Alexandria Engineering Journal (2022) 61, 1319–1334

HOSTED BY

ELSEVIER

Alexandria University

Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com

Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath

Kranthi Kumar Lella [*], Alphonse Pja

Department of Computer Applications, NIT Tiruchirappalli, Tamil Nadu 620015, India

Received 1 February 2021; revised 25 May 2021; accepted 15 June 2021
Available online 19 June 2021

KEYWORDS

Artificial Intelligence;
Deep Convolutional Networks;
COVID-19;
Respiratory Sounds

Abstract The problem of respiratory sound classification has received good attention from the clinical scientists and medical researcher's community in the last year to the diagnosis of COVID-19 disease. The Artificial Intelligence (AI) based models deployed into the real-world to identify the COVID-19 disease from human-generated sounds such as voice/speech, dry cough, and breath. The CNN (Convolutional Neural Network) is used to solve many real-world problems with Artificial Intelligence (AI) based machines. We have proposed and implemented a multi-channeled Deep Convolutional Neural Network (DCNN) for automatic diagnosis of COVID-19 disease from human respiratory sounds like a voice, dry cough, and breath, and it will give better accuracy and performance than previous models. We have applied multi-feature channels such as the data De-noising Auto Encoder (DAE) technique, GFCC (Gamma-tone Frequency Cepstral Coefficients), and IMFCC (Improved Multi-frequency Cepstral Coefficients) methods on augmented data to extract the deep features for the input of the CNN. The proposed approach improves system performance to the diagnosis of COVID-19 disease and provides better results on the COVID-19 respiratory sound dataset.

1. Introduction:

As of 24th May 2021, the COVID-19 epidemic was declared a pandemic by the World Health Organization (WHO) in [34] the year 2020 of March 11, it claiming over 3,459,996 lives worldwide. The global situation as of 24th, May 2021, there have been 166,860,081 confirmed cases of COVID-19 which includes 3,459,996 deaths were reported to WHO is shown in Fig. 1. Experts in microbiology believe that data collection is critical for isolating infected people, tracing connections, and slowing the spread of the virus. Although advancements in testing have made these methods more common in recent months and the need for affordable, fast, and scalable screen-

[*] Corresponding Author at: Department of Computer Applications, NIT Tiruchirappalli, Tamil Nadu - 620015, India.
E-mail addresses: kranthi1231@gmail.com (K.K. Lella), alphonse@nitt.edu (A. Pja).

https://doi.org/10.1016/j.aej.2021.06.024
1110-0168

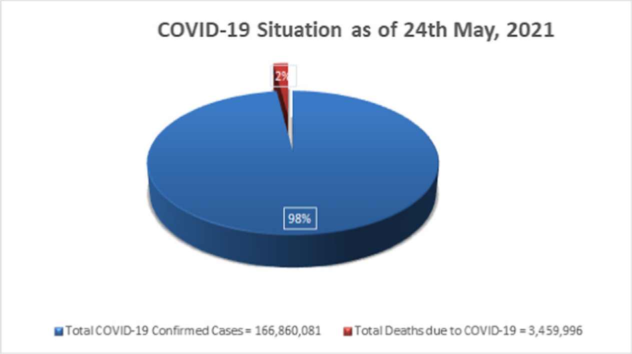


Fig. 1 The global situation were reported to WHO as of 24th May 2021.

ing technology for COVID-19 is very much needed. The seriousness of COVID-19 disease is classified into three categories, namely extreme, middle/moderate, and mild. The problem of respiratory sound classification and diagnosis of COVID-19 disease has received good attention from the clinical scientists and researchers community in the last year. In this situation, many AI-based models [4,8,10] entered into the real-world to solve such problems; and researchers have provided different machine learning, signal processing, and deep learning techniques to solve the real-world problem [5,37].

Nowadays, the COVID-19 pandemic is existed in the entire real-world and getting fear into the people to communicate physically. So, we have many ways to diagnose COVID-19 disease, and one of them is through human respiratory sounds [18,19,35]. Respiratory sounds generated by the human body like vibration, voice, lung sound, heart, food absorption, breathing, cough, and sighs have been used by clinical experts to diagnose the disease [18,20,21]. Until recently, such signals were normally obtained during scheduled visits via manual auscultation. Technological researchers and medical scientists have now begun using electronic technologies to collect sound from the human body (like Digital- Stethoscopes) [21,26] and carry out an automated examination of the human sounds data, e.g., recognition of wheeze in asthma. Researchers have also piloted the use of the human voice to aid early detection of several diseases: Alzheimer's disease corresponds with normal to slur [25], stammer, repeat, and use incomplete phrases and words (The beloved one can have trouble forming clear phrases or recognizing conversations), Parkinson's Disease (PD) may have many effects on the voice (several people with PD talk softly and they don't show enough feeling in one tone, speaking voice breathy or hoarse occasionally, and the end of a phrase, people with Parkinson's could slur sentences, mumble or trail off) [32,33], frequency of speech with coronary heart disease (people can develop neck pain, fatigue, voice disorder) [27–29], invisible disorders such as battle fatigue, brain trauma, and psychological situations correlate with sound pitch, vocal tone, speech rhythm & frequency, and voice sound volume.

The use of human respiratory sound as a diagnostic tool for different diseases presents tremendous potential for early detection and inexpensive solutions that could be rolled out to the masses if incorporated in commodity products. It is true for people if the solution can be unobtrusively tracked in different people during their everyday lives. The efficiency of the respiratory sound classification on the COVID-19 sounds dataset has been enhanced in the last year by applying different ML (Machine Learning) techniques such as SVM (Support Vector Machine), LVQ (Learning Vector Quantization), and MLR (Multivariate Linear Regression). The human respiratory sounds can be obtained by an electronic/digital stethoscope [20,21]; the presented model is applicable and demonstrates favorable robustness. Recent research has begun to investigate how respiratory sounds recorded by smart devices from patients confirmed positive cases of COVID-19 in the clinic varies with the healthy people's respiratory sound signs (e.g., breath sound, cough sound, and voice). Lung auscultation digital stethoscope data can be used for diagnosis of COVID-19 disease; a COVID-19-related cough detection analysis obtained with iOS/Android phones is presented using a group of forty-eight (48) patients with COVID-19 symptoms other clinical coughs trained in a series of models [1]. In order to automatically classify patients' health fitness and analyzed COVID-19 affected patient's respiratory sound signs from clinical patients. Our study includes studying the use of human respiratory sounds or respiratory sounds in crowd-sourced, unregulated data to diagnosis COVID-19 disease.

The DCNN (Deep Convolutional Neural Network) model is proposed and implemented in the present research that specifically classifies respiratory sound signs into usual and unusual independently of cough, voice, and breathing sounds. In addition, the model is constructed with a multi-feature channel by using a De-noising Auto Encoder (DAE) technique, GFCC (Gamma-tone Frequency Cepstral Coefficients), and IMFCC (Improved Multi-frequency Cepstral Coefficients) [30] to obtaindepthfeatures of respiratory sounds/voice is the basic input function to the deep convolution model. The functionality of DCNN is subsequently transforming the depth input features derived by the DAE, GFCC, and IMFCC methods and execute pooling activity [27]. Finally, using a "Softmax" classifier, the processed signals are categorized. The DAE is derived in-depth features of sound signals by removing mummer sounds from the background, the IMFCC is used to provide rich features from respiratory sounds, and GFCC is used to provide transient respiratory sound features [6,26,30]; the usefulness of the fundamental function of this analysis throughout the classification of respiratory sounds is thus demonstrated [21]. The classification accuracy with deep CNN has a high and F-score slightly increased for diagnosing COVID-19 disease with human respiratory sounds.

Specifically, we have collected COVID-19 sounds from Cambridge University with mutual agreement, and the name of the dataset is COVID-19 crowd-sourced sounds data. The dataset was gathered through an android/ iOS mobile app and the internet that collected voluntarily speech, cough, and breathing sound samples as well as their clinical samples to the background and signs. This iOS/Android app also collected whether the user has tested with positive COVID case before or not and basic information related to users. To date, this COVID-19 sounds dataset has collected around 9k unique users on the order of 13k samples. Although there are other attempts to gather user-related data, they are mostly narrow in reach or size. To our understanding, this is the world's largest unregulated crowdsourced data set of COVID-19 based sounds. This is a distinctive reason to collect the COVID-19 sounds dataset and perform our deep convolutional with multi-feature channels on augmented data to diagnose the COVID-19 disease. In this paper, we briefly described the

COVID-19 data, COVID-19 respiratory sound analysis and proposed a deep CNN approach in the introduction; background related works illustrate literature reviews of each author and background study for this research; methods describes the proposed deep CNN convolutional method, dataset collection, augmentation process, and multi-feature channels used to extract the deep features; experimental results and discussion section outlines the result analysis and discussions concerning the proposed model, and we have concluded this research with the best accuracy.

2. Background related work

Researchers and scientists have long recognized the utility of sound as a potential predictor of actions and health. For example, independent audio recorders were used for the reason in digital stethoscopes to identify sounds from the human respiratory [21]. These also involve highly trained clinicians to listen and interpret and are freshly and quickly being replaced with various methodologies such as MRI (Magnetic Resonance Imaging), sonography with which it is simple to examine and interpret. However, the recent works in automated sound modeling and interpretation can face these methods and give respiratory sound as an alternative that is inexpensive and easily distributed. More recently, microphones have been exploited for sound processing on goods and product-based machines such as android/iOS devices (smartphones) and wearables technologies. In [1], Brown C. et al. proposes an Android/iOS app to collect COVID-19 sounds data from crowdsourced sounds respiratory data of more than 200 positives for COVID-19 from more than 7 k unique users; Brown c. et al. has taken many general parameters and 3 major set COVID-19 tasks based on breath and cough sound. Here parameters are, i) positive COVID-19/negative COVID-19, ii) positive COVID-19 with cough/negative COVID-19 with cough, iii) positive COVID-19 with cough/non-COVID asthma cough; the Task-1 achieved 80% of accuracy for 220 users with modality is cough + breath; the Task-2 achieved 82% of accuracy for 29 users with modality is cough only; finally in Task-3 achieved 80% of accuracy for 18 users with modality is breath. Recall function is slightly lower (72%) because of the not specialized net to detect every COVID-19 cough. The authors have been improved accuracy for Task-2 and Tas-3 using VGG (Visual Geometry Group) Net with augmentation approach. In [2], Kun Qian et al. proposed a study of intelligent analysis on COVID-19 speech data by considering four parameters: i. Sleep Quality, ii. Severity, iii. Anxiety, iv. Fatigue. Kun Qian et al. collected data from the “COVID-19 sounds app” has launched by scientists and researchers from Cambridge University, and the “Corona voice detect App” has launched by researchers from Mellon University. After data processing, these people have obtained 378 total segments; from this preliminary study, they have taken 260 recordings for future analysis. These 256 sound pieces have been collected from 50 COVID-19 infected patients; for future study, poly impulses with such a sample rate of 0.016 MHz are converted. They have considered two acoustic feature sets in this study, namely ComParE&eGeMAPS; both feature sets were achieved 69% accuracy.

In [3], Lara O et al. implemented the “COUGHVID” crowdsourced dataset for cough analysis in COVID-19 symptom; More than 20,000 crowdsourced cough recordings reflecting a broad range of topic gender, age, geographic locations, and COVID-19 status are given in the COUGHVID dataset. They have collected a series of 121 cough sounds and 94 no-cough sounds first-hand to train the classifier including voice, laughter, silence, and various background noises. They have taken self-reported status variables (25% of recording sounds with healthy values, 25% sound recordings with COVID values, 35% sound recordings with symptomatic value, and 15% sounds recordings with non-reported status; It ensured that authors labeled 15% of cough sounds) for the selection of the recordings to be labeled. The percentage of COVID positive, Symptoms of COVID, and healthy subjects were 7.5%, 15.5%, and 77% from the subject of 65.5% males and 34.5% females respectively. In [18], Wang Y. et al. proposed a method to classify large-scale screening of people infected with COVID-19 differently; this work can be used to identify various breathing patterns and we can bring this tool for practical use in the real world. In this paper, first, a new and strong RS (Respiratory Simulation) model is introduced to fill the gap between a huge amount of training data and inadequate actual data from the real-world to considering the characteristics of real respiratory signals. To identify six clinically important respiratory patterns, they first applied bidirectional neural networks like the GRU (Gated Recurrent Units) network attentional tool (BI_at_-GRU) (Tachypnea, Eupnea, Biots, Cheyne-Stokes, Bradypnea, and Central-Apnea). In comparative studies, the acquired BI_at_GRU specific to the classification of respiratory patterns outperforms the existing state-of-the-art models.

In [4] Ali Imran et al. implemented an AI (Artificial Intelligence) based screening solution to detect COVID, transferable through a smart mobile phone application was suggested, developed, and finally tested. The mobile app called AI4COVID-19 records and sends to AI-based clouds running in the cloud triple 3-second cough sounds and comeback reaction within two minutes. Generally, cough is a basic indication of over 30 medical conditions associated with non-COVID-19. In [6], M Bader et al. proposed the significant model with the combination of Mel-Frequency Cepstral Coefficients (MFCCs) and SSP (Speech Signal Processing) to the extraction of samples from non-COVID and COVID and it finds the person correlation from their relationship coefficients. These findings indicate a high similarity between various breathing respiratory sounds and COVID cough sounds in MFCCs, although MFCC sound is more robust between non-COVID-19 samples and COVID-19 samples. Jiang X et al. [7]; This research shows that crowdsourced cough audio samples collected worldwide on smartphones; various groups have gathered several COVID-19 cough recording datasets and used them to train machine learning models for COVID-19 detection. However, each of these models has been trained on data from a variety of formats and recording settings; collected additional counting and vocal recordings, others exclusively collect cough recordings.

M. Al Ismail et al. [9] proposed a model with an analysis of vocal fold oscillation to detect COVID-19; because most symptomatic COVID-19 patients have mild to extreme impairment of respiratory functions, the model hypothesizes that through analyzing the movements of the vocal folds, and COVID-19 signatures might be detectable. Experimental findings on COVID-19 positive and negative subjects on a scientifically selected dataset show deep feature patterns of vocal fold oscillations associated with COVID-19. J Laguarta et al. [12] proposed an AI (Artificial Intelligence) model from cough sound recordings to detect the COVID symptoms; this model allows a solution to prescreen COVID-19 sound samples countrywide with no cost. A Hassan et al. [14] implemented a system to diagnose COVID-19 positive by using the RNN model; authors illustrated the major impact of RNN (Recurrent Neural Network) with the use of SSP (Speech Signal Processing) to detect the disease and this exiting model used to evaluate the acoustic characteristics of patients' cough, breathing, and voice, in the process of early screening and diagnosing the COVID-19 virus. Thomas F. Q. et al. [17]; proposed a framework structure to identify COVID symptomatic condition with Signal Processing (SP) and speech modeling techniques; this technique relies on the complexity of neuromata synchronization over speech/sound respiratory subsystem inside in the articulation, breathing, and phonation, driven by the existence of COVID symptom involving in upper inflammation versus lower respiratory inflammation tract.

J. Sharma et al. proposed [11] the attention-based deep CNN (Convolutional Neural Network) and multi-feature channels (GFCC, MFCC, Chromagram, QCT (Constant Q-Transform)) model for environmental sound classification [24]and obtained good accuracy on US8K and ECS-10 datasets. The CNN model for lung sound classification [13], which performed on collected respiratory sounds through a digital stethoscope and obtained around 80% accuracy for respiratory-based sound classification and 62% for audio-based classification. From all these background work senses, there is no accurate model for diagnosing COVID-19 disease symptoms. So, we are implementing the deep CNN model along with multi-feature channels (De-noising Auto Encoder, GFCC, and IMFCC) to perform better in the COVID-19 sounds crowdsourced dataset to the diagnosis of COVID-19 disease, and it improves to achieve better results on this dataset.

2.1. Summary Table

Table 1 represents the summary analysis of previous works to identify the COVID-19 symptoms for earlier detection of the disease. J. Sharma et al. proposed [11] the attention-based deep CNN (Convolutional Neural Network) and multi-feature channels (GFCC, MFCC, Chromagram, QCT (Constant Q-Transform)) model for environmental sound classification [24]and obtained good accuracy on US8K and ECS-10 datasets. The CNN model for lung sound classification , which performed on collected respiratory sounds through a digital stethoscope and obtained around 80% accuracy for respiratory-based sound classification and 62% for audio-based classification. From all these background work senses, there is no accurate model for diagnosing COVID-19 disease symptoms. So, we are implementing the deep CNN model along with multi-feature channels (De-noising Auto Encoder, GFCC, and IMFCC) to perform better in the COVID-19 sounds crowdsourced dataset to the diagnosis of COVID-19 disease and it improves to achieve better results on this dataset.

3. Materials and methods

3.1. Dataset collection

We have collected the COVID-19 sounds dataset from Cambridge University with mutual agreement for a research purpose; this dataset is approved at Cambridge University, Dept. of Computer Science and Tech., by following all ethics from the ethics committee. Brown C. and team [1] have implemented one android app and a web-based application to collect COVID-19 sounds; the main attributes of these applications are mostly equal. They have collected the past health history of a user for those who have been admitted before into the clinic. Users then enter their symptoms and report breathing sounds (if there are any): they collected cough three times sounds, breath heavily via their mouth 3–5 times, and read a brief statement within 30 s duration on the mobile/-computer screen. Finally, if they have been checked for COVID-19, users are questioned, and a position sample is obtained with consent. Besides, the iOS and Android applications [1] prompts users every two days to input additional sounds and symptoms, offering a particular chance to examine the breakthrough of sound-based patient well-being. This data is very securely encrypted in Cambridge University servers; the collected data is stored locally until connected to Wi-Fi, data is transmitted from the telephones. The data is deleted from the system if the correct data reaches and not collected any personal information.

3.1.1. COVID-19 sounds from crowdsourced dataset

At the end of May 2020, University of Cambridge researchers have been specially collected around 4.5 k unique from a web-based application and 2.5k samples collected uniquely from the android based application [1]; they have collected around 5k and 6k samples from different countries, respectively. Among these, around 300 users declared COVID-19 positive from both web-based and android applications. This android app is collecting more than one sample from the different users, and it leads to redundancy and causes a large dataset, so our future work is to remove this redundancy to improve performance. They have collected and analyzed general data (past and current medical history, age, gender) along with three different respiratory sounds (voice sound, cough, breath) from the unique uses in both web-based and android applications. A dry cough is the most commonly involved symptom identified in this category while coughing and sore throat are the most common signs. Interestingly, the most commonly affected signs are wet and dry cough and loss of ability to smell, and chest tightness (breathing) is the most common combined symptoms. This is consistent with knowledge from the COVID-19 symptom monitor. The reality provides further motivation for the use of sounds as a specific symptom that cough is among the most recorded signs of COVID-19, but it is also a common sign of many other diseases. So, the DCNN model is to classify and diagnosis the COVID-19 from these all symptoms.

3.1.2. Augmentation of the data

We have experimented with five augmentation sets, and we have given the details below. Before translating it into the

Table 1 The summary table for background analysis to diagnose COVID-19 disease.

Author	Dataset	Tasks		Accuracy (%)
Brown C et al., [1]	COVID-19 Crowd-sourced Sounds Dataset	1. COVID-19 positive/ COVID-19 Negative		80
		2. Positive with cough/ Negative with Cough		82
		3. Positive with cough/ Negative with asthma cough		80
Quian J. el al., [2]	COVID-19 Audio Data	Sleep		55
		Fatigue		42
		Anxiety		49
Lara O. et al., [3]	COUGHVID Dataset	Wheezing		90
		Audible_Dyspnea		93
		Stridor Sound		98
		Nasal-Congestion		99
		Choking		99
		Labeled as COVID-19 mild		86
Ali Imran et al., [4]	Data collected through mobile app with name of COVID-19 Samples	Speech		92
		Cough		92
		Overall		88
Bader M. et al., [6]	Own data set collected from hospital with 14 patients	COVID-19 Negative Vs Positive	Cough	42
		COVID-19	Breath	43
		COVID-19 with Cough Vs COVID-	Voice	79
		19 without Cough	Cough	65
			Breath	58
Jiang Z. et al., [7]	Data is collected from Ruijin Hospitals	COVID-19 Detection from Thermal videos and breathing Patterns.		83
Al Ismail M. et al., [9]	Data set collected from 521 individuals	Linear Regression for COVID detection		82
Lella Krnthi Kumar and Alphonse [39]	COVID-19 Crowd-sourced Sounds Dataset	Positive COVID-19/Negative COVID-19		90
		Positive cough of COVID-19/Negative cough of COVID-19		88
		Positive cough of COVID-19/Negative COVID-19 Asthma Cough		88
		Asthma Breath/Healthy Breath		84
		Asthma cough/Normal Cough		86
Hassan A. et al., [14]	Data collected from COVID affected 14 patients	COVID-19 Sample Cough		97
		COVID-19 Breath Sound		98
Shui-Hua Wang et al., [35]	CT Scan Image Dataset	COVID-19 Detection		97
Shui-Hua Wang et al., [36]	296 Chest CT Images	COVID-19 Detection		86
Sree Jagadeesh M. and Alphonse PJA [38]	COVID-19 English Labelled Tweets Dataset	Prediction of COVID-19 from Tweets		91
Laguarta J. et al., [12]	MIT open voice data set	COVID-19 Positive Symptoms		79

input representation, often used to train the neural network, each deformation is directly applied to the respiratory sound signal [23]. Notice that it is essential that we have chosen the deceleration parameters for each augmentation so that the functional validity of the mark is preserved. The following defines the deformations and the associated augmentation sets:

i. Stretching Time (ST): Increase or reduce the sample sound signal (to unchanged running pitch). Based on the four factors {0.80, 0.94, 1.06, and 1.24} the duration is stretched.

ii. Shift Pitch1 (SP1): Sound/audio samples can be increase or decrease (to unchanged running pitch), and every sample can be shifted differently by four values (-1,-2,-2,-1).

iii. Shift Pitch2 (SP2): We wanted to build a second augmentation package because our initial tests showed that pitch shifting was an especially beneficial increase. Every sample was pitch moved by four higher values (in different sizes and shapes) this time (-2.5,-3.5, 3.5, and 2.5).

iv. Compression of Range Dynamically (CRD): We have compressed these 4 parameters online, one taken from the 'ICECAST' streaming server (it is a free software server for streaming multimedia), and three from standardDolby E (it is a digital sound/audio stream can be processed a regular stereo pair of digital sound/audio tracks).

v. Background of Noise (BN): Pair the sample with some other sequence of various kinds of audio scenes containing background noises; four sound scenes were combined for each sample (while taking respiratory sounds – we have combined environmental sound noise). The mixed or combined value is generated as 'c'.

So, c = ((1-r) × a) + (r$\cdot$b). Where, a- audio signal original sample, b- background noise signal, r- random weight parameter (0.10, 0.50). Using the MUDA library, the augmentations were added, to which the reader is referred for more information on the execution of each deformation. MUDA selects the audio recording and the accompanying JAMS format annotation directory and produces the deformed audio along with the improved JAMS data containing all the deformation parameters used. In this study (see below), we have ported the original annotations given with the dataset used for evaluation into JAMS files and made them accessible on the internet together with the JAMS files after deformation.

3.2. Multi-Feature channels preparation

In this work, we have implemented three multi-feature channels (DAE (De-noising Auto Encoder), GFCC (Gamm-atone Frequency Cepstral Coefficients) filter bank, and IMFCC (Improved Mel-frequency Cepstral Coefficients)) to acquire deep features from the input respiratory sounds (voice, cough, and breathing sound) to the diagnosisof COVID-19 disease. The CNN provides better distinguishable features and comparable feature representations for the accurate classification of sound signals by integrating various signal processing techniques that obtain various kinds of information. The DAE, GFCC, and IMFCC features are placed together to establish a multi-channel input for the DCNN (Deep Convolutional Neural Network). The DAE acts to extract sound features by removing noise from background sound signals, GFCC provides transient respiratory sound features, and IMFCC is the backbone to extract the deep features in this work; Section 3.2.1–3.2.3 shows the working process of multi-feature channels.

3.2.1. Data De-noising Auto Encoder (DAE)

In addition, a De-noising Auto Encoder (DAE) algorithm is used as one of the channels for the input function of the deep convolution network to obtain the deep feature of respiratory sounds by removing background noise. We have used this DAE method to remove the background noise and provide noisy-clean respiratory sounds on COVID-19 sounds crowd-sourced data. The De-noising Auto Encoder is used to extract deep features of COVID-19 respiratory sounds from input data by removing the noise. It includes both single linear decoding stage and non-linear encoding stage for the real value respiratory COVID-19 sounds as:

$$\langle(X_i) = \sigma(WD_1 X_i + c) \tag{1}$$

$$\widehat{Y}_i = WD_2 \cdot \ \langle(X_i) + b. \tag{2}$$

Where WD_1 is an encoding neural network weight connection, WD_2 is the decoding matrix of neural network weight; usually, one weighted type of regularization matrix is$WD = WD_1 + WD_2{}^T$. Input and output layer vector biases are b,c.The logistic function is defined for nonlinear hidden neuron as$\sigma(Y) = \left(1 + e^{(-Y)}\right)^{-1}$. The following object function in Eq. (3) is determined the parameters of respiratory sound features.

$$\mathcal{L}(\theta) = \sum_i \| Y_i - \widehat{Y}_i \|_2^2 \tag{3}$$

Where $\theta = \{WD, b, c\}$ is a set of parameters, and Y_i is the clean sound related to the version of X_i input sound. In addition to using connected weights, the introduction of weight normalization and hidden neural output can help to prevent overfitting in order to improve the generalized statement. Weight decline and infrequent batch normalization on hidden neuron output instances are formulated as:

$$K(\theta) = \mathcal{L}(\theta) + \alpha \| WD \|_2^2 + \beta\rho(\langle(X)) \tag{4}$$

Where $\| WD \|_2^2 = \sum_{i,j} WD_{i,j}^2 \cdot \ \rho(\langle(X))$ is hidden layer outputs of regularization function weight coefficients of regularizations are ‘α’ and ‘β’. In this work, we have set ‘α’ and ‘β’ values are 0.002 and 0 (zero), we have not considered infrequent regularization in this work (we will consider it in future work). Then the set of parameters can be collected as in Eq. (5).

$$\theta \triangleq \begin{matrix} argmin \\ \theta \end{matrix} K(\theta) \tag{5}$$

To estimate(WD^*, b^*, c^*), we have used the Quasi-Newton optimization algorithm based on linear search in this work [15]. We have used a greedy-based pre-training approach with fine-tuning to train the De-noising Auto Encoder. The first auto-encoder training pair is(X, Y) and the next auto-encoder learning pair will become $\langle(X_i)$ and $\langle(Y_i)$ is depicted in Fig. 2. The final solution is likely to be better than training the De-noising Auto Encoder with batch normalization.

3.2.2. Gamma-tone Frequency Cepstral Coefficients (GFCC) filter bank

The Gamma-tone filter-bank is a collection of cochlear simulation filters. The sampling rate of a gamma filter is very near to the features of the magnitude of a human ear filter. It is able to represent the comprises of different motions with the gamma-tone filter-bank. A gamma-tone filter's sampling rate is the result of a spectrum of gamma and a sinusoidal sound, whose core frequency is f_c, which can be represented as Eq. (6).

$$g(x) = \left(Ax^{n-1}\right) \cdot \ e^{2\pi Bx} \cos\left(2\pi f_c x + \varphi\right) \tag{6}$$

Where A is the amplitude gain; B is the bandwidth of the filter $(B = ERB(f_c) + 1.019)$, ERB-Equivalent Rectangular Bandwidth; f_c is the core frequency in Hz; n is the order of the filter and the phase shift is φ. We have set the order of the filter as four because the gamma filter of the fourth-order is identical to the feature used to describe the human auditory filter [16]. The ERB at each point along the cochlea is a frequency modulation indicator of the auditory filter distance. An impractical but useful simplification of modeling estimation, cubic band-pass filters, is implemented when determining the bandwidth of human hearing. The bandwidth of the auditory filter is the value of an ERB centered at frequency f. The association between ERB and f with scale is depicted in Eq. (7).

$$ERB_s(f) = 21.40\log_{10}((0.0043 * f) + 1) \tag{7}$$

The filter bank should cover each point in the ERB_s space in order to model the human sound frequency spectrum accurately. In this work, a linear partition model is utilized for core frequencies of each gamma-tone filter are scaled Eq. (7), which is defined in Eq. (8).

$$f_{c_i} = ERB_s{}^{-1}\left(\left(k_i \times \frac{ERB_s\left(f_{high}\right) - ERB_s(f_{low})}{N}\right) + ERB_s(f_{low})\right). \tag{8}$$

where the inverse of ERB_s is defined as $ERB_s{}^{-1}$, lowest frequency (0.01 kHz) considered as f_{low}, highest frequency (20 kHz) considered as f_{high}, the total number of the gamma-tone filter is N and the filter index is k_i. We can calculate GFCC with a gamma-tone filter bank, there is a way of measuring GFCC similar to an extraction method of IMFCC.

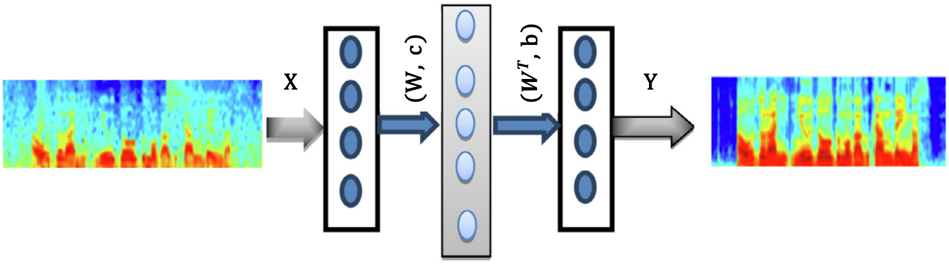


Fig. 2 The training of De-noising Auto Encoder with noisy clean respiratory sound pairs.

Firstly, small frames are decomposed into the primary audio signals. In this work, the frame length is fixed to 25ms by default. Fast Fourier Transformation (FFT) is then used for each frame to evaluate the frame response. The gamma bandpass filter is consequently used for the FFT of the signal to reaching the sub-band spectrum. Each sub-band filter is represented as Y_n to calculate energy, the log function and the discrete cosine transformation are applied in modeling the perception of human loudness and not correlated to the outputs of the logarithmically compressed filters in the last step. It is possible to calculate the GFCC as Eq. (9).

$$GFCC_n = \sqrt{\frac{2}{M}}\left(\sum_{m=1}^{M} \log_{10}(Y_n)\cdot \cos\left[\frac{\pi m}{M}\left(n-\frac{1}{2}\right)\right]\right), 1 \le n \le N \tag{9}$$

Whereas the number of gamma-tone filters is defined as M, the number of GFCC filters is indicated as N, and m^{th} sub-band energy is defined as Y_n. The block diagram to calculate GFCC, where the number of the filter bank is m, extracted coefficients are defined by N, and frame signal is the respiratory sound signal in the frame is shown in Fig. 3.

3.2.3. Improved Mel-frequency Cepstral Coefficients (IMFCC)

Improved Mel-frequency Cepstral Coefficients (IMFCC) is used to extract automatic deep features from input sound or voice and it will scale frequency in logarithmic nature. The IMFCC method has five basic steps as follows. The structure of Improved MFCC is represented in Fig. 4.

Step I: In this stage, because of sudden growth in the sound signal, sound signals are frames in short frames. This isn't much shorter and not much longer to have an excellent spectral estimate frame. And then it's done to eliminate the interruptions at the beginning and the end of the frame window. The window is $W_j(n), 0 \le n \le M_n - 1$, where M_n is every frame of quantity samples, the output will be presented as $X(n) = Y(n) \times W_j(n)$, where $0 \le n \le M_n - 1$, we are getting output signal as $X(n)$ by multiplying window $W_j(n)$ and the signal of input $Y(n)$. The representation of $W_j(n)$ is denoted in Eq. (10).

$$W_j(n) = 0.54 - 0.46\sin\left(\left(\frac{\pi}{2}\right) - \left(\frac{2\pi n}{M_n - 1}\right)\right), 0 \le n \le M_n - 1 \tag{10}$$

Step II: The next step is FFT (Fast Fourier Transform) for converting from the spatial time domain to the frequency domain. Each frame has M_n samples that are translated to the domain of frequency. The representation of $\sum_{n=0}^{M_n} t(n)$ shows sound frame, a number of M_n samples are around 160, and f&n are the frequency and time indexes.

$$|T(f)|^2 = \left|\sum_{n=0}^{M_n-1} t(n)\cdot\left(\left(\frac{-j2\pi nf}{M_n}\right)\right)\right|^2 \tag{11}$$

Step III: We have to estimate the power spectrum of each frame by calculating the periodogram after making the frames with the following Eq. (12).

$$P_j(i) = \frac{1}{M} \times |S_j(i)|^2 \tag{12}$$

The power spectral approximation still includes a large amount of information not needed for Automated Sound Recognition (ASR) for the respiratory system, so the difference between these two widely spaced frequencies is not seen. In different frequency areas, the Mel-filter bank can be used to get an indication of how much periodogram exists. The very first filter in the Mel-filter bank shows how often energy exists near zero Hz, and the first filter is very small. The Mel-scale offers details on how to arrange the filter banks and it also informs how large to build them is represented in Eq. (13). Where 'f_i' is the power spectrum filter function.

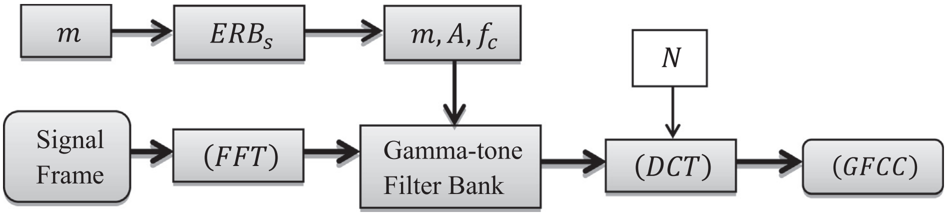


Fig. 3 The block diagram to compute GFCC(The number of the filter bank is m, extracted coefficients are defined by N, and frame signal is the respiratory sound signal in the frame).

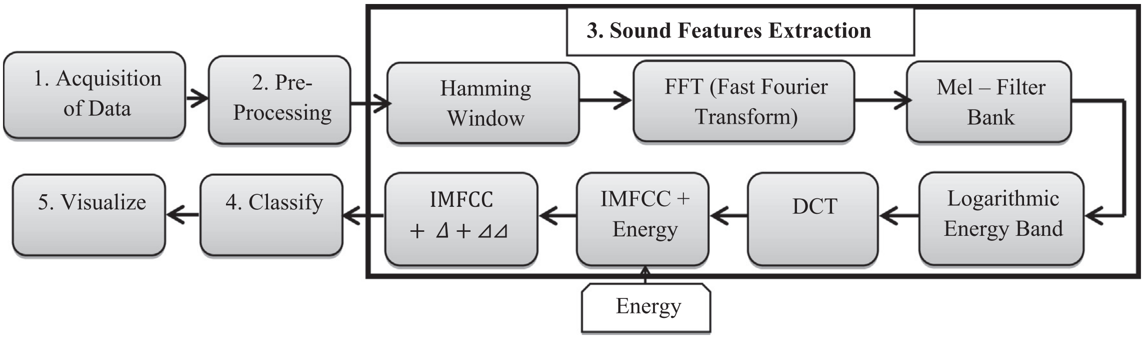


Fig. 4 The structural diagram of features extraction method using IMFCC.

$$f_i(x) = \begin{cases} 0, x < f_i(n-1) \\ \frac{x - f_i(\mathrm{n}-1)}{f_i(\mathrm{n}) - (f_i(\mathrm{n}-1))}, f_i(n-1) \le x \le f_i(n) \\ \frac{f_i(\mathrm{n}+1) - x}{(f_i(\mathrm{n}+1)) - f_i(\mathrm{n})}, f_i(n) \le x \le f_i(n+1) \\ 0, x > f_i(n=1) \end{cases}. \tag{13}$$

Step IV: In this step, we have to calculate the energy band logarithmic value by using *Eq.*(14) and it shows conversion of the frequency spectrum to Mel-filter scale.

$$M_f(x) = 2595 \times \log_{10}\left(1 + \left(\frac{x}{100}\right)\right) \tag{14}$$

Step V: In the final step, we have to measure the DCT (Discrete Cosine Transformation) for the energy of the log filter bank. Since the energies of the filter banks are connected with each other, and the proposed system filter banks all overlap with one another. As well as a result is known as IMFCC (Improved Mel-frequency Cepstral Coefficient)) after DCT. So, the final cepstral is calculated as C_n is represented in Eq. (15).Where $S = 1, 2, 3 \cdots, n$, $k = 0$to $S - 1$ and 'c' constant factor value for discretization.

$$C_n = \sqrt[2]{\frac{2}{S}} \sum_{k=0}^{S-1} \left((\log_{10}[c \times (k+1)]) \cdot \sin\left(\frac{\pi}{2}\right) - \left[n \times \left(\frac{2k-1}{2}\right) \cdot \frac{\pi}{S} \right] \right) \tag{15}$$

3.3. The proposed deep convolutional Neural network with Multi-Feature channels

The deep convolutional network classifies the respiratory sound according to the respiratory diseases (COVID-19 positive, COVID-19 negative, normal flue, asthma, healthy sound) after acquiring features from multi-feature channels (DAE, GFCC, and IMFCC). The following Fig. 5 shows the proposed deep convolutional architecture with multi-feature channels. In this work, we have implemented a deep convolutional network for the diagnosis of COVID-19 disease. The DCNN model that was introduced consists of three convolutional network layers, two pooling operations along two dense (fully connected) layers to secure convolutional layers.

The data to the system composed of TF-P (Time Frequency-Patch) obtained from the log-scaled Mel-spectrogram model of the audio signal representation, related to the originally approved feature learning methods adapted to COVID-19 sound classification. To obtain log-scaled Mel-spectrograms with 256 bands (components) representing the recognizable frequency spectrum (0–22.05 kHz), using a frame rate of 25ms (the frequency range of 0.44 MHz to 1024 samples) and a hop size with the same length, we used the python library (Essentia). We obtain a total of 733 feature dimensions along with 256-dimensions of convolutional model feature vectors, handicraft feature vectors of 477-dimensions, and the combined modality (voice, cough, breath) feature vectors.

The evaluation dataset (described in Section 3.1.1) is varying input respiratory sound duration time from zero to 30s; we fix this input respiratory sound data TF-P (Time-Frequency-Patch) T to 5s (256 frames), i.e. $T \in \mathbb{R}^{256 \times 256}$. As mentioned further below, TF-P is extracted automatically within a time from the complete log-Mel-spectrogram from each sound excerpt during training. For our given input T, the system is designed to learn the parameters θ of a generalized non-linear function F(..|θ) that maps T to the output (estimation) Y is represented in *Eq.*(16).

$$Y = F(T|\theta) = f_K(\ldots . f_2(f_1(((T|\theta_1)|\theta_2)|\theta_K) \tag{16}$$

Where a layer of a network for each operation is represented as $f_k(..|\theta_k)$, and $K = 5$ in our proposed work. The first three layers of this architecture are, $k \in \{1, 2, 3\}$are convolutional layers and the expression is represented in Eq. (17)

$$Y_k = f_k(T_k|\theta_k) = h(W * T_k + b), \qquad \theta_k = [W, b] \tag{17}$$

Where W is a sequence of three-dimensional filters (also called as kernels) of N, T_k is a three-dimensional vector input made up with M feature maps, '$*$' represents the convolutional operator, the activation function with point-wise is $h(\cdot)$, and b is the bias for the function. The network shape of W, X_k, Y_k are $(M, \mathrm{d}_0, \mathrm{d}_1), (N, M, \mathrm{n}_0, \mathrm{n}_1), (andN, \mathrm{d}_0 - \mathrm{n}_0 + 1, \mathrm{d}_1 - \mathrm{n}_1 + 1)$. The first layer of the network is $\mathrm{d}_0 = \mathrm{d}_1 = 256$ is the input dimensions of the TF-P. We have applied the max-pooling layer after the first two convolutional layers $k \in \{1, 2\}$ with equivalent pooling dimensions, this decreases the sizes of the output map function and thus improves training performance and provides certain classification accuracy in the network. The final 2 layers are the dense (fully-connected) layers $k \in \{4, 5\}$ and it consists of a matrix product instead of a convolution. This is shown in Eq. (18)

$$Y_k = f_k(T_k|\theta_k) = h(W\theta T_k + b), \theta_k = [W, b] \tag{18}$$

Where the column vector representation of T_k is flattened to M, the activation function is represented a point-wise as $h(\cdot)$, W is the dimension shape (N, M), and vector of N is 'b'. Fig. 6 depicts the layered architecture of the proposed deep convolutional network model.

The proposed deep convolutional network is modeled as follows:

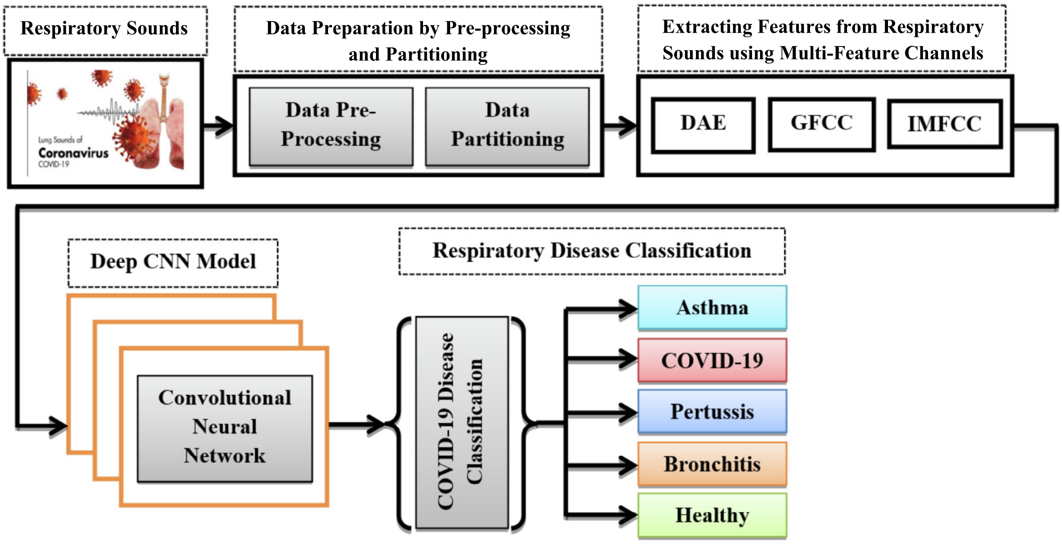

Respiratory Sounds
Lung Sounds of
Coronavirus
COVID-19
Data Preparation by Pre-processing and Partitioning
Data Pre-Processing
Data Partitioning
Extracting Features from Respiratory Sounds using Multi-Feature Channels
DAE
GFCC
IMFCC
Deep CNN Model
Convolutional Neural Network
Respiratory Disease Classification
COVID-19 Disease Classification
Asthma
COVID-19
Pertussis
Bronchitis
Healthy

Fig. 5 The architecture of proposed deep convolutional neural network with multi-feature channels.

k_1: In this level, the convolutional kernel is generating with the field of $(5,5)$ for 24 filters, therefore the shape of W is $(24,1,5,5)$ maximum pooling stride of the last two dimensions is $(4,2)$ with respect to the time and frequency, and the activation function (ReLU – Rectified Linear Unit) for the next layer is $h(y) = \max(y,0)$.

k_2 : In this level, the convolutional kernel is generating with the field of $(5,5)$ for 48 filters, therefore the shape of W is $(48,24,5,5)$ as like k_1, the maximum pooling stride is $(4,2)$, and the activation function (ReLU – Rectified Linear Unit) for the next layer.

k_3 : In this level, the convolutional kernel is generating with the field of $(5,5)$ for 48 filters, therefore the shape of W is$(48,48,5,5)$, and the activation function ReLU(Rectified Linear Unit) for the next layer.

k_4 : In this level, the convolutional model is generating 64 are hidden units, therefore the shape of W is$(2400,64)$, and the activation function (ReLU – Rectified Linear Unit) for the next layer.

k_5 : In this level, the convolutional model is generating 5 are output units, therefore the shape of W is$(64,5)$, and the activation function is “Softmax”.

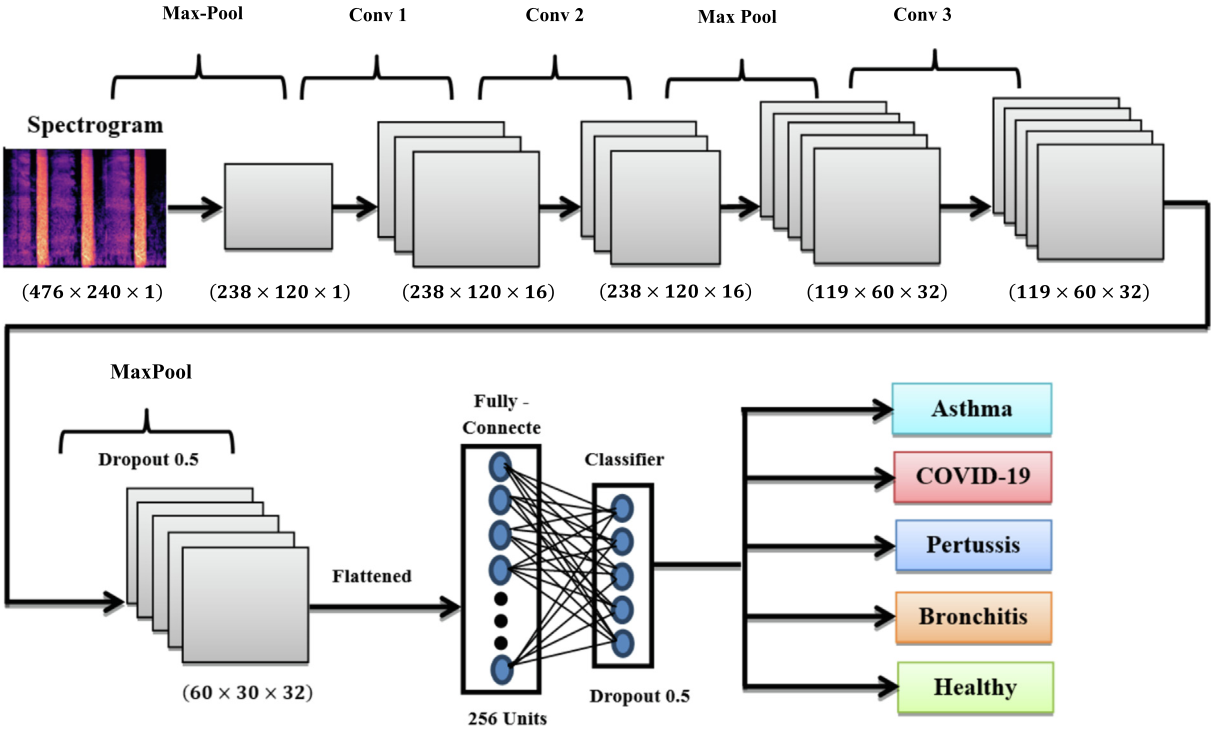
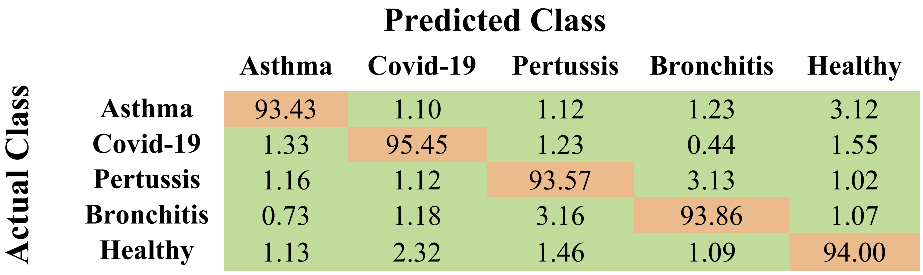


Fig. 6 The layered architecture of proposed deep convolutional network for COVID-19 disease classifier.



Actual Class	Predicted Class Asthma	Covid-19	Pertussis	Bronchitis	Healthy
Asthma	93.43	1.10	1.12	1.23	3.12
Covid-19	1.33	95.45	1.23	0.44	1.55
Pertussis	1.16	1.12	93.57	3.13	1.02
Bronchitis	0.73	1.18	3.16	93.86	1.07
Healthy	1.13	2.32	1.46	1.09	94.00

Fig. 7 The average confusion matrixes to the diagnosis of disease from respiratory sounds (cough, breath, voice) for different classes.

Table 2 Classification Results of five tasks for respiratory COVID-19 sound base para*meters (Cough, Breath, Voice)*.

Task	Modality	Accuracy (%)	F_1– Score (%)
Asthma/Non-Covid-19	Breath + Cough	93.43	94.69
Non-Coid-19/Covid-19	Breath + Cough + Sample Voice	95.45	96.96
Covid-19/Pertussis	Cough	93.57	94.13
Pertussis/Bronchitis	Cough	93.86	94.13
Bronchitis/Healthy	Breath + Cough	94.00	94.89

Notice that we have used a small feature map $(5,5)$ in k_1 proportional to the input sizes $(256,256)$ is made to enable the systems to learn small, scattered signals which can be combined at successive layers to collect arguments to confirm larger 'time–frequency signs' that indicate the influence of various sound groups, even though acoustic masking occurs. We have normalized the model loss cross-entropy through a mini-batch stochastic gradient descent optimizer. Each and every batch contains 100 randomly chosen TF-Patches from the training outcomes without repetition. Every 5 seconds TF-P is extracted from the completed frequency Mel- spectrogram model of every training data sample from different positions in time. We have applied dropout in the last two layers$k_1 \in \{4,5\}$ with 0.5(50%)probability, and the learning rate of the convolutional model is 0.01 as a constant. The finality factor of 0.001 is applied to the last two-layer weights with L_2-Normalization. The convolutional model is check-pointed for every epoch and trained for 100 epochs. The classification of the test-set achieves the highest prediction accuracy, and the test dataset is used where classification is carried out by splitting the test sample into overlapping TF-Ps, and finally selecting the experiment level prediction as the highest probability mean output activation over all frames.

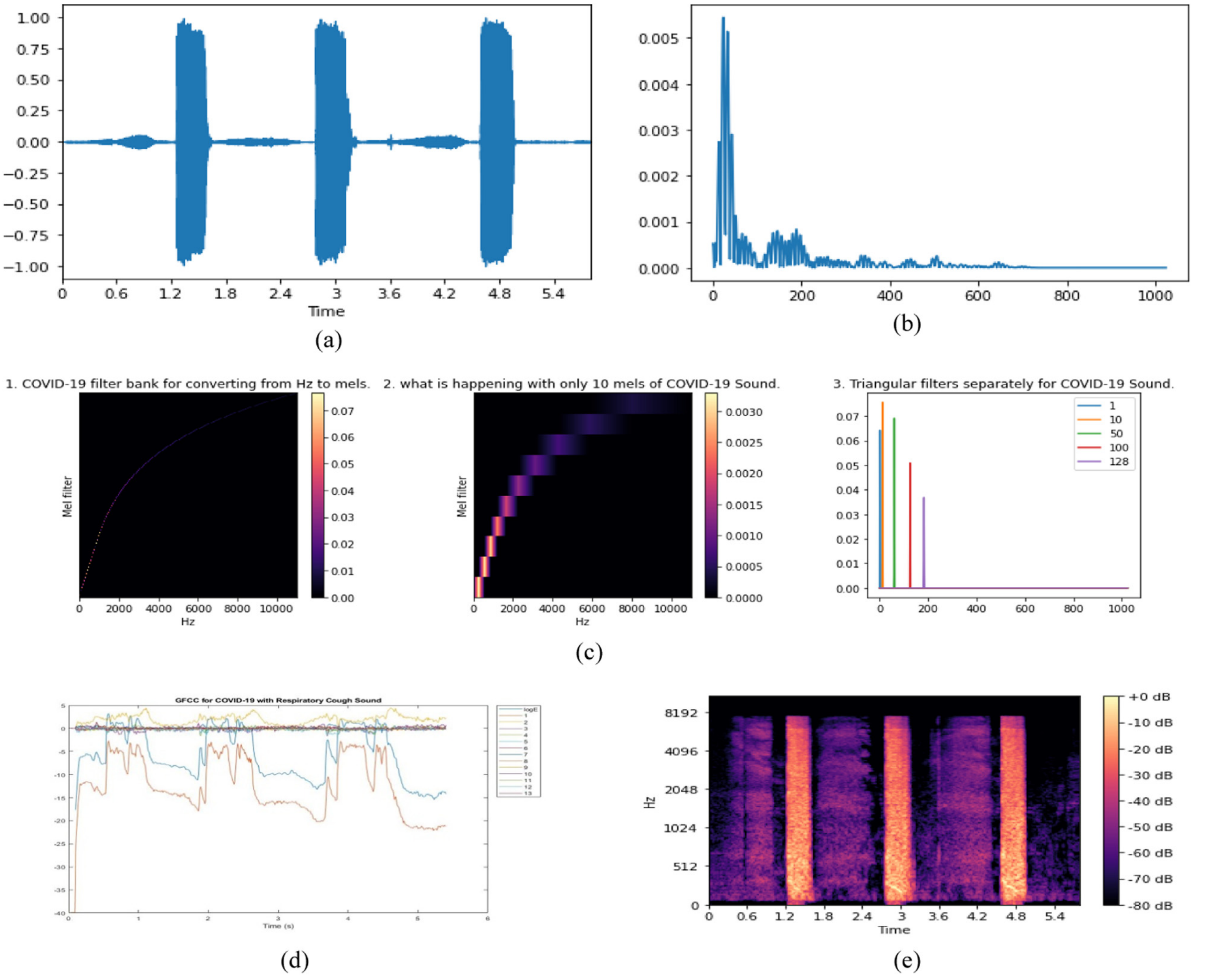
3.3.1. Evaluation

The Crowdsourced COVID-19 sounds dataset is used to test the proposed Deep CNN architecture and the effect of the various augmentation sets. The dataset is collected of 1539 audio sounds (voice, breath, and cough) tracks is upto 30s from individuals. The COVID-19 crowdsourced respiratory sound dataset consisting of different classes are Asthma breath, Asthma with cough, Asthma with cough and breath, non-COVID with the breath, non-COVID from cough, COVID-19 from breath, COVID-19 from cough, COVID-19 from cough and breath, healthy symptoms from cough, healthy symptoms from breath. This model result is compared with previously developed approaches on this dataset with the testing accuracy. In terms of testing accuracy, the new framework is evaluated and compared with the previous models. The data is categorized into five different folders with label numbers and analyzed all the data labels with the model with the higher accuracy. The data were divided into training and testing sets; we use one of five training folders in each division to train the latest deep convolutional network architecture as a test set to identify the learning epoch that provides the best result when working with the remaining four folders.

4. Results and discussion

We use the performance measures of precision, recall, specificity, accuracy $(Eq.(20))$;F_1-score $(Eq.(19))$ on the test collection and also out-of-sample testing the results to evaluate the model. The precision of the model refers to the model's total accuracy. The k-fold classification technique has been used, which is well-suited for evaluating the performance on restricted data of machine learning and deep learning models. The average confusion matrixes from out-of-sample testing are focused on these performance metrics is shown in Fig. 7. Besides, to avoid the issue of over-fitting, we also use regularization techniques, such as tuning the model parameters of the Support Vector Machine (SVM) against the precision of cross-validation and choosing several parameters that gave us the best generalizability of the models. Based on out-of-sample testing, tuning of the different hyper-parameters (number of hidden layers, activation functions, learning rate, and dropout rate) of deep convolutional network-based models has also been carried out. Besides, to eliminate the possibility of over-fitting, the decay of system loss compared to the number of epochs has been examined. From our observations on the crowdsourced COVID-19 sounds dataset, the benefits of deep convolutional network architecture using multiple feature channels on augmented data are shown in Table 2.

$$F_1\ \ Score = \frac{\frac{PT}{PT+PF} \times \frac{PT}{PT+NF}}{\frac{PT}{PT+PF} \times \frac{PT}{PT+NF}} \tag{19}$$



(a)

(b)

(c)

(d)

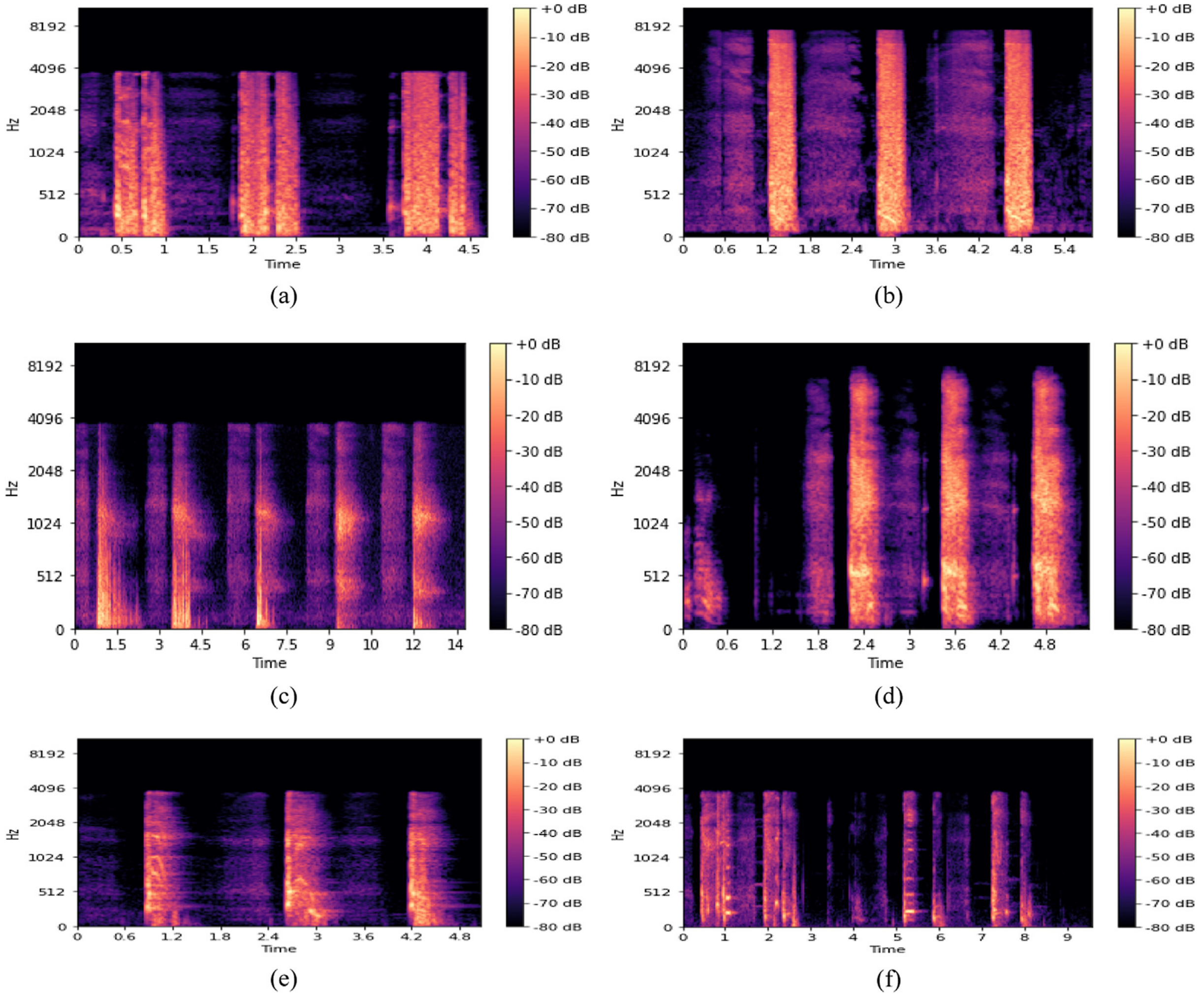
(e)

Fig. 8 The process to extract DCNN input features using multi-feature channels for COVID-19 respiratory sounds ((a) – sample COVID-19 sound, (b) – Data De-noising from COVID-19 cough sound sample and extracting soft signal from the input, (c) – log Mel-Spectrum, (d) – GFCC features, (e) – IMFCC feature extraction).

$$Accuracy = \frac{PT + NT}{PT + NT + PF + NF} \times 100 \tag{20}$$

We have used multi-feature channels to extract better features from the respiratory sound data for multi classes [31] and the system has plotted different spectrograms for multiple classes is depicted in Fig. 8, the system may perform better with multi-feature channels. The comparison of IMFCC features with different respiratory sounds (Asthma Cough Sound, COVID-19 Cough Sound, Pertussis Cough Sound, Bronchitis Cough Sound, Healthy Cough Sound, and COVID-19 without Cough Sound) is depicted in Fig. 9. The system can sense the cepstral coefficient values of each sound and identifies the different cepstral values from every respiratory input sound. The system made the changes with respect to the vocal fold and pitch frequency sound to extract the deep features.

The system can normalize the frequency of every respiratory sound with leakage factor to improve the system performance and it can be boxplot the different respiratory sounds shown in Fig. 10. It represents the normalized frequency of every respiratory sound sample (COVID-19, Asthma, Pertussis, Bronchitis, Healthy, and COVID-19 with no cough) with respect to the time and frequency domains. The system can sense the cepstral coefficient values of each sound and identifies the different cepstral values from every respiratory input sound. The changes were made by the system with respect to the vocal fold and pitch frequency sound to extract the deep features. The changes in pitch and frequency sound are represented in Fig. 10. It shows the normalized factor for all respiratory sounds (COVID-19, Asthma, Bronchitis, and Pertussis). The normalized frequency with leakage factor and boxplot representation for different respiratory sounds are normalized frequency sound of the covid-19 sample, boxplot diagram for normalized covid-19 sample sound, normalized frequency sound of asthma, boxplot diagram for normalized frequency sound of asthma, normalized frequency sound of pertussis, boxplot diagram for normalized frequency sound of pertussis, normalized frequency sound of Bronchitis, boxplot diagram for normalized frequency sound of Bronchitis, normalized



(a)

(b)

(c)

(d)

(e)

(f)

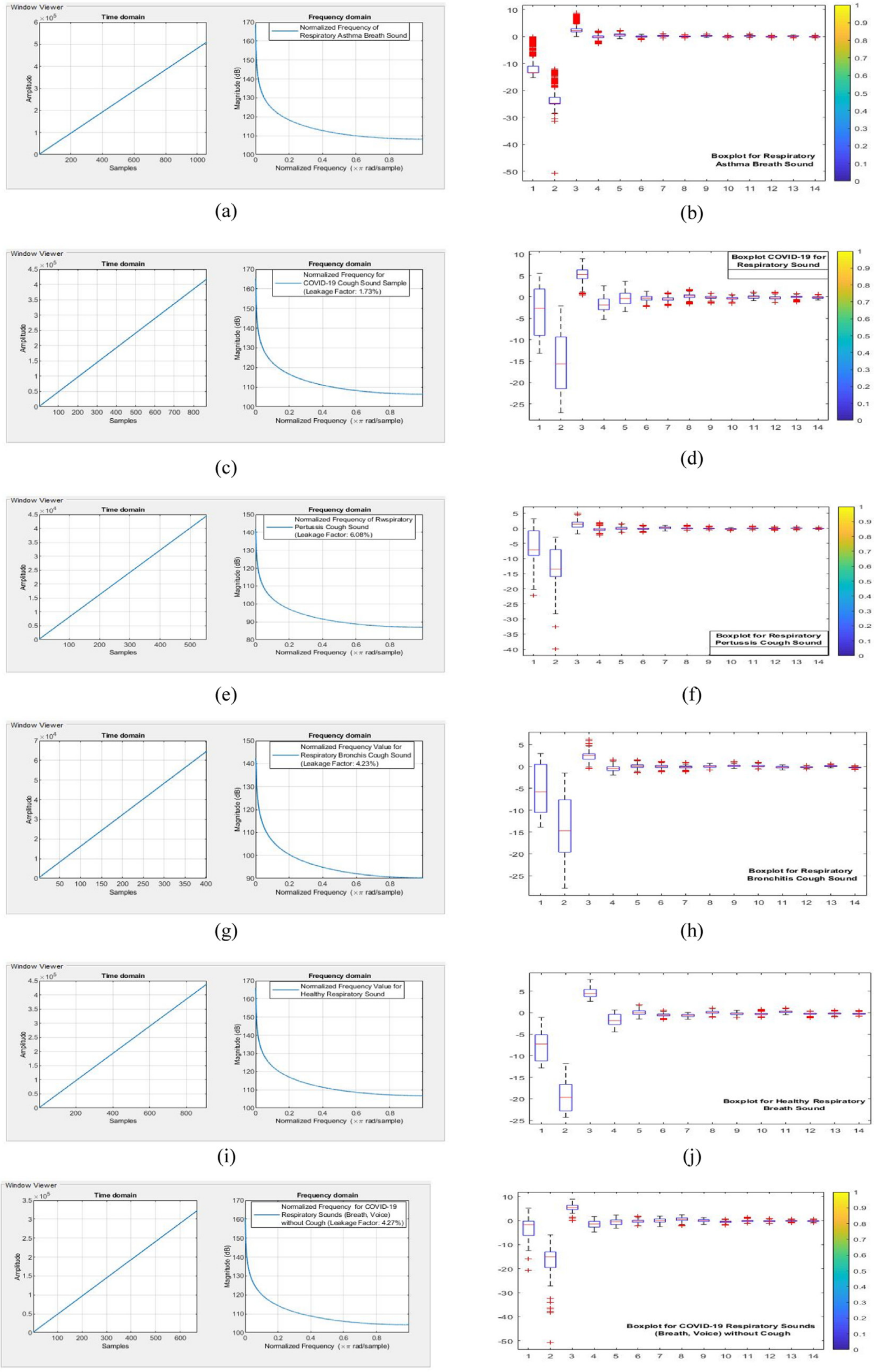
Fig. 9 The Comparison of IMFCC Feature with different Respiratory Sounds ((a) – Asthma Cough Sound, (b) – COVID-19 Cough Sound, (c) – COVID-19 without Cough Sound, (d) – Pertussis Cough Sound, (e) – Bronchitis Cough Sound, (f) –Healthy Cough Sound).

frequency signal for healthy sound, boxplot diagram for healthy sound, normalized frequency sound of covid-19 without cough, boxplot diagram for covid-19 without cough sound.

4.1. The comparison of convolutional network architecture

We compare the implemented DCNN(Deep Convolutional Neural Networks) with the architecture of the VGG network with the same network depth. With our implemented DCNN, this VGG framework has the same network parameters except for the substitution of 5 × 5 convolution filters and 2 × 2 stride pooling, and we refer to this model as VGG Net. In Table 3, the accuracy of the implemented DCNN and VGG classification on the crowdsourced COVID-19 sound dataset is given. The experiment results show that on the COVID-19 sound dataset, the proposed DCNN model often works better than the VGG Net.

The collection of the deep sound features does not need preprocessing in comparison with the conventional input for any duration of COVID-19 based respiratory sounds. The respiratory sounds can thus be specifically entered into the classification methods. Applying the DCNN model instead of VGG in this study greatly enhances the ability of the entire system model to interpret COVID-19 sound signals without ever using X-ray, CT scanning, and any other reports. The proposed method achieved reliable identification efficiency based on acoustic features [22]. The framework developed in this work is being used for regular research in everyday life or treatment modalities for the general public. If a disorder happens, then it is best to visit the clinic for more examination. This technique can only distinguish COVID-19 positive, Non-COVID-19, Asthma, Pertussis, Bronchitis, and Healthy from respiratory sounds (cough, voice sample, breath). At present, the technique could be succeeded with a limited number of samples. Therefore, in clinical practice, it has not been implemented but it could be used for advanced detection of illness with COVID-19. This COVID-19 crowdsource data was applied in clinical practice to diagnose COVID-19 disease by Cambridge University research people [1] and with the obtained dataset from the Cambridge University, an initial attempt was made to improve the performance. Further, develop the algorithms will be developed and sample selection for clinical practice will be made in the near future.



(a) (b)

(c) (d)

(e) (f)

(g) (h)

(i) (j)

(k) (l)

Fig. 10 Normalized Frequency with Leakage Factor and Boxplot representation for different Respiratory Sounds ((a) –Normalized Frequency Sound of COVID-19 Sample, (b) – Boxplot Diagram for Normalized COVID-19 Sample Sound, (c) –Normalized Frequency Sound of Asthma, (d) – Boxplot Diagram for Normalized Frequency Sound of Asthma, (e) –Normalized Frequency Sound of Pertussis, (f) – Boxplot Diagram for Normalized Frequency Sound of Pertussis, (g) –Normalized Frequency Sound of Bronchitis, (h) – Boxplot Diagram for Normalized Frequency Sound of Bronchitis, (i) –Normalized Frequency Signal for Healthy Sound, (j) – Boxplot Diagram for Healthy Sound, (k) – Normalized Frequency Sound of COVID-19 without Cough, (l) – Boxplot Diagram for COVID-19 without Cough Sound).

Table 3 The comparison of proposed convolutional network architecture with VGG Net model.

Task	Model	Modality	Accuracy (%)	F_1- Score (%)
Asthma/Non-Covid-19	DCNN	Breath + Cough	93.43	94.69
	VGG Net		88.36	89.78
Non-Coid-19/Covid-19	DCNN	Breath + Cough + Sample Voice	95.45	96.96
	VGG Net		87.48	90.26
Covid-19/Pertussis	DCNN	Cough	93.57	94.13
	VGG Net		86.86	89.13
Pertussis/Bronchitis	DCNN	Cough	93.86	94.13
	VGG Net		87.96	88.62
Bronchitis/Healthy	DCNN	Breath + Cough	94.00	94.89
	VGG Net		89.89	89.43

Table 4 The Classification Results Comparison of three COVID-19 tasks on respiratory COVID-19 sounds.

Task	Modality	Accuracy(%)	F_1- Score(%)
Positive cough of COVID-19 / Negative cough of COVID-19	Cough	93.12	94.13
Positive COVID-19/ Negative COVID-19	**Breath + Cough + Voice Sample**	**95.45**	**96.96**
Positive cough of COVID-19 / Negative COVID-19 Asthma Cough	Cough	93.43	94.69
COVID-19 positive / Pertussis	Cough	93.57	94.23
COVID-19 positive / Bronchitis	Cough	94.00	95.16

Table 5 Comparison of Proposed Model with Previous Models on this dataset.

Model	Dataset	Accuracy(%)
SVM + PCA [1]	COVID-19 Sound Crowdsourced Data	Task – I: 80 Task – II: 82 Task – III: 80
VGG Net withAugmentation [1]	COVID-19 Sound Crowdsourced Data	Task – II: 87 Task – III: 88
DCNN with Multi-Feature Channels and Augmentation	COVID-19 Sound Crowdsourced Data	Task – I: 93.43 ***Task – II: 95.45*** Task – III: 93.57 Task – IV: 93.86 Task – V: 94.00

4.2. Differentiating COVID-19 positive users with Non-COVID-19 users from respiratory sounds

Table 4 represents the classification results of five different tasks (positive cough of COVID-19/ negative cough of COVID-19, positive symptoms of COVID-19/negative symptoms of COVID-19, positive cough of COVID-19/negative COVID-19 cough with asthma) for five classes (Asthma, COVID-19, Pertussis, Bronchitis, Healthy). A classification network was developed that classified five tasks from five class categories. The first task represents binary classification used for discriminating whether the user is a COVID-19 positive or Non-COVID-19 user, the second task discriminates positive cough of COVID-19 from negative cough symptoms of COVID-19, and the third task is for discriminating positive cough symptoms of COVID-19 from negative COVID-19 asthma cough. The fourth and fifth tasks are discriminating COVID-19 positive / Pertussis, COVID-19 positive / Bronchitis. The exact accuracy comparison for classification results of five different tasks for five classes on respiratory sounds is shown in Fig. 11.

Table 5 displays the comparison results for different tasks for the proposed DCNN model with the existing SVM and VGG Net classifier [1]. The proposed approach achieves around 7% more accuracy than previous models. The accuracy comparison of the proposed model with different tasks and the previous VGG Net model is shown in Fig. 11. The accuracy comparison of the proposed model with the previous two models (SVM, VGG Net) on crowdsourced COVID-19 benchmark dataset is shown in Fig. 12.

SVM + PCA [1] Model (Task-I is Positive COVID-19/ Negative COVID-19, Task-II is Positive cough of COVID-19 / Negative cough of COVID-19, Task-II is Positive cough of COVID-19 / Negative COVID-19 Asthma Cough), VGG Net with Augmentation [1] Model (Task-II is Positive cough of COVID-19 / Negative cough of COVID-19, Task-II is Positive cough of COVID-19 / Negative COVID-19 Asthma Cough), DCNN with Multi-Feature Channels and Augmentation Model (Task-I is Asthma/Non-Covid-19, Task-II is Non-Coid-19/Covid-19, Task-III is Covid-19/Pertussis, Task-IV is Pertussis/Bronchitis, Task-V is Bronchitis/Healthy).

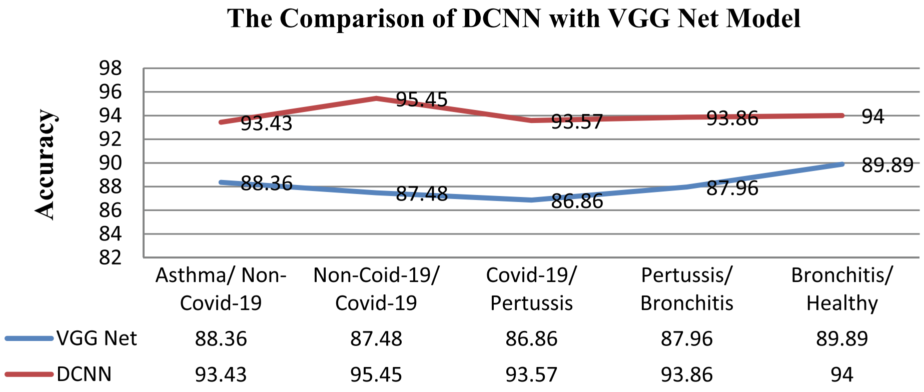


Fig. 11 The Comparison of Proposed DCNN Model with VGG Net for Five Tasks (Task1:Asthma/Non-Covid-19, Task 2:Non-Coid-19/Covid-19, Task 3:Covid-19/Pertussis, Task 4:Pertussis/Bronchitis, Task 5: Bronchitis/Healthy).

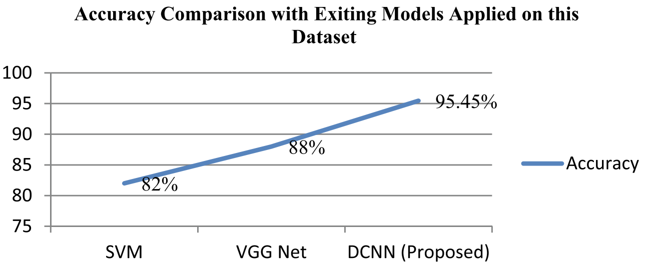


Fig. 12 The Accuracy Comparison of Proposed Model with Previous Two Models (SVM, VGG Net) on Crowdsourced COVID-19 Benchmark Dataset.

Dataset is obtained from Cambridge University on mutual agreement and the dataset has a fixed number of fewer samples. The benchmark data shows that there exist some discriminatory indications in the data while testing for COVID-19, user coughs mixed with breathing may be a good predictor. Specifically, test accuracy for Task-II is around 95% by combining breath sounds with cough and voice sounds, Task-I, Task-III, Task-IV achieves around 93% of accuracy, and Task-V achieves around 94% of test accuracy on COVID-19 crowdsourced data. We can diagnose COVID-19 disease with 95.45% accuracy on the COVID-19 crowdsourced dataset by using DCNN with multi-feature channels. TheDCNN is a proposed model for the classification of COVID-19 disease from the user samples.

5. Conclusion

The Artificial Intelligence (AI) based models entered into the real-world to diagnosis the COVID-19 symptoms from human-generated respiratory sounds such as voice/speech, cough, and breath. The Convolutional Neural Network (CNN) is used to solve many real-world problems with Artificial Intelligence (AI) based machines. In this work, the implemented DCNN (Deep Convolutional Neural Network) model to diagnose COVID-19 disease with human respiratory sounds collected from the COVID-19 sounds crowdsourced dataset. We have applied multi-feature channels to extract deep features of the acoustic respiratory sound signal instead of traditional approaches and perform around 7% better accuracy on the COVID-19 crowdsourced benchmark dataset. The model is classified as Asthma sounds, COVID-19 sounds, Pertussis, Bronchitis, and regular Healthy sounds using a DCNN classifier and shown around 95.45% of accuracy to detect the COVID-19 disease from respiratory sounds. The proposed model improves efficiently to classify COVID-19 sounds to detect COVID-19 positive symptoms. This work is intended to provide a working prototype to encourage support from the community accompanied by large-scale trials with more labeled results.

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.

References

[1] C. Brown et al, Exploring Automatic Diagnosis of COVID-19 from Crowd-Sourced Respiratory Sound Data", Accepted to KDD'20 (Health Day), ACM, ISBN 978 USA (Virtual Event) (2020), https://doi.org/10.1145/3394486.3412865.

[2] Jing Quian et al., "An early study on intelligent analysis of speech under COVID-19: Severity, Sleep Quality, Fatigue, and Anxiety", arXiv:2005.00096v2[eess.AS], 2020.DOI: 10.21437/Inter speech.2020-2223.

[3] Lara O. et al., "The COUGHVID crowdsourcing dataset: A corpus for the study of large scale cough analysis algorithms", arXiv:2009.11644v1[cs.SD], 2020. DOI: 10.5281/zenodo.4048312.

[4] Ali Imran et al., "AI4COVID-19: AI-Enabled Preliminary Diagnosis for COVID-19 from Cough Samples via an App", arXiv:2004.01275v6 [eess.AS], 2020. https://doi.org/10.1016/j.imu.2020. 100378.

[5] Tarik Alafif et al, Machine and Deep Learning towards COVID-19 Diagnosis and Treatment: Survey, Challenges, and Future Directions, J. LATEX Class Files 14 (8) (2020), https://doi.org/10.13140/RG.2.2.20805.47848/1.

[6] M. Bader et al., "Studying the Similarity of COVID-19 Sounds based on Correlation Analysis of MFCC", arXiv:2010.08770v1 [cs. SD], 2020. DOI: 10.1109/CCCI49893.2020.9256700

[7] X. Jiang et al., "Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough", December 2020. arXiv:2011.13320v2[cs.SD]

[8] J. Shuja et al, "COVID-19 open-source data sets: a comprehensive survey, Appl Intell. page 1–30 (2020), https://doi.org/10.1101/2020.05.19.20107532, PMC7503433.

[9] M. Al Ismail et al., "Detection of COVID-19 through the Analysis of Vocal Fold Oscillations", arXiv:2010.10707v1 [eess. AS],2020.

[10] J. Rasheed et al, A Survey on Artificial Intelligence Approaches in Supporting Frontline Workers and Decision Makers for COVID-19 Pandemic, Chaos, Solitons Fractals 141 (2020) 110337, https://doi.org/10.1016/j.chaos.2020.110337.

[11] J. Sharma et al., "Environment Sound Classification using Multiple Feature Channels and Attention-based Deep Convolutional Neural Network", arXiv:1908.11219 [cs.SD], 2020.

[12] J. Laguarta, F. Hueto, B. Subirana, COVID-19 Artificial Intelligence Diagnosis using only Cough Recordings, IEEE Open J. Eng. Med. Biol. (2020), https://doi.org/10.1109/OJEMB.2020.3026928, Sept.

[13] M. Aykanat, Ö. Kılıç, B. Kurt, et al, Classification of lung sounds using convolutional neural networks, J. Image Video Proc. (2017) 65, https://doi.org/10.1186/s13640-017-0213-2.

[14] A. Hassan, I. Shahin, and M. B. Alsabek, "COVID-19 Detection System using Recurrent Neural Networks," 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI), Sharjah, United Arab Emirates, pp. 1-5, Nov. 2020. DOI: 10.1109/CCCI49893.2020.9256562.

[15] Schmidt M., Berg E., Friedlander, M. & Murphy K., "Optimizing Costly Function with Simple Constraints: A Limited Memory Projected Quasi-Network Algorithm". Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, in PMLR 5:456-463. (2009)

[16] M. Slaney, An efficient implementation of the Patterson-Holdsworth auditory filter bank, Apple Comput Percept. Group Tech. Rep (1993).

[17] T.F. Quartieri, T. Talker, J.S. Palmer, A Framework for Biomarkers of COVID-19 Based on Coordination of Speech-Production Subsystems, IEEE Open J. Eng. Medi. Biol. 1 (2020) 203–206, https://doi.org/10.1109/OJEMB.2020.2998051.

[18] Y. Wang et al., "Abnormal Respiratory Patterns Classifier May Contribute to Large-Scale Screening of People Infected With COVID-19 in an Accurate and Unobtrusive Manner", arXiv:2002.05534 [cs. LG], Feb. 2020.

[19] Zheng Jiang et al., "Combining Visible Light and Infrared Imaging for Efficient Detection of Respiratory Infections Such as Covid-19 on Portable Device", arXiv:2004.06912 [cs.CV], Apr. 2020. Bibcode:2020arXiv200406912J.

[20] R. Liu, S. Cai, K. Zhang, and N. Hu, "Detection of Adventitious Respiratory Sounds based on Convolutional Neural Network," 2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Shanghai, China, 2019, pp. 298-303, DOI: 10.1109/ICIIBMS46890.2019.8991459.

[21] H. Pasterkamp, S.S. Kraman, G.R. Wodicka, Respiratory sounds: Advances beyond the stethoscope, Am. J. RespirCrit. Care Med. 156 (3 Pt 1) (1997 Sep) 974–987, https://doi.org/10.1164/ajrccm.156.3.9701115, PMID: 9310022.

[22] Kimberly L. Dahl et al, Acoustic Features of Transfeminine Voices and Perceptions of Voice Femininity, J. Voice 34 (6) (2019) 961.e19–961.e26, https://doi.org/10.1016/j.jvoice.2019.05.012.

[23] J. Salamon, J.P. Bello, Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification, IEEE Signal Process Lett. 24 (3) (March 2017) 279–283, https://doi.org/10.1109/LSP.2017.2657381.

[24] A. Khamparia, D. Gupta, N.G. Nguyen, A. Khanna, B. Pandey, P. Tiwari, Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network, IEEE Access 7 (2019) 7717–7727, https://doi.org/10.1109/ACCESS.2018.2888882.

[25] J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, Multimodal Neuroimaging Feature Learning With Multimodal Stacked Deep Polynomial Networks for Diagnosis of Alzheimer's Disease, IEEE J. Biomed. Health. Inf. 22 (1) (Jan. 2018) 173–183, https://doi.org/10.1109/JBHI.2017.2655720.

[26] Yinghui Huang, SijunMeng, Yi Zhang, et al. 2020. The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods.medRxiv (2020). https://doi.org/10.1101/2020.04.07.20051060 12 pages.

[27] Li, F., Liu, M., Zhao, Y. et al. 'Feature extraction and classification of heart sound using 1D convolutional neural networks', EURASIP J. Adv. Signal Process. 2019, 59 (2019). https://doi.org/10.1186/s13634-019-0651-3

[28] V. Klára, I. Viktor, M. Krisztina, Voice Disorder Detection on the Basis of Continuous Speech, in: Á. Jobbagy (Ed.), 5th European Conference of the International Federation for Medical and Biological Engineering, Springer, Berlin, Heidelberg, 2011, https://doi.org/10.1007/978-3-642-23508-5_24.

[29] L. Verde, G. De Pietro, G. Sannino, Voice Disorder Identification by Using Machine Learning Techniques, IEEE Access 6 (2018) 16246–16255, https://doi.org/10.1109/ACCESS.2018.2816338.

[30] Md. Sahidullah, GoutamSaha, "Design, analysis and experimental evaluation of block-based transformation in MFCC computation for speaker recognition", Speech Communication, Volume 54, Issue 4, 2012, Pages 543-565, ISSN 0167-6393, https://doi.org/10.1016/j.specom.2011.11.004.

[31] Srinivasamurthy, RavisuthaSakrepatna, "Understanding 1D Convolutional Neural Networks Using Multiclass Time-Varying Signals", All Theses. 2911.https:// tigerprints.clemson.edu/all_theses/2911, (2018).

[32] L Brabenec, J Mekyska, Z Galaz, and Irena Rektorova. 2017. Speech disorders in Parkinson's disease: Early diagnostics and effects of medication and brain stimulation. Journal of Neural Transmission 124, 3 (2017), 303–334.

[33] BetulErdogdu Sakar, GorkemSerbes, and C. Okan Sakar. 2017. Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. PloS One 12, 8 (Aug. 2017), 1–18. https://doi.org/10.1371/journal.pone.0182428

[34] World Health Organization, 2020. Coronavirus disease 2019 (covid-19): https://www.who.int/.

[35] Kranthi Kumar Lella, Alphonse PJA. (2021) A literature review on COVID-19 disease diagnosis from respiratory sound data[J]. AIMS Bioengineering, 2021, 8(2): 140-153. doi: 10.3934/bioeng.2021013

[36] Shui-Hua Wang, Vishnu Varthanan Govindaraj, Juan Manuel Go´rriz, Xin Zhang, Yu-Dong Zhang, Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network, Info. Fusion 67 (208–229) (2021) 1566–2535, https://doi.org/10.1016/j.inffus.2020.10.004.

[37] Shui-Hua Wang and Xiaosheng Wu and Yu-Dong Zhang and Chaosheng Tang and Xin Zhang. (2020) Diagnosis of COVID-19 by Wavelet Renyi Entropy and Three-Segment Biogeography-Based Optimization, International Journal of Computational Intelligence Systems, volume-13, issue-1, pages:1332-1344, issn: 1875-6883, doi: https://doi.org/10.2991/ijcis.d.200828.001

[38] S.J. Malla, P.J.A. Alphonse, COVID-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets, Applied Soft Computing, Volume 107, ISSN 107495 (2021) 1568–4946, https://doi.org/10.1016/j.asoc.2021.107495.

[39] Lella Kranthi Kumar, and Alphonse PJA, 2021. Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: cough, breath, and voice, AIMS public health, vol. 8,2 240-264. 10 Mar. 2021, doi:10.3934/publichealth.2021019.