A Multi-Branch Deep Learning Network for Automated Detection of COVID-19

Ahmed Fakhry¹, Xinyi Jiang, Jaclyn Xiao², Gunvant Chaudhari³, Asriel Han⁴

¹University of Alexandria, Department of Electronics and Communication, Egypt
²Duke University, Department of Biomedical Engineering, USA
³University of California San Francisco, School of Medicine, USA
⁴Stanford University, School of Humanities and Sciences, USA

ahmedfakhry805@gmail.com, xinyi.jiang2@outlook.com, jaclyn.xiao@duke.edu, gunvant.chaudhari@ucsf.edu, asrielhan@stanford.edu

Abstract

Fast and affordable solutions for COVID-19 testing are necessary to contain the spread of the global pandemic and help relieve the burden on medical facilities. Currently, limited testing locations and expensive equipment pose difficulties for individuals seeking testing, especially in low-resource settings. Researchers have successfully presented models for detecting COVID-19 infection status using audio samples recorded in clinical settings, suggesting that audio-based Artificial Intelligence models can be used to identify COVID-19. Such models have the potential to be deployed on smartphones for fast, widespread, and low-resource testing. However, while previous studies have trained models on cleaned audio samples collected mainly from clinical settings, audio samples collected from average smartphones may yield suboptimal quality data that is different from the clean data that models were trained on. This discrepancy may add a bias that affects COVID-19 status predictions. To tackle this issue, we propose a multi-branch deep learning network that is trained and tested on crowdsourced data where most of the data has not been manually processed and cleaned. Furthermore, the model achieves state-of-art results for the COUGHVID dataset. After breaking down results for each category, we have shown an AUC of 0.99 for audio samples with COVID-19 positive labels.

Index Terms: cough analysis, disease detection, COVID-19

1. Introduction

With the spread of coronavirus disease 19 (COVID-19), over 103M positive cases have been reported by the date of February 3rd, 2021 [1]. Many clinical and medical efforts have been exerted to contain the crisis, such as the creation of reverse transcription-polymerase chain reaction (RT-PCR) tests that the gold standard for detecting COVID-19 in clinical practice due to their high sensitivity and specificity [2]. Despite the reliability of RT-PCR tests, they require costly reagents and tools that have prevented ubiquitous access globally [3]. Furthermore, administering and processing the test comes with concerns for exposure, and the test results are returned in hours to days [2]. These issues prevent RT-PCR tests from being used as primary screening for COVID-19 status. Although vaccination efforts globally are underway, distribution efforts have been impeded in low- and middle-income countries [4]. Therefore, a fast, accurate, low-cost, and accessible screening test for COVID-19 is necessary to help limit its spread.

Emerging Artificial Intelligence (AI) technologies show promise in allowing the creation of such a solution. Deep learning and machine learning algorithms could be used to analyze cough sounds of infected patients and infer predictions. Multiple research groups have been dedicated to gathering sound recordings for COVID-19 patients of all ages, in various settings, symptomatic or asymptomatic, and at different time periods prior to symptoms onset. This ability allows AI algorithms to learn audio patterns particular to the disease for patients with different sets of demographic and medical characteristics. The most commonly collected audio recordings for detecting COVID-19 are coughs. Some groups, such as Coswara [5] and Virufy [6], collect additional counting and vowel data along with cough recordings.

The state-of-the-art performance of AI algorithms for detecting COVID-19 ranges from 0.68 AUC to 0.97 AUC [2, 15], based on several factors, such as the quality of audio files used to train the learning algorithm, preprocessing methodology, and the algorithm structure. In our previous work [6], our group developed a deep learning model with an average AUC of 0.77 across several datasets, including Coswara [19], COUGHVID [9], and Virufy [6] datasets. In this paper, we build on our previous approach and present the state-of-the-art deep learning model on the COUGHVID dataset.

We propose a multi-branch deep learning network based on fusing heterogeneous input features. Our model achieves the new State-Of-Art for the COUGHVID dataset. Our proposed network successfully learns patterns from mel-frequency cepstral coefficients, mel-spectrum images, and other clinical information to detect COVID-19 from cough audio submissions. It also distinguishes COVID-19 positive donors from healthy symptomatic donors. Our algorithm scores 0.99 AUC for detecting COVID-19 with a precision of 98.4% and a recall of 92.15% across different sections of age and gender.

2. Background

Since the outbreak of COVID-19, groups like Coswara [19] and COUGHVID [9] have focused on collecting high-quality biometric data. Furthermore, research has been conducted to train cough-based machine learning models to detect COVID-19 [8, 10, 11, 12, 13]. The high AUC exhibited by such studies has demonstrated the potential of COVID-19 detection from cough data. For example, the model developed by Cambridge University had an AUC of 0.82 and used mel-frequency cepstral coefficients (MFCC) and other audio features [10]. Similarly, MIT researchers used a biomarker layer with
ResNet-50 [8] based models [13] that input mel-frequency cepstral coefficients (MFCC) and output biomarkers that were used as inputs to the three parallel ResNet-50 convolutional neural networks. This model was trained on 4,256 subjects and resulted in an AUC of 0.97. Despite markedly high AUC values, previous models rely largely on manually filtered data and analysis of MFCC and spectrogram inputs. The consideration of other features may result in more clinically applicable and accurate results.

Many studies have shown that cough symptoms from multiple respiratory conditions have distinguishable features [14-17]. COVID-19 has been noted to share similar latent cough traits as other respiratory conditions [18-20]. For instance, a patient with non-COVID-19-associated pneumonia may be falsely identified by a machine learning model as COVID positive due to the large prevalence of pneumonia that develops in severe COVID-positive patients [21]. Thus, classification of a patient with pneumonia as purely COVID negative may lead to erroneous identification of traits during model training.

In this paper, we propose a COVID-19 diagnostic model that integrates heterogeneous feature inputs. We adopt a three-class diagnostic system previously created by Agbley et. al [7]: COVID-19 positive, symptomatic COVID-19 negative, and asymptomatic COVID-19 negative. We show that our model achieves a state-of-the-art performance at identifying COVID-19 positive patients, regardless of symptom status.

3. Methods

3.1. Data

The COUGHVID dataset, a publicly available dataset of globally crowdsourced cough audio recordings, was utilized to train our network. COUGHVID includes a total of 20,072 cough recordings labeled with positive COVID-19, symptomatic COVID-19 negative, and asymptomatic COVID-19 negative, along with other clinical information and metadata [9]. One metadata parameter calculated for each audio sample is the confidence of whether the given recording constitutes a cough sound, based on a pre-trained automatic classifier. We filtered out all audio recordings that have a degree of certainty below 0.9, resulting in 5,749 total audio recordings of which 380 (6.6%) were COVID-positive.

3.2. Data Augmentation

We applied data augmentation techniques to create a more balanced training dataset. By adding gaussian noise, pitch shifting, shifting the time signal and stretching the time signal, we increased the number of COVID-19 labeled samples from 380 to 750. Adding 750 randomly selected samples for each one of the remaining two labels led to a dataset of 2,250 audio samples. Prior to the augmentation phase, we split the data into training, validation and test datasets. For all data splits, each class is represented by one third of the number of split samples, resulting in a balanced distribution.

Table 1: Distribution of samples with augmentation (total samples/ original COUGHVID dataset samples)

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Validation</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>COVID-19 positive</td>
<td>600/304</td>
<td>75/38</td>
<td>75/38</td>
</tr>
<tr>
<td>Symptomatic COVID-19 negative</td>
<td>600/600</td>
<td>75/75</td>
<td>75/75</td>
</tr>
<tr>
<td>Asymptomatic COVID-19 negative</td>
<td>600/600</td>
<td>75/75</td>
<td>75/75</td>
</tr>
</tbody>
</table>

3.3. Audio and Clinical Features

We train our network on mel-frequency cepstral coefficients (MFCCs) and mel-frequency spectrograms, both commonly used audio features for audio classification and speech-recognition [22]. In our network illustrated below in Figure 1, we use two classifiers, one for mel-spectrograms and the other for MFCCs. Each cough audio file was downsampled to 22kHz and split into audio chunks. The first 13 MFCCs were extracted from the preprocessed audio chunks using the librosa package in python. The mel-spectrograms were extracted using the librosa package for the same parameters used to extract MFCCs. Each mel-spectrogram color image was reshaped to the size of (224,224,3), the original input size of the ResNet-50 convolutional neural network.

To supplement features from cough audio, we add clinical information in the COUGHVID dataset, including history of respiratory conditions and symptoms such as fever. We pass this clinical information in a one-dimensional vector of binary numbers as each binary number represents the presence or absence of a symptom or a condition. All data was randomly grouped into train-validation-test sets using a 80-10-10 split.

We also do slice-based analysis and divided the test dataset into groups based on age and gender. According to age, we split the test dataset into four groups. First group is for patients below 20 years old, second is for patients between 20 & 40 years old, the third group is for patients between 40 & 60 years old while the last group is for elderly above 60 years old. For gender, we have split the test dataset into two groups, one for male and the other for females. Table 2 and Table 3 show the histogram of the number of audio records for each group.

Table 2: Distribution of test samples in each age group.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Cough files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≤ 20</td>
<td>35</td>
</tr>
<tr>
<td>20 &lt; Age ≤ 40</td>
<td>125</td>
</tr>
<tr>
<td>40 &lt; Age ≤ 60</td>
<td>50</td>
</tr>
<tr>
<td>60 &lt; Age</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 3: Distribution of test samples in each gender group.

<table>
<thead>
<tr>
<th>Gender Group</th>
<th>Cough files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>125</td>
</tr>
<tr>
<td>Female</td>
<td>92</td>
</tr>
</tbody>
</table>

3.4. Model

The model is a multi-branch ensemble learning architecture based on a ResNet-50 convolutional neural network that is pre-trained on ImageNet dataset and stripped of the top layer (classification layer) [23]. The input for the CNN is a mel-spectrogram heatmap of size (224,224,3) and the output of the
CNN is passed to both, a global average pooling layer and a global maximum pooling layer in two separate and parallel links. Each of them is followed by batch normalization and dropout layers before concatenated together in a single dense layer to make the first branch.

The second branch is a multi-layer feedforward neural network consisting of two dense layers that are of 8 nodes and 64 nodes, respectively. And each layer is followed by a batch normalization and dropout layers. The input for the first branch is a 1D vector of binary numbers. Each binary number represents one of the clinical information associated with the patient record, such as history of respiratory diseases, type of cough and whether the patient has fever or not. This branch is meant to incorporate the clinical information.

The third branch is a double parallel feedforward neural network which takes a vector of mel-frequency cepstrum coefficients as an input vector of size (13,1). Each of the two parallel links is a multi-layer feed forward neural network which consists of two layers as each layer is followed by a batch normalization and dropout layers. The high ends of both links are concatenated together in a single dense layer.

The extracted high-level features at the high end of the three branches are fused together before being passed to a Static Feed Forward Network (SFFN) in Figure 1 that is followed by a softmax layer for a multi-label classification task. The three labels are: asymptomatic Covid-19 negative, symptomatic Covid-19 negative and positive Covid-19.

The proposed network architecture enabled us to take advantage of several heterogeneous classifiers and fuse together the extracted high-level features from spectrogram image using ResNet-50 CNN, and from MFCCs using a DNN. The network architecture, number of hidden layers for each branch and number of units per each layer were hyperparameters and were decided on using grid-search.

### 3.5. Statistics

Area under the ROC curve (AUC) is calculated for each class using one vs. all methodology and the average AUC for all classes using the micro-averaging scoring metric. For sensitivity and specificity analysis, a prediction threshold value of 0.9 was used.

### 4. Results

After the success of the multi-branch model, which we have proposed in our previous work [6] with a maximum value of 0.80 AUC for COUGHVID and Coswara datasets combined, we now propose the state of the art deep learning model for COUGHVID dataset with micro-average AUC of 0.91, as shown in Figure 2.

Table 4 contains the results of our proposed network in Figure 1 on the test dataset. Our network scores an AUC value of 0.99 for detecting positive COVID-19 coughs, implying that our model has extremely low false negative and false positive rate. As illustrated in Table 4, the sensitivity and specificity of detecting COVID-19 are 85% and 99.2%, respectively, which shows the high diagnostic performance of our network. In addition, despite the low prevalence of the COVID-19 disease in the test set (33.3%), the network has scored high positive-predictive value (98.4%). Such high precision and sensitivity ensures that our network will have very low false-negative results COVID-19, making it an appropriate screening for population-scale testing.

In Table 4, we show an ablation analysis comparing the performances of the Multi-Branch model to a single ResNet-50 model using only mel-spectrogram features.

![Figure 1: Multi-Branch network architecture](image)

<table>
<thead>
<tr>
<th>Model</th>
<th>Class 1 AUC</th>
<th>Class 2 AUC</th>
<th>Class 3 AUC</th>
<th>Micro-avg AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Branch</td>
<td>0.84</td>
<td>0.82</td>
<td>0.99</td>
<td>0.91</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>0.73</td>
<td>0.69</td>
<td>0.97</td>
<td>0.84</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Sens.</th>
<th>Spec.</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-Branch</td>
<td>85%</td>
<td>99.2%</td>
<td>98.4%</td>
<td>92.2%</td>
</tr>
<tr>
<td>ResNet-50</td>
<td>64%</td>
<td>97.1%</td>
<td>92.3%</td>
<td>83.6%</td>
</tr>
</tbody>
</table>

In Table 5, we compare the performance of our network in detecting COVID-19 to another deep learning model [7] trained on the COUGHVID dataset, along with medical experts’ prediction performances. Our network outperforms both experts and Agbley et al. [7] in detecting COVID-19 from cough audio files.
Successfully detects COVID-19 patients, which opens the door for using our algorithm to detect a variety of respiratory diseases.

All training data for our model was derived from crowdsourced sources. Thus, that data’s quality and character closely resemble the input to the AI system when it will be used in smartphone applications. This characteristic distinguishes our approach from previous approaches that have utilized manually cleaned data. While clean data has the potential advantage of yielding more accurate results, AI models intended for public use should be trained on the same type of potentially noisy data that is collected in practice to ensure consistent predictions. From our experiments, we have found that the inclusion of all three features: MFCC, mel-spectrogram, and clinical features combined with a heavy model, results in picking up more wholesome information for final decision making but a future study with an independent test set would be ideal to determine the level of overfitting of the model.

We would also like to acknowledge a few limitations of this study. First, there has been a common issue for crowdsourced data that it is difficult to verify the authenticity of the labels of the data. As a result, not all the labels from the COUGHVID dataset have been verified. Second, due to the difference between data collection methods and intention to maintain anonymity, COUGHVID dataset does not have some of the metadata that may be documented in other datasets [6], for example, race information.

6. Conclusions

Our developed model has achieved high performance on the COVID-19 positive class coughs, proving its power to capture the viral feature in audio files. Subsequently, it can theoretically be extended to other applications such as identifying other respiratory diseases with enough patient cough samples. Furthermore, as more data from the under-represented areas such as South America is collected, our model can further generalize to more crowdsourced datasets and improve the model performance for multiple datasets from various races, regions, and populations.

7. References


