



AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app

Ali Imran^{a,b}, Iryna Posokhova^{b,c}, Haneya N. Qureshi^a, Usama Masood^a,
Muhammad Sajid Riaz^a, Kamran Ali^d, Charles N. John^a, MD Iftikhar Hussain^{b,e},
Muhammad Nabeel^{a,*}

^a AI4Networks Research Center, Dept. of Electrical & Computer Engineering, University of Oklahoma, USA

^b AI4Lyf LLC, USA

^c Kharkiv National Medical University, Ukraine

^d Dept. of Computer Science & Engineering, Michigan State University, USA

^e Allergy, Asthma & Immunology Center PC, USA

ARTICLE INFO

Keywords:

Artificial intelligence
COVID-19
Preliminary medical diagnosis
Pre-screening
Public healthcare

ABSTRACT

Background: The inability to test at scale has become humanity's Achilles' heel in the ongoing war against the COVID-19 pandemic. A scalable screening tool would be a game changer. Building on the prior work on cough-based diagnosis of respiratory diseases, we propose, develop and test an Artificial Intelligence (AI)-powered screening solution for COVID-19 infection that is deployable via a smartphone app. The app, named AI4COVID-19 records and sends three 3-s cough sounds to an AI engine running in the cloud, and returns a result within 2 min.

Methods: Cough is a symptom of over thirty non-COVID-19 related medical conditions. This makes the diagnosis of a COVID-19 infection by cough alone an extremely challenging multidisciplinary problem. We address this problem by investigating the distinctness of pathomorphological alterations in the respiratory system induced by COVID-19 infection when compared to other respiratory infections. To overcome the COVID-19 cough training data shortage we exploit transfer learning. To reduce the misdiagnosis risk stemming from the complex dimensionality of the problem, we leverage a multi-pronged mediator centered risk-averse AI architecture.

Results: Results show AI4COVID-19 can distinguish among COVID-19 coughs and several types of non-COVID-19 coughs. The accuracy is promising enough to encourage a large-scale collection of labeled cough data to gauge the generalization capability of AI4COVID-19. AI4COVID-19 is not a clinical grade testing tool. Instead, it offers a screening tool deployable anytime, anywhere, by anyone. It can also be a clinical decision assistance tool used to channel clinical-testing and treatment to those who need it the most, thereby saving more lives.

1. Introduction

By April 28, 2020, there were 3,024,059 confirmed cases of corona virus disease 2019 (COVID-19), leading to 208,112 deaths and disrupting life in 213 countries and territories around the world [1]. The losses are compounding everyday. Given no vaccination or cure exists as of now, minimizing the spread by timely testing the population and isolating the infected people is the only effective defense against the unprecedentedly contagious COVID-19. However, the ability to deploy this defense strategy at this stage of pandemic hinges on a nation's ability to timely test significant fractions of its population including

those who are not contacting medical system yet. The capability for agile, scalable and proactive testing has emerged as the key differentiator in some nations' ability to cope and reverse the curve of the pandemic, and the lack of the same is the root cause of historic losses for others.

1.1. Why might not clinic visit based COVID-19 testing mechanisms alone sufficiently control the pandemic at this stage?

The "Trace, Test and Treat" strategy succeeded in flattening the pandemic curve (e.g., in South Korea, China and Singapore) in its early

* Corresponding author.

E-mail address: muhmd.nabeel@ou.edu (M. Nabeel).

<https://doi.org/10.1016/j.imu.2020.100378>

Received 4 May 2020; Received in revised form 19 June 2020; Accepted 19 June 2020

Available online 26 June 2020

2352-9148/© 2020 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

stages. However, in many parts of the world the pandemic has already spread to an extent that this strategy is not proving effective anymore [2]. Recent studies show that it is virus often transmitted when an undiagnosed population coughs, that contributes to its much rapid and covert spread [3]. Data shows that 81% of COVID-19 carriers do not develop severe enough symptoms for them to seek medical help, and yet they act as active spreaders [4]. Others develop symptoms severe enough to prompt medical intervention only after several days of being infected. These findings call for a new strategy centered on “Pre-screen/test proactively at population scale, self-isolate those tested positive for self-healing without further spreading and channel medical care towards the most vulnerable”.

As per World Health Organization (WHO) guidance, Nucleic Acid Amplification Tests (NAAT) such as real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) should be used for routine confirmation of COVID-19 cases by detecting unique sequences of virus ribonucleic acid (RNA). This test method, while being the current gold standard, is not an adequate way to control the pandemic for reasons that include but are not limited to:

- 1) The limited availability of testing due to geographical and temporal factors.
- 2) The scarcity and expense of clinical tests needed to cover the massive time-sensitive demand.
- 3) The requirement of in-person visits to a hospital, clinic, lab or mobile lab. Such visits expose more members of the public to COVID-19. This is not a trivial problem given the recent studies that show how highly stable and hence contagious COVID-19 appears to be. For example [5], shows that the aerosol stability of COVID-19 is up to 3 h in aerosols and up to seven days on different surfaces.
- 4) The turnaround time for current tests is several days, recently stretching to 10 days in some countries as labs are becoming overwhelmed [6,7]. By the time a patient is diagnosed using current methods, the virus has already been passed to many.
- 5) The in-person testing methods put the medical staff, particularly those with limited protection, at serious risk of infection. The inability to protect our medics can lead to further shortage of medical care and increased distress on the already stressed medical staff.

To make tests more readily accessible, on March 28th the United States Food and Drug Administration (FDA) approved a faster test that can yield results in 15 min [8]. The test works similar to Polymerase Chain Reaction (PCR) by identifying a portion of the COVID-19 RNA in the nasopharyngeal or oropharyngeal swab. The FDA also recently approved another rapid molecular-based test, which delivers positive results in as little as 5 min and negative results in 13 min [9]. However, the FDA warns that there is a high probability of false negative results using this test [10]. While a leap forward, this test still requires an office visit and thus the breaching of social distancing and self-isolation. Though much faster, the newly approved test still does not solve many of the aforementioned problems. Furthermore, emerging reports of shortages of critical equipment used to collect patient specimens, like masks and swabs, could blunt its impact on controlling the pandemic [11,12]. In order to protect others from potential exposure, the FDA has also approved at-home sample collection [13]. However, once a patient collects a nasal sample, they need to put it in a saline solution and ship it overnight to a certified lab authorized to run specific tests on the kit. Hence, this approach also introduces delays and could compromise on the quality of samples if the sample is stored for too long. In addition, it could also introduce the chances of errors while collecting the sample, since the patients collect the sample themselves, rather than trained doctors or healthcare professionals.

More recently, two alternative approaches for COVID-19 infection diagnosis leveraging analysis of either X-ray [14–26] or CT Scan [27–33] images have been proposed in the literature. These techniques, either through an examination by a radiologist, or when combined with

AI-based image processing, are able to diagnose COVID-19 with even higher accuracies, and in some cases even better than the rRT-PCR based test. Recent studies report a pooled sensitivity of 94% (95% confidence interval: 91%–96%), but a low specificity of 37% (95% confidence interval 26%–50%) for CT-based diagnosis [34]. Therefore, CT based diagnosis may help to overcome the sub-optimal sensitivity of PCR tests [35]. However, while both of these approaches reduce the burden on radiologists to perform the diagnosis, they still require a visit to a well-equipped clinical facility. As a result, these approaches also inherit the issues of office visit based tests that are highlighted above.

It is mainly due to the inability to test large swaths of populations timely, safely and cost effectively and exactly track the actual spread that even the richest nations on earth are finding it difficult to contain the pandemic.

1.2. Proposed cough based COVID-19 screening approach

The idea of using cough for possible preliminary diagnosis of COVID-19, and the need to investigate its feasibility is motivated by the following key findings:

- 1) Prior studies have shown that cough from distinct respiratory syndromes have distinct latent features [36–43]. These distinct features can be extracted by appropriate signal processing and mathematical transformations of the cough sounds. The features can then be used to train a sophisticated AI engine for performing the preliminary diagnosis solely based on cough. Our in-depth analysis of the pathomorphological alternations caused by COVID-19 in the respiratory system (reported in Section 2.1), shows that the alternations are distinct from those caused by other common non-COVID-19 respiratory diseases. This finding is corroborated by the meta-analysis of several recent independent studies (reported in Section 2.1) that show that COVID-19 infects the respiratory system in a distinct way. Therefore, it is logical to hypothesize that cough caused by COVID-19 is also likely to have distinct latent features and the risk of these features overlapping with those associated with other respiratory infections is low. These distinct latent features can be exploited to train a domain aware AI engine to differentiate COVID-19 cough from non-COVID-19 cough. Our experiments (Fig. 1, Section 2.2) show that this is indeed possible.
- 2) Cough manifests as a symptom in the majority (e.g., 67.7% as per [44]) but not all COVID-19 carriers. However, studies show that coughing is one of the key mechanisms for the social spreading of COVID-19 [3]. Droplets containing the virus emitted through cough landing on surfaces where the virus has been shown to survive for long periods of time has been reported as the most prolific

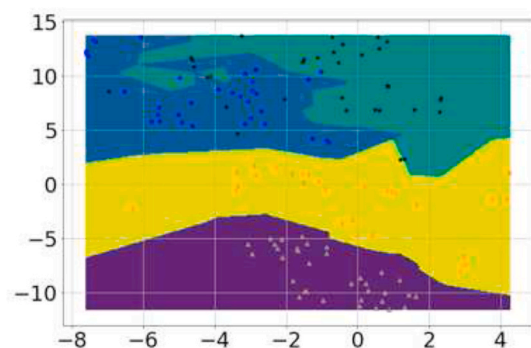


Fig. 1. Visualization of features for the four classes via t-SNE (gray triangles correspond to normal, blue circles correspond to bronchitis, black stars correspond to pertussis and orange diamonds represent COVID-19 cough. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

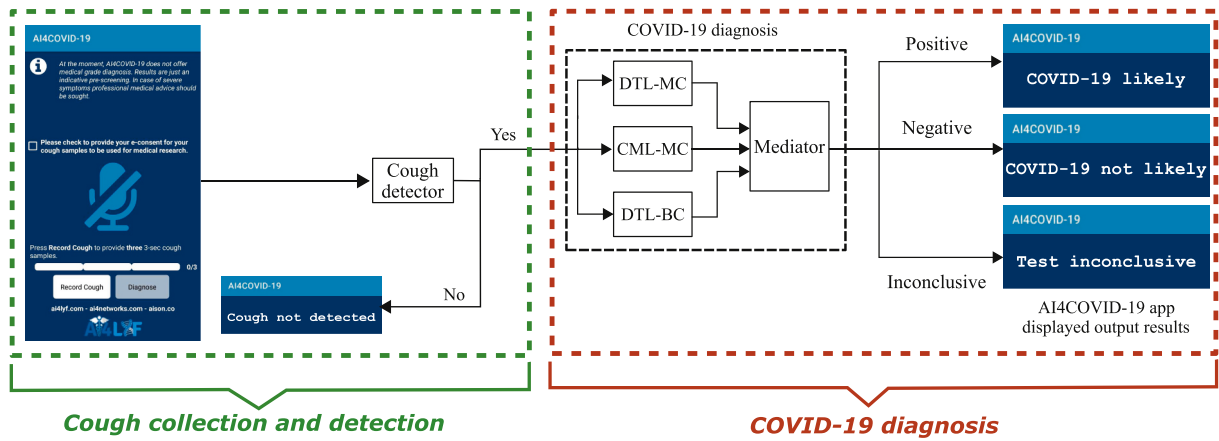


Fig. 2. Proposed system architecture and flow diagram of AI4COVID-19, showing snapshot of Smartphone App at user front-end and back-end cloud AI-engine blocks consisting of Cough Detector block (further elaborated in Fig. 4 and Section 2.3) and COVID-19 diagnosis block containing Deep Transfer Learning-based Multi-Class classifier (DTL-MC), Classical Machine Learning-based Multi-Class classifier (CML-MC) and Deep Transfer Learning-based Binary-Class classifier (DTL-BC) (further elaborated in Fig. 5 and Section 2.3).

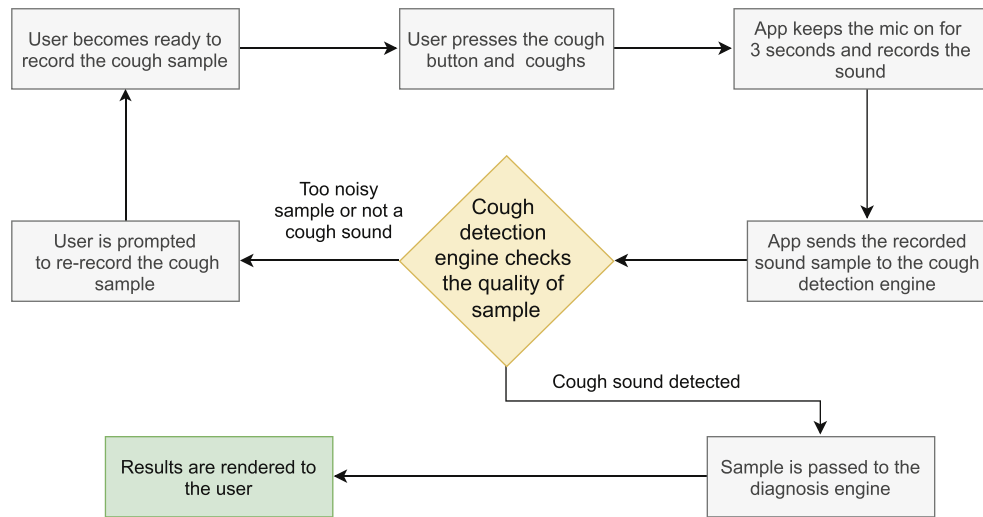


Fig. 3. A flow chart highlighting the steps of the proposed system.

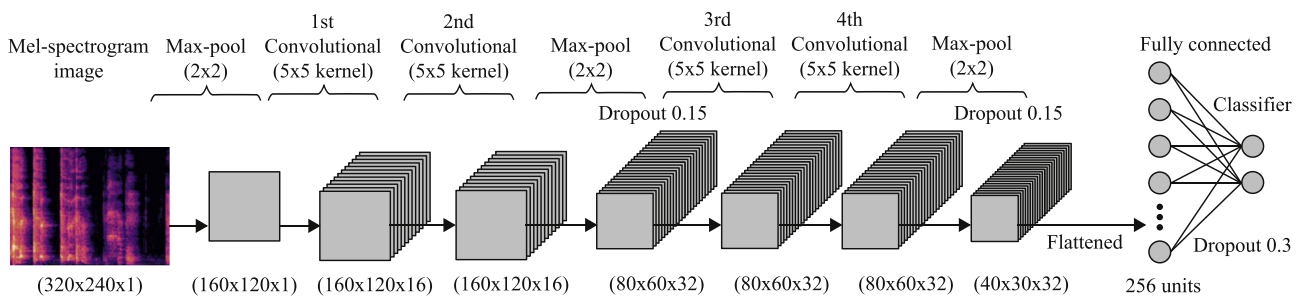


Fig. 4. Cough detection classifier.

mechanism of spreading the COVID-19 [45]. Hence, if a COVID-19 patient is not showing cough as a symptom, the patient is most likely not spreading as actively as a coughing COVID-19 patient. In other words, cough-based testing, even if far from being as sensitive as clinical testing, can actually directly help in reducing R_0 [4].

3) Due to the ease of measurement, a temperature scan is currently the predominant screening method for COVID-19, e.g., used at the

airports. However, between cough and fever, the number of non-COVID-19 medical conditions that can cause fever are much larger than the non-COVID-19 conditions that can cause cough. Our analysis shows that cough contains COVID-19 specific features even if it is non-spontaneous, i.e., when the COVID-19 patient is asked to cough. This means cough can be used as a pre-screening method by asking the subject to simulate cough.

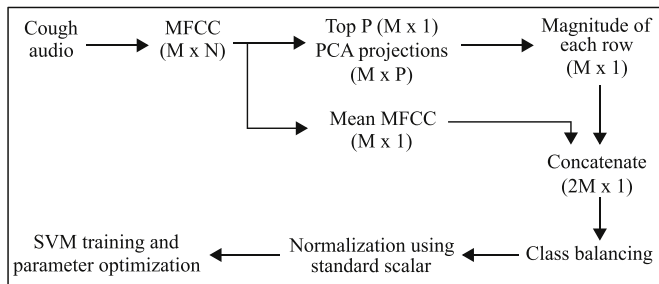


Fig. 5. Classical Machine Learning-based Multi-Class classifier (CML-MC).

1.3. Contributions and paper contents

The contributions and contents of this paper are outlined below:

- 1) We analyze the pathomorphological changes caused by COVID-19 in the respiratory system from the studies examining X-rays and CT scans of alive COVID-19 patients. Our analysis also includes the autopsy report studies of deceased patients. The purpose of this analysis is to apply first principle-based approach. The goal is to see if the pathomorphological alterations caused by COVID-19 in the respiratory system (i.e., the part of body that produces a cough sound) are different from those caused by other common bacterial or viral infections. This is to determine if it is even theoretically possible for the COVID-19 cough to have any distinct latent features. The in-depth study of pertinent pathomorphological alterations suggests that it is possible.
- 2) Building on the insights from first principle-based approach and our prior work [46], as well as several other independent studies [36–43] that suggest distinct latent features in cough sounds can be used for successful AI-based diagnosis of several respiratory diseases, we hypothesize that “Cough sound can be used at least for preliminary diagnosis of the COVID-19 by performing differential analysis of its unique latent features relative to other non-COVID-19 coughs.”
- 3) Continuing the medical literature review, we further identify and shortlist the non-COVID-19 respiratory syndromes that are relatively common and are known to cause similar-sounding cough as that of COVID-19 patients. The shortlist includes pertussis, bronchitis, influenza, asthma, pneumonia, bronchiolitis and croup.
- 4) Given that even the shortlist is too long to gather reliable data for this time sensitive project, we reduce the size of our data gathering campaign to a manageable one by leveraging the findings from literature which show that cough caused by the last five medical conditions in the shortlist above does have features unique to each condition. Therefore, in the interest of time, we go on to focus on the differential analysis of COVID-19 cough, and coughs associated with pertussis and bronchitis as these two conditions are not examined earlier.
- 5) We gather cough data of COVID-19, pertussis and bronchitis patients. Cough samples from COVID-19 patients include both spontaneous cough (symptomatic) and non-spontaneous (i.e., when the patient is asked to cough). This is to make the test applicable to those who may not be showing cough as a symptom yet but are already infected. We also gather cough samples from otherwise healthy individuals with no known medical condition, hereafter referred to as a normal cough. The normal cough is included in the analysis to see if it can be differentiated from the simulated cough produced by the COVID-19 patients. Using these data, we test the hypothesis using a variety of data analysis and pre-processing tools. Multiple alternative analysis approaches show that COVID-19 associated cough does have certain distinct features, at least when compared to pertussis, bronchitis and a normal cough.

- 6) Building on the insights from medical domain knowledge and cough data analysis, we develop an AI engine for preliminary diagnosis of COVID-19 from cough sounds. This engine runs on a cloud server with a front-end programmed as a simple user-friendly mobile app called AI4COVID-19. The app listens to cough when prompted, and then sends it to the AI engine wirelessly. The AI engine first runs a cough detection test to see if the recorded sound is a cough or not a cough. In case the sound is not a cough, it commands the app to indicate so. The cough detection part of the AI engine is designed to detect cough even in the presence of background noise. This is to make the app a useful screening tool even at public places such as airports and crowded shopping malls. If a cough is detected, it is passed on to the diagnosis part of the AI engine. After the AI engine completes the analysis, the app renders the result with three possible outcomes:

- COVID-19 likely.
- COVID-19 not likely.
- Test inconclusive.

- 7) To make the results as reliable as possible with the limited data available at the moment, we propose and implement a risk-averse architecture for the AI engine. It consists of three parallel classification solutions designed independently by three teams. The classifiers’ outcomes are consolidated by an automated mediator. In the current design, each classifier has veto power, i.e., if all three classifiers do not agree, the app returns ‘Test inconclusive’. This architecture employs the “2nd opinion” practice in medicine and reduces the rate of misdiagnosis, compared to stand alone classifiers with binary diagnosis, albeit at a cost of an increased rate of returning ‘Test inconclusive’ result.

2. Methodology

2.1. Hypothesis formulation and the devising a manageable validation strategy guided by relevant clinical findings

Our hypothesis in question is: “Cough sounds of COVID-19 patients contain unique enough latent features to be used as a diagnosis medium”. In this section, we describe our first principle-based approach that established the theoretical possibility of our hypothesis to be true. Then we describe the deep domain knowledge-based approach we take to reduce the amount of data required to test this hypothesis, thereby making this project feasible in a constrained time.

- 1) *Is COVID-19 cough unique enough to yield AI-based diagnosis?* Unfortunately, cough is a very common symptom of over a dozen medical conditions caused by either bacterial or viral respiratory infections not related to COVID-19 [47–49]. Several non-respiratory conditions can also cause cough. Table 1 summarizes the non-COVID-19 medical conditions which are known to cause cough. Theoretically, a cough based COVID-19 diagnosis, therefore, must take into account the cough sound data associated with all of the conditions listed in Table 1.

Trained physicians have been using cough sounds to perform a differential diagnosis among several respiratory conditions such as pneumonia, asthma, COPD, laryngitis and Tracheitis [49–54]. This is possible because in all these diseases the nature and location of the underlying irritant in the respiratory system is quite different leading to audibly distinct cough sounds. However, an unaided human ear is not capable of differentiating coughs caused by the conditions listed in Table 1. Even with AI, in case there are no unique latent features in the cough sound of COVID-19 patients, there is a risk for a cough-based AI diagnosis tool to confuse the cough caused by any of the diseases identified in Table 1 with the cough caused by COVID-19. A brute force-based approach to evaluate this risk would require gathering cough data from a large number of patients for each of the conditions listed in Table 1. This deluge of data can be

then used to train a powerful AI engine, such as very deep neural network to see if it can differentiate COVID-19 cough from those caused by all of the other medical conditions listed in Table 1. This approach is not practical at the moment given that the gathering such all-encompassing data will take too much time, rendering this approach of no help for the current pandemic.

To ensure that our developed solution works in practice with useful accuracy while being trainable with timely available data, we take another approach that we call domain-aware AI-design. Domain-aware here refers to the fact that the proposed AI engine does not solely rely on blind big data churning, e.g., through a deep neural network. Instead it relies on the deep domain knowledge of medical researchers trained in respiratory and infectious diseases to assess and narrow down the hypothesis testing scope, and to minimize the amount of data needed to test our hypothesis. By deep domain knowledge of medical researchers, we mean the use of medical knowledge of medical experts in this field to analyze pathomorphological changes caused by COVID-19 in the respiratory system and thus to evaluate the feasibility of an AI-based approach using cough-based analysis. It also means identifying the location of irritant in different types of coughs and using that information for smart feature extraction and faster training.

To this end, the medical researchers in our team began with an in-depth analysis of the pathomorphological changes caused by COVID-19 in the respiratory system by examining the data reported in numerous recent X-rays and CT-scans based studies of COVID-19 patients. The goal here is to see if the pathomorphological alterations caused by COVID-19 are distinct from that of other common medical conditions, particularly the ones identified in Table 1, that are well known to cause cough. If this turns out to be the case, then in cough caused by COVID-19 we should have latent features distinct from the cough caused by the other medical conditions. An appropriately designed AI should then be able to pick these cough feature idiosyncratic to COVID-19 infection and yield a reliable diagnosis, given enough labeled data. In the case of no such differences at pathomorphological level, the idea of cough based COVID-19 diagnosis should be dropped. In that case, any AI-based diagnosis yielded from cough is more likely to be a frivolous correlation and not a meaningful causal relationship. Such AI-based diagnosis will be an artifact of the training data rather than unique latent features of COVID-19 caused cough. Such a domain oblivious solution irrespective of its performance in lab will not be useful in practice.

2) Distinct pathomorphological alternations in respiratory system caused by COVID-19: In a recent study, it has been discovered that in COVID-19

infected people, there are distinct early pulmonary pathological signs even before the onset of the symptoms of COVID-19, such as dry cough, fever and some difficulty in breathing [55]. Early histological changes include evident alveolar damage with alveolar edema and proteinaceous exudates in alveolar spaces, with granules; inflammatory clusters with fibrinoid material and multinucleated giant cells; vascular congestion. Reactive alveolar epithelial hyperplasia and fibroblastic proliferation (fibroblast plugs) were indicative of early organization.

Contrary to the above observation of no early symptoms, it has also been noted that in some patients, COVID-19 leads to onset of pneumonia and pneumonia is marked by a peculiar cough [44]. However, pneumonia can also be caused by many other factors including non-COVID-19 viral or bacterial infections. Therefore, the question arises: is there a difference between COVID-19 caused pneumonia and other types of pneumonia that can be expected to translate into a difference in associated cough's latent features? Recent study in Ref. [56] shows that compared to non-COVID-19 related pneumonia, COVID-19 related pneumonia on chest CT scan was more likely to have a peripheral distribution (80% vs. 57%), ground-glass opacity (91% vs. 68%), vascular thickening (59% vs. 22%), reverse halo sign (11% vs. 9%) and less likely to have a central + peripheral distribution (14% vs. 35%), air bronchogram (14% vs. 23%), pleural thickening (15% vs. 33%), pleural effusion (4% vs. 39%) and lymphadenopathy (2.7% vs. 10.2%). Hence, these findings clearly suggest that cough sound signatures with COVID-19 caused pneumonia are likely to have some idiosyncrasies stemming from the distinct underlying pathomorphological alterations.

Moreover, CT scan-based studies also show that in the early stage of COVID-19 disease, it mainly manifests as an inflammatory infiltration restricted to the subpleural or peribronchovascular regions of one lung or both lungs, exhibiting patchy or segmental pure ground-glass opacities (GGOs) with vascular dilation. There is an increasing range of pure GGOs and the involvement of multiple lobes of the lung, consolidation of lesions, and crazy-paving patterns during the progressive stage. There are diffuse exudative lesions and lung "white-out" during an advanced stage [57]. Furthermore, AI-based analyses of X-ray [14–17] and CT scan [27,28] of the respiratory system have also shown to exploit the differences in pathomorphological alternations caused by COVID-19 to perform differential diagnosis among bacterial infection, non-COVID-19 viral infection and COVID-19 viral infection, with good accuracy. This further implies that COVID-19 affects the respiratory system in a fairly distinct way compared to other respiratory infections. Therefore, it is logical to hypothesize and investigate that the sound waves of cough produced by the COVID-19 infected respiratory system may also have distinct latent features.

The feasibility of diagnosing several common respiratory diseases using cough is not only supported by prior studies [58–60] but also in a recent clinically validated and widely publicized study [61]. In Ref. [61], a large team of researchers showed that cough alone can be used to diagnose asthma, pneumonia, bronchiolitis, croup and lower respiratory tract infections with over 80% sensitivity and specificity.

Recently, many machine learning teams around the world have started working on the idea of using cough sound for possible diagnosis of COVID-19, some interdependently and others inspired by our preliminary results in Ref. [46] and pre-print version of this work¹. However, to the best of authors' knowledge, this is the first work to propose and evaluate the feasibility of this idea, and develop and test the prototype of an AI engine powered mobile app based solution for anytime, anywhere tele-testing and pre-screening for COVID-19.

Table 1

Non-COVID-19 medical conditions that can cause cough.

RESPIRATORY	NON-RESPIRATORY
Upper respiratory tract infection (mostly viral infections)	Gastro-esophageal reflux
Lower respiratory tract infection (pneumonia, bronchitis, bronchiolitis)	Drugs (angiotensin converting enzyme inhibitors; beta blockers)
Upper airway cough syndrome	Laryngopharyngeal reflux
Pertussis, parapertussis	Somatic cough syndrome
Tuberculosis	Vocal cord dysfunction
Asthma and allergies	Obstructive sleep apnea
Early interstitial fibrosis, cystic fibrosis	Tic cough
Chronic obstructive pulmonary disease (emphysema, chronic bronchitis)	Smoking
Postnasal drip	Foreign body
Croup	Mediastinal tumor
Laryngitis	Air pollutants
Tracheitis	Tracheo-esophageal fistula
Lung abscess	Left-ventricular failure
Lung tumor	Congestive heart failure
Pleural diseases	Psychogenic cough
Interstitial lung disease	Idiopathic cough

2.2. Data description and practical viability of the solution with available data

As mentioned earlier, ideally cough data associated with all diseases listed in Table 1 is desirable for such a project. However, gathering such mammoth data is not possible in this time-constrained project, as the COVID-19 pandemic needs rapid response. To achieve meaningful results in the constrained time, we leverage domain knowledge, instead of just seeking big data. From Table 1, using the insights from Section 2.1, we shortlist cough causing infections that are most likely to confuse our AI engine due to similar pathomorphological changes in the respiratory system as of COVID-19 and, hence, similar cough signatures. The shortlist includes pertussis, bronchitis, asthma, pneumonia, bronchiolitis, croup and influenza. We further note that the prior study [61] has shown that cough associated with all of these seven medical conditions, except pertussis and bronchitis, have unique latent features. We use findings from this earlier study to reduce the scope of our data gathering campaign and differential analysis to only the respiratory diseases, the cough for which has not been analyzed before for having unique features, i.e., pertussis and bronchitis.

- 1) *Data used for training cough detector*: In order to make AI4COVID-19 app employable in a public place or where various background noises may exist (e.g., airport), we design and include a cough detector in our AI-Engine. This cough detector acts as a filter before the diagnosis engine and is capable to distinguish cough sound from 50 types of common environmental noises. To train and test this detector, we use the ESC-50 dataset [62] and the cough and non-cough sounds recorded from our own smartphone app. The ESC-50 dataset is a publicly available dataset that provides a huge collection of human and environmental sounds. This collection of sounds is categorized into 50 classes, one of these being cough sounds. We have used 1838 cough sounds and 3597 non-cough environmental sounds for training and testing of our cough detection system.
- 2) *Data used for training COVID-19 diagnosis engine*: To train our cough diagnosis system, we collected cough samples from COVID-19 patients as well as pertussis and bronchitis patients. We also collected normal coughs, i.e., cough sounds from healthy people. At the time of writing, we had access to 96 bronchitis, 130 pertussis, 70 COVID-19, and 247 normal cough samples from different people, to train and test our diagnosis system. Obviously, these are very small numbers of samples and more data is needed to make the solution more generalizable. New COVID-19 cough samples are arriving daily, and we are using these unseen samples to test the trained algorithm.
- 3) *Data pre-processing and visualization to evaluate the practical feasibility of AI4COVID-19*: In Section 2.1, by applying medical domain knowledge, we analyzed the theoretical viability of our hypothesis. However, in AI-based solutions, theoretical viability does not guarantee practical viability as the end outcome depends on the quantity and quality of the data, in addition to the sophistication of the machine learning algorithm used. Therefore, here we use the available cough data from the four classes, i.e., bronchitis, pertussis, COVID-19 and normal, to first evaluate the practical feasibility of a cough based COVID-19 diagnosis solution.

All audio files used in our study are in uncompressed PCM 16-bit format with a sampling rate of 44.1 kHz and a fixed 3-s length. We convert the cough audio samples for all four classes into the Mel scale for further processing. The Mel scale is a pitch categorization where listeners judge changes in pitch to be equal in distance from one another along this scale. It is meant to make changes in frequency, such as with a spectrogram, more closely reflect audible changes. We used the Mel spectrogram over a typical frequency spectrogram because the Mel scale in the Mel spectrogram has unequal spacing in the frequency bands and provides a higher resolution (more informative) in lower frequencies and vice versa, as compared to equally spaced frequency bands in normal spectrogram [63]. Since cough

sounds are known to have more energy in lower frequencies therefore, the Mel spectrogram is a naturally suitable representation for cough sounds. There are several methods for converting the frequency scale to the Mel. Here, we convert frequency f into Mel scale m as:

$$m = 2595 \times \log_{10} \left(1 + \frac{f}{700} \right) \quad (1)$$

We perform Cepstral analysis on the Mel spectrum of audio cough samples to compute their Cepstral coefficients, commonly known as Mel Frequency Cepstral Coefficients (MFCC) [64]. The extracted MFCC features for every sample result in an $M \times N$ matrix, where each column represents one signal frame and each row represents extracted MFCC features for a specific frame. The number of frames N can vary from sample to sample. There are several possible ways to use these extracted features for classification. In our approach, we extract two $M \times 1$ MFCC based feature vectors for each input cough sample and concatenate them into a single final $2M \times 1$ feature vector for that sample. For the first feature vector, we take the mean of MFCC features corresponding to all the frames. For the second feature vector, we take the top P $M \times 1$ Principle Component Analysis (PCA) projections [65] of the MFCC features across all the frames and combine them into a single $M \times 1$ vector by taking their magnitude. Finally, we concatenate both feature vectors into a single $2M \times 1$ feature vector. This approach is further illustrated in Fig. 5 in Section 2.3.

Since the features extracted from cough audio are multi-dimensional, in order to visualize the features, a nonlinear dimensionality reduction technique, t-distributed Stochastic Neighbor Embedding (t-SNE) [66] is applied, as it is well-suited for embedding high-dimensional data in a low-dimensional space of two-dimensions. In particular, this technique models each high-dimensional object by a two-dimensional point such that similar objects are modeled by nearby points and dissimilar objects are modeled by distant points with high probability. This visualization allows us to interpret the features in the form of clusters or classes with classification decision boundaries. Fig. 1 illustrates the 2-D visualization of these features for the four classes through t-SNE with classification decision boundaries/contours. It can be observed from the figure that different cough types possess features distinct from each other, and the features for COVID-19 are different from other cough types, such as bronchitis and pertussis. Hence, this observation suggests the practical viability of AI-powered cough based preliminary diagnosis for COVID-19 encouraging us to proceed towards an AI-engine design for maximum accuracy and efficient implementation to enable app-based deployment.

2.3. The AI4COVID-19 AI-Engine

In this section we explain the system architecture and the details of a two-stage solution that we developed for: 1) detection of cough sound from mixed cough, non-cough and noisy sounds; and 2) diagnosis of COVID-19 from the cough sound.

The training data is used to train different variants of deep learning and one classical machine learning algorithm as described in this section. After these models are trained, the pre-trained models for both cough detection and COVID-19 diagnosis are then implemented at the cloud server. The app then provides a user interface for using these pre-trained models. Another advantage of cloud-based implementation is the possibility of refining the model continuously as more data becomes available, as no update in the app is required for the refinement in the back-end AI-based diagnosis engine.

- 1) *System architecture*: The overall system architecture is illustrated in Fig. 2 and a flow chart highlighting the complete steps is shown in Fig. 3. The smartphone app records sound/cough when prompted by the press and release button. The recorded sounds are forwarded to the server when the diagnosis button is pressed. At the server, the sounds are first fed into the cough detector. In case, the sound is not detected as cough, the server commands the app to prompt so. In case, the sound is detected as a cough, the sound is forwarded to three parallel, different classifier systems, i.e., Deep Transfer Learning-based Multi Class classifier (DTL-MC), Classical Machine Learning-based Multi Class classifier (CML-MC) and Deep Transfer Learning-based Binary Class classifier (DTL-BC). The results of all these three classifiers are then passed on to a mediator. The app reports a diagnosis only if all three classifiers return identical classification results. If the classifiers do not agree, the app returns 'test inconclusive'. This tri-pronged mediator centered architecture is designed to minimize the probability of misdiagnosis. With this architecture, results show that AI4COVID-19 engine predicting 'COVID-19 likely' when the subject is not suffering from COVID-19 or vice-versa is extremely low when validated on the testing data available at the time of writing. The multi-pronged architecture is inspired by the "second opinion" practice in health care. The added caution here is that the three (diagnosis) opinions are solicited, each with veto power. How this architecture manages to reduce the overall misdiagnosis rate of the AI4COVID-19 despite the relatively higher misdiagnoses rate of individual classifiers is further explained in Section 3.3 through (4) and (5).

For app implementation in real-time, to ensure stricter quality control, we plan to run these pre-trained algorithms on at least three cough samples from the same patient and then make a preliminary diagnosis based on majority voting. Also, the cough detector, implemented before COVID-19 diagnosis (see Figs. 2 and 3) is ensuring some quality control by passing only those cough samples to the COVID-19 diagnosis engine that are of satisfactory quality. If samples are of poor quality, for example, a lot of background noise or the sound is too low, it rejects those samples by not detecting them as cough and therefore, not passing them on for diagnosis. In this case, the user is prompted to re-record the cough sample.

The details of detection and diagnosis classifiers are presented below.

- 2) *Cough detection*: The recorded cough sample is forwarded to our cloud-based server where the cough detector engine first computes its Mel-spectrogram (as explained in Section 2.2) with 128 Mel-components (bands). This image is then resized and converted into grayscale to unify the intensity scaling and reduce the image dimensions, resulting in a $320 \times 240 \times 1$ dimensional image. The resultant image is then fed into our Convolutional Neural Network (CNN) based classifier to decide whether the recorded input sound is of cough or not.

An overview of our used CNN structure is shown in Fig. 4. As the input Mel spectrogram image is of high dimensions, it's first passed through a 2×2 max-pooling layer to reduce the overall model complexity before proceeding. This is followed by two blocks of layers, each block comprising two convolutional layers followed by a 2×2 max pooling layer and a 0.15 dropout. Convolutional layers in first block use 16 filters and a 5×5 kernel size, whereas the second block uses 32 filters each in both convolutional layers. The learned complex features from these 4 convolutional layers are flattened and then passed to a fully connected layer of 256 neurons followed by a 0.30 dropout layer to prevent overfitting. Finally, the output layer with 2 neurons and a softmax activation function is used to classify between cough and not cough for the given input. ReLU is used as the activation function for all convolutional layers in this model, while Adam [67] is used as the optimizer due to its relatively better

efficiency and flexibility. A binary cross entropy loss function completes the detection model.

- 3) *COVID-19 diagnosis*: When the input sound is detected to be cough by the cough detection engine, it is forwarded to our tri-pronged mediator-centered AI engine to diagnose between COVID-19 and non-COVID-19 coughs. In order to produce results with maximum reliability, with the limited data available at the moment, the three classifiers used in the system use different approaches and are designed independently by three teams and cross-validated [68].

The three classification approaches are described below.

- a) *Deep Transfer Learning-based Multi Class classifier (DTL-MC)*: The first solution leverages a CNN-based four class classifier, using Mel spectrograms (described above) as input. The four classes here are cough caused by 1) COVID-19, 2) pertussis, 3) bronchitis or 4) normal person with no known respiratory infection. Similar CNN architecture used for cough detection is used here with a slight modification to make it a four class classifier (instead of binary classifier previously) by changing the number of neurons in output layer to four neurons for classifying the input between four possible output classes. Deep transfer learning [69] is used here to transfer the knowledge (features) learned by cough detection model (trained using relatively more data) to the similar diagnosis model. This allowed us to train a deep architecture (see Fig. 4) using limited amount of training data, as the basic features of the input Mel-spectrogram characterizing cough are already learned and only fine-tuning is required to learn more subtle features using new disease data. In this DTL-MC model, we froze the initial weights of the first convolutional layer, as the initial layers learn low-level latent features, and only allowed other layers to fine-tune their weights. This transfer learning approach allowed us to get better performance (reported in Section 3) than training on disease data from scratch.
- b) *Classical Machine Learning-based Multi Class classifier (CML-MC)*: A second parallel diagnosis test uses classic machine learning instead of deep learning. This to mitigate the over-fitting that may still be happening in the deep learning-based classifier due to small amount of training data. To maximize independence among the classifiers that together constitute the AI diagnosis engine, the 2nd classifier begins with a different pre-processing of cough sounds. Instead of using a spectrogram like the first classifier, it uses MFCC and PCA based feature extraction as explained in Section 2.2. These smart features are then fed into a multi-class support vector machine (SVM) for classification. Class balance is achieved by sampling from each class randomly such that the number of samples equals to the number of minority class samples, i.e., class with the lowest number of samples. Using the concatenated feature matrix (of mean MFCC and top few PCAs) as input, we perform SVM with k-fold validation for 100,000 iterations. This approach is illustrated in Fig. 5.
- c) *Deep Transfer Learning-based Binary Class classifier (DTL-BC)*: The third parallel diagnosis test also uses deep transfer learning based CNN on the Mel spectrogram image of the input cough samples, similar to the first branch of the AI engine, but performs only binary classification of the same input, i.e., is the cough associated COVID-19 or not. The CNN structure used for this technique is similar to the one used for the cough detector (see Fig. 4).

3. Results

In order to evaluate the model we use the performance metrics of *accuracy, specificity, sensitivity/recall, precision, F1-score* on validation set and also cross-validate the models. The *accuracy* here refers to the overall accuracy of the model. We use k-fold cross validation methodology, that is well-suited to evaluate the performance of machine learning models on limited data [68]. These performance metrics are based on mean confusion matrices from cross-validation. In addition, we

		Predicted class	
		Cough	No cough
True class	Cough	96.01	3.99
	No cough	4.81	95.19

Fig. 6. Normalized mean confusion matrix for cough detection (in percentage) using 5-fold cross validation.

have used regularization techniques to prevent the problem of over-fitting, for example, we tune the regularization parameter of SVM against the cross-validation accuracy and choose those parameters that gave us the best generalizability of the models. Tuning of the various hyper-parameters (number of hidden layers, learning rate, activation functions, dropout rate) of deep neural network-based models has also been performed, based on the cross-validation accuracy. Furthermore, the decay of model loss versus the number of epochs has been investigated to rule out the possibility of over-fitting.

3.1. Cough detection

The confusion matrix and performance metrics for detection algorithm are reported in Fig. 6 and Table 2, respectively. Results demonstrate that our cough detection algorithm can classify between cough event and no cough event with an overall accuracy of 95.60%.

The error graph of mean loss versus epochs of this neural network based model, for both training and validation data sets is shown in Fig. 7. The decay of both the training and testing curve shows that this model has not been over-fitted.

3.2. COVID-19 diagnosis

The performance metrics for the first classifier, that is DTL-MC classifier are reported in Table 3. At the moment, with limited data available, the overall accuracy of deep transfer learning based multi-class classifier is 92.64%. The mean normalized confusion matrix resulting from this approach is shown in Fig. 8. Future work will continue to improve this model as more training data becomes available for CNN. Fig. 9 shows the mean loss versus epochs of the DTL-MC classifier, for both training and validation data sets. Both the curves start to saturate after around 25 epochs, indicating a reasonable learning time, without over-fitting.

For the second classifier, i.e., CML-MC classifier, the normalized mean confusion using 5-fold cross validation is shown in Fig. 10 and the CDF of overall accuracy with varying k's in k-fold cross validation is shown in Fig. 11. Table 4 reports the performance metrics for this approach, utilizing data available at this moment. Results indicate an overall accuracy of 88.76%.

Performance metrics for the third approach, that is DTL-BC are

Table 2
Performance metrics for cough detection.

F1-Score (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
95.61	96.01	95.19	95.22	95.60

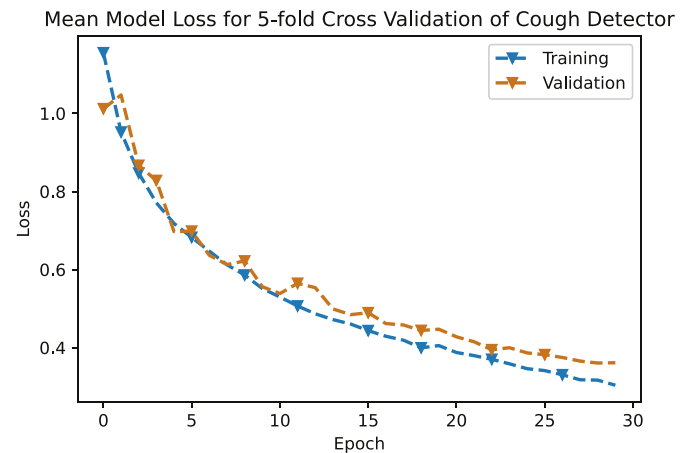


Fig. 7. Mean model loss for 5-fold cross validation of cough detector.

reported in Table 5, with the normalized mean confusion matrix shown in Fig. 12. The classification accuracy with this approach is 92.85%. The loss versus number of epochs for both training and validation is illustrated in Fig. 13. Here, both the curves start to level off after [20] epochs, hence depicting a reasonable training time, while avoiding over-fitting. Currently, the number of non-COVID cough samples are much larger than COVID-19 cough samples when binary classification is chosen. Once more data becomes available, the current classification accuracy using DTL-BC is likely to increase.

The performance of the two deep learning-based classifiers (DTL-MC and DTL-BC) is superior than the manual feature extraction based classic machine learning classifier (CML-MC). This is expected, because with shortage of training data circumvented via transfer learning, considerable amount of training data and automatic feature extraction capability of the deep neural network are expected to extract even more subtle distinct features hidden in the data than the manual feature extraction used in the second classifier, i.e., CML-MC.

3.3. Overall performance under independence assumption

After analyzing the performance of the three different classifiers, we now analyze the overall performance of AI4COVID-19 AI engine that utilizes a mediator-based architecture. This architecture will yield optimal performance when its prongs (i.e., the classifiers) are fully independent.

The independence of the three classifiers depends on dependence among the training data fed into these classifiers, as well as the similarity among the classifier's internal architectures. Therefore, in reality, the classifiers will never be truly independent, because of following key reasons: (i) Even if we use unique training data for each classifier, there will be some dependence (e.g., correlation introduced by age group, gender, native language etc). (ii) Even if we manage to choose fully independent training data for each classifier, the similarities in the architectures of classifiers would introduce some degree of dependence.

However, lacking absolute independence does not completely

Table 3
Performance metrics for DTL-MC.

	F1-Score (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
Overall	–	–	–	–	92.64
COVID-19	89.52	89.14	96.67	89.91	–
Pertussis	94.04	93.57	98.19	94.51	–
Bronchitis	91.63	93.86	96.33	89.50	–
Normal	95.43	94.00	99.00	96.90	–

		Predicted class			
		Normal	Bronchitis	Pertussis	COVID-19
True class	Normal	94.00	1.86	0.57	3.57
	Bronchitis	0.28	93.86	1.86	4.00
	Pertussis	0.43	3.57	93.57	2.43
	COVID-19	2.29	5.57	3.00	89.14

Fig. 8. Normalized mean confusion matrix for cough diagnosis (in percentage) for DTL-MC using 5-fold cross validation.

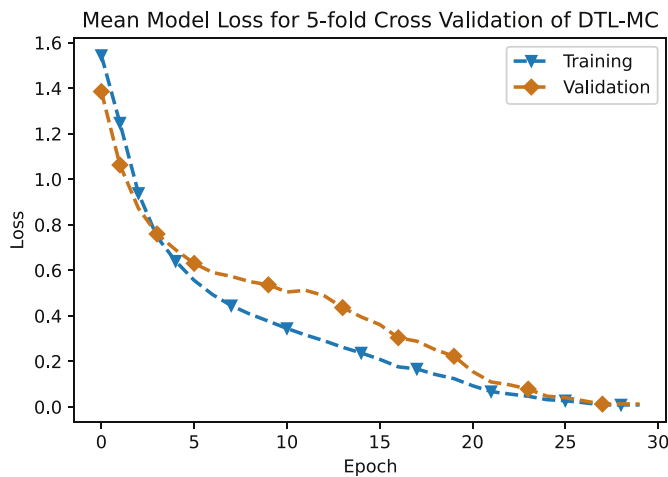


Fig. 9. Mean model loss for 5-fold cross validation of DTL-MC.

eliminate the advantages of proposed multi-pronged architecture. This is similar to a scenario when a second diagnosis sought from a physician, who has the same speciality, reads same medical literature, and has correlated neuroanatomy as the first physician, is considered an independent opinion for all practical purposes and is known to reduce misdiagnosis rates, though in strict theoretical sense it is not fully

		Predicted class			
		Normal	Bronchitis	Pertussis	COVID-19
True class	Normal	84.40	3.92	4.29	7.39
	Bronchitis	1.74	91.29	5.33	1.64
	Pertussis	2.65	4.55	87.64	5.16
	COVID-19	7.27	0.99	0.03	91.71

Fig. 10. Normalized mean confusion matrix for cough diagnosis (in percentage) for CML-MC using 5-fold cross validation.

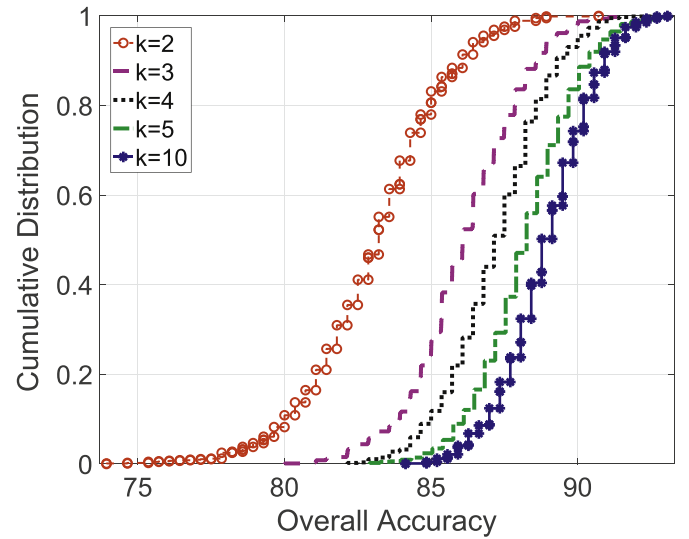


Fig. 11. Overall accuracy CDF for varying k-fold experiments in CML-MC approach.

Table 4

Performance metrics for CML-MC.

	F1-Score (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
Overall	–	–	–	–	88.76
COVID-19	89.08	91.71	95.27	86.60	–
Pertussis	88.84	87.64	96.78	90.08	–
Bronchitis	90.94	91.29	96.84	90.61	–
Normal	86.09	84.40	96.11	87.86	–

independent diagnosis.

Acknowledging that the three classifiers are not fully independent but they will become almost independent by using unique training data when more COVID-19 cough data becomes available, in the following, we analyze the performance of overall AI architecture under the independence assumption. This is to compare the misdiagnosis rate of individual classifier decisions versus the mediator's decision.

Let k_1, k_2, k_3 be the predicted class labels for the three classifiers, DTL-MC, CML-MC and DTL-BC, respectively and k_f be the predicted diagnosis result of the app. The possible values that k_f can take are 'COVID-19 likely' (C), 'COVID-19 not likely' (C') and 'test inconclusive' (I). Then, the probability that the app predicts 'COVID-19 likely', when the patient actually has COVID-19, can be calculated as:

$$P(k_f = C|C) = P(k_1 = C|C) \cdot P(k_2 = C|C) \cdot P(k_3 = C|C) \quad (2)$$

$$P(k_3 = C|C) = 0.891 \cdot 0.917 \cdot 0.946 = 0.773$$

The probability that the app predicts 'COVID not likely' when the subject actually does not have COVID-19 can be represented as:

$$P(k_f = C'|C') = P(k_1 = C'|C') \cdot P(k_2 = C'|C') \cdot P(k_3 = C'|C') \\ = 0.966 \cdot 0.952 \cdot 0.911 = 0.838 \quad (3)$$

The app can also predict 'COVID-19 likely' when the subject is not suffering from COVID-19 or vice-versa. In these cases, we can write the

Table 5

Performance metrics for DTL-BC.

F1-Score (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Accuracy (%)
92.97	94.57	91.14	91.43	92.85

		Predicted class	
		COVID-19	No COVID-19
True class	COVID-19	94.57	5.43
	No COVID-19	8.86	91.14

Fig. 12. Normalized mean confusion matrix for cough diagnosis (in percentage) for DTL-BC using 5-fold cross validation.

probabilities as:

$$P(k_f = C|C^*) = P(k_1 = C|C^*) \cdot P(k_2 = C|C^*) \cdot P(k_3 = C|C^*) \\ = 0.033 \cdot 0.047 \cdot 0.088 = 1.365 \times 10^{-4} \quad (4)$$

$$P(k_f = C^*|C) = P(k_1 = C^*|C) \cdot P(k_2 = C^*|C) \cdot P(k_3 = C^*|C) \\ = 0.108 \cdot 0.082 \cdot 0.054 = 4.782 \times 10^{-4} \quad (5)$$

Equations (4) and (5) signify the importance of the mediator in our proposed architecture and show how this risk-averse architecture is able to reduce the overall misdiagnosis rate of AI4COVID-19. From (4) and (5), both the false negative as well as the false positive rate of the overall architecture are near zero. Note that none of the classifiers have near zero misdiagnosis rate simultaneously for both healthy and COVID-19 cases. For a given classifier low False Positive Rate (FPR) is at the cost of high False Negative Rate (FNR) and vice versa. Mediator counters the over sensitivity or under sensitivity of the individual classifiers by masking it with the ‘Test inconclusive’ result. i.e., from (f), the lowest false positive rate of DTL-MC classifier is the most contributing factor in the near-zero probability the app will predict ‘COVID-19 likely’ when the subject is not suffering from COVID-19. The most contributing factor in the near-zero probability that the app will predict ‘COVID-19 not likely’ when the subject is actually suffering from COVID-19, is the lowest false negative rate of DTL-BC classifier, as observed from (5). In other words, the mediator in AI4COVID-19 architecture complements the weakness of one classifier with the strength of other and vice versa, resulting in reduced misdiagnosis rate as compared to using these classifiers independently, i.e., without the proposed mediator.

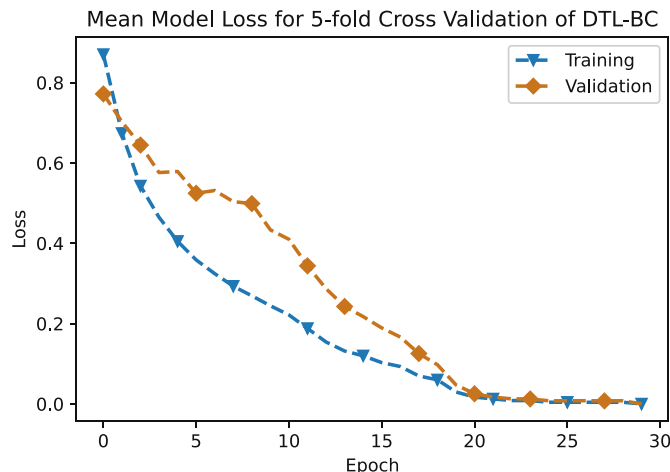


Fig. 13. Mean model loss for 5-fold cross validation of DTL-BC.

Table 6

The overall current performance of AI4COVID-19 AI engine.

Event	Probability
App reports ‘COVID-19 likely’ when the subject actually has COVID-19	0.773d ₁
App reports ‘COVID-19 likely’ when the subject actually does not have COVID-19	1.365 × 10 ⁻⁴ d ₂
App reports ‘COVID-19 not likely’ when the subject actually does not have COVID-19	0.838d ₃
App reports ‘COVID-19 not likely’ when the subject actually has COVID-19	4.782 × 10 ⁻⁴ d ₄
App reports ‘test inconclusive’ when the subject actually has COVID-19	0.226d ₅
App reports ‘test inconclusive’ when the subject actually does not have COVID-19	0.161d ₆

In the cases where the reports ‘Test inconclusive’, the test subject can either have COVID-19 or not, in reality. The respective probabilities for those cases are:

$$P(k_f = I|C) = 1 - [P(k_f = C|C) + P(k_f = C^*|C)] \\ = 1 - [0.773 + 4.782 \times 10^{-4}] = 0.226 \quad (6)$$

$$P(k_f = I|C^*) = 1 - [P(k_f = C|C^*) + P(k_f = C^*|C^*)] \\ = 1 - [1.365 \times 10^{-4} + 0.838] = 0.161 \quad (7)$$

Currently, the app would predict an inconclusive test result 38.7% of the time ($P(k_f = I) = P(k_f = I|C) + P(k_f = I|C^*)$). This percentage can be reduced by switching to a mediation scheme where app result reflects simple or weighted majority of the N number of classifiers. This scheme will be explored once more data becomes available. The results are summarized in Table 6. The numbers here just indicate how including the proposed mediator-based architecture may reduce the misdiagnosis rate compared to using individual classifiers. These probabilities are under independence assumption and can change depending on the degree of dependence between the training data and architectures of the individual classifiers, as explained earlier. We can capture this dependency factor by introducing a co-efficient, d_i in each of the above six calculated probabilities, where $i = 1 \dots 6$. The values of d_i ’s can be estimated empirically once more data becomes available in the future and can in turn be used to determine the weights to be assigned to each classifier in weighted average based mediator design.

4. Discussion

4.1. Potential utilities of the AI4COVID-19

The AI4COVID-19 app based on preliminary diagnosis is not meant to replace or compete with the medical grade testing by any means. Instead, the proposed solution offers the following complementing use cases to control the pandemic.

- 1) Enabling tele-screening for anyone, anywhere, anytime.
- 2) Addressing the shortage of testing facilities. This is particularly useful in remote areas of the world where medics have no option but to rely on phone based or questionnaire based tele-screening. In such places, the app can act as a clinical decision assistance tool.
- 3) Opportunity to protect medics from unnecessary exposure, particularly for non-critical patients where the medical advice for whom anyway would be ‘stay at home’ or ‘self-isolate’ to wait for self-healing.
- 4) Minimizing covert spread that happens to be the biggest problem.
- 5) Tracing and monitoring the spread. This is particularly easy with AI4COVID-19 as the cough samples can be spatio-temporally tagged anonymously, without having to compromise the patient’s privacy.
- 6) AI4COVID-19 can be used as a low cost screening tool, instead of or in addition to the temperature scanner at the airports, borders or

elsewhere as needed. This is possible because our tests show that the app can diagnose COVID-19 even in a non-spontaneous cough of COVID-19 positive people. The cost of using such an app-based solution would be significantly low, since it can be readily installed on any existing smartphone using the existing internet connections, by a large number of people simultaneously.

- 7) The app can help in enabling and maintaining informed social distancing and self-isolation.
- 8) By default the app can provide centralized record of tests with spatial and temporal stamps. Thus, the data gathered from the app can be used for long term planning of medical care and policy making.

4.2. Comparison and contrast of AI4COVID-19 with existing studies

Existing methods to screen COVID-19 patients include Nucleic Acid Amplification Tests (NAAT), such as real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR). While far more sensitive than proposed method, these methods are marked by limitations identified in Section 1.1 that includes limited geographical and temporal availability, high cost, large turnaround time, requirement of in-person visits to hospitals or mobile labs and the need and shortage of protective equipment. In contrast, AI4COVID-19 is useable anywhere, anytime for anyone.

Recent AI-based studies towards COVID-19 preliminary diagnosis include the use of either X-ray [14–17] or CT Scan [27–29]. These methods demonstrate comparable or higher sensitivities, ranging from 72% to 96%, compared to proposed approach. However, both of these approaches still require a visit to a well-equipped clinical facilities and does not meet the utilities identified in Section 4.1. In contrast, AI4COVID-19 is the only screening method proposed in the literature so far that can be used in-situ and eliminates the need for an in-person visit to the testing facility or getting out of homes or places of self-isolation, thereby meeting all use cases identified in Section 4.1.

4.3. Key limitations of current version of AI4COVID-19

At the time of writing, the performance of AI4COVID-19 app is limited by the following factors:

- 1) The quantity of the training and testing data. Due to time constraints and difficulty of getting cough data, we could gather data only from a small number of patients for each of the four groups. We tried to minimize the impact of this limitation by combining data hungry approaches that are capable of extracting more hidden features i.e., deep learning, with the ML approaches that can work with a small amount of data through manual feature extraction. The shortage of training data was also to some extent circumvented by using transfer learning in the deep learning based classifiers. Still, the need for more data cannot be overemphasized.
- 2) The quality of the training and testing data: We have strived to ensure that the data is correctly labeled. However, any error in the labeling of the data that managed to slip through our scrutiny is likely to impact reported performance. Such impact can be particularly pronounced when the data is not that big in the first place.
- 3) Our in-depth medical differential analysis suggested that COVID-19 associated pathomorphological alternations are fairly distinct, and hence cough of COVID-19 patients is likely to have at least some distinct latent features. However, this does not guarantee the absence of overlap in COVID-19 cough features and those of diseases not included in the training and testing. The approach we used to combat this issue is the clever mediator-based architecture that practically eliminates misdiagnosis by declaring test to be inconclusive if the cough samples are even slightly confusing i.e., lying very close to decision boundaries. Still, we are working to address this limitation in future releases of AI4COVID-19 by incorporating cough associated with other non-COVID-19 medical conditions identified in Table 1 as

well as including other dimensions such as age, gender, smoking or non-smoking status and certain bio markers.

- 4) Large scale trial-based validation to test the generalization capability: In the end, the only way to evaluate the generalization capability and practical performance of the proposed AI4COVID-19 based testing is a large scale medically supervised validation in real world. The findings of this paper provide promising enough preliminary results and proof of concept to encourage first systematic large-scale cough data gathering campaigns followed by large scale trials. Once the testing of the prototype app on a much larger data set is completed, the provision of automatic updates will also be enabled.
- 5) In the current prototype design, all AI processing happens at the cloud. The app is just a thin client that records and sends the audio data to the server where the AI engine resides. Due to low complexity, the app does not have stringent CPU and RAM requirements and it can run on most smartphones. This cloud-based design allows the screening to be done not only via commodity smartphones but also via a web portal link accessible in any browser. In the future, to enable offline screening using an edge device such as smart phone, we plan to investigate edge-based implementation of the modified lightweight version of the proposed AI-engine. This will be done by edge AI techniques such as distilled deep learning. The potential of distilled deep learning for enabling edge device based medical diagnoses has been verified in our recent work [70].

4.4. Planned future upgrades of AI4COVID-19

AI4COVID-19 accuracy can be improved by incorporating other acoustic data such as breathing sound and speech. Moreover, for higher accuracy and better generalization across larger populations, we also plan to investigate the impact of incorporating meta-data such as age, gender, smoking, non-smoking, ethnicity and medical history. The accuracy is also likely to improve by including multi-sensory data instead of relying on only acoustic data and meta-data. For example, recent studies show that in a small fraction of COVID-19 patients, cutaneous anomalies are part of the symptoms [71]. Therefore, including skin images in addition to acoustic data may help improve the diagnosis. Another planned upgrade is the inclusion of bio-markers that can be measured by wearable sensors such as wristbands, rings and skin patches or ambient sensors such as infrared cameras or wireless sensors, which can also lead to more reliable results. The examples of bio-markers that are worthy of investigation that can be easily collected via aforementioned wearable or ambient sensors include respiration rate, temperature, blood oxygen saturation, pulse rate, heart rate variability, resting heart rate, blood pressure, mean arterial pressure, stroke volume, sweat level, systematic vesicular resistance, cardiac output, pulse pressure and cardiac index.

5. Conclusion

Scarcity, cost and long turnaround time of clinical testing are key factors behind covert rapid spread of the COVID-19 pandemic. Motivated by the urgent need, this paper presents a ubiquitously deployable AI-based preliminary diagnosis tool for COVID-19 using cough sound via a mobile app. The core idea of the tool is inspired by our independent prior studies that show cough can be used as a test medium for diagnosis of a variety of respiratory diseases using AI. To see if this idea is extendable to COVID-19, we perform in-depth differential analysis of the pathomorphological alternations caused by COVID-19 relative to other cough causing medical conditions. We note that the way COVID-19 affects the respiratory system is substantially unique and hence, cough associated with it is likely to have unique latent features as well. We validate the idea further by the visualization of latent features in cough of COVID-19 patients and two common infections, pertussis and bronchitis as well as non-infectious coughs. Building on the insights from the medical domain knowledge, we propose and develop a tri-

pronged mediator centered AI-engine for the cough-based diagnosis of COVID-19, named AI4COVID-19. The results show that the AI4COVID-19 app is able to diagnose COVID-19 with negligible misdiagnosis probability thanks to its risk-avert architecture.

Despite its impressive performance, AI4COVID-19 is not meant to compete with clinical testing. Instead, it offers a unique functional tool for timely, cost-effective and most importantly safe monitoring, tracing, tracking and thus, controlling the rampant spread of the global pandemic by virtually enabling testing for everyone. While we are working on improving the AI4COVID-19, this paper is meant to present a proof of concept to encourage community support for more labeled data followed by large scale trials. We hope that the AI4COVID-19 app can be leveraged to pre-screen for COVID-19 at a population scale, particularly in regions around the world where the pandemic is spreading covertly due to the lack of testing. The AI4COVID-19 enabled tele-screening can alleviate the crushing burden on the overwhelmed medical systems around the world and help save countless lives.

Ethical statement

“Nothing declared”.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is dedicated to those affected by the COVID-19 pandemic and those who are helping to fight this battle in anyway they can.

References

- [1] World Health Organization. Coronavirus disease (COVID-19) outbreak situation [Online]. Available, <https://www.who.int/emergencies/diseases/novel-coronavirus-us-2019>. [Accessed 28 April 2020].
- [2] BBC. Coronavirus in South Korea: how ‘trace, test and treat’ may be saving lives. Accessed on: Mar. 31, 2020. [Online]. Available, <https://www.bbc.com/news/world-asia-51836898>; 2020.
- [3] Science. Not wearing masks to protect against coronavirus is a ‘big mistake,’ top Chinese scientist says. Accessed on: April. 1, 2020. [Online]. Available, <https://www.sciencemag.org/news/2020/03/not-wearing-masks-protect-against-coronavirus-big-mistake-top-chinese-scientist-says>; 2020.
- [4] Cascella M, Rajnik M, Cuomo A, Dulebohn SC, Di Napoli R. “Features, evaluation and treatment coronavirus (COVID-19),” in *StatPearls [Internet]*. StatPearls Publishing; 2020.
- [5] Van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. “Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *N Engl J Med* 2020;382(16):1564–7.
- [6] Tribune. Coronavirus test results in Texas are taking up to 10 days. Accessed on: Mar. 31, 2020. [Online]. Available, https://tylerpaper.com/covid-19/coronavirus-test-results-in-texas-are-taking-up-to-10-days/article_5ad6c9cc-4fa9-573a-bb4d-ada1b97acf42.html; 2020.
- [7] Post Washington. Hospitals are overwhelmed because of the coronavirus. Accessed on: Mar. 31, 2020. [Online]. Available, <https://www.washingtonpost.com/opinions/2020/03/15/hospitals-are-overwhelmed-because-coronavirus-heres-how-help/>; 2020.
- [8] CNN. FDA authorizes 15-minute coronavirus test. Accessed on: Mar. 31, 2020. [Online]. Available, <https://www.cnn.com/2020/03/27/us/15-minute-coronavirus-test/index.html>; 2020.
- [9] Abbott. Detect COVID-19 in as little as 5 minutes. Accessed on: May. 30, 2020. [Online]. Available, <https://www.abbott.com/corpnnewsroom/product-and-innovation/detect-covid-19-in-as-little-as-5-minutes.html>; 2020.
- [10] STAT. FDA says Abbott’s 5-minute COVID-19 test may miss infected patients [Online]. Available, <https://www.statnews.com/2020/05/15/fda-says-abbotts-5-minute-covid-19-test-may-miss-infected-patients/>. [Accessed 30 May 2020].
- [11] The New York Times. The latest obstacle to getting tested? A shortage of swabs and face masks. Accessed on: Mar. 31, 2020. [Online]. Available, <https://www.nytimes.com/2020/03/18/health/coronavirus-test-shortages-face-masks-swabs.html>; 2020.
- [12] Post Washington. Shortages of face masks, swabs and basic supplies pose a new challenge to coronavirus testing. Accessed on: Mar. 31, 2020. [Online]. Available, <https://www.washingtonpost.com/climate-environment/2020/03/18/shortages-face-masks-cotton-swabs-basic-supplies-pose-new-challenge-coronavirus-testing/>; 2020.
- [13] A FD. Coronavirus (COVID-19) update: FDA authorizes first diagnostic test using at-home collection of saliva specimens [Online]. Available, <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-diagnostic-test-using-home-collection-saliva>. [Accessed 30 May 2020].
- [14] Wang L, Wong A. “COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images,”. 2020. *arXiv preprint arXiv:2003.09871*vol. 1.
- [15] Zhang I, Xie Y, Li Y, Shen C, Xia Y. “COVID-19 screening on chest X-ray images using deep learning based anomaly detection,”. 2020. *arXiv preprint arXiv:2003.12338*.
- [16] Narin A, Kaya C, Pamuk Z. “Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks,”. 2020. *arXiv preprint arXiv:2003.10849*.
- [17] Hemdan EE-D, Shouman MA, Karar ME. COVIDX-Net: a framework of deep learning classifiers to diagnose COVID-19 in x-ray images,”. 2020. *arXiv preprint arXiv:2003.11055*.
- [18] Deshpande G, Schuller B. “An overview on audio, signal, speech, & language processing for COVID-19,”. 2020. *arXiv preprint arXiv:2005.08579*.
- [19] Wong HYF, Lam HYS, Fong AH-T, Leung ST, Chin TW-Y, Lo CSY, Lui MM-S, Lee JCY, Chiu KW-H, Chung T, et al. “Frequency and distribution of chest radiographic findings in COVID-19 positive patients.” *Radiology*; 2020. p. 201160.
- [20] Apostolopoulos ID, Mpesiana TA. COVID-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* 2020;1.
- [21] Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A, “COVID-CAPS. A capsule network-based framework for identification of COVID-19 cases from X-ray images,”. 2020. *arXiv preprint arXiv:2004.02696*.
- [22] Pham Q-V, Nguyen DC, Hwang W-J, Pathirana PN, et al. “Artificial intelligence (AI) and big data for coronavirus (COVID-19) pandemic.”: A Survey on the State-of-the-Arts; 2020.
- [23] Li X, Li C, Zhu D. COVID-MobileXpert: on-device COVID-19 screening using snapshots of chest X-ray. 2020. <https://arxiv.org/pdf/2004.03042v2.pdf>.
- [24] N. Tsiknakis, E. Trivizakis, E. E. Vassalou, G. Z. Papadakis, D. A. Spandidos, A. Tsatsakis, J. Sánchez-García, R. López-González, N. Papanikolaou, A. H. Karantanias et al., “Interpretable artificial intelligence framework for COVID-19 screening on chest X-rays,” *Experimental and Therapeutic Medicine*.
- [25] Ahammed K, Satu MS, Abedin MZ, Rahaman MA, Islam SMS. “Early detection of coronavirus cases using chest X-ray images employing machine learning and deep learning approaches.” *medRxiv*; 2020.
- [26] Albahli S. “Efficient GAN-based Chest Radiographs (CXR) augmentation to diagnose coronavirus disease pneumonia. *Int J Med Sci* 2020;17(10):1439–48.
- [27] Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Chen Y, Su J, Lang G, et al. “Deep learning system to screen coronavirus disease 2019 pneumonia,”. 2020. *arXiv preprint arXiv:2002.09334*.
- [28] Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, et al. “Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest ct.” *Radiology*; 2020. 200905.
- [29] Zhao W, Zhong Z, Xie X, Yu Q, Liu J. “Relation between chest ct findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study.” *American Journal of Roentgenology* 2020;1–6.
- [30] Han Z, Wei B, Hong Y, Li T, Cong J, Zhu X, Wei H, Zhang W. “Accurate screening of covid-19 using attention based deep 3d multiple instance learning,”. *IEEE Transactions on Medical Imaging*; 2020.
- [31] Li Y, Xia L. “Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management.” *American Journal of Roentgenology* 2020;214(6):1280–6.
- [32] Ye Z, Zhang Y, Wang Y, Huang Z, Song B. “Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review,”. *European Radiology*; 2020. p. 1–9.
- [33] Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, Bai J, Lu Y, Fang Z, Song Q, et al. “Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT.” *Radiology*; 2020. 200905.
- [34] Adams HJ, Kwee TC, Kwee RM. “COVID-19 and chest CT: do not put the sensitivity value in the isolation room and look beyond the numbers. *Radiology* 2020. 201709.
- [35] Gietema HA, Zelis N, Nobel JM, Lambriks LJ, van Alphen LB, Lashof AMO, Wildberger JE, Nelissen IC, Stassen PM. “CT in relation to RT-PCR in diagnosing COVID-19 in The Netherlands: a prospective study.” *medRxiv*; 2020.
- [36] Thorpe W, Kurver M, King G, Salome C. “Acoustic analysis of cough,” in *the IEEE seventh Australian and New Zealand intelligent information systems conference*. 2001. p. 391–4.
- [37] Chatzarrin H, Arcelus A, Goubran R, Knoefel F. “Feature extraction for the differentiation of dry and wet cough sounds. In: *IEEE international Symposium on medical Measurements and applications*; 2011. p. 162–6.
- [38] Song I. “Diagnosis of pneumonia from sounds collected using low cost cell phones. In: *international joint Conference on neural networks. IJCNN*; 2015. p. 1–8.
- [39] Infante C, Chamberlain D, Fletcher R, Thorat Y, Kodgule R. “Use of cough sounds for diagnosis and screening of pulmonary disease. In: *2017 IEEE global humanitarian technology conference. GHTC*; 2017. p. 1–10.
- [40] You M, Wang H, Liu Z, Chen C, Liu J, Xu X-H, Qiu Z-M. “Novel feature extraction method for cough detection using NMF. *IET Signal Process* 2017;11(5):515–20.
- [41] Pramon RXA, Imtiaz SA, Rodriguez-Villegas E. “Automatic cough detection in acoustic signal using spectral features. In: *2019 41st annual international*

- Conference of the IEEE Engineering in Medicine and biology society. EMBC; 2019. p. 7153–6.
- [42] Miranda ID, Diacon AH, Niesler TR. "A comparative study of features for acoustic cough detection using deep architectures. In: 41st annual international Conference of the IEEE Engineering in Medicine and biology society. EMBC; 2019. p. 2601–5.
- [43] Soliński M, Lepek M, Kołtowski Ł. "Automatic cough detection based on airflow signals for portable spirometry system. In: Informatics in medicine unlocked; 2020. p. 100313.
- [44] World Health Organization. "Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). 2020. 2020.
- [45] National Institute of Health. New coronavirus stable for hours on surfaces. Accessed on: Mar. 31, 2020. [Online]. Available, <https://www.nih.gov/news-events/news-releases/new-coronavirus-stable-hours-surfaces>; 2020.
- [46] Bales C, Nabeel M, John CN, Masood U, Qureshi HN, Farooq H, et al. "Can machine learning Be used to recognize and diagnose coughs?". 2020. arXiv:2004.01495, <https://arxiv.org/pdf/2004.01495.pdf>.
- [47] Irwin RS, Madison JM. "The diagnosis and treatment of cough. N Engl J Med 2000; 343(23):1715–21.
- [48] Chang AB, Landau LI, Van Asperen PP, Glasgow NJ, Robertson CF, Marchant JM, Mellis CM. "Cough in children: definitions and clinical evaluation. ' Medical Journal of Australia 2006;184(8):398–403.
- [49] Gibson PG, Chang AB, Glasgow NJ, Holmes PW, Kemp AS, Katelaris P, Landau LI, Mazzone S, Newcombe P, Van Asperen P, et al. "Cicada: cough in children and adults: diagnosis and assessment. australian cough guidelines summary statement. ' Medical Journal of Australia 2010;192(5):265–71.
- [50] Irwin RS, French CL, Chang AB, Altman KW, Adams TM, Azoulay E, Barker AF, Birring SS, Blackhall F, Bolser DC, et al. "Classification of cough as a symptom in adults and management algorithms: chest guideline and expert panel report. Chest 2018;153(1):196–209.
- [51] Korpás J, Sadloňová J, Vrabec M. "Analysis of the cough sound: an overview. Pulm Pharmacol 1996;9(5-6):261–8.
- [52] Diehr P, Wood RW, Bushyhead J, Krueger L, Wolcott B, Tompkins RK. "Prediction of pneumonia in outpatients with acute cough—a statistical approach. J Chron Dis 1984;37(3):215–25.
- [53] Irwin RS, Baumann MH, Bolser DC, Boulet L-P, Braman SS, Brightling CE, Brown KK, Canning BJ, Chang AB, Diczpinigaitis PV, et al. "Diagnosis and management of cough executive summary: acrp evidence-based clinical practice guidelines. Chest 2006;129(1).
- [54] Heckerling PS, Tape TG, Wigton RS, Hissong KK, Leikin JB, Ornato JP, Cameron JL, Racht EM. "Clinical prediction rule for pulmonary infiltrates. Ann Intern Med 1990;113(9):664–70.
- [55] Tian S, Hu W, Niu L, Liu H, Xu H, Xiao S-Y. "Pulmonary pathology of early phase 2019 novel coronavirus (COVID-19) pneumonia in two patients with lung cancer. J Thorac Oncol 2020;15(5).
- [56] Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, Pan I, Shi L-B, Wang D-C, Mei J, et al. "Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT23. ' Radiology; 2008. 2020.
- [57] Dai W-c, Zhang H-w, Yu J, Xu H-j, Chen H, Luo S-p, Zhang H, Liang L-h, Wu X-l, Lei Y, et al. "CT imaging and differential diagnosis of COVID-19. ' Canadian Association of Radiologists Journal; 2020. 0846537120913033.
- [58] Al-khassaweneh M, Bani Abdelrahman R. "A signal processing approach for the diagnosis of asthma from cough sounds. J Med Eng Technol 2013;37(3):165–71.
- [59] Amrulloh Y, Abeyratne U, Swarnkar V, Triasih R. "Cough sound analysis for pneumonia and asthma classification in pediatric population. In: 2015 IEEE 6th international conference on intelligent systems. Modelling and Simulation; 2015. p. 127–31.
- [60] Pramono RXA, Imtiaz SA, Rodriguez-Villegas E. "A cough-based algorithm for automatic diagnosis of pertussis. PLoS One Sep. 2016;11(9):e0162128.
- [61] Porter P, Abeyratne U, Swarnkar V, Tan J, Ng T-w, Brisbane JM, Speldewinde D, Choveaux J, Sharan R, Kosasih K, et al. "A prospective multicentre study testing the diagnostic accuracy of an automated cough sound centred analytic system for the identification of common respiratory disorders in children. Respir Res 2019;20(1): 81.
- [62] Piczak KJ. "ESC: dataset for environmental sound classification. In: 23rd ACM international Conference on multimedia, ser. MM'15. Brisbane, Australia: ACM; Oct. 2015. p. 1015–8.
- [63] Stevens SS, Volkman J, Newman EB. "A scale for the measurement of the psychological magnitude pitch. J Acoust Soc Am 1937;8(3):185–90.
- [64] Patel K, Prasad R. "Speech recognition and verification using MFCC. Int J Emerg Sci Eng 2013;1(7).
- [65] Wold S, Esbensen K, Geladi P. "Principal component analysis. Chemometr Intell Lab Syst 1987;2(1-3):37–52.
- [66] Maaten Lvd, Hinton G. "Visualizing data using t-SNE. J Mach Learn Res 2008;9 (Nov):2579–605.
- [67] Kingma DP, Ba J. "Adam: a method for stochastic optimization. In: international Conference on learning representations. San Diego, CA: ICLR; May 2015.
- [68] Refaellizadeh P, Tang L, Liu H. "Cross validation," in *Encyclopedia of database systems*. Springer US; 2009. p. 532–8. https://doi.org/10.1007/978-0-387-39940-9_565 [Online]. Available:.
- [69] Pan SJ, Yang Q. "A survey on transfer learning. IEEE Trans Knowl Data Eng 2009; 22(10):1345–59.
- [70] Hartmann M, Farooq H, Imran A. Distilled deep learning based classification of abnormal heartbeat using ECG data through a low cost edge device. In: 2019 IEEE symposium on computers and communications. ISCC; 2019. p. 1068–71.
- [71] S. Recalcati, "Cutaneous manifestations in COVID19: a first perspective [published online ahead of print march 26, 2020]," *J Eur Acad Dermatol Venereol*, vol. 16387.