Eur. Phys. J. Spec. Top. (2022) 231:3329–3346
https://doi.org/10.1140/epjs/s11734-022-00432-w

THE EUROPEAN
PHYSICAL JOURNAL
SPECIAL TOPICS

Regular Article

COVID-19 disease diagnosis with light-weight CNN using modified MFCC and enhanced GFCC from human respiratory sounds

Lella Kranthi Kumar[a] and P.J.A. Alphonse[b]

Health Analytics Research Labs, Department of Computer Applications, NIT Tiruchirappalli, Tiruchirappalli, Tamil Nadu 620015, India

Received 15 September 2021 / Accepted 18 December 2021 / Published online 24 January 2022
© The Author(s), under exclusive licence to EDP Sciences, Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract In the last 2 years, medical researchers and clinical scientists have paid close attention to the problem of respiratory sound classification to classify COVID-19 disease symptoms. In the physical world, very few AI-based (Artificial Intelligence) techniques are often used to detect COVID-19/SARS-CoV-2 respiratory disease symptoms from the human respiratory system-generated acoustic sounds such as acoustic voice sound, breathing (inhale and exhale) sounds, and cough sound. We propose a light-weight Convolutional Neural Network (CNN) with Modified-Mel-frequency Cepstral Coefficient (M-MFCC) using different depths and kernel sizes to classify COVID-19 and other respiratory sound disease symptoms such as Asthma, Pertussis, and Bronchitis. The proposed network outperforms conventional feature extraction models and existing Deep Learning (DL) models for COVID-19/SARS-CoV-2 classification accuracy in the range of 4–10%. The model's performance is compared with the COVID-19 crowdsourced benchmark dataset and gives a competitive performance. We applied different receptive fields and depths in the proposed model to get different contextual information that should aid in classification. And our experiments suggested 1 × 12 receptive fields and a depth of 5-Layer for the light-weight CNN to extract and identify the features from respiratory sound data. The model is also trained and tested with different modalities of data to showcase its effectiveness in classification.

1 Introduction

The World Health Organization (WHO) announced COVID-19/SARS-CoV-2 epidemic was the biggest pandemic in the entire world in March 2020. It claims over 4,539,723 lives worldwide as of September 3rd, 2021. As of September 3, 2021, there have been 218,946,863 confirmed COVID-19 cases, 4,539,723 deaths, and 5,289, 724,918 are completed vaccinations as of August 31st [1]. According to the biomedical experts, data collection and tracing contacts is very difficult for spreading the COVID-19 disease. Although advancements in testing have increased the popularity of these methods in recent months, there is an essential need for COVID-19 screening technology that is inexpensive, quick, and flexible. The severity of the COVID-19 virus is categorized into three parts: mild, moderate, and extreme. Over the last year, biomedical scientists and researchers have paid special attention to identifying abnormalities in the classification of respiratory sounds [2] and COVID19 disease diagnosis [3]. Many AI-based frameworks have joined the real world to solve this problem [4–6]. Biomedical researchers and scientists have proposed various Deep-Learning (DL), Signal-Processing (SP), and Machine-Learning (ML) approaches to diagnosis various diseases from human respiratory-generated sounds [7–9]. Recent research focused on prognostic models diagnostic models, and pertained ensemble models for the first wave, second wave, and third waves in India to identify SARS-CoV-2/COVID-19 disease symptoms [10–13].

The COVID-19 epidemic is now widespread in reality, creating fear in people's opportunity to connect physically. As a result, several techniques are used to detect respiratory COVID-19 disease symptoms with respiratory generated sounds [14]. Biomedical experts used respiratory sounds (lung sounds, cough, breath, voice, food absorption, body vibration, sighs, and heart sound) to diagnose the different human diseases (Asthma, Bronchitis, Pertussis, and SARS-CoV-2/COVID-19) [15]. In recent times, such signals were commonly extracted during clinical interactions through manual auscultation. Biomedical scientific and public health researchers have already officially started to use electronic methods for collecting audio sounds from the human body and perform automated analyses of infection of the respira-

[a] e-mail: kranthi1231@gmail.com (corresponding author)
[b] e-mail: alphonse@nitt.edu

tory sound data, such as recognizing wheeze in patients with Asthma [16].

Researchers have also performed experiments with using a person's speech to aid in the early diagnosis of many illnesses: Alzheimer's, Parkinson's can have many effects on the voice, the intensity of speech with cardiovascular disease, unidentifiable abnormalities like fatigue, head injuries, and psychological conditions directly relate with voice tone, voice style, voice rhythm, and intensity [17–19]. The use of such a respiratory system sounds like early diagnosis treatment for future diseases holds enormous promise for early identification and low-cost response that can be made available to the general public if integrated into primary commodities. It is common for individuals if the remedy can be discreetly monitored in different individuals during their daily lives [20]. Over the last year, the performance of respiratory acoustic sound classification of abnormalities on the respiratory COVID-19 sounds dataset can be built using different Artificial Intelligence (AI) techniques.

Recent studies have started to investigate how respiratory sound records can be classified using smart devices from patient respiratory sound data like breath auscultation sound, cough sound, heartbeat sound, and lung auscultation [21,22]. Deep learning and Machine Learning methods with various feature extraction techniques such as Data De-noising auto-encoder, Mel-frequency coefficients, LSF (Line Spectral Frequency), DCT (Discrete Wavelet Transformation) and Gamatone frequency [23]. Till then, the framework can only automatically identify the patient's condition and analyzes the illnesses of the concerned patients from human respiratory sounds. This is not the scenario in our research, which involves studying human respiratory sounds in unregulated crowdsourced records to identify COVID-19 disease. This research majorly focuses on respiratory sound classification and diagnosis of different human respiratory diseases (Asthma, COVID-19, Pertussis, and Bronchitis) using light-weight CNN model with Modified-Mel-frequency Cepstral Coefficients (M-MFCC) and Enhanced-Gamma-tone Frequency Cepstral Coefficients (EGFCC) feature extraction technique.

This research work is arranged in the following sections: Sect. 2 demonstrates the existing approaches on the identification of COVID 19 respiratory disease. Sect. 3 describes the dataset preparation and preprocessing methods such as Modified-Mel-frequency Cepstral Coefficients (M-MFCC) and Enhanced-Gamma-tone Frequency Cepstral Coefficients (EGFCC) implemented using the proposed light-weight CNN model to extract the best features from respiratory sound data. Sect. 4 summarized the performance of the proposed light-weight CNN model and compared it with current techniques. And finally, we have concluded this research with preliminary findings using the proposed model performance.

2 Background work

Both extractions of audio features and sound extraction have a long history. As a result, many other studies covering such essential issues were published. The more relevant studies focus on a single sound problem domain, including COVID-19 and other respiratory diseases with human respiratory sound data and encompass a small set. Following are summaries of significant studies in the ground of sound features extraction.

Brown Chloe and the team implemented an Android application to gather respiratory COVID-19 sound records from human respiratory sound data of over 200 COVID-19 positives from over 7000 distinctive users. The authors have implemented majorly three binary tasks (Task-I: COVID-19_Positive/Asthma_Cough, Task-II: COVID-19_Positive_Cough/COVID-19_Nega tive _Cough, Task-III: COVID-19_Positive_Cough/Heal thy_Cough) on a crowdsourced respiratory dataset. Task-I achieves around 80% precision from 220 unique users with breath and cough modalities using the SVM model; Task-II and Task-III achieve approximately 88% accuracy with VGGNet [24].

Lara et al. [25] proposed an Artificial Intelligence (AI) based framework for analysis of COVID-19 symptoms with cough sounds from a crowdsourced "COU GHVID" dataset. The COUGHVID dataset contains over 20k user-generated cough audio representing a wide scope of subject age, geographical area, gender, and SARS-CoV-2/COVID-19 prior medical history. The authors collected around 120 cough sound signals and 95 other respiratory sound signals first-hand, including speech, laughing, acquiescence, and various other background murmuring sound noises to improve the classification performance. Menghan et al. [26] implemented a model framework to detect COVID-19 disease from extensive scale screening of the different peoples with different breathing auscultation patterns that helps in the physical world. In this research, an accurate model for the respiratory system is implemented to bridge between a massive volume of training input data and insufficient real-world existing data to identify the COVID-19 features from human respiratory sound data.

Imran et al. proposed Artificial Intelligence (AI) based framework with the name AI4COVID-19 [27] to identify COVID-19 symptoms from the cough sound signals. In this work, the authors have been implemented a mobile application to collect the respiratory sound data; after that, they have applied AI techniques to classify COVID-19 symptoms from the respiratory sound data. Badar et al. [28] introduced a substantial framework that uses Sound Signal Processing techniques and Mel-frequency coefficients to extract the best sound features from COVID cough sounds and

Non-COVID cough sounds. This research shows a high correlation between COVID cough samples and Non-COVID cough samples using the MFCC technique. The authors proved that this MFCC method is better for extracting the features from COVID respiratory cough sound Non-COVID respiratory cough sounds.

Gunvant et al. [29] implemented one mobile application to collect the respiratory sound data to recognize SARS-CoV-2 symptoms. The authors collected the data using a mobile application and then applied Deep Learning (DL) models to classify COVID-19 symptoms from the cough sounds dataset. However, most of the Deep Learning (DL) networks have been trained on respiratory sound data from various file structures and audio data configurations in multiple environments. Rita et al. [30] implemented a framework for analyzing vocal fold parameters to identify SARS-CoV-2 symptoms because most of the symptoms of COVID-19 are concerned with respiratory sounds. The COVID-19 symptoms can be detectable by analyzing vocal fold parameters, and authors have contributed their innovations to detecting COVID-19 symptoms with vocal fold parameters from voice and sound data.

Lagarta et al. [31] introduced a Deep Learning (DL) framework to classify respiratory COVID-19 symptoms from human cough sound samples. The model enables an alternative to the initial screening of SARS-CoV-2 symptoms from human respiratory sound samples across the world at no cost. Hasan et al. [32] proposed a Recurrent Neural Network (RNN) model using Signal Processing (SP) techniques for the classification of SARS-CoV-2 symptoms from pulmonary sounds like a cough, speech, and breathing sounds. This pertained model is used to identify SARS-CoV-2 disease symptoms by evaluating the sound features from the respiratory cough sounds.

Tanya et al. [33] introduced a biomarkers framework for classifying SARS-CoV-2 /COVID-19 symptoms using speech processing techniques and Signal Processing (SP) techniques from respiratory sound data. This existing pertained model extract features from lower inflammation and upper respiratory inflammation parameters, which helps detect COVID-19 symptoms. Lella and Alphonse [34,35] implemented an Artificial Intelligence (AI) framework to automatically classify COVID-19 symptoms using a 1D CNN network and Deep CNN from respiratory sound data. The 1D CNN model is implemented using a Data De-noising Auto-Encoder (DDAE) mechanism, which helps to give good performance to identify SARS-CoV-2 disease symptoms from various respiratory sounds (cough, voice, and breath). All of these background studies indicate that there is no appropriate prediction method for classifying SARS-CoV-2 disease symptoms.

In recent years, there has been a lot of research in the area of human respiratory sound analytics. However, we discovered that far too many studies concentrated on a very small no. of regularly utilized basic features, while the most newly respiratory sound features are discussed very infrequently. In contrast to previous studies, we focus on deep sound extracting features, covering a broader range of respiratory sound features, and incorporating some framework into the field. As per previous studies, we introduce the light-weight network model with modified Mel-Frequency coefficients and enhanced Gama-tone frequency techniques to function more effectively on SARS-CoV-2/COVID-19 sounds data to classify SARS-CoV-2 disease symptoms, and it enhances to perform much better on the crowdsourced dataset.

3 Materials and methods

3.1 Dataset

The respiratory COVID-19 sounds crowdsourced dataset was collected with mutual agreement from Cambridge University for research purposes. It was authorized at Cambridge University in the department of computer science and technology by adhering to all ethical committee guidelines. Brown et al. [24] created an android based mobile application and a browser-based web application to obtain respiratory COVID-19 sounds; the fundamental features of this application are nearly identical. The user's past medical background has been compiled for those who have previously been admitted to the treatment center. Users are then asked to enter their symptoms and collect the three respiratory audio sounds (breath sound for 30 s, cough sound for 3–5 intervals, sample voice of reading a single sentence). Further to that, the users have to enter additional symptoms and sounds to provide a unique opportunity to investigate and discover the previous medical history of the user. This data is extremely secure in the local data centers of the University of Cambridge; the acquired information is stored internally until it is linked with the Wi-Fi network. The data is removed from the database if the required data is received from the user's device. If a user requests that information be removed from the server, it will be updated.

Kun et al. implemented an intelligent technology on voice data by taking different considerations like the quality of sleep, anxiety, severity, and fatigue [36]. The Cambridge University researchers have acquired data from the "COVID-19 sounds" applications, and Mellon University research scientists have been obtained data from the "COVID-19" application. This preliminary study yielded 378 components and 256 audio files for feature analysis, according to the authors. These 256 audio sound features were gathered from approximately fifty COVID-19 infected patients. In this work, the poly signals with a sampling frequency rate of 0.016MHz have been converted for two acoustic sound feature sets.

3.1.1 *Collection of human respiratory COVID-19 sounds*

The Cambridge University researchers gathered around 25,000 samples from an android-based application and 45,000 samples from a web-based application. The researchers have collected around 6000 and 5000 samples from various countries. Approximately about 325 users reported COVID-19 positivity from both datasets. The Android app collects multiple samples from different users. It creates redundant information and vast data, and further work will eliminate this to improve performance. They gathered and analyzed basic details (historical and contemporary health records, age, and gender) and three different audio respiratory data (speech sound, cough sound, breath sound) from distinctive uses through web-based and android-based applications. In this classification, the most widely accepted symptoms to identify COVID-19 disease are dry cough and sore throat while coughing. Surprisingly, the most normal observed symptoms are wet-cough, dry-cough, as well as loss of smell, and tightness of the chest being the most widely known mixed symptom. These symptoms will match with information obtained from the respiratory COVID-19 disease monitor. The truth is that the human respiratory cough sound is one of the commonly identified symptoms of a respiratory disease like COVID-19 ads to the case for using respiratory sound as the distinct symptom of the human being. Nonetheless, it is a common symptom of a variety of many other diseases. As a result, the light-weight CNN network model is used to identify and diagnose respiratory COVID-19 disease based on all of these symptoms.

3.2 Feature extraction method representation

The model have been developed with two feature extraction methods such as MMFCC (Modified Multi-frequency Cepstral Coefficients) and EGFCC (Enhanced Gamma-tone Frequency Cepstral Coefficients) to extract in-depth features from human respiratory sound data (breathing sounds, cough sounds, and voice sounds) [37,38]. The system model compared with extracted deep features according to the system metrics. The light-weight CNN gives distinct features and similar features for efficiently identifying respiratory sound acoustic signals by incorporating different Signal Processing (SP) methods to gather various types of feature attributes. The EGFCC method gives transient respiratory audio sound features, while MMFCC serves as the base for extracting in-depth features in this work. We have compared the light-weight CNN model with EGFCC features and MMFCC deep features concerning the model performance.

3.2.1 *Enhanced Gamma-tone filter bank implementation*

The Gammatone frequency filter banks are indeed a group of cochlear modeling filter banks [39,40]. A Gammatone filter bank frequency selection is very close to the characteristics of an average human ear filter. It can depict the components of various actions in the Gammatone filter banks. A gammatone determines the sampling frequency rate of a Gammatone filter and a sinusoidal sound, the primary frequency of that is ‘f_c’, which can be depicted as Eq. (1).

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

where B is the bandwidth ($B = ERB\left(f_c\right) + 1.019$), amplitude gain is A, core frequency is f_c, the order of the filter depends on n value, and φ is the shift-phase value. We built the model with a fourth-order filter because the fourth-order gamma-tone frequency filter is similar to the features that are being used to represent the human respiratory sound auditory filters [20].

The model has created by setting the order of a filter bank as four because the fourth-order Gamma-tone filter banks are identical to the features that are being used to represent human respiratory sound auditory filters [20]. The ERB is just a frequency modulation index that indicates the frequency spectrum band of the human respiratory sound acoustic filter range at each point along the cochlea. The cubic band-pass frequency filters are used to define human standard hearing frequency bandwidth, which is a very unrealistic and correct simplified representation of cubic modeling prediction. The acoustic frequency filters bandwidth is the frequency band value of an Equivalent Rectangular Bandwidth (ERB) primarily focused at frequency f, and the relation between the f and ERB is the particular scale factor as shown in Eq. (2).

$$ERB_s\left(f\right) = 21.40 \log_{10}\left(\left(0.0043 * f\right) + 1\right). \tag{2}$$

The Gamma tone frequency filter banks should contain every feature to model the Gamma tone frequency sound band accurately. The fundamental frequency bands of each Gamma tone standard filters are adjusted in this work using Eq. (2), and Gamma tone features can be calculated with Eq. (3) from the human respiratory sounds.

$$f_{c_i} = ERB_s^{-1}\left(\left(k_i \times \frac{ERB_s\left(f_{high}\right) - ERB_s\left(f_{low}\right)}{N}\right)\right. \\ \left. + ERB_s\left(f_{low}\right)\right). \tag{3}$$

The inverse of the ERB_s is described as ERB_s^{-1}, f_{high} is the highest frequency (20000 Hz), f_{low} is considered as lowest frequency (10Hz) i is the index number, k_i filter index and N is the total no. of gamma-tone filters. We can compute the enhanced gamma-tone frequency coefficients with gamma-tone filter banks. A technique to measure the enhanced GFCC relates to a feature extraction technique of modified Mel-frequency coefficient. The length of the respiratory sound frame is defined as 25ms by default in this work. The frame response is then evaluated using the Fast Fourier Transformation (FFT) for every frame. The Gamma tone, band-pass frequency filters can be assessed based on frame reactions at each point of the frequency filter. As a result, the gamma-tone frequency band-pass filters are used as an input signal for FFT to reach the sub-band

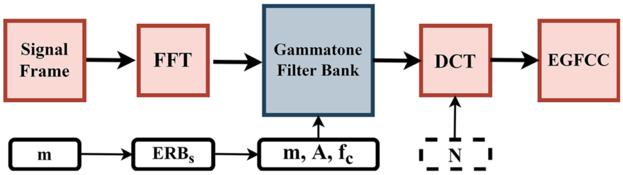


Fig. 1 The block diagram of enhanced gamma-tone frequency filter bank (N-extracted coefficients, m-filterbanks, and the human respiratory sound signal is the frame signal)

spectrogram. The sub-band filter of the spectrogram is depicted as Y_n to compute sub-band energy. The Mel-frequency logarithmic function and DCT (Discrete-Cosine-Transformation) are used to emulate the perception of cough sound wave (loudness). They are unrelated to the outcome of the frequency compact filters in the final step, and the EGFCC can be calculated using Eq. (4).

$$GFCC_n = \sqrt{\frac{2}{M}} \left(\sum_{m=1}^{M} \log_{10}(Y_n) \cdot \cos\left[\frac{\pi m}{M} \left(n - \frac{1}{2} \right) \right] \right), \quad 1 \le n \le N, \tag{4}$$

M defines as total no. of gamma-tone filters, sub-band energy is m, $GFCC_n$ defines the no. of gamma-tone frequency filters and it always depends upon the range of gamma-tone frequency ($1 \le n \le N$), Y_n is define as m^{th} sub-band energy. The enhanced gamma-tone frequency filer block diagram (N-extracted coefficients, m-filterbanks, and the human respiratory sound acoustic signal is the model window frame signal) is represented in Fig. 1.

3.2.2 *Preparation of Modified-Mel-Frequency Cepstral Coefficients (M-MFCC)*

The Modified-Mel-Frequency Cepstral Coefficients (MMFCC) is applied to obtain acoustic depth respiratory sound features from the input human respiratory audio sound data (cough sound, breath sound, or acoustic voice sound) it exponentially scales the frequency. This MMFCC model consists of seven different steps, which are represented below. The modified Mel-Frequency model with scale frequency structural framework is illustrated in Fig. 2.

Step-1: Because of the rapid increase in the acoustic sound signals, the acoustic sound spectrogram signal is structured in short frames at this stage. This is not much smaller or much bigger than having a perfect spectral estimate window frame. It is then done to remove the disruptions at the start and the end of a window frame. $W_j(n), 0 \le n \le M_n - 1$ is the window, quantity samples of each frame are 'M_n', i and j represents the index numbers, $X(n) = Y(n) \times W_j(n)$ will be represented as an output frame with the interval of $\le n \le M_n - 1$. Finally, we will get the '$X(n)$' output signal by multiplying frame window of '$W_j(n)$' and '$Y(n)$' input signal. The frame window of '$W_j(n)$' representation is shown in the Eq. (5).

$$W_j(n) = 0.54 - 0.46 \sin\left(\left(\frac{\pi}{2} \right) - \left(\frac{2\pi n}{M_n - 1} \right) \right), \quad 0 \le n \le M_n - 1. \tag{5}$$

Step-2: The FFT (Fast Fourier Transform) is used to convert the frequency/spectral domain from the spatial or time domain of the input respiratory input acoustic sound signal. Each and every frame contains 'M_n' respiratory sound samples are transformed to the frequency or spectral domain. The respiratory sound frame representation shows as $\sum_{n=0}^{M_n} t(n)$, 'M_n' represents the number of samples (around 160 samples), and the frequency and time domain indexes 'f' and 'n' are shown in the Eq. (6).

$$|T(f)|^2 = \left| \sum_{n=0}^{M_n - 1} t(n) \cdot e^{\left(\left(\frac{-j2\pi n f}{M_n} \right) \right)} \right|^2. \tag{6}$$

Step-3: After creating the frames, we must estimate the power spectral density for every frame by determining the periodogram with Eq. (7), where i represents the power spectral density for every frame and j represents the window index numbers.

$$P_j(i) = \frac{1}{M} \times |S_j(i)|^2. \tag{7}$$

The spectral power estimation still contains essential data not required for ASR (Automated Sound Recognition) of a human respiratory sound acoustic audio system, so the variation in the two frequently separated frequency ranges is not visible. The Mel-frequency filters are being used to determine how often periodogram will exist in various frequency areas in the respiratory sound spectrogram. The very first filter bank of the Mel frequency can show how frequently the power spectrum exists near to zero Hertz (minimal). The first filter is straightforward, and the Mel-scale provides information about establishing the frequency filters and how substantial to build them, as shown in Eq. (8). The spectrogram filter frequency function is denoted by 'f_a'.

$$f_a(y) = \begin{cases} 0, & y < f_a(n-1) \\ \frac{y - f_a(n-1)}{f_a(n) - (f_a(n-1))}, & f_a(n-1) \le y \le f_a(n) \\ \frac{f_a(n+1) - y}{(f_a(n+1)) - f_a(n)}, & f_a(n) \le y \le f_a(n+1) \\ 0, & y > f_a(n=1) \end{cases} \tag{8}$$

Step-4: In this case, we must determine the logarithmic significance of the active layer using Eq. (9), which demonstrates the transformation of the acoustic sound frequency band to the Mel- frequency filter spectrum scale.

$$M_f(y) = (2595) \times Log._{10}\left(1 + \left(\frac{y}{100} \right) \right). \tag{9}$$

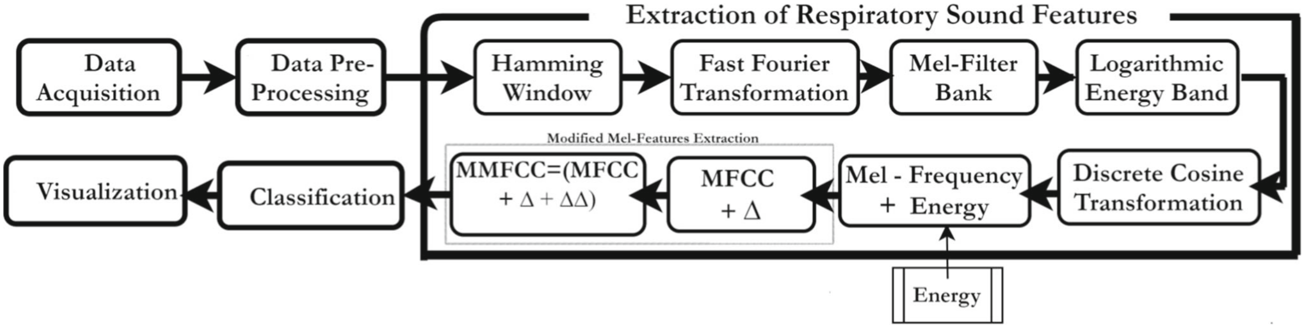


Fig. 2 Modified Mel Frequency Cepstral Coefficient (MMFCC) structural frequency framework

Step-5: In this process, we must have to calculate the DCT (Discrete-Cosine-Transformation) for the logarithmic filter band energy of the power spectrum. The frequency-filter spectrum power bank energies are linked with Mel-frequency filter spectrum scale. The DCT (Discrete-Cosine-Transformation) can be calculated using Eq. (10). As 'C_n' determined in Eq. (10), the frequency cepstral sound features will be identified with the range of $k = 0$ to $S - 1$, whereas $S = 1, 2, 3 \ldots, n$ and constant factor value is 'c' for discretization.

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

Step-6: In this process, we have to calculate the MFCC (Mel-Frequency Coefficients) for the discretized logarithmic filter band energy of the power spectrum. The frequency-filter spectrum power bank energies are linked, and the newly implemented framework filter banks will all be overlapped. So, the frequency Mel-spectrogram can be calculated using Eq. (11), the frequency cepstral sound features will be identified with the range of $k = 0$ to $S - 1$, whereas $S = 1, 2, 3, \ldots .m$ and constant factor value is 'c' for discretization.

$$C_n = \left(2\sqrt{\left(\frac{2}{S}\right)}\right) \sum_{k=0}^{S-1} \left((\log_{10}[c \times (k+1)]) \cdot \sin\left(\frac{\pi}{2}\right) - \left[n \times \left(\frac{2k-1}{2}\right) \cdot \frac{\pi}{S}\right]\right). \tag{11}$$

Step-7: At the end of the process, we have to calculate the Mel-frequency (MMFCC) with logarithmic function for the logarithmic filter band energy of the power spectrum. The frequency-filter spectrum power bank energies are linked, and the newly implemented framework filter banks will all be overlapped. Then the processed input sound of the Modified-Mel-frequency (MMFCC) is identified after DCT. As 'C_n' determined in Eq. (12), the frequency cepstral sound features will be identified with the range of $k = 0$ to $S - 1$, whereas $S = 1, 2, 3 \ldots, n$ and constant factor value is 'c' for discretization.

$$C_n = \left(2\sqrt{\left(\frac{2}{S}\right)}\right) \sum_{k=0}^{S-1} \left((\log_{10} . (\log_{10}[c \times (k+1)])) \cdot \sin\left(\frac{\pi}{2}\right) - \left[n \times \left(\frac{2k-1}{2}\right) \cdot \frac{\pi}{S}\right]\right). \tag{12}$$

3.3 The CNN model

The light-weight CNN (Convolutional Neural Network) classifies the human respiratory audio sounds based on the different respiratory diseases (Normal Flue, Asthma, Pertussis, Negative_ COVID-19, Positive_CO VID-19, Bronchitis, and Healthy human respiratory sound features) using enhanced gamma-tone frequency filter banks and modified Mel-frequency feature extraction channels. Figure 3 depicts the proposed convolutional model architecture with two feature extraction techniques (EGFCC and MMFCC). In this work, we have implemented a light-weight convolutional model framework to diagnose human respiratory disease with human respiratory-generated sound data. The model shows comparative performance analysis using EGFCC and MMFCC methods. The proposed light-weight CNN model includes two pooling layer operations, three convolutional (kernel) layers, and two fully connected (dense) layers to protect the convolutional network layers.

Figure 4 depicts the light-weight CNN network flow structure, which accepts time-stamps as input from the input layer kernel filter values. The model consists of pooling and convolutional layers to extract the deep respiratory sound layer-by-layer. The light-weight CNN layers are made up of various convolution kernels of different sizes. The results are categorized by the fully connected layers after max-pooling using the batch normalization technique. Our idea is to build a light-weight CNN network that is fully connected.

In this work, we have used the cross-entropy loss function on the one-hot encoded output to calculate model training loss. The single respiratory sound spectrogram one-hot encoded loss function looks like in the below equation (Eq. (13)). Where N is the number of classes, c represents class value, x is the vector value of input respiratory sound spectrogram, x_c is either 1 or 0

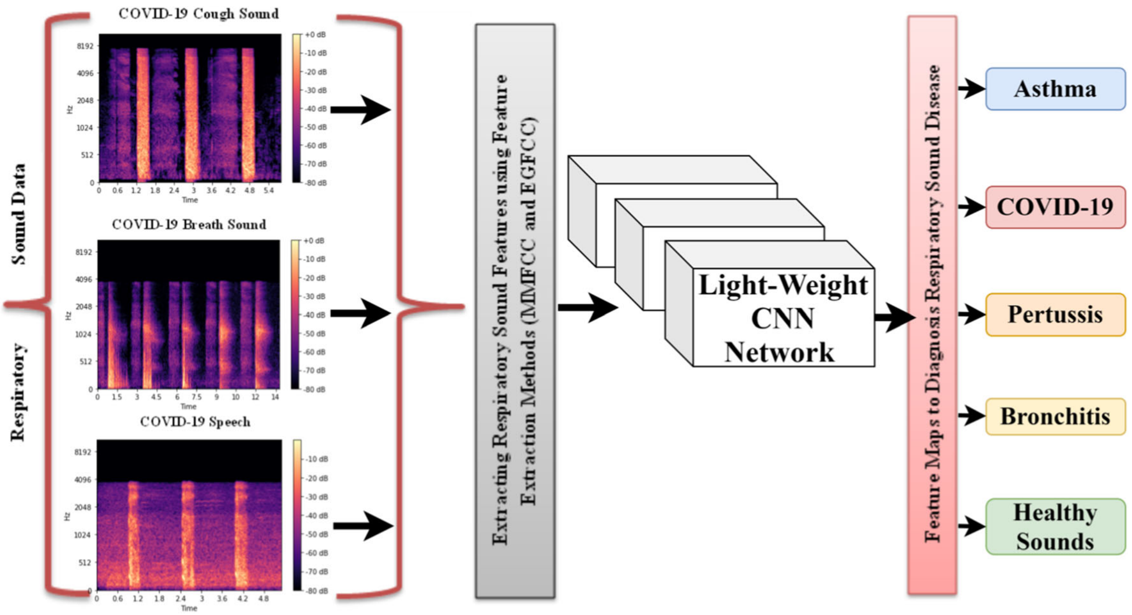


Fig. 3 The proposed convolutional model architecture with two feature extraction techniques (EGFCC and MMFCC)

(we have given x_c = 1), $\hat{x}_c$ is model prediction class (it represents the output of the “Softmax’ function for ‘c’ classes).

$$L_i = -\sum_{c=1}^{N} x_c \times \log_{10} \hat{x}_c \,. \tag{13}$$

In this case, the activation function is ‘Softmax’. This function ensures that all output nodes have values between 0 and 1, and that the sum of all output node values is always 1. The ‘Softmax’ function is represented in Eq. (14).

$$Softmax\,(x_c) = \frac{e^{x_c}}{\sum_{c=1}^{N} e^{x_c}} \,. \tag{14}$$

The light-weight CNN network model is developed with a fully connected Neural Network (NN) model to prevent various parameters of 5-layered depth features for training and test sets. The model has been built with a batch normalization function to normalize the output of each light-weight CNN layer. For network sparsity, the activation function Rectified Linear Unit (ReLU) is being used to reduce parameter interconnection and reduce the likelihood of overfitting issues. In this model, the convolutional layers are divided into five levels, and feature contracts are encoded into a single column matrix to fit them into the fully connected layer just after the fifth convolutional layer output. Following that, the layer is flattened with a single dense cell and one dropout layer. Finally, an activation function of “Softmax” is applied to classify the COVID-19/SARS-CoV-2-positive symptoms and COVID-19/SARS-CoV-2-negative symptoms. The exact process is applied to compute the remaining classes (Asthma, Bronchitis, Pertussis, and Healthy Symptoms).

The model’s input is given as 1 × 477 × 256 dimensional vectors to the light-weight CNN, and the model is trained with a two-dimensional convolutional network. The light-weight CNN model has been implemented with a sigmoid transfer function, and the pooling sheet with stride one is generated using the max-pooling layer. Finally, the “Softmax” classifier is used to classify each class’s prior disease probability based on respiratory sound data. The light-weight CNN structure is illustrated in Fig. 4 using a 5-layer network with different kernel sizes like 1 × 12, 1 × 24, 1 × 36, 1 × 48, and 1 × 60. In the 2D convolutional network, the convolutional kernel dimensions are increased based on column transformation. While maintaining the same size convolutional filters for the network model, the outcome of various kernel sizes (1 × 12, 1 × 24, 1 × 36, 1 × 48, and 1 × 60) of a convolutional network is tested on convolutional frameworks’ output. A model’s framework is implemented with multiple hyper-parameters such as categorical cross-entropy, batch size 32, ’Adam’ optimizer, three hidden layers, activation function (ReLU), dropout (0.01), max pooling, the classifier (Softmax), and the number of epochs (100). There are 122,112 trainable parameters and 0 (zero) non-trainable parameters. The hyper-parameter objective function was updated based on the initial analysis results, and a subsequent better model framework was examined.

The system acquired data consist of T-FP (Time-Frequency Patches) collected values from the logarithmic scaled Mel-spectrum framework of the respiratory sound signal features, connected to the first accepted feature training techniques adapted respiratory COVID-19 identification. The ‘Essentia’ python library is used to generate a log-scaled Mel-spectrum

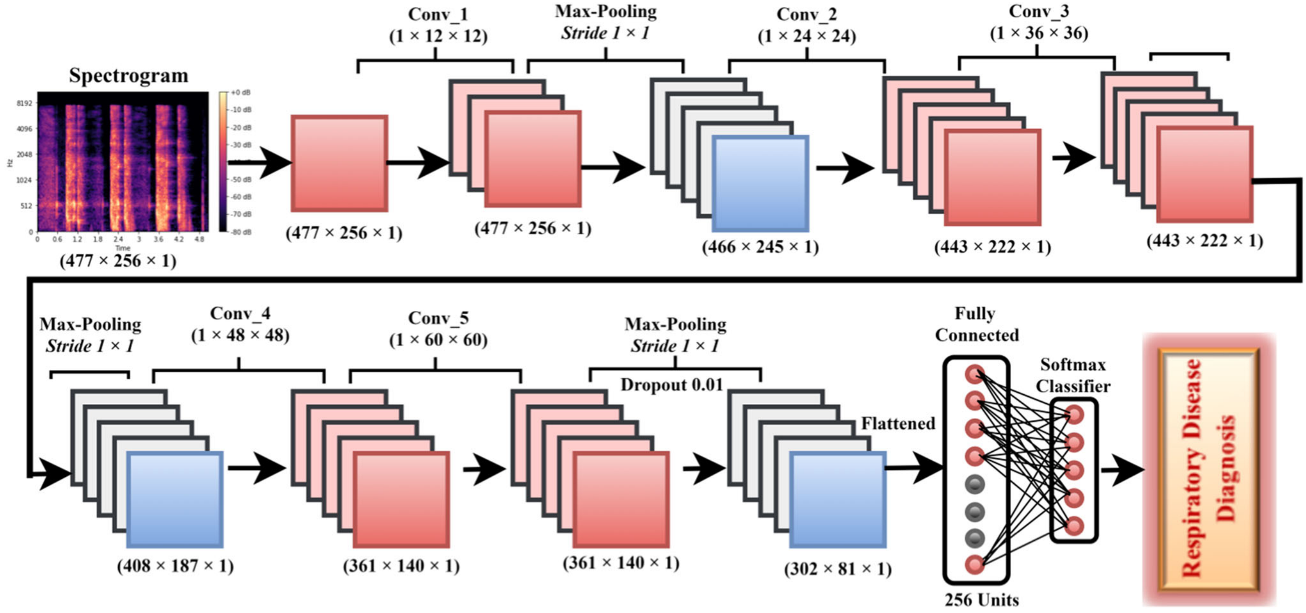


Fig. 4 The proposed layered light-weight CNN architecture for respiratory disease classification

with two fifty-six bands. We have used the 'Essentia' python library to generate a logarithmic scaled Mel-spectrum with two fifty-six bands chosen to represent the recognizable respiratory sound frequency spectrogram range of 0–22 kHz at a frequency frame rate of 25 ms (the capacity of the frequency range is 0.440 MHz for 1024 different samples) with the same length of the hopping spectrum. The model has acquired various types of feature vectors along with 256 convolutional features vectors. The combined modality (cough, voice, breath) of feature vectors is 256, and 477-dimensions of feature vectors of handicrafts. The dataset evaluation is described in Sect 3.1.1, and it varies the input respiratory acoustic sound period between 0 and 30 s.

We have fixed this input time duration T to 05 s to 256 frames for respiratory sound data T-FP (Time-Frequency Patches), i.e., $T \in R^{256\times256}$. T-FP is derived instantly during training from the full logarithmic Mel-spectrum of each respiratory sound excerpt. The model is developed to learn Θ parameters to derive non-linear function $F(..|\Theta)$, which maps time T (given input value) to the estimated value Y is depicted in Eq. (15).

$$Y = F\left(T|\Theta\right) = f_K(\ldots..f_2(f_1\left(\left(\left(T|\theta_1\right)|\theta_2\right)|\theta_K\right). \quad (15)$$

The layer of a convolutional network of every operation can be depicted as $f_k\left(..\,|\,\theta_k\right)$, $K = 5$ in this proposed framework. The proposed model architecture has constructed with three convolutional layers ($k \in \{1, 2, 3\}$) that can be expressed as Y_k in Eq. (16).

$$Y_k = f_k\left(T_k|\Theta_k\right) = h\left(W * T_k + b\right), \theta_k = [W, b]. \quad (16)$$

The sequence of 3-dimensional convolutional filters have been represented as 'W' of 'N', the 3-dimensional input vector is 'T_k' constructed with M feature maps, $h(\cdot)$ is the point-wise activation function, the convolutional operation is '$*$', and the biased function is 'b'. The shape of the network can be represented as W, X_k, Y_k are $\left(M, d_0, d_1\right)$, $\left(N, M, n_0, n_1\right)$, and $(N, d_0 - n_0 + 1, d_1 - n_1 + 1)$. The T-FP input layer dimensionality of the first layer of the network is represented as $d_0 = d_1 = 256$. We have applied max-pooling after every convolutional layer $k \in \{1\}$, $k \in \{3\}$, $k \in \{6\}$, to reduce the dimensionality. This reduces the size of the resultant feature map function, which enhances model training performance and provides a certain level of classification accuracy rate in the light-weight CNN network. The matrix product of fully connected layers $k \in \{4, 5\}$ is represented in Eq. (17). Whereas T_k is the vector representation that can be flattened to M, the shape of the dimensionality is $W\ (N, M)$, function activation is $h\left(\cdot\right)$, and 'b' is the vector of 'N'. The layered architecture of this proposed light-weight convolutional model is represented in Fig. 4.

$$Y_k = f_k\left(T_k|\Theta_k\right) = h\left(W \cdot T_k + b\right), \Theta_k = [W, b]. \quad (17)$$

3.3.1 *Model analysis*

The human respiratory COVID-19 crowdsourced dataset is implemented to test the light-weight CNN model with various data augmentation sets. The total dataset consists of 1541 respiratory sound audio files (including breath, voice, and cough) collected from different users with a track size of 30 seconds. The model has created various classes while training the dataset (COVID-19_with_Cough, COVID-19_with_Breath, Non-COVID-19_with_Cough, Non-COVID-19_with_Breath, Asthma_with_Cough, Asthma_with_Breath, Pertussis_with_Cough, Bronchitis_with_Cough, COVID-19_with_Cough_and_Breath, COVID-19_with_Cough_and_Breath_and_Voice, Healthy_Cough sounds, Healthy_Breath sounds, Healthy_Speech sounds) to identify

disease symptoms. The data are organized into five distinct folders, each with its own label number (001 Asthma, 002 COVID-19, 003 Pertussis, 004 Bronchitis, 005 Healthy Symptoms), and the labeled data are analyzed using a higher accuracy model. With testing precision, the model results obtained are compared to previous works on this dataset. The newly implemented model is tested and compared to earlier models to test accuracy. The data have been partitioned into two datasets as testing data and training data. We have selected one of the five training folders in each section to train the most recent light-weight CNN network structure as a test set to determine the learning epoch that produces the best results when working with the four remaining respiratory sound audio data folders.
The novel points of the research:

1. We have implemented new feature extraction methods (MMFCC and EGFCC) to extract depth features from the respiratory sound data.
2. The model has made a comparison between MMFCC and EGFCC for all respiratory sound diseases (COVID-19, Pertussis, Bronchitis, and Asthma) and regular respiratory sounds.
4. We made a comparative analysis of abnormalities between COVID-19 diseases Vs. Other respiratory sound diseases.
5. Customized light-weight convolutional framework with input data using Modified Mel-frequency Cepstral Coefficient (MMFCC) and EGFCC optimized for automated feature learning rather than individual feature extraction technique.
6. Self-adapting light-weight CNN framework is suggested for automated parameter determination rather than depending on individual experience.
7. An average pooling layer accompanies the feature vector in the light-weight CNN framework to read the performance in each level to prevent various implications of features between training and validation results.
8. We applied different receptive fields and depths in the proposed model to get different contextual information that should aid in classification. And our experiments suggested 1 × 12 receptive fields and depth (5 Layer) for the light-weight CNN to extract and identify the features from respiratory sound data.
9. The proposed network outperforms conventional feature extraction models and existing Deep Learning (DL) models for COVID-19/SARS-CoV-2 classification accuracy in the range of 4–10%.

4 Results and discussion

The model compares COVID-19 sounds with other respiratory sounds and identifies the relationship between the COVID-19 symptoms with other respiratory diseases. The model is identified deep features based on audio sound parameters such as frequency, loudness, air volume, subglottic pressure, acoustic signal, cough peak flow rate, cough expire volume, peak velocity-time, pitch, duration, intensity, sound quality, Signal to Noise Ratio (SNR), Voice Activity Detection (VAD), and Strength of Lombard Effect (SLE) for native speakers. Figure 5 shows the comparisons for different respiratory diseases among MMFCC and EGFCC.

In this section, we have conducted three relative experiments. The first compares different kernel convolution forms, the second different feature channels, and the third other network-layer numbers. In the statistical study of multi-classification, the 'F_1' score indicates the specificity of a test. The mean average score of recall and precision is defined as 'F_1' score, with 100 percent being the best and zero percent being the worst. The accuracy rate and 'F_1' score for recognition is used in this analysis to assess the method's efficiency. The 'F_1' score and accuracy are calculated using the equations (Eqs. (18) and (19)). TP—true positive, FP–false positive, FN—false negative, and TN–true negative. Table 1 compares the accuracy and F_1 score of different kernel types and the light-weight CNN kernel shape and size. Table 1 shows a comparison of convolution with varying channels in terms of accuracy.

$$F_1 Score = \frac{\frac{True\,Positive\,(TP)}{True\,Positive\,(TP)+False\,Positive\,(FP)} \times \frac{True\,Positive\,(TP)}{True\,Positive\,(TP)+False\,Negative\,(FN)}}{\frac{True\,Positive\,(TP)}{True\,Positive\,(TP)+False\,Positive\,(FP)} + \frac{True\,Positive\,(TP)}{True\,Positive\,(TP)+False\,Negative\,(FN)}} \tag{18}$$

$$Accuracy = \frac{(True\,Positive\;(TP) + True\,Negative\;(TN))}{\begin{pmatrix} True\,Positive\;(TP) + True\,Negative\;(TN) + \\ False\,Positive\;(FP) + False\,Negative\,(FN) \end{pmatrix}} \times 100\,. \tag{19}$$

The comparisons of various respiratory sounds are represented in Figs. 5 and 6. It analyzes each respiratory sound's significant parameters and shows the variations of the pitch of sound and synthesis value of each respiratory sound. The comparison is made up with different modalities of respiratory sounds like frequency, loudness, air volume, subglottic pressure, acoustic signal, cough peak flow rate, cough expires volume, peak

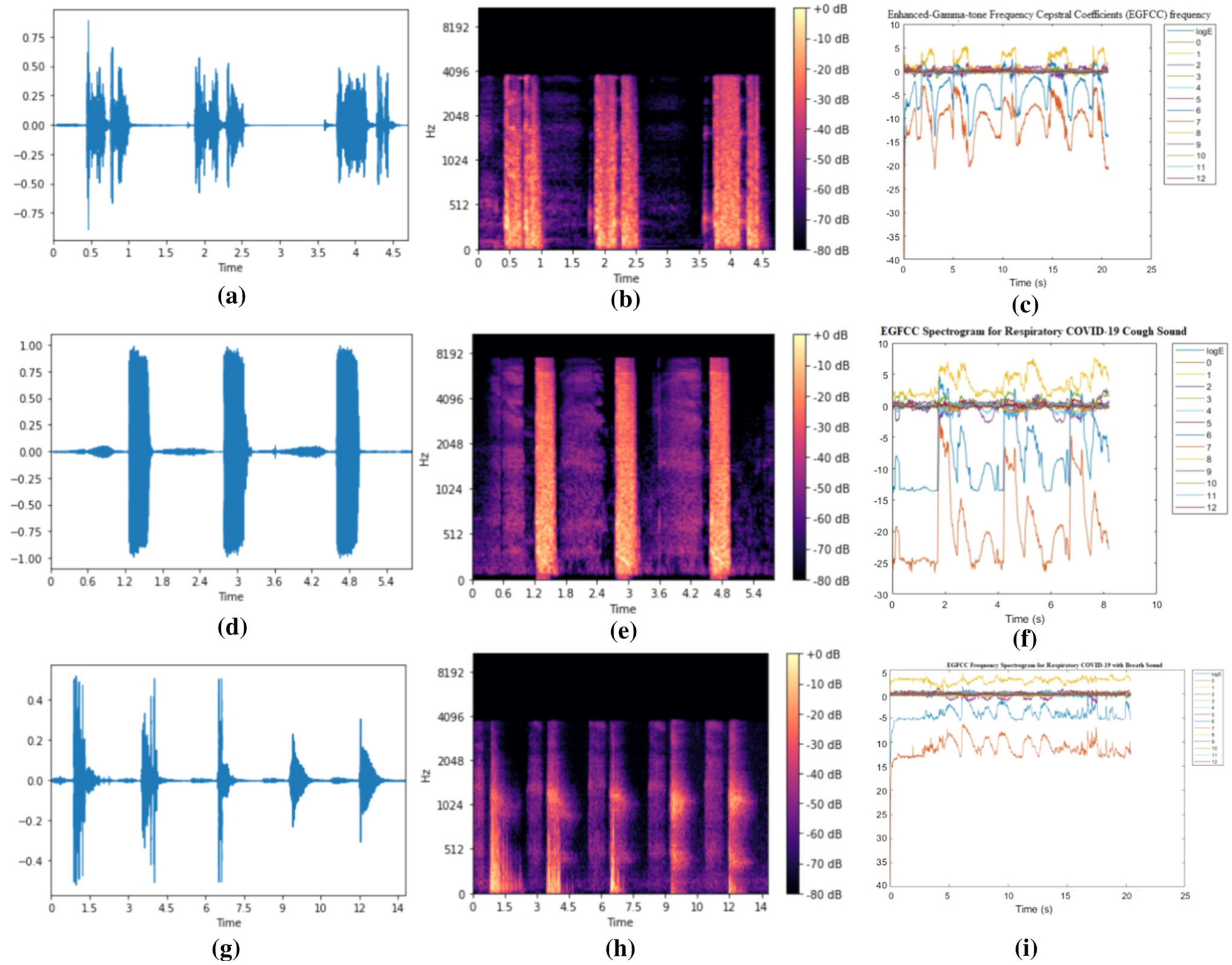


Fig. 5 Comparing sample respiratory diseases (Asthma, COVID-19 with cough, COVID-19 without cough) with MMFCC and EGFCC feature spectrograms. **a** sample asthma sound signal, **b** MMFCC feature spectrogram for respiratory asthma cough sound, **c** EGFCC spectrogram for respiratory asthma cough sound, **d** sample COVID-19 sound signal, **e** MMFCC feature spectrogram for respiratory COVID-19 cough sound, **f** EGFCC spectrogram for respiratory COVID-19 cough sound, **g** sample COVID-19 with breath sound signal, **h** MMFCC feature spectrogram for respiratory COVID-19 with breath sound, **i** EGFCC spectrogram for respiratory COVID-19 with breath sound

velocity–time, pitch, duration, intensity, sound quality, Signal to Noise Ratio (SNR), Voice Activity Detection (VAD), and Strength of Lombard Effect (SLE) for native speakers.

The comparison is made up for different diseases like Asthma, COVID-19 with cough, COVID-19 without cough, Pertussis, and Bronchitis, along with healthy symptoms for cough, breath, and voice using MMFCC (Modified Mel-frequency Cepstral Coefficient) and EGFCC (Enhanced Gama-tone Frequency Cepstral Coefficient). Finally, the model is trained with MMFCC generated feature spectrum and performs the analysis for each respiratory sound-based disease. Figure 7 depicts the comparative analysis of abnormalities in COVID-19 disease with other respiratory infections for different classes like COVID-19_Cough Vs. Asthma_Cough, COVID-19_Cough_Breath Vs. Bronchitis_Cough_Breath, COVID-19_Cough Vs. Pertussis _Whooping_Cough, COVID-19_Cough Vs. Healthy _Cough symptoms. The convolutional kernels of lightweight CNN play a vital role in detecting abnormalities from human-generated respiratory sounds. The total number of layers required for the convolutional level in the model architecture is calculated with an observational analysis of respiratory sound data. The approximately 13,000 input respiratory sound data samples are divided into 10% training data and 10% test data, and the remaining 80% can be used for model training with a 32-batch size of 70 epochs.

The researchers have been used prognostic models, pertained ensemble models, and different feature extraction techniques to identify respiratory sound features from respiratory sound data to detect COVID-19/SARS-CoV-2 symptoms. Table 1 shows the new

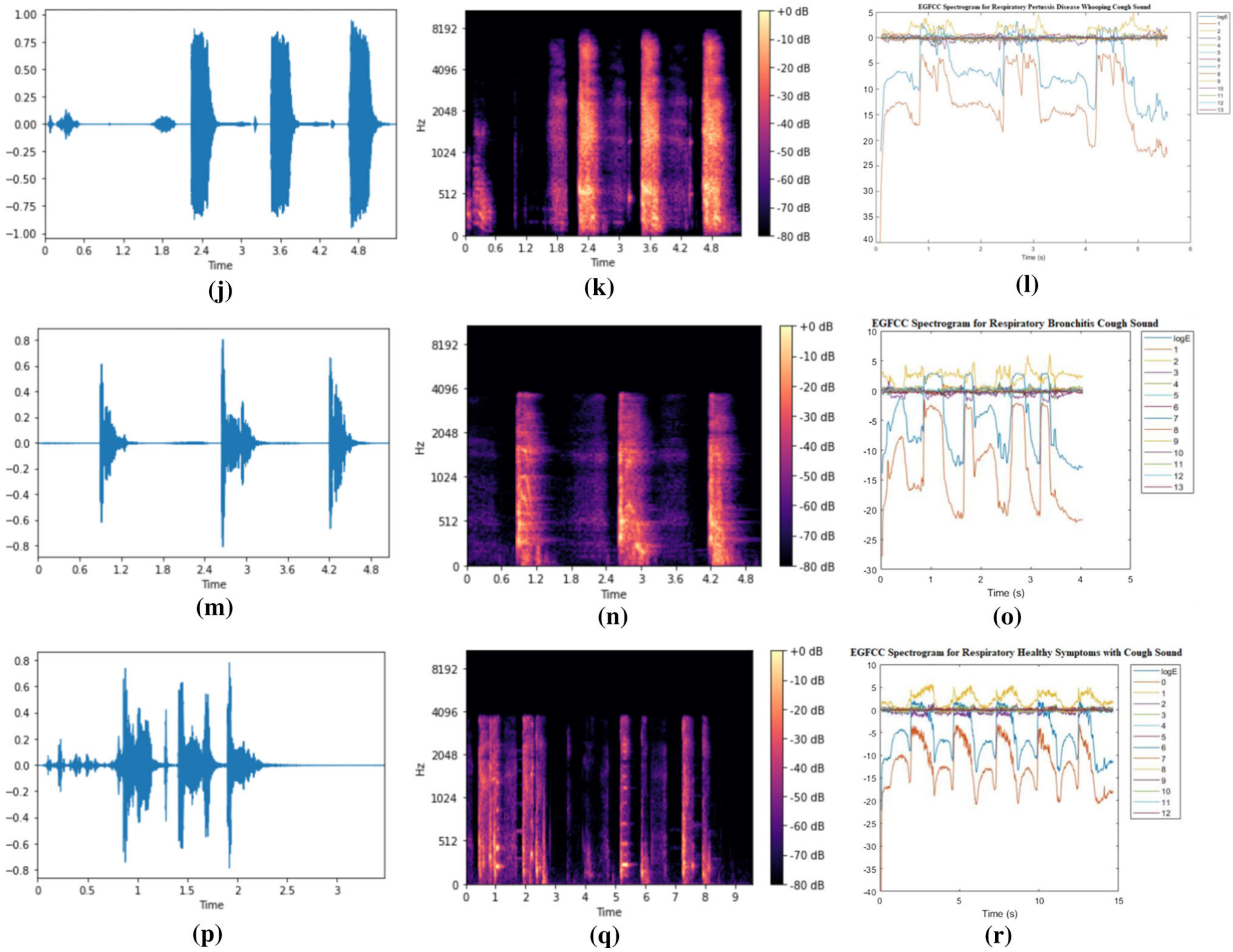
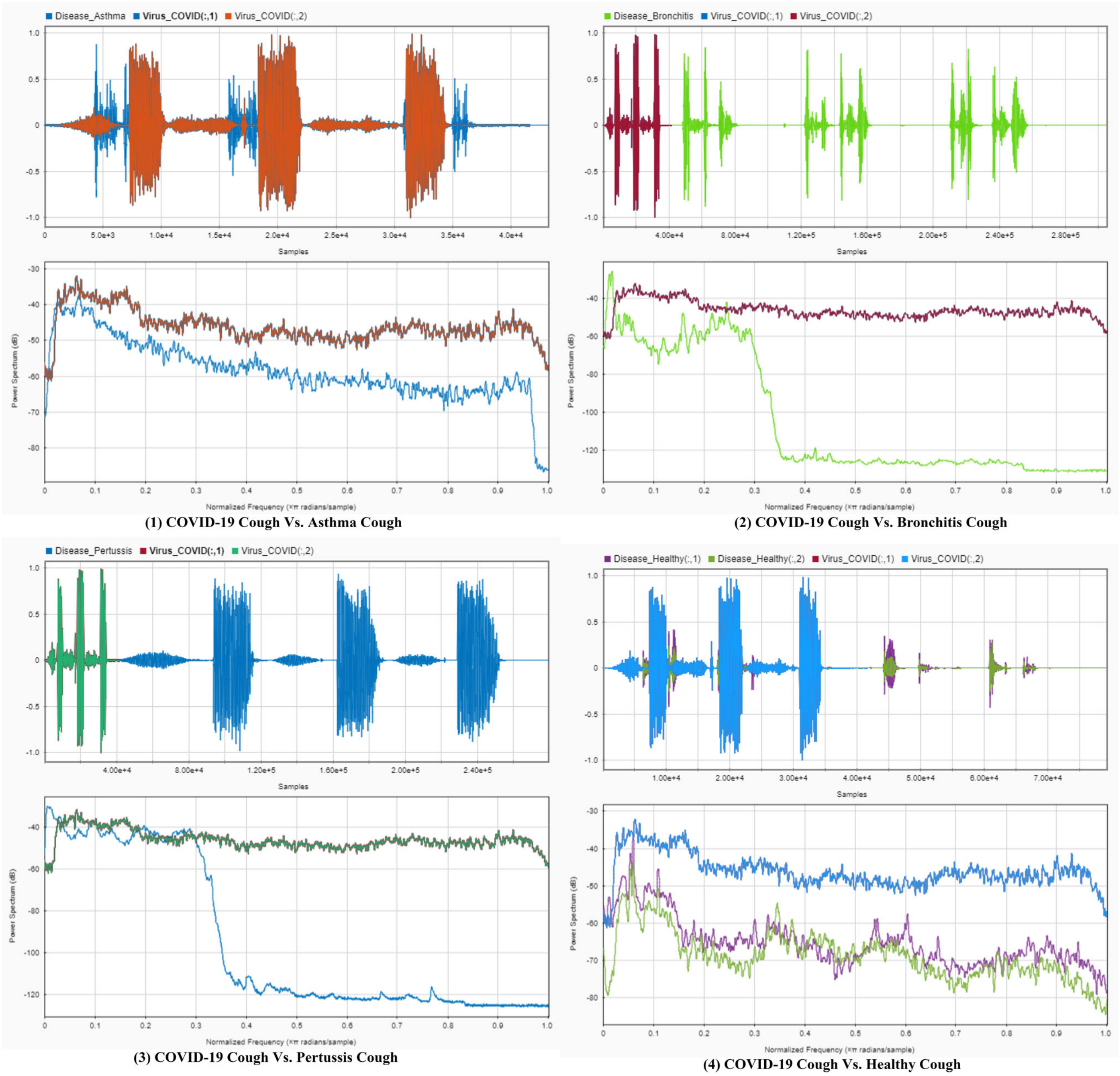


Fig. 6 Comparing sample respiratory diseases (Pertussis, Bronchitis, and Healthy Symptoms) with MMFCC and EGFCC feature spectrograms. **j** Sample pertussis whooping cough sound signal, **k** MMFCC feature spectrogram for respiratory pertussis whooping cough sound, **l** EGFCC spectrogram for respiratory pertussis disease whooping cough sound, **m** sample bronchitis disease sound signal, **n** MMFCC feature spectrogram for respiratory bronchitis cough sound, **o** EGFCC spectrogram for respiratory bronchitis cough sound, **p** sample cough signal for healthy symptoms, **q** MMFCC feature spectrogram for respiratory healthy cough sound, **r** EGFCC spectrogram for respiratory healthy symptoms with cough sound

findings of this research for different feature depths and various kernel shapes. The model's performance is compared with the COVID-19 crowdsourced benchmark dataset and gives a competitive performance. We applied different receptive fields and depths in the proposed model to get additional contextual information that should aid in classification. And our experiments suggested 1 × 12 receptive fields and a depth of 5-Layer for the light-weight CNN to extract and identify the features from respiratory sound data is represented in Fig. 8. The light-weight CNN layers are entirely relevant to obtain the required performance after adding convolutional and pooling layers. When the hidden layers exceed five, the network output remains nearly constant.

Here, we have compared the results for different channels (MFCC, EGFCC, and MMFCC) with various kernel sizes (1 × 12, 1 × 24, 1 × 36, 1 × 48, 1 × 60). The model achieves around 78.12% using MFCC with 1 × 60 kernel size, 82.27% using EGFCC (Enhanced Gammatone Frequency) with 1 × 60 kernel size, 84.13% using MMFCC (Modified Mel-Frequency) with 1 × 60 kernel size. The model achieves around 79.38% using MFCC with 1 × 48 kernel size, 83.16% using EGFCC (Enhanced Gamma-tone Frequency) with 1 × 48 kernel size, 85.42% using MMFCC (Modified Mel-Frequency) with 1 × 48 kernel size. The model achieves accuracy around 81.26% using MFCC with 1 × 36 kernel size, 83.97% using EGFCC (Enhanced Gamma-tone Frequency) with 1 × 36 kernel size, 86.18% using MMFCC (Modified Mel-Frequency) with 1 × 36 kernel size. The model achieves around 82.87% using MFCC with 1 × 24 kernel size, 84.68% using EGFCC (Enhanced Gammatone Frequency) with 1 × 24 kernel size, 88.34% using



(1) COVID-19 Cough Vs. Asthma Cough

(2) COVID-19 Cough Vs. Bronchitis Cough

(3) COVID-19 Cough Vs. Pertussis Cough

(4) COVID-19 Cough Vs. Healthy Cough

Fig. 7 The comparative analysis of abnormalities in COVID-19 disease with other respiratory diseases

MMFCC (Modified Mel-Frequency) with 1 × 24 kernel size. The model achieves accuracy around 83.264% using MFCC with 1 × 12 kernel size, 85.68% using EGFCC (Enhanced Gamma-tone Frequency) with 1 × 12 kernel size, 91.32% using MMFCC (Modified Mel-Frequency) with 1 × 12 kernel size.

The new findings of various feature channels for different layers are shown in Table 2. The results suggest that the neural network's growth in this research enhanced the identification, including the accuracy of the sound signals for various tests. The three-layered and five-layered light-weight CNN model comparison results with different light-weight CNN kernel sizes are depicted in Fig. 9. The experimental findings indicate that the present model's in-depth features preserve additional information on COVID-19 sound signals by enhancing classification accuracy. Table 2 compares the model with different channels of various light-weight CNN layers for five different light-weight CNN kernel shapes. Here, we have compared the results for different channels (MFCC (Mel-Frequency), EGFCC (Enhanced Gamma-tone), MMFCC (Modified Mel-Frequency)) for 3 layered architecture and 5-layered architecture of light-weight CNN model with various kernel sizes like 1 × 60, 1 × 48, 1 × 36, 1 × 24, 1 × 12. The proposed method achieves approximately 4% to 10% accuracy improvement than that of the existing MFCC technique for the classification of COVID-19 disease. And it gives a state-of-the-art performance with other existing models for detecting respiratory diseases. In the future,

Table 1 The comparison of results with different kernel shapes and feature types

Name of the channel	CNN Kernel size	Accuracy (%)	$\mathbf{F_1}$ Score
MFCC (Mel-Frequency)	1 × 60	78.12	80.18
EGFCC (Enhanced Gamma-tone Frequency)		82.27	83.12
MMFCC (Modified Mel-Frequency)		84.13	85.13
MFCC (Mel-Frequency)	1 × 48	79.38	80.14
EGFCC (Enhanced Gamma-tone Frequency)		83.16	84.31
MMFCC (Modified Mel-Frequency)		85.42	86.18
MFCC (Mel-Frequency)	1 × 36	81.26	82.13
EGFCC (Enhanced Gamma-tone Frequency)		83.97	85.12
MMFCC (Modified Mel-Frequency)		86.18	87.87
MFCC (Mel-Frequency)	1 × 24	82.87	83.26
EGFCC (Enhanced Gamma-tone Frequency)		84.68	86.18
MMFCC (Modified Mel-Frequency)		88.34	89.45
MFCC (Mel-Frequency)	1 × 12	83.64	84.42
EGFCC (Enhanced Gamma-tone Frequency)		85.68	86.92
MMFCC (Modified Mel-Frequency)		91.32	92.48

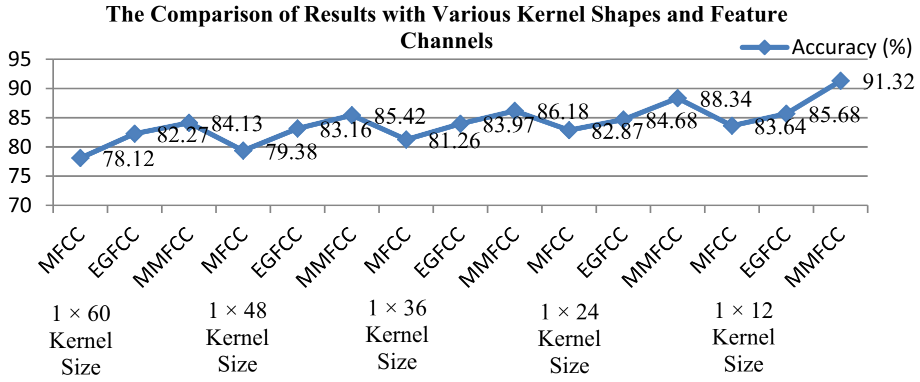


Fig. 8 The accuracy comparison for various feature channels with different kernel sizes

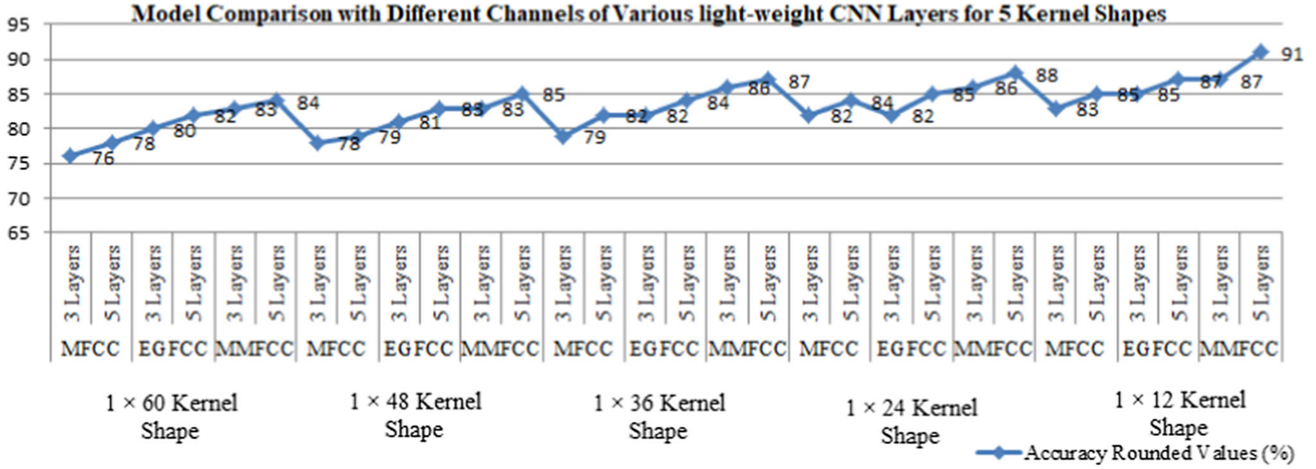


Fig. 9 The model comparison with different channels for five kernel shapes with different CNN layers (5-layer and 3-layer approach)

we will develop multi-feature channel-based CNN to improve the performance of respiratory disease diagnosis on the crowdsourced COVID-19 sounds dataset.

We have observed in this research that the model performance is raised when the kernel size is 1 × 12, and the model performance varies 4% to 10% by comparing with existing models on the benchmark dataset. The new findings represent that the proposed approach is better to identify COVID-19/SARS-CoV-2 symptoms on respiratory sounds data. Compared to regular conventional input, the selection of deep respiratory sound features does not require pre-processing for any time frame of COVID-19-based sounds. As a result, the respiratory COVID-19 sounds can be decided to enter into the classification techniques. Using the light-weight CNN model network instead of other models

Table 2 The comparison of the model with different channels of various light-weight CNN layers for five different light-weight CNN kernel shapes

Name of the channel	No. of CNN layers	CNN Kernel size	Accuracy (%)	**F$_1$** Score (%)
MFCC (Mel-Frequency)	5L	1 × 60	78.12	80.18
	3L		76.42	77.23
EGFCC (Enhanced Gamma-tone)	5L		82.27	83.12
	3L		79.82	81.23
MMFCC (Modified Mel-Frequency)	5L		84.13	85.13
	3L		82.92	83.48
MFCC (Mel-Frequency)	5L	1 × 48	79.38	80.14
	3L		77.92	78.83
EGFCC (Enhanced Gamma-tone)	5L		83.16	84.31
	3L		80.76	81.52
MMFCC (Modified Mel-Frequency)	5L		85.42	86.18
	3L		83.46	84.31
MFCC (Mel-Frequency)	5L	1 × 36	82.26	83.13
	3L		79.42	80.62
EGFCC (Enhanced Gamma-tone)	5L		83.97	85.12
	3L		81.63	83.34
MMFCC (Modified Mel-Frequency)	5L		87.18	88.87
	3L		85.63	86.58
MFCC (Mel-Frequency)	5L	1 × 24	83.87	84.26
	3L		81.74	82.92
EGFCC (Enhanced Gamma-tone)	5L		84.78	86.18
	3L		82.23	84.06
MMFCC (Modified Mel-Frequency)	5L		88.34	89.45
	3L		85.69	86.36
MFCC (Mel-Frequency)	5L	1 × 12	84.64	85.42
	3L		82.76	83.93
EGFCC (Enhanced Gamma-tone)	5L		86.68	87.92
	3L		84.82	86.06
MMFCC (Modified Mel-Frequency)	5L		91.32	92.48
	3L		87.38	89.13

Table 3 The respiratory disease classification with five different class labels (COVID-19, Bronchitis, Pertussis, Asthma, and Healthy sound classes) for various sound modalities (Voice, Cough, and Breath)

Class	Sound modality	Accuracy (%)	**F$_1$**-Score (%)
Asthma Disease	Cough + Breath	89.78	91.12
COVID-19 Disease	Cough + Breath + Voice	92.32	93.48
Pertussis Disease	Whooping Cough	89.69	90.31
Bronchitis Disease	Cough + Breathing shortness	88.74	89.14
Healthy Symptoms	Cough + Breath + Voice	93.65	94.83

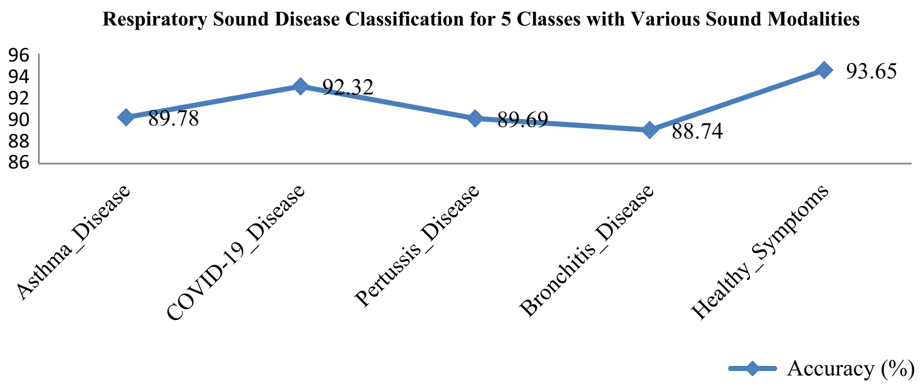


Fig. 10 The respiratory sound disease diagnosis for five different classes (COVID-19, Bronchitis, Pertussis, Asthma, and Healthy sound classes) for various sound modalities (Voice, Cough, and Breath)

Table 4 Proposed light-weight CNN model accuracy comparison with existing methods on COVID-19 respiratory sounds dataset

Model	Dataset	Accuracy (%)
SVM with PCA	Crowdsourced COVID-19 Sounds Dataset	Task-1: 80 Task-2: 82 Task-3: 80
VGG-Net with Augmentation	Crowdsourced COVID-19 Sounds Dataset	Task-2: 87 Task-3: 88
1DCNN with DDAE	Crowdsourced COVID-19 Sounds Dataset	Task-1: 90 Task-2: 88 Task-3: 88 Task-4: 84 Task-5: 86
Light-Weight CNN with MMFCC (Proposed)	Crowdsourced COVID-19 Sounds Dataset	Class-1: 89.78 Class-2: 92.32 Class-3: 89.69 Class-4: 88.74 Class-5: 93.65

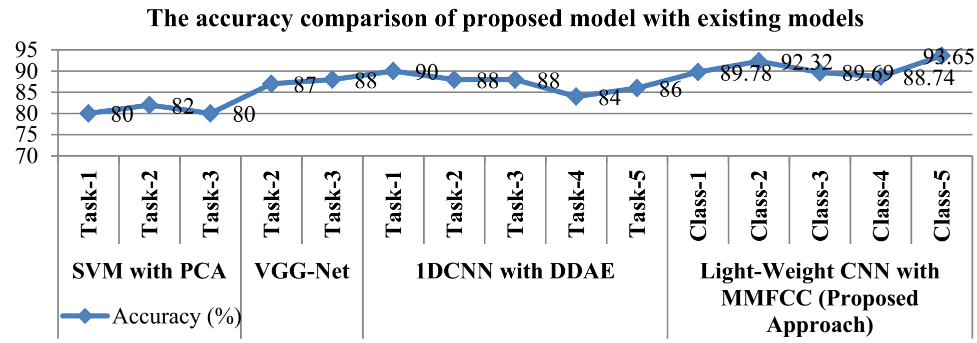


Fig. 11 Proposed light-weight CNN model accuracy comparison with existing methods

frameworks in this study greatly improves the overall system model's ability to interpret respiratory COVID-19 respiratory sound acoustic signals without using an X-ray, pulse oximeter, or CT-scanning, and some other disease diagnosis techniques.

4.1 Difference between COVID-19-positive symptoms versus other respiratory diseases with sounds

The respiratory sound classification model for five different classes (Asthma, Bronchitis, Pertussis, COVID-19, Healthy sounds class) is represented in Table 3. The first class discriminates the patient is having asthma problem from respiratory cough and breathing sounds. Class second discriminating users have bronchitis disease from shortness of breath sound and dry cough. The third class is for discriminating the symptoms of pertussis disease with respiratory whooping cough sound.

The fourth category denotes the classification used to determine whether the user is having respiratory COVID-19-positive symptoms or COVID-19-negative symptoms. Class five represents the healthy respiratory symptoms from respiratory sounds (Voice, Cough, and Breath). Table 3 illustrates the comparative results for the proposed light-weight CNN model versus the existing classification techniques for various classes. The proposed framework shows around 4% more accuracy than existing methods, as shown in Fig. 10.

4.2 Proposed light-weight CNN model accuracy comparison with existing methods on COVID-19 respiratory sounds dataset

The benchmark dataset shows few biased values identification in the respiratory sound data while examining for different respiratory disease user's cough combined with breath and voice may be the best predictor. The network performs better results for class two (COVID-19 disease) and class five (Healthy symptoms). The model achieves around 92% accuracy to predict the user has COVID-19 respiratory disease with three different modalities (voice, breath, and cough). The model can diagnosis asthma disease using cough and breath sounds with around 90% of best accuracy. The system may diagnosis pertussis disease using whooping cough with an accuracy of approximately 89%. The model may predict bronchitis disease from user respiratory cough and breathing shortness with around 89% of good accuracy. The model performs better for identifying healthy symptoms from cough, breath, and voice with approximately 93% of accuracy. The proposed light-weight CNN model accuracy comparison

Table 5 The proposed model comparison with existing methods with different modalities

Model	Dataset	Modality	Accuracy (%)
MFCC with DCT [28]	Clinical sample dataset1	Only Cough	Class 1: 59
	Clinical sample dataset1		Class 2: 72
ResNet50 with MFCC [31]	Open voice data by MIT	Speech and Cough	79
MFCC with DCT [28]	Android App developed for COVID-19 sound samples.	Speech, Cough, and Breath	79
DTL + MC [4]	Android App developed for COVID-19 sound samples.	Speech and Cough	88
SVM with PCA [24]	Crowdsourced COVID-19 Sounds Dataset	Speech, Cough, and Breath	Task-1: 80
			Task-2: 82 Task-3: 80
VGG Net with Augmentation [24]	Crowdsourced COVID-19 Sounds Dataset	Voice, Cough, and Breath	Task-2: 87
			Task-3: 88
1DCNN with DDAE [34]	Crowdsourced COVID-19 Sounds Dataset	Voice, Cough, and Breath	Task - 1: 90
			Task-2: 88 Task-3: 88 Task-4: 84 Task-5: 86
Light-Weight CNN with MMFCC (Proposed)	Crowdsourced COVID-19 Sounds Dataset	Voice, Cough, and Breath	Class-1: 89.78
			Class-2: 92.32 Class-3: 89.69 Class-4: 88.74 Class-5: 93.65

with existing methods on COVID-19 crowdsourced data is represented in Table 4 and Fig. 11.

The primary benchmark dataset was collected from Cambridge University (around 13,000 samples) on mutual agreement, and data related to pertussis disease and bronchitis disease is collected from different crowdsourced sound datasets. We are planning to collect more samples to improve the light-weight CNN model's performance to identify COVID-19/SARS-CoV-2 disease symptoms. The plan will be accomplished to neural network model in future works to identify SARS-CoV-2 disease with different modalities such as body temperature, pulse rate, cough sound, voice, and breathing sounds. Table 5 depicts the proposed model's comparative analysis with current Deep Learning (DL) models on respiratory sound data with different modalities.

In this analysis, the light-weight CNN convolutional model network's implementation rather than the other Deep Learning (DL) models dramatically enhances the entire model performance to detect the COVID-19 and other respiratory acoustic sound signals. Without X-ray and CT-scan images, the suggested framework will produce nearly accurate results to detect COVID-19 and other respiratory diseases. The light-weight CNN model with Enhanced-GFCC and Modified-MFCC network will help the general public to prescreen the COVID-19 and other respiratory disease symptoms before visiting the hospital. It is necessary to consult the hospital for any additional tests if a disease is identified. The model is restricted to identifying respiratory sound diseases such as Asthma, Bronchitis, Pertussis, COVID-19, and healthy respiratory sound symptoms from human respiratory cough, voice, and breathing sound data. Thus, it has never been used in clinical practice but can be used to detect human respiratory disease. We will develop Deep Learning (DL) model in future works to identify COVID-19 disease with different modalities like body temperature, pulse rate, cough sound, voice, and breathing sounds.

5 Conclusion

We proposed and implemented a light-weight CNN model using Modified-Mel-frequency Cepstral Coefficient (M-MFCC) to identify symptoms of respiratory diseases such as Asthma, Bronchitis, Pertussis, COVID-19, and Healthy symptoms in this study. The proposed deep learning model classifier outperforms feature-based methods in a wider band with different datasets. The model's adaptiveness with respect to different modalities shows the applicability of the model in different scenarios. The key advantage of this model is to unambiguously extract deep respiratory sound features and classify it accordingly. The model's performance is compared with MFCC and EGFCC (a customized version of GFCC) for various kernel sizes and depths. The proposed method achieves approximately 4–10% accuracy improvement than that of the existing MFCC technique for the classification of COVID-19 disease. And it gives a state-of-the-art performance with other existing models for detecting respiratory diseases. In the future, we will develop multi-feature channel-based CNN to improve the performance of respiratory disease diagnosis on the crowdsourced COVID-19 sounds dataset. The experiment also leads to the importance of Neural Architecture Search (NAS) for finding better architectures for classification.

Acknowledgements We would really like to thank Prof. Cecilia Mascolo, a medical scientist at the University of Cambridge for sharing the data. We thank each and everyone who is working to stop spreading COVID-19/SARS-CoV-2 epidemic.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: All data included in this manuscript are not publicly available because the dataset is collected from cited sources in references [14,24,35] for research purposes but is available from the corresponding author on reasonable request. The sample analysis of this work is available in figshare [41] repository at https://doi.org/10.6084/m9.figshare.18666419.v2.]

Declarations

Conflict of interest The authors have no competing interests.

References

1. World Health Organization. Coronavirus disease (COVID-19) (2019), https://www.who.int/
2. Y. Wang, M. Hu, Q. Li et al., Abnormal respiratory patterns classifier may contribute to large-scale screening of people infected with COVID-19 in an accurate and unobtrusive manner (2020). arXiv:2002.05534 [cs.LG]
3. Z. Jiang, M. Hu, F. Lei et al., Combining visible light and infrared imaging for efficient detection of respiratory infections such as COVID-19 on portable device (2020). arXiv:2004.06912 [cs.CV]
4. A. Imran, I. Posokhova, H.N. Qureshi et al., AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform. Med. Unlocked **20**, 100378 (2020)
5. J. Shuja, E. Alanazi, W. Alasmary et al., COVID-19 open source data sets: a comprehensive survey. Appl. Intell. **21**, 1–30 (2020)
6. J. Rasheed, A. Jamil, A.A. Hameed et al., A survey on artificial intelligence approaches in supporting frontline workers and decision makers for the COVID-19 pandemic. Chaos Solitons Fractals **141**, 110337 (2020)
7. T. Alafif, A.M. Tehame, S. Bajaba et al., Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. Int. J. Environ. Res. Public Health **18**, 1117 (2021)
8. K.V.S. Ritwik, B.K. Shareef, V. Deepu, COVID-19 patient detection from telephone quality speech data (2020). arXiv:2011.04299v1 [cs.SD]
9. L. Kranthi Kumar, P.J.A. Alphonse, A literature review on COVID-19 disease diagnosis from respiratory sound data. AIMS Bioeng. **8**(2), 140–153 (2021). https://doi.org/10.3934/bioeng2021013
10. D. Easwaramoorthy, A. Gowrisankar, A. Manimaran, S. Nandhini, L. Rondoni, S. Banerjee, An exploration of fractal-based prognostic model and comparative analysis for second wave of COVID-19 diffusion. Nonlinear Dyn. **2021**, 1–21 (2021). https://doi.org/10.1007/s11071-021-06865-7
11. C. Kavitha, A. Gowrisankar, S. Banerjee, The second and third waves in India: when will the pandemic be culminated? Eur. Phys. J. Plus **136**, 596 (2021). https://doi.org/10.1140/epjp/s13360-021-01586-7
12. A. Gowrisankar, L. Rondoni, S. Banerjee, Can India develop herd immunity against COVID-19? Eur. Phys. J. Plus **135**(6), 526 (2020). https://doi.org/10.1140/epjp/s13360-020-00531-4
13. M. SreeJagadeesh, P.J.A. Alphonse, COVID-19 outbreak: an ensemble pre-trained deep learning model for detecting informative tweets,. Appl. Soft Comput. **107**, 107495 (2021). https://doi.org/10.1016/j.asoc.2021.107495 (**ISSN 1568-4946**)
14. L. KranthiKumar, P.J.A. Alphonse, 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 **8**(2), 240–264 (2021). https://doi.org/10.3934/publichealth2021019
15. Y. Huang, S. Meng, Y. Zhang et al., The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods (2020). https://doi.org/10.1101/2020.04.07.20051060
16. N. Rebecca et al., Symptom-based screening tool for asthma syndrome among young children in Uganda. NPJ Prim. Care Respir. Med. **30**, 1 (2020). https://doi.org/10.1038/s41533-020-0175-1

17. J. Shi, X. Zheng, Y. Li et al., Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J. Biomed. Health Inform. **22**, 173–183 (2018)
18. L. Brabenec, J. Mekyska, Z. Galaz et al., Speech disorders in Parkinson's disease: early diagnostics and effects of medication and brain stimulation. J. Neural. Transm. (Vienna) **124**, 303–334 (2017)
19. S.B. Erdogdu, G. Serbes, C.O. Sakar, Analyzing the effectiveness of vocal features in early telediagnosis of Parkinson's disease. PLoS ONE **12**, e0182428 (2017)
20. V. Klára, I. Viktor, M. Krisztina, in *Voice Disorder Detection on the Basis of Continuous Speech*, ed by Á. Jobbágy, 5th European Conference of the International Federation for Medical and Biological Engineering. IFMBE Proceedings, Springer, Berlin, Heidelberg. https://doi.org/10.5220/0010193101350141 (2011)
21. R. Liu, S. Cai, K. Zhang, N. Hu, in*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. https://doi.org/10.1109/ICIIBMS46890.2019.8991459
22. H. Pasterkamp, S.S. Kraman, G.R. Wodicka, Respiratory sounds: advances beyond the stethoscope. Am. J. Respir. Crit. Care Med. **156**(3 Pt 1), 974–87 (1997). https://doi.org/10.1164/ajrccm.156.3.9701115 (**PMID: 9310022**)
23. L. KranthiKumar, P.J.A. Alphonse, A literature review on COVID-19 disease diagnosis from respiratory sound data. AIMS Bioeng. **8**(2), 140–153 (2021). https://doi.org/10.3934/bioeng.2021013
24. C. Brown, J. Chauhan, A. Grammenos et al., in *Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data*, Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. https://doi.org/10.1145/3394486.3412865 (2020)
25. O. Lara et al., The COUGHVID crowdsourcing dataset: A corpus for the study of large scale cough analysis algorithms, arXiv:2009.11644v1 [cs.SD] (2020). https://doi.org/10.5281/zenodo.4048312
26. 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 (2020). arXiv:2002.05534 [cs. LG]
27. A. 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
28. M. Bader et al., Studying the Similarity of COVID-19 Sounds based on Correlation Analysis of MFCC, arXiv:2010.08770v1 [cs. SD] (2020). https://doi.org/10.1109/CCCI49893.2020.9256700
29. X. Jiang et al., Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough (2020). arXiv:2011.13320v2 [cs.SD]
30. M. Al Ismail et al., Detection of COVID-19 through the Analysis of Vocal Fold Oscillations (2020). arXiv:2010.10707v1 [eess. AS]
31. J. Laguarta, F. Hueto, B. Subirana, COVID-19 Artificial Intelligence Diagnosis using only Cough Recordings. IEEE Open Journal of Engineering in Medicine and Biology (2020). https://doi.org/10.1109/OJEMB.2020.3026928
32. A. Hassan, I. Shahin, 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 (2020), pp. 1–5. https://doi.org/10.1109/CCCI49893.2020.9256562
33. 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. Med. Biol. **1**, 203–206 (2020). https://doi.org/10.1109/OJEMB.2020.2998051
34. K.K. Lella, A. Pja, 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 **8**(2), 240–264 (2021). https://doi.org/10.3934/publichealth.2021019
35. K.K. Lella, A. Pja, Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath. Alexandria Eng. J. ISSN 1110-0168, (2021). https://doi.org/10.1016/j.aej.2021.06.024
36. J. Han, K. Qian, M. Song et al., An Early Study on Intelligent Analysis of Speech under COVID-19: Severity, Sleep Quality, Fatigue, and Anxiety (2020). arXiv:2005.00096v2 [eess.AS]
37. Md. Susanta Sarangi, G.S. Sahidullah, Optimization of data-driven filterbank for automatic speaker verification. Dig. Signal Process. **104**, 102795 (2020). https://doi.org/10.1016/j.dsp.2020.102795 (**ISSN 1051-2004**)
38. M. Dua, R.K. Aggarwal, Performance evaluation of Hindi speech recognition system using optimized filterbanks. Eng. Sci. Technol. Int. J. **21**(3), 389–398 (2018). https://doi.org/10.1016/j.jestch.2018.04.005 (**ISSN 2215-0986**)
39. A. Adiga, M. Magimai, C.S. Seelamantula, *Gammatone wavelet Cepstral Coefficients for robust speech recognition*, 2013 IEEE International Conference of IEEE Region 10 (TENCON 2013) (Xi'an, China, 2013), pp. 1-4. https://doi.org/10.1109/TENCON.2013.6718948
40. A. Krobba, M. Debyeche, S.A. Selouani, Mixture linear prediction Gammatone Cepstral features for robust speaker verification under transmission channel noise. Multimed. Tools Appl. **79**, 18679–18693 (2020). https://doi.org/10.1007/s11042-020-08748-2
41. L. Kranthi Kumar, COVID-19 disease diagnosis with light-weight CNN. figshare. Journal contribution (2022). https://doi.org/10.6084/m9.figshare.18666419.v2