains prevalent. Many individuals with mild symptoms go undiagnosed, either to avoid long hospital wait times [3] or to sidestep the financial burdens incurred through higher insurance premiums resulting from ordering at-home PCR testing kits [4]. The neglect of symptoms due to external factors can have major effects on the health and lifestyles of individuals, and may even lead to severe consequences in the future.

Despite numerous studies on COVID-19 classification using cough data—many of which report high accuracy—few user-facing applications have been developed to provide widespread public access to AI-driven COVID-19 screening. Research published by the National Institute of Health emphasizes the potential impact of such tools, stating that,

> Mobile app technology, biosensors (for rapid diagnosis), and AI methods (for diagnosis in the early and acute stages of the disease) can reduce high mortality rates and minimize the consumption of hospital resources [5].

While accurate models exist and experts advocate for AI-powered COVID-19 screening, real-world deployment remains limited, with only a handful of publicly available applications. With RespiraCheck, we aim to bridge the gap between research and real-world application by developing an accurate, compact model capable of detecting COVID-19, while also ensuring accessibility to the general public through an intuitive public-facing website.

### B. Related Works

In a paper published by Loey and Mirjalili in 2021 [6], the authors tackle the problem of COVID-19 classification by using image representations of audio to train several deep learning models to detect COVID-19. Instead of directly analyzing sound waves, the researchers converted cough signals into scalogram images—a transformation technique that represents time-frequency information. Using these as input, they trained six pre-trained deep learning models (Google Net, ResNet18/50/101, MobileNetV2 and NasNetMobile) to differentiate between COVID-19 and non-COVID coughs. Their best-performing model was ResNet18, which achieved an accuracy of 94.9%, with a sensitivity of 94.44% and a specificity of 95.37% using the SGDM optimizer. This paper demonstrates the feasibility of deep learning-based COVID-19 detection from cough sounds, and highlights the effectiveness of fine-tuning ResNet18.

Another paper published by Pahar, Klopper et al. in 2021 uses transfer learning and bottleneck features for COVID-19 classification [7]. This study leverages five large, unlabeled audio datasets containing cough, sneeze, speech, and non-vocal sounds to pre-train CNN, LSTM, and ResNet50 models. The pre-trained networks are then either fine-tuned using smaller datasets of cough, breath, and speech audios with COVID-19 labels or used as feature extractors for shallow classifiers. Using this double-tiered approach, the authors aimed to mitigate the effects of overfitting the models on the small amount of COVID-19 labelled data available. This method achieved an AUC of 0.98 when trained on cough sounds, suggesting that cough signals contain the strongest COVID-19 signatures as opposed to breath and speech. This study also emphasizes the importance of data augmentation techniques, including time and frequency masking, to enhance model generalization.

## II. METHODOLOGY

### A. Training Data

The goal of our model is to classify cough audio into positive or negative for COVID-19. For this, we use the crowdsourced COUGHVID [8] and Coswara [9] datasets containing cough audio samples labeled by clinicians as positive or negative. The use of crowdsourced data allows our model to train on data that is reflective of the data that users will be recording on their own through our website.

TABLE I
TRAINING DATA CLASSES

| Data Source | Positive Samples | Negative Samples |
| --- | --- | --- |
| Coswara Light Coughs | 1477 | 591 |
| Coswara Heavy Coughs | 1477 | 591 |
| Coughvid Coughs | 4661 | 1578 |

Since the dataset is unbalanced, we apply data augmentation techniques to increase the number of positive samples. Our data augmentation methods include time shifting, pitch shifting, time masking, and frequency masking. To maintain the integrity of our validation process, we ensure that augmented data points are prevented from being used in the validation set if the original sample was part of the training dataset, as this would lead to an inflated test accuracy.

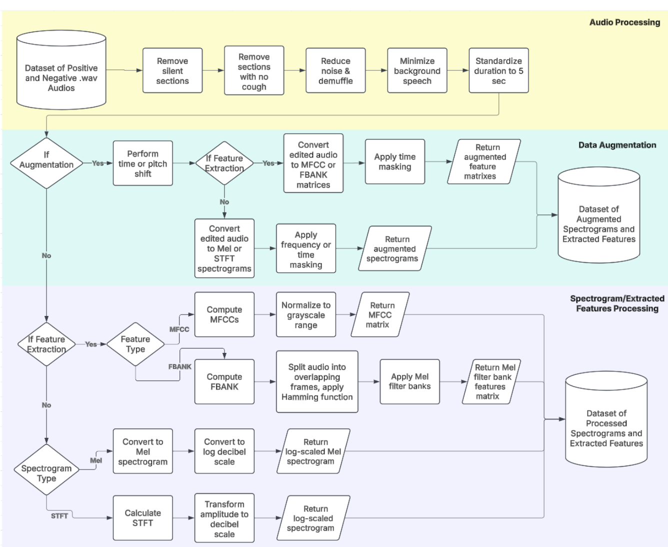


Fig. 1. Full data processing pipeline.

### B. Audio Processing

Each audio sample used for training and inference undergoes a processing pipeline, where it is standardized and cleaned. First, we standardize audio inputs into .wav format with a standardized sample rate of 48 kHz to ensure consistency across different recording devices and environments. To enhance the quality of cough recordings, we apply spectral gating-based noise reduction, which suppresses background noise while preserving the integrity of the cough signal. Next, we apply a Butterworth bandpass filter to remove unwanted low-frequency noise, specifically targeting background speech. We then use adaptive thresholding to detect and remove silent sections longer than 800 ms, isolating the active cough portion. Finally, we trim or pad the audio to ensure a uniform input length of 5 seconds.

### C. Spectrogram Processing

For Mel spectrogram extraction, we apply the Mel filter bank to the power spectrum of the signal, using 128 Mel bands with a maximum frequency of 8000 Hz to closely mimic human auditory perception. The spectrogram is then converted to a log scale using power-to-decibel transformation, emphasizing subtle variations in cough intensity.

For Short-Time Fourier Transform (STFT) spectrograms, we compute the STFT with a Hanning window, using a 2048-point FFT and a 512-sample hop length, capturing both temporal and frequency domain information. The resulting magnitude spectrogram is then transformed into a log-scaled representation to enhance feature differentiation.

### D. Feature Extraction

In addition to spectrograms, we explored two alternative feature representations. First, we calculated Mel-Frequency Cepstral Coefficients (MFCC), which provide a compact representation of an audio clip's spectral envelope using the same Mel scale as Mel spectrograms. We also implemented filter bank (FBANK) feature extraction, which allows us to map the sample's frequency set onto a filter bank feature set. We then converted both MFCCs and FBANK features into grayscale images to train separate models.

### E. Model Framework

We selected ResNet18 as our model due to its lightweight architecture and strong transfer learning capabilities. ResNet18 is a residual-based CNN for image classification with 18 convolutional layers. It was originally trained on ImageNet and generalizes well to image classification tasks [10]. ResNet18 has also demonstrated strong performance on image-based audio classification tasks [6], making it a good fit for our application. Due to our limited dataset size, we kept the earlier convolutional layers frozen and fine-tuned only the final convolutional and fully connected layers. This preserves the pre-trained ImageNet features, mitigating overfitting while allowing the model to specialize in cough classification.

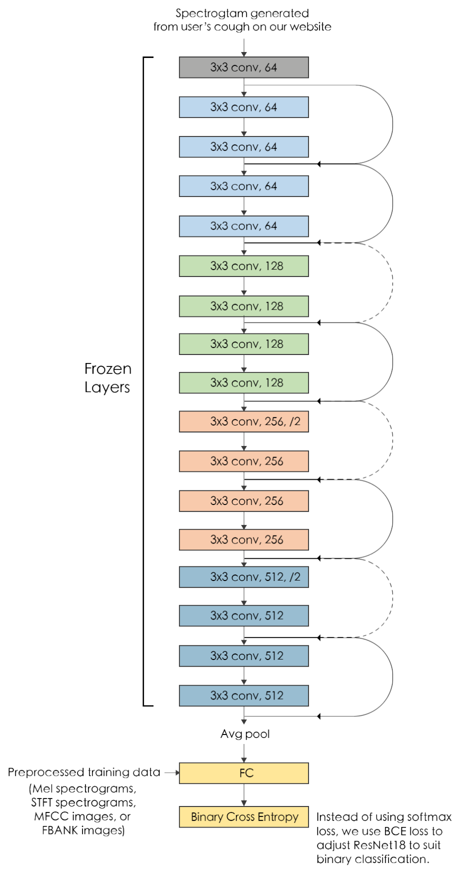


Fig. 2. ResNet18 transfer learning architecture.

Including both original and augmented data, the final model was trained on about 4000 negative and 4000 positive samples to ensure that accuracy was not biased towards either class. Our final model was trained on 30 epochs, at which point we observed a plateau of both train and validation loss. A batch size of 32 was used to best pass data into the model. Both Adam and stochastic gradient descent (SGD) were utilized as the training optimizer, and despite the faster convergence of Adam, we found SGD outperformed Adam on average.

## III. Results

This project is still ongoing, as it began in February 2025. We initially trained four separate models, each respectively utilizing Mel-spectrograms, STFT spectrograms, MFCC features, and FBANK features. Test results currently indicate the Mel-spectrograms performed best out of the three data types, with an accuracy of 64%.

## IV. Conclusion

RespiraCheck represents a significant step forward in using AI-driven solutions for accessible COVID-19 diagnosis. Due to our fully audio-based analysis of cough samples, our website provides a convenient, non-invasive method for COVID testing. Our model, trained on the COUGHVID dataset and optimized using various audio feature representations, demonstrates promising results after only one month of development. Although further improvements are necessary to improve accuracy, RespiraCheck currently stands as a proof of concept to bridge the gap between clinical research and real-world application. By offering an accessible alternative to traditional testing methods, it paves the way for broader, at-home COVID-19 screening, with the possibility of extending the model to diagnose other respiratory illnesses.

## References

[1] J. P. A. Ioannidis, "Global perspective of covid-19 epidemiology for a full-cycle pandemic," *European Journal of Clinical Investigation*, December 2020.

[2] P. H. A. of Canada, "Covid-19; descriptive epidemiology of deceased cases of covid-19 reported during the initial wave of the epidemic in canada, january 15 to july 9, 2020," 2020.

[3] S. Talic *et al.*, "Effectiveness of public health measures in reducing the incidence of covid-19, sars-cov-2 transmission, and covid-19 mortality: Systematic review and meta-analysis," November 2021.

[4] T. Halliday *et al.*, "Financial implications of covid-19 polymerase chain reaction tests on independent laboratories," *Journal of General Internal Medicine*, August 2022.

[5] M. Gheisari *et al.*, "Mobile apps for covid-19 detection and diagnosis for future pandemic control: Multidimensional systematic review," February 2024.

[6] M. Loey and S. Mirjalili, "Covid-19 cough sound symptoms classification from scalogram image representation using deep learning models," *Computers in Biology and Medicine*, December 2021.

[7] M. Pahar *et al.*, "Covid-19 detection in cough, breath and speech using deep transfer learning and bottleneck features," *Computers in Biology and Medicine*, vol. 141, 2022.

[8] L. Orlandic, T. Teijeiro, and D. Atienza, "The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Scientific Data*, vol. 8, no. 1, p. 156, 2021.

[9] D. Bhattacharya, N. K. Sharma, D. Dutta, S. R. Chetupalli, P. Mote, S. Ganapathy, C. Chandrakiran, S. Nori, K. K. Suhail, S. Gonuguntla, and M. Alagesan, "Coswara: A respiratory sounds and symptoms dataset for remote screening of sars-cov-2 infection," *Scientific Data*, vol. 10, no. 1, p. 397, 2023.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.