COVID-19 Cough Recognition

Jiaxuan Guo
Department of Applied Physics
Stanford University
guojx@stanford.edu

Shuo Xin
Department of Physics
Stanford University
xinshuo@stanford.edu

Abstract

Fast diagnosis of COVID-19 is important in stopping the spread of the epidemic. Using machine learning techniques, we may be able to recognize COVID-19 status through cough recordings. Virufy is a team led by Stanford students in collecting data and building network models to achieve COVID-19 cough recognition. In this report we made several improvement to the original model of Virufy by processing mel-spectrogram with DenseNet, as well as other adjustments. Further, we augment the dataset by splitting individual coughs. The model improvement and data augmentation lead to better performance.

1 Introduction

There have been over 120 million people infected by COVID-19. Fast diagnosis will be extremely helpful for stopping the spread of pandemic areas. Cough is a common symptom of COVID-19. Machine learning turns out to be a good approach for diagnosing COVID-19 using cough audios. Previous studies [1, 2, 3, 4, 5] have shown it promising to automatically diagnose COVID-19 through cough and other audio data.

Spectrogram is commonly used for audio processing. For human sound, another quantity, mel-frequency cepstrum (MFC), is widely used as features for speech recognition. The information in MFC is contained in mel-frequency cepstrum coefficients (MFCCs). There have been standard algorithms for computing MFCCs of an audio script in audio processing libraries.

2 Related work

The recent works widely uses the MFCC features and convolutional neural networks (CNNs) in the machine learning model. Laguarta et. al [3] used ResNet50 to process the MFCC inputs. Bagad et. al [4] used spectrogram as input data and ResNet18 as key structures in their model. Brown et. al [1] apply VGGish model [6], a CNN for audio classification, and used the output feature of VGGish together with the features in MFCC for their model.

3 Dataset and Features

3.1 Public datasets

Detection of COVID-19 through cough and other audio signals have been an active area of study. There have been several public datasets available now:

CS230: Deep Learning, Winter 2021, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

- Cosvara by Indian Institute of Science (IISc) Bangalore uses respiratory, cough and speech sounds for diagnostics. [7] Datasets are available at https://github.com/iiscleap/Coswara-Data
- COUGHVID dataset provided by researchers at EPFL contains over 2000 expert-labeled audio scripts. [8]
- Virufy[9] is a Stanford led project on COVID cough detection. The datasets are available at https://github.com/virufy/virufy_data

In our project, we mainly used the Coswara and Coughvid data, because they have the highest data volume among publicly available datasets. After cleaning up the data and removing invalid items, we reached a total number of 2871 cough clips, among which 528 are positive. The dataset is still very limited in terms of size and is also very imbalanced between positive and negative classes. These characteristics of the dataset would have significant impacts when we train our model. Considering the small dataset, we split the data into 70-15-15 for the train, valid, and test sets, and we will use the area under the ROC curve for metrics, which is a common practice for imbalanced data.

3.2 Data augmentation

The problem of small and imbalanced dataset is the biggest challenge. Only making use of these original data didn't bring significant improvements comparing to the original model proposed by Virufy. We further incorporate heavy cough data (the audio files named `cough-heavy.wav` in Coswara data) in dataset2 in addition to shallow cough.

Further, we found that many audio scripts contain several repeated coughs while others contain only one cough. So we split the audio by silence, separate each individual coughs as independent data points. The dataset obtained in this way (dataset3) is around 5 times larger than the original dataset.

	description	positive coughs	total coughs
dataset1	COUGHVID and Coswara (only shallow coughs)	528	2871
dataset2	adding heavy coughs as independent data in dataet1	622	4348
dataset3	Splitting repeated coughs in dataset2	2520	17084

3.3 Splitting individual coughs

We used a simple criterion to separate individual coughs: Split by silence. Set a threshold of 0.8 times the maximum amplitude of the whole recording. Then for each non-silent sections, if the maximum amplitude is larger than the threshold, we cut it into clips of 0.3s length. This is shown in Fig. 1 for some recordings in dataset2.

Because of the limited number of sample in the dataset, it is relatively difficult to reach an good performing end to end neural network for COVID cough detection. In the future, it is worth considering use another neural network to do the recording preprocess here completed by simple algorithm. After the cough recordings being well processed into standard clips, the neural network might do better in terms of giving more accurate result.

4 Methods

The Virufy team uses the mel-frequency spectrogram and reshape it to a 64×64 image as the input of CNN. Parallel to that, the first 39 MFCC coefficients, after a time average, is the input to a second network with two fully connected layers. Also, the Coswara and COUGHVID datasets contain several labels. The Fever/myalgia and medical history of respiratory disease are used as inputs to the third net. The three models merged and after two fully connected layers leads to the output.

In order to make use of the additional metadata of cough clips, we decided to base our network on Virufy's model consist of 3 separate networks. By running Virufy's model on our dataset, we get a baseline performance as shown in FIG.3. Our goal is to improve AUC by adjusting network architectures and hyperparameters. The AUC is often regarded as an intrinsic metric of the model, while the accuracy depends on the decision thresholds of the final binary classification and can

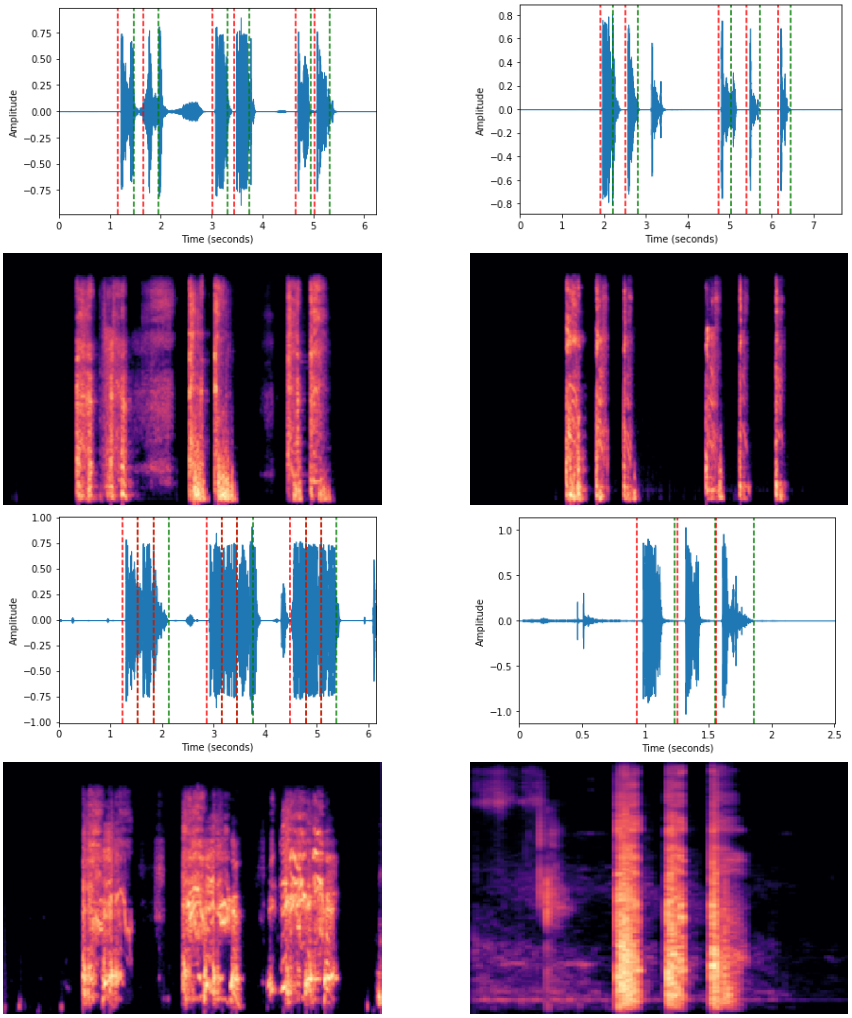

Figure 1: Processing the waveform of cough recordings to separate individual coughs, for recordings No. 16 (top left), 199 (top right), 1049 (bottom left) and 2546 (bottom right) in dataset2. Red vertical lines indicate the starting time of an identified cough and green lines indicate the end. Corresponding spectrograms are shown below each waveform.

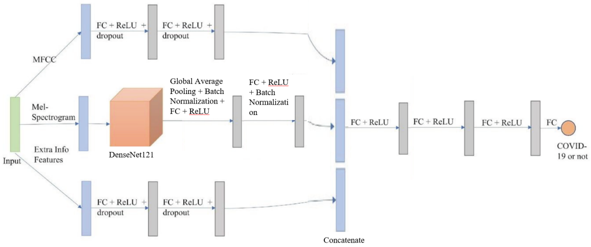

Figure 2: Model structure using DenseNet

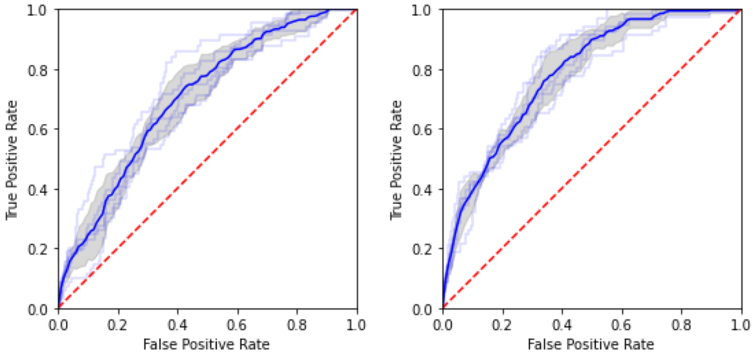

Figure 3: ROC curves for Virufy original model (left) and improved model using DenseNet (right) trained on dataset1.

be misleading when the data is highly imbalanced. When training the neural network, we use the cross-entropy loss function which is a common choice for binary classification problems. We also tried the class-weighed cross-entropy loss function to address our problem of imbalanced training data.

As for the network architecture, we chose DenseNet among other pre-trained networks such as VGG, ResNet, etc. In DenseNet, each layer obtains additional inputs from all preceding layers and passes on its output to all subsequent layers. This means DenseNet is able to utilize features of all complexity levels, while traditional CNN only makes decisions based on the most complex features. DenseNet tends to give more smooth decision boundaries and generally performs well when training data is relatively insufficient.[10]

5 Experiments and results

5.1 Architectures

The inputs to the model include three parts: time averaged MFCC, Mel-frequency spectrogram and two extra labels: Fever/myalgia and medical history of respiratory disease. The model have three networks, the first pass the time averaged MFCC through two fully connected layers with dropouts, the second pass the spectrogram through DenseNet (instead of ordinary CNN as [9]), the third pass two extra features through two fully connected layers with dropouts.

5.2 Experiments

Based on the model of Virufy team [9], we came up with several methods:

- Variant1: Use early stopping in the original model.
- Variant2: Add a fourth parallel network to process MFCC, with three convolutional layers, each followed by an average pooling layer.
- Variant3: Replace the shallow CNN by a bigger one.
- Variant4: Replace the shallow CNN by `DenseNet121` with usual cross-entropy loss function.
- Variant5: Replace the shallow CNN by `DenseNet121` with class weighted loss function to address the class imbalance problem of our dataset.

We then run the variant models and monitor their ROC curves. By comparing the original model and variant 1, we get a improvement of AUC from 0.70 to 0.746 just by reduce training epoch from 20 to 12. This means the overfitting problem is quite serious, which may be the consequence of small training set. By using more complex network architectures, the model has a little improvement in AUC (0.745, 0.742, 0.766 for variant 2,3,4 respectively), which implies getting to subtle features may helps but the improvement is not satisfying. By changing from usual cross-entropy loss function to class-weighted loss function, we see a improvement from 0.766 in variant 4 to 0.782 in variant 5. This

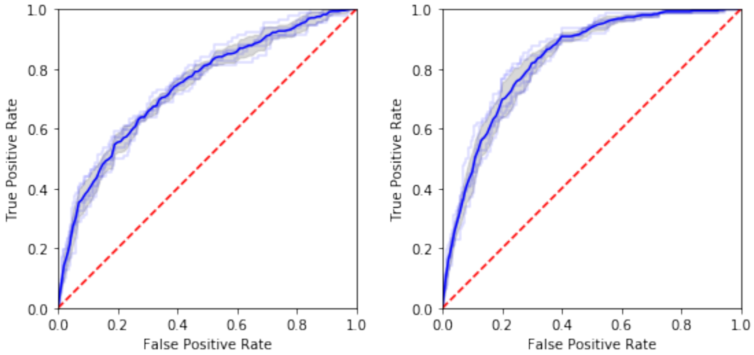

Figure 4: Same as Fig. 3 but trained on dataset2.

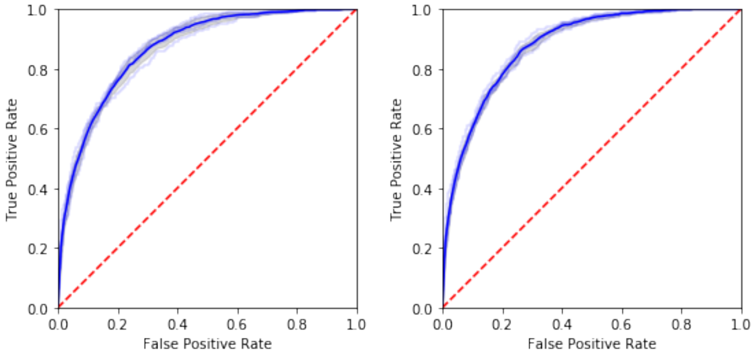

Figure 5: Same as Fig. 3 but trained on dataset3.

modification is effective to deal with the data imbalanced problem in our dataset. We also tried to include new features of MFCC as input to the model. Inspired by [1], we use 10 additional statistical quantity as the input to the first network, instead of just the time average of MFCC coefficients. But no major improvement is achieved.

Variant5 is our best improved model, whose structure is shown in Fig. 2. We present the comparison of ROC curves between that and the original Virufy model in Fig. 3.

We then focus on data augmentation. First we tried to make use of all data available. In Coswara data, each case is associated with two cough recordings, a shallow one and a heavy one, along with other audios like speaking and breathing. In the previous Virufy model [9], only the shallow one is used. Now we also include the heavy one as dataset2. By training the same model as Variant5 on this dataset, the AUC is improved to 0.83. The ROC curve is shown in Fig. 4. The original Virufy model trained on the same dataset is also shown for comparison. Finally by further augmenting the data and train the model on dataset3 (Variant8), we got an AUC of 0.88 (right panel of Fig. 5). This is our current best result. Notably, training Virufy original model on dataset3 also gives much better result, with an AUC of 0.87.

6 Discussion and Future Work

To summarize, we have made some improvement to the virufy model using `DenseNet` and data augmentation. The amount of data is still limiting the performance. Currently the number of "authentic" recordings in our data sets is less than 5000 with highly unbalanced distribution (only 622 positive cases) and some of them are coughs of same persons, while the data MIT team [6] collected has 2660 positive cases and an AUC of 0.97 is achieved. In the future we'll try looking for more data and better data augmentation techniques, as well as further improving the model.

7 Contributions

As a team, both of us contributed equally to the planning, experimenting, coding, and every other aspects of the project. Specifically, we are grateful to the open source effort on COVID-19 cough detection. Their excellent work provide us with accessible dataset and valuable experiences.

References

[1] Chloë Brown, Jagmohan Chauhan, Andreas Grammenos, Jing Han, Apinan Hasthanasombat, Dimitris Spathis, Tong Xia, Pietro Cicuta, and Cecilia Mascolo. Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3474–3484, 2020.

[2] Rob Dunne, Tim Morris, and Simon Harper. High accuracy classification of covid-19 coughs using mel-frequency cepstral coefficients and a convolutional neural network with a use case for smart home devices. 2020.

[3] Jordi Laguarta, Ferran Hueto, and Brian Subirana. Covid-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, 1:275–281, 2020.

[4] Piyush Bagad, Aman Dalmia, Jigar Doshi, Arsha Nagrani, Parag Bhamare, Amrita Mahale, Saurabh Rane, Neeraj Agarwal, and Rahul Panicker. Cough against covid: Evidence of covid-19 signature in cough sounds. *arXiv preprint arXiv:2009.08790*, 2020.

[5] Jing Han, Kun Qian, Meishu Song, Zijiang Yang, Zhao Ren, Shuo Liu, Juan Liu, Huaiyuan Zheng, Wei Ji, Tomoya Koike, et al. An early study on intelligent analysis of speech under covid-19: Severity, sleep quality, fatigue, and anxiety. *arXiv preprint arXiv:2005.00096*, 2020.

[6] Shawn Hershey, Sourish Chaudhuri, Daniel PW Ellis, Jort F Gemmeke, Aren Jansen, R Channing Moore, Manoj Plakal, Devin Platt, Rif A Saurous, Bryan Seybold, et al. Cnn architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp)*, pages 131–135. IEEE, 2017.

[7] https://coswara.iisc.ac.in/.

[8] Lara Orlandic, Tomas Teijeiro, and David Atienza. The coughvid crowdsourcing dataset: A corpus for the study of large-scale cough analysis algorithms. *arXiv preprint arXiv:2009.11644*, 2020.

[9] Gunvant Chaudhari, Xinyi Jiang, Ahmed Fakhry, Asriel Han, Jaclyn Xiao, Sabrina Shen, and Amil Khanzada. Virufy: Global applicability of crowdsourced and clinical datasets for ai detection of covid-19 from cough. *arXiv preprint arXiv:2011.13320*, 2020.

[10] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.