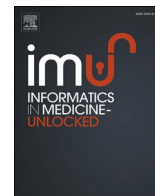




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A novel deep learning model to detect COVID-19 based on wavelet features extracted from Mel-scale spectrogram of patients' cough and breathing sounds

Mohammed Aly^{a,*}, Nouf Saeed Alotaibi^b

^a Department of Artificial Intelligence, Faculty of Computers and Artificial Intelligence, Egyptian Russian University, Badr City, 11829, Cairo, Egypt

^b Department of Computer Science, College of Science, Shaqra University, Shaqra City, 11961, Saudi Arabia

ARTICLE INFO

Keywords:

Covid-19
Breathing and coughing sound
Deep learning
Mel-scale spectrogram
Sound classification

ABSTRACT

The goal of this paper is to classify the various cough and breath sounds of COVID-19 artefacts in the signals from dynamic real-life environments. The main reason for choosing cough and breath sounds than other common symptoms to detect COVID-19 patients from the comfort of their homes, so that they do not overload the Medicare system and therefore do not unwittingly spread the disease by regularly monitoring themselves. The presented model includes two main phases. The first phase is the sound-to-image transformation, which is improved by the Mel-scale spectrogram approach. The second phase consists of extraction of features and classification using nine deep transfer models (ResNet18/34/50/100/101, GoogLeNet, SqueezeNet, MobileNetv2, and NasNetmobile). The dataset contains information data from almost 1600 people (1185 Male and 415 Female) from all over the world. Our classification model is the most accurate, its accuracy is 99.2% according to the SGDM optimizer. The accuracy is good enough that a large set of labelled cough and breath data may be used to check the possibility for generalization. The results demonstrate that ResNet18 is the best stable model for classifying cough and breath tones from a restricted dataset, with a sensitivity of 98.3% and a specificity of 97.8%. Finally, the presented model is shown to be more trustworthy and accurate than any other present model. Cough and breath study accuracy is promising enough to put extrapolation and generalization to the test.

1. Introduction

Wavelet coherence (WC) is still a time-frequency approach for estimating the phase lag between two different time series in terms of wavelet transform (WT) [1,2]. Although electroencephalogram (EEG) series are non stationary, nonlinear approaches based on Wavelet Transform (WT) are suitable. To generate high-resolution phase coherence in both time and frequency, wavelet correlation (WC) was used to five frequency band activities (fba) of single-trial non-averaged EEG series transmitted by effective images [3]. Through a 16-channel EEG cap, experimental data were gathered from both healthy individuals and patients diagnosed with first-episode psychosis. To separate normal from patients, WC estimates were generated for eight electrode pairs in line with each given fba and emotional state and classified using Least Squares Support Vector Machines with tenfold cross-validation [4]. The results indicate that Gamma has the best classification accuracy [5]. For the first time, quantifiable relationships between Cognitive Emotion

Regulation techniques (CERs) and EEG synchronisation levels have been explored. To that end, spectral coherence (COH), phase locking value, as well as mutual information were applied to brief segments of 62-channel resting state eyes-opened EEG data gathered from healthy people with varying emotion regulation strategies. Instant classification using full-band COH calculations yielded high classification accuracies (CA) of 99.44% and 98.33% [6].

The researchers use of deep learning models can be clarified in both neuroscience and medicine. The results indicating a strong relationship between emotional arousal and neuro-functional brain connectivity measurements. Support Vector Machines (SVMs) driven by Graph Theoretical segregation measurements of the brain network are used to classify different distinct emotional states for this goal. The findings show that distinct emotional states are defined by balanced network measures, yet when both segregation and integration measures differ based on the arousal ratings of audio-visual stimuli owing to neurotransmitter release throughout video viewing [7].

* Corresponding author.

E-mail address: mohammed-alsalem@eru.edu.eg (M. Aly).

<https://doi.org/10.1016/j.imu.2022.101049>

Received 29 June 2022; Received in revised form 8 August 2022; Accepted 9 August 2022

Available online 13 August 2022

2352-9148/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The suggested Deep Learning (DL) technique was incorporated as the major core of a DSS enabled by Internet of Things (IoT) design that may assist the human operator by lowering the time required for qualitative analysis carried out manually in this specific area by approximately to 90%. The research offers a DL ordinal technique for the categorization of aesthetic quality control (AQC). Unlike previous deep ordinal approaches, the conventional categorical cross-entropy is used in conjunction with the cumulative link model and the ordinal restriction enforced by the thresholds and slope parameters. Experiment results for completing an AQC task on a novel picture dataset derived from a certain company's requirement were obtained [8].

COVID-19 is a novel infection of SARS (SARS-Cov-2), first appeared in November 2019 and has since grown over the world, causing a global pandemic issue [9]. According the World Health Organization's (WHO) April 2021 report [10] there are already roughly 150 million confirmed illnesses and more than 3 million people die. Furthermore, with over 32.5 million cases as well as 500,000 deaths, the United States (USA) has recorded the greatest number of total diseases and deaths. These high numbers have put a strain on several healthcare services, especially given the virus's potential to create more genetic variations and spread more quickly among persons.

Currently, numerous studies have used new developing artificial intelligence (AI) methods to identify and classify COVID-19 in CT and X-ray images [11]. Some research (using CT scans as input data) applied machine and deep learning algorithms with a discriminate accuracy of more than 95% between infected and uninfected participants [12–17]. The key contribution of these studies is the ability of classification model such like support vector machines (SVM) as well as convolutional neural networks (CNN) to detect COVID-19 in CT images with minimal pre-processing steps. Furthermore, to identify COVID-19 in CT images, some research have used deep learning with additional feature fusion approaches [18–21].

Motivated by the foregoing, this paper presents a comprehensive deep learning approach for COVID-19 detection via cough and breathing sounds (Fig. 2). The proposed approach may be implemented as a rapid, low-cost, and commonly accessible COVID-19 pre-screening tool, especially in areas where the pandemic has expanded widely.

The reverse-transcription polymerase chain reaction (RT-PCR) test is the current standard for diagnosing COVID-19. There are some inexpensive accurate rapid tests that are used more and more now instead of expensive PCR test. The AI methods or machine and deep learning algorithms are only aiding approaches, which need their infrastructures and professionals and never can be replaced instead of clinically accepted methods such as radiology examinations, biopsy, etc. As an outcome, the development of a deep learning model eliminates the bulk of these limits, leading to greater revival in numerous nations' healthcare and financial sectors.

Machine learning, as well as deep learning, are used in all of the sounds classifying approaches. Support vector machines [22], as well as decision trees [23], are examples of machine learning classifiers, whereas Convolutional Neural Network (CNN) models are examples of deep learning classifiers (AlexNet [24], VGGNet [25], GoogleNet [26], ResNet [27]). The CNN image classification models family designed for speed and efficiency. The following are the research's key contributions:

- A new deep learning approach for detecting COVID-19 from a selection of tones.
- Using a Mel-scale spectrogram approach to transform sound into image, the suggested model improves sound detection efficiency.
- To achieve maximum efficiency, nine deep learning training models are used.

This part integrates available information on the application of deep learning and machine learning in sound detection. The sound classification stages may be divided into three distinct stages: pre-processing, extraction, as well as classification. The majority of sound detection

researches [28–30] concentrate on sound construction and recognition using traditional machine learning methods. The current study concentrates on the classification and recognition of cough and breath tones generated by those infected with the COVID-19 virus.

In the study of Schuller et al. [31] the deep learning algorithm based on Convolutional Neural Network (CNN) used to detect raw cough and breath of COVID-19 patients. They modified the CNN approach, which uses breath and cough sounds to detect if a patient is diagnosed with COVID-19 or healthy. The suggested method is better than the standard baseline. Although the CNN model reached an accuracy of 80.7%, a deep learning model can produce the greatest performance with the present data provided.

In the study of Bansal et al. [32], a CNN suggested model for COVID-19 sound classification rely on Mel-frequency cepstral coefficients (MFCC). Two learning-dependent techniques based on the Visual Geometry Group (VGG16) architecture were used. The presented model achieved 70.58% test accuracy and 81% sensitivity using a high-quality outcomes method.

In study of Imran et al. [33] the suggested model used to differentiate COVID-19 sounds from other non-COVID-19 tones. For training and testing, they utilized 1838 cough and 3597 non-cough tones divided into 50 classes. According to the study, the total accuracy of the deep learning-based multi-class classification achieved 92.6%.

In the study of Liu et al. [34] used a transfer machine learning approach to detect cough sound. The pre-training and fine-tuning are two main stages of Neural network (NN) models, and also the decoded data are gathered by a Hidden Markov Model (HMM). Three cough HMMs as well as only one non-cough HMM are introduced to the suggested model in this study. The experiments were carried out on a dataset collected from 22 persons suffering from a variety of respiratory illnesses. Their suggested method demonstrates that the suitable deep model can achieve a high precision of 90%.

In the study of Hee et al. [35] suggested a machine learning classification for asthmatic as well as healthy kids. The dataset contained 1192 cough cases from asthmatic kids as well as 1240 cough cases from healthy kids. The audio was used to generate features like MFCC. A Gaussian Mixture Model-Universal Context Model was used to construct the learned machine learning method. According to the findings, the total sensitivity and specificity of the classifier using machine learning were 82.81% and 84.76%, respectively.

In the study of Amrulloh et al. [36] the authors suggested a classification approach for pneumonia as well as asthma. The suggested method achieved 89% sensitivity and 100% precision. The findings demonstrate how their approach could be used to discriminate between pneumonia and asthma in open areas.

In the study of Loay et al. [37] the authors presented a classification model to detect COVID-19. The dataset presented includes 1457 wave cough sounds (755 of COVID-19 and 702 of healthy). The scalogram technique was used to transform sound to image. According to the findings, the total sensitivity and specificity of the suggested model were 94.44% and 95.37%, respectively.

In the study of Rahman et al. [38] the authors presented a novel machine learning method to detect COVID-19. Cambridge and Qatari are two breath and cough datasets used to examine this method. Cambridge dataset consists of 582 healthy and 141 COVID-19 patients. Qatari dataset consists of 245 healthy and 96 COVID-19 patients. According the results, the total accuracy, sensitivity and specificity of the presented approach were 96.5%, 96.42% and 95.47%, respectively.

In the study of Liu et al. [39] the purpose of this study is to explore at the virus's involvement in the respiratory system by examining the sounds of breathing and coughing. To identify COVID-19, this study introduced a unique multi-type feature fusion technique in a CNN-based architecture. The dataset presented consists of 282 (62 of COVID-19 and 220 of healthy). The scalogram technique was used to transform sound to image. According the results, the total accuracy of the presented technique were 85.4%.

In the study of Mehrotra et al. [40] this study used the DenseNet201 architecture to develop a three-dimensional deep convolutional neural network (D-CNN) relying on the transfer learning technique for the diagnosis of chronic pulmonary diseases. This study used two classification models. Model one attempts to determine if a given Chest X-Ray (CXR) image is normal or diseased. Model two explores the exact type of Chronic Pulmonary Disorder (CPD) with which the CXR image is infected if model one determines that it is. The dataset presented includes 1949 CXR images. According to the results, the total accuracy of model one and model two are same 96.8%.

The aim of this paper is to classify the various cough and breath sounds of COVID-19 artefacts in the signals from dynamic real-life environments. Also, this paper helps to detect COVID-19 patients from the comfort of their homes, so that they do not overload the Medicare system and therefore do not unwittingly spread the disease by regularly monitoring themselves. The majority of the research studies mentioned above applied mathematical analysis as well as machine learning to properly diagnose COVID-19 infections. Fewer research were identified to use transfer learning and CNN of cough and breath tone datasets for the variables of COVID patients and normal patients. As a result, more research on deep learning with simplified efficiency measurements were required. According to the literature review introduced here, cough and breath sounds should be used to diagnose COVID-19. The new paradigms are more effective and faster in fighting the COVID-19 pandemic. The novelty of this paper includes the use of a Mel-scale spectrogram to show signal characteristics as well as its ability to discriminate biometrically.

The current research organized as follows: Section 2 of this research presents the materials and methods. Section 3 describes the results of this paper. Section 4 provides the discussion of the suggested COVID-19 coughing and breathing sound model. Section 5 shows the conclusions of this paper.

2. Materials and methods

2.1. Dataset characteristics

The dataset utilized in this paper was collected from Coswara [41], this dataset aims to provide an open-access database for respiratory sounds of healthy and ill people, includes those infected with COVID-19. Since then, it has collected data from over 1600 people from all around the world (Male: 1185, Female: 415; mostly Indian population). Breath, cough, and speech sounds were collected by crowdsourcing via an interactive internet application designed for smartphone devices. All sounds were captured using a smartphone microphone as well as sampled with 48 kHz frequency range. All audio samples (in WAV format) were individually selected to use an online interface that allows many authors to examine each audio file as well as enhance the performance and accuracy of labeling. The database currently contains 120 COVID-19 cases, which is practically a one-to-ten ratio when compared to healthy (control) people. To establish a balanced dataset, all COVID-19 participants' data were evaluated, as well as the same number of tests were allocated randomly from the control participants' data. In addition, just two main types of breathing sounds, shallow as well as deep, were recorded from each patient and used for further study.

The suggested model was created to give breath and cough classification to present it in a public dataset. The diagnostic engine uses this breathing and cough classifiers to detect whether one sound is associated with COVID-19. To evaluate the classifier, we used data from the Coswara dataset as well as the dataset contains COVID-19 and healthy sounds.

2.2. The proposed model

Early detection of Covid-19 in both symptomatic and asymptomatic patients demonstrated its potential to identify Covid-19 cases at an early

stage. It will give enough time to implement Covid-19 therapies and to decrease death rates through early care. Early detection of Covid-19 can assist to limit propagation rates of $SARS - Cov - 2 \leq 1$, hence reducing infection propagation in contacts. Furthermore, early diagnosis of Covid-19 improves the capacity of healthcare workers in third - world countries lacking sufficient resources to detect Covid-19.

The origin of cough tones remains unknown since laryngeal structures as well as nasal and thoracic cavity resonance are all engaged in cough and their functions in cough are unclear. The normal cough sound is commonly separated into three parts (shown in Fig. 1) [42].

- (1) An explosive expiration caused by the glottis quickly opening,
- (2) An intermediate phase with cough sound attenuation,
- (3) A voiced phase caused by the vocal cord shutting. In fact, coughing patterns vary; for instance, certain cough tones only have two main phases (the intermediate phase and the voiced phase), and the explosive phase is generally longer due to specific diseases.

Cough frequency assessment is regarded as the gold standard for objectively assessing cough. Besides it, the intensity of coughing, coughing pattern, and acoustic features of cough tones can be utilized as clinical endpoints. Coughing may be measured in four distinct ways:

- (1) Explosive cough sounds: the amount of distinct explosive cough impulses,
- (2) Cough seconds: the length of seconds and hours with at least one explosive phase,
- (3) Cough breaths: breathing rates that include at least one coughing phase,
- (4) Cough epochs: the quantity of cough sounds separated by no more than 2 s.

The proposed approach consists of two main components. The first stage converts the cough or breath sound-to-image, which is enhanced using a Mel-scale spectrogram approach. The next stage is to extract and classify features using nine deep transfer models (GoogLeNet, ResNet18/34/50/100/101, SqueezeNet, MobileNetv2, and NasNetmobile). Cough and breath patterns and the clarity of a patient's cough sound might provide important information about their health. So, in this paper a Chebyshev I filter is used to reduce low and high-frequency noise.

Fig. 3 shows the suggested Deep Learning cough-breathing classification model's architecture diagram. Pre-processing, feature extraction, and classification are all required by the proposed Deep Learning cough classification model. The proposed model includes 2 main steps. Step 1 is the feature extraction, which converts sound to image using a Mel-spectrogram, as well as the second step is the extraction of features and classifying model. The following Deep Learning models (GoogLeNet, ResNet18, ResNet34, ResNet50, ResNet100, ResNet101, Mobile-Netv2, NasNetmobile, and SqueezeNet) are used in the extraction of features and classification. The much more widely applied Deep Learning transfer learning models are GoogLeNet, ResNet, Mobile-Netv, NasNetmobile, and SqueezeNet. Deep Learning models were used in the feature extraction and classification steps of the proposed model's learning, validating, and assessment.

Breathing and coughing sounds were obtained from an open-access database for respiratory tones (Coswara). The dataset included data from almost 1600 persons (1185 Male and 415 Female) from all across the world. The study included 240 individuals, 120 of which were infected with COVID-19 and the surviving 120 were healthy. Then, a deep learning framework was used based on features obtained from nine deep transfer models and deep-activated features obtained from a convolutional and recurrent neural network combination. The efficiency of using artificial intelligence (AI) as a suitable pre-screening method for COVID-19 was then assessed and further explained.

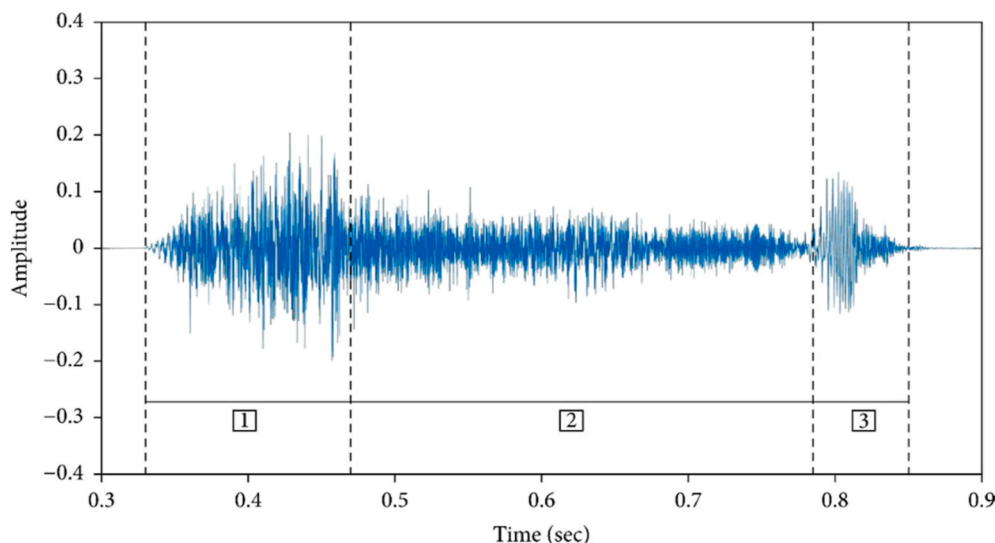


Fig. 1. Shows typical cough sound (1: explosive phase, 2: intermediate phase, and 3: voiced phase).

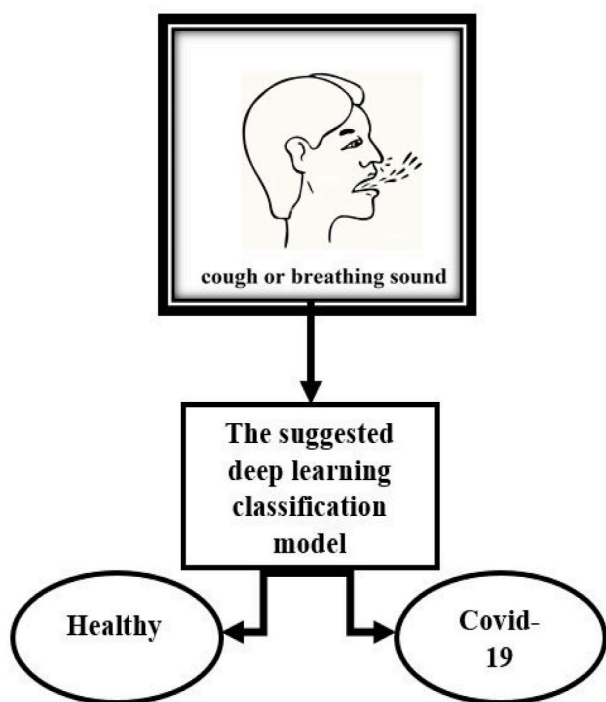


Fig. 2. The suggested Deep Learning classification model.

2.2.1. Mel-Scale spectrogram

Human ears, on the other hand, do not perceive variances in all frequency ranges in the same way. It becomes more difficult for humans to differentiate between distinct frequencies as the frequencies increase. We use the Mel scale to measure frequencies in order to successfully replicate human ear behavior using deep learning models. For human hearing, every equal distance between frequencies on the Mel-scale sounds similarly different. The Mel-scale uses the following equation to convert frequency from Hertz (f) to Mel (m):

$$m = 2595 * \log(1 + f / 700) \tag{1}$$

A spectrogram with frequencies calculated in Mel is known as a Mel-scale spectrogram. A Mel-scale spectrogram is a value of the short-time Fourier transform (STFT) [42]. This paper uses the Mel-scale

spectrogram technique in two measurements. The one dimension (1-D) signals are first preprocessed for denoising. Second, 2-D Mel-scale spectrogram applying CWT are implemented with the preprocessed signals. The CWT is applied to convert the signal from the time domain to the frequency domain. Convolution using a Chebyshev I filter removes low and high frequency noise. The CWT employs internal results to assess the similarity between a wave and an evaluation function like the Fourier transform (FT). The (CWT) is a time-frequency analysis approach that varies from the more standard (STFT) in that it allows indefinitely high-frequency signal feature localization in time. The CWT will do this by employing a flexible window width proportionate to the observer scale—flexibility that allows for the isolation of high-frequency characteristics. The CWT differs from the STFT in it is not restricted to the use of sinusoidal analyzing functions. The CWT of function $x(t)$ is measured using equation (2). Where, $\beta(t)$ is father signal, mostly in the time and frequency domains, $\beta(t)$ is a continuous function. (x) is the scale parameter's continuously varying values, and (y) is the position parameter's continuously varying values.

$$CWT(x, y) = (\sqrt{x})^{-1} \int_{-\infty}^{\infty} f(t)\beta(t - y/x)dt \tag{2}$$

The coefficients of CWT coefficients generate a matrix packed with wavelets that are positioned and scaled. The objective of the father signal is to give the production core feature of the child signals. Fig. 4 illustrates cough and breath sound waveforms, as well as a Mel-Scale spectrograms for COVID-19 and Non-COVID-19 patients.

2.2.2. Data augmentation

Because of the limited sample available, it is necessary for deep learning apps to integrate augmented data. Rather than training the model only on the current dataset, data augmentation enables the creation of fresh modified duplicates of the given samples. These new copies contain comparable properties to the original data, but they have been carefully modified as though they came from a different source (subject). This procedure is required in order to present the deep learning model to additional variances in the training data. It was vital to avoid the model from over-fitting, which occurs when the model learns just the input data with extremely limited generalisation capabilities for unseen data.

2.2.3. Deep learning framework

Many effective pre-train CNNs can pass learning. They do, however,

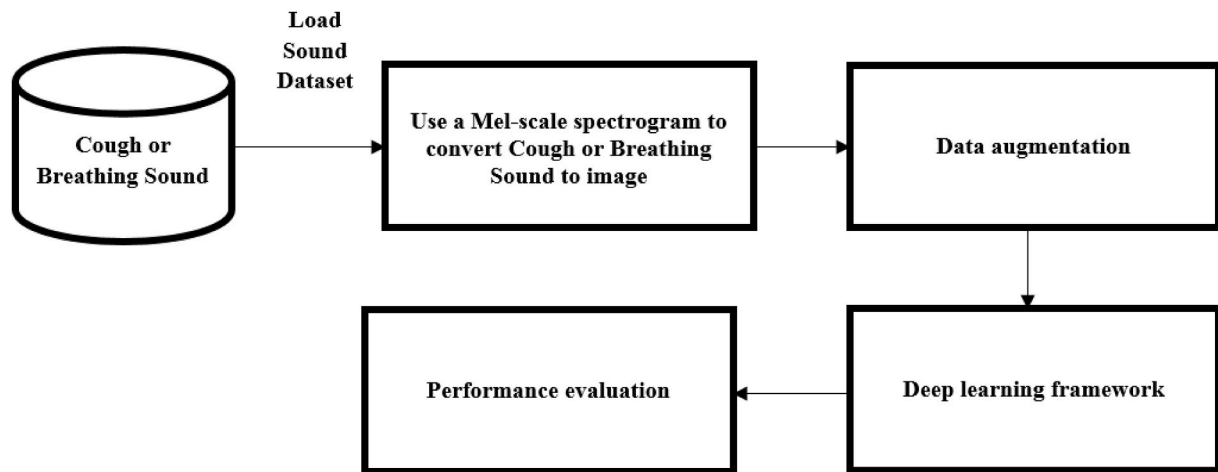


Fig. 3. The suggested cough and breathing Deep Learning classification model.

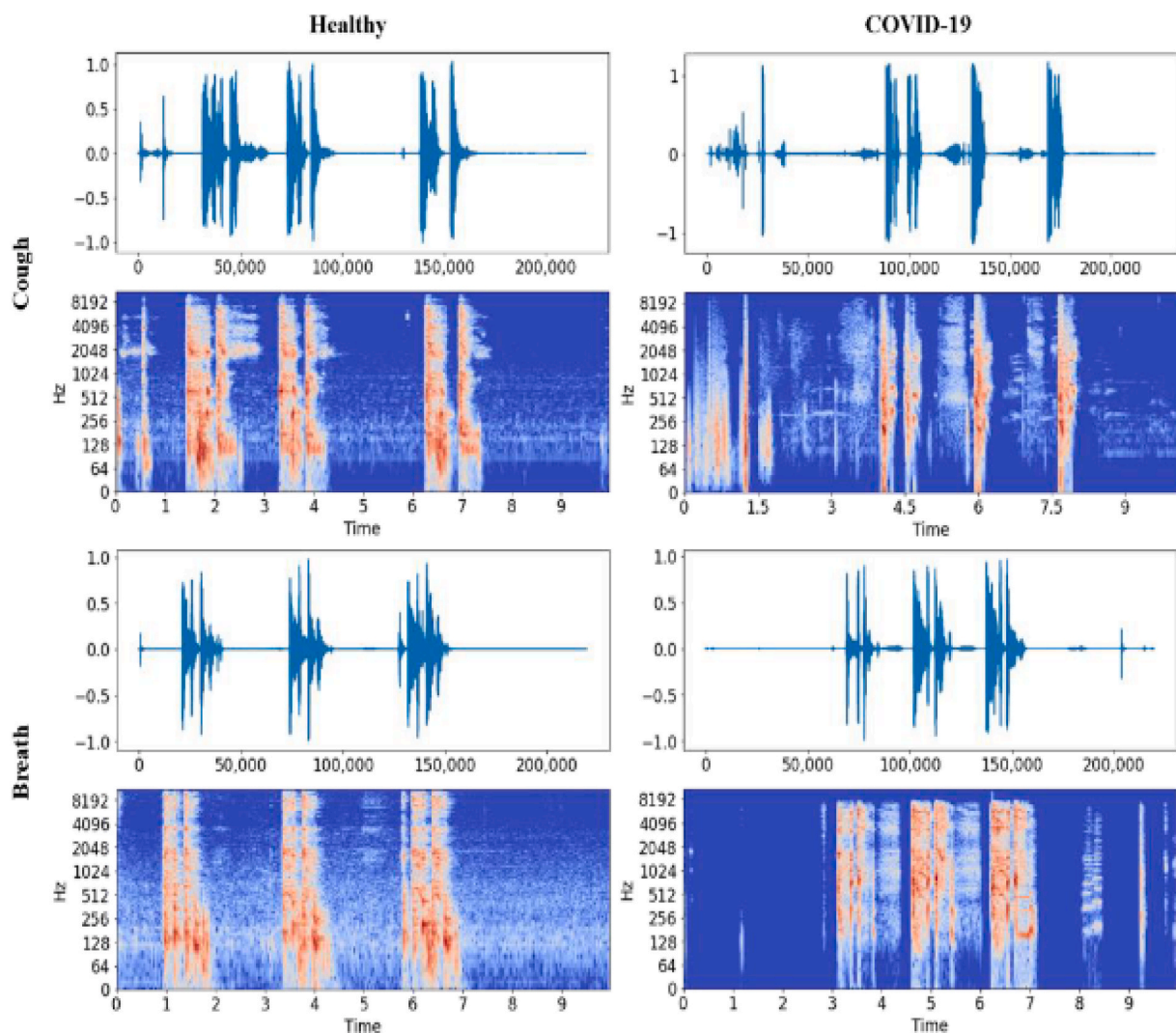


Fig. 4. Displays cough and breath sound waveforms, as well as a Mel-Scale spectrograms for COVID-19 and Non-COVID-19 patients.

need dataset training and analysis at the input layer. The networks are constructed using a variety of strategies and combinations.

Deep Learning models for mobile devices include MobileNetV2 and NasNetMobile. MobileNetV2 has a total of 155 layers and 164

connections in its architecture [43,44]. MobileNetV2 is built on a mobile architecture that uses depth-wise separable convolutions to generate light weight deep neural networks. NasNetMobile’s mobile edition has Twelve partitions. NasNet is a modular CNN made up of

basic construction components, which have been refined through recurrent neural networks [34]. A cell consists of simply a few operations, which are repeatedly copied owing to the network's needed scale. The layer Global Average Pooling [45] was employed, which considerably reduces the failure of forwarding error prediction.

SqueezeNet is a smaller network that was created to be a more compact alternative to AlexNet [46]. It has over 50 times fewer parameters than AlexNet yet executes three times faster.

The following are the main concepts of SqueezeNet:

- Approach one is to use 1×1 filters instead of 3×3 filters.
- Approach second: decrease the input channels to 3×3 filters.
- Approach three: reduce the network late in the process such that the convolution layers have huge activation maps.

The SqueezeNet architecture has 15 layers with 5 different layers as 2 convolution layers.

One of the most well-known deep learning models is the Residual Network (ResNet). The introduction of these Residual blocks alleviated the challenge of training very deep networks, and the ResNet model is built up of these blocks. ResNet offers a variety of models, including 18/34/50/101/152. ResNet18 has 72-layer architecture with 3×3 filter layers and eighteen deep layers. This network's architecture was designed to allow a high number of convolutional layers to function effectively. Therefore, adding several deep layers to a network frequently degrades the output. The ResNet-34 architecture included inserting shortcut connections onto a plain network to transform it into its residual network counterpart. The plain network was influenced by VGG neural networks (VGG-16, VGG-19) in this case, with a 3×3 filter in the convolutional networks. The ResNet-34 architecture has 34 convolution layers. Despite the fact that the Resnet50 design is based on the above model, it does vary in one important manner. Large Residual Networks, such as ResNet101, which has 101 layers, and ResNet152, which has 152 layers, are built using additional 3-layer blocks. The 152-layer ResNet has substantially reduced complexity even when the network depth is increased.

To substantially minimize the number of parameters, GoogLeNet is built on multiple very small convolutions. The design of GoogLeNet has 22 layers, however, parameters has dropped from 60 million (AlexNet) to 4 million. There are nine inception modules in GoogLeNet in order to examine clustering and network inside a network.

The module range is computed during the inception modules, and the completely linked layers are removed. Pooling in the inception modules, reduces the amount of parameters. In addition, a shadow network, as well as an auxiliary classifier, have been implemented to improve the results [47].

A Mel-scale spectrogram applied to convert a time-domain signal into a frequency-domain signal, which is then analyzed on multi-resolution. Considering this, the signal processing system keeps its morphological difficulty. This suggests that machine learning based on fundamental classifiers may be ineffective in identifying complicated signals. We presented an image to CNN's Deep Learning, which proved to be the most efficient for visual morphology detection. The output of Deep Learning models has not been equated. As an outcome, the current research aimed to develop the most representative deep learning models (GoogLeNet, ResNet18/34/50/100/101, MobileNetv2, SqueezeNet, and NasNetmobile) for image classification. The suggested classification model was assessed using an image as an original signal, since it is typical for Deep Learning models, as mentioned in the Experimental results section.

2.2.4. Network training

The network's construction begins with 1D convolutional layers. Convolutions in DL relate to a large number of dot products used to 1D data on pre-defined segments. The network collects deep features (activation) from the input data using sequential convolutions to produce an overall feature map. In this paper, there are 3 convolutional layers were utilized to build the initial stage of the deep neural network. every convolutional layer was pursued by a max-pooling layer, to decrease the model's dimensions and complexity.

We used the dropout strategy [48] to avoid over-fitting concerns with the Deep Learning net. The teaching criteria were the loss function $L(x, t)$, which itself is defined as the summation of binary plus box loss functions, as shown in eq. (3). Also, Eqs. (5) and (6) are used to determine the regression loss L_{re} :

$$L(x, t) = L_{cl}(x_c) + \delta[b > 0]L_{re}(z, z^*) \quad (3)$$

where, (z_a, z_b, z_w, z_h) indicates the bounding boxes of z and z^* , w as well as h signify the box's width and height respectively, and x_c denotes the predicted score class c . Non-background boxes at zero are defined by $\delta[b > 0]$. The bounding box, as well as the classification loss L_{cl} , are involved in the regression loss, as seen in eq. (4).

$$L_{cl}(x_{c^*}) = -\log(x_{c^*}) \quad (4)$$

$$L_{re}(k, k^*) = \sum_{i \in (a,b,w,h)} V_{LI}(k_i - k_i^*) \quad (5)$$

where,

$$V_{LI}(p) = \begin{cases} 0.5p^2, & \text{if } |p| < \text{zero} \\ |p| - 0.5, & \text{otherwise} \end{cases} \quad (6)$$

2.2.5. Performance evaluation

Testing might yield a positive result, demonstrating the stability of the DL models. The confusion matrix is a method for calculating the statistical performance of research. Accuracy, sensitivity, specificity, precision, F1 score, as well as the Matthews Correlation Coefficient (MCC), are 6 statistical evaluation. Fig. 5 and Fig. 6 illustrate the confusion matrices for the 2 categories (COVID-19 as well as Healthy). To get as close to the truth as possible, Eq. (7) was applied.

$$\text{accuracy} = TP + TN / (TP + TN + FP + FN) \quad (7)$$

where, TP, FN, TN, and FP are number of correctly labelled, mislabelled, clearly labelled instances of the remaining classes and incorrectly labelled instances of the remaining classes respectively.

The quantitative accuracy of the DL models' predictions was assessed. Sensitivity and precision are two extensively used classification efficiency measures. Eq. (8) and Eq. (9) are used to measure Sensitivity and precision respectively. The Sensitivity, Specificity, Precision, F1 Score, and MCC of the 9 Deep Learning models, are shown in Fig. 7. Eqs. (10)–(12) are used to calculate the Precision, F1 score, and MCC respectively.

$$\text{Sensitivity} = TP / (TP + FN) \quad (8)$$

$$\text{Specificity} = TN / (TN + FP) \quad (9)$$

$$\text{Precision} = TP / (TP + FP) \quad (10)$$

$$\text{F1 score} = 2(\text{Precision} \times \text{Sensitivity} / (\text{Precision} + \text{Sensitivity})) \quad (11)$$

$$\text{MCC} = TN \times TP - FP \times FN / \left(\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)} \right) \quad (12)$$

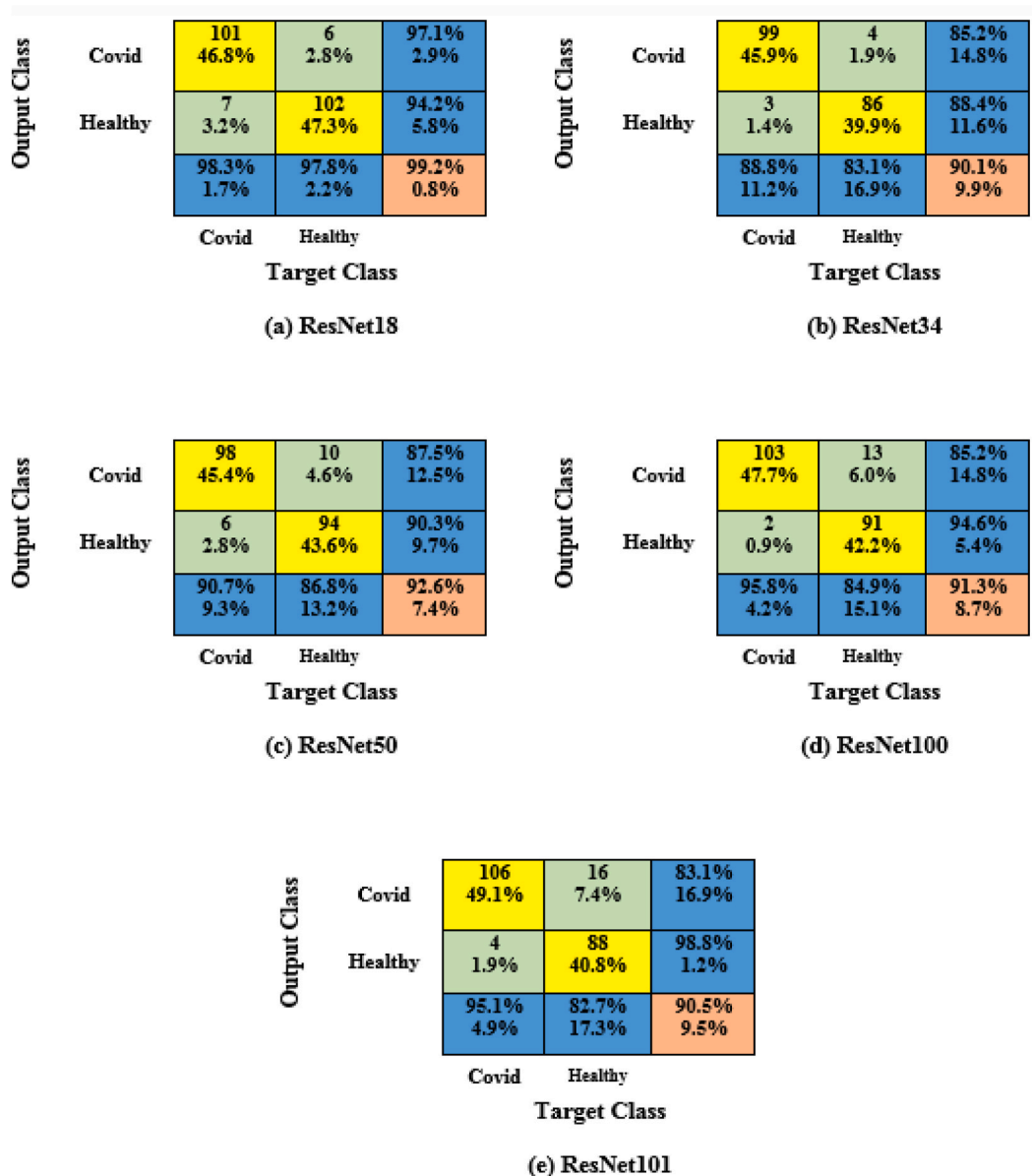


Fig. 5. Shows confusion matrix of ResNet18/34/50/100/101.

3. Results

The provided DL model is performed in transfer mode using the suggested basic training setup (batch norm epsilon, weight decay are equal e^{-3} , and batch norm decay, and dropout are equal $1/2$). The batch size is equal to 8, as well as the learning rate is equal to 0.02, which was lowered till it reached e^{-5} automatically. The Deep Learning models are tested for 20 h on a DELL PC with a 2.4 GHz Intel Core (TM) i7, MATLAB R2016 64-bit, and 16 GB RAM running Windows 10 as well as tensorflow's Deep Neural Network library (CuDNN).

The dataset separated to three parts: 80 % training images, 10% validation images, and 10% test images. In our study, we employed both labelled and assessment data. Validation accuracy is a classifying score, which is applied to evaluate the learning method as it progresses. The split ratio is determined by the size of the dataset. An optimal balance between training and testing must be achieved to guarantee the highest degree of model efficiency. Additionally, a direct reaction to the procedure or parameter is not available, which pushes one over the edge.

Similar training settings and stopping criteria as shown in Table 1 were used to train nine of deep learning transfer models. Table 1 contains the findings of each deep learning transfer model, with an initial learning rate of 0.02, number of epochs at 22, and epoch patience at 11. The batch size was fixed at 8, and stopping criteria was set to 11. It was discovered that when additional samples were utilized, the model output increased [49]. The optimizer approach used in this work to increase detector performance was Stochastic Gradient Descent with Momentum (SGDM) [50].

The performance of the 5 ResNet models (ResNet 18/34/50/100/101) are showed in Fig. 5 (a,b,c,d,e) as well as the total accuracy is 99.2%, 90.1%, 92.6%, 91.3%, and 90.5% respectively. Because of the limited dataset, Resnet18 obtained the maximum accuracy. Also, the accuracy of the GoogLeNet, MobileNetv2, NasNetMobile, and SqueezeNet models are showed in Fig. 6 (a,b,c,d) as well as the total accuracy is 89.8%, 88.1%, 89.4%, and 87.6% respectively. The precision, F1 score, and MCC for the 9 Deep Learning models are shown in Fig. 7.

Coughing is a symptom of over 30 medical conditions other than COVID-19. This makes detecting COVID-19 infection by coughing alone a significant challenge for a variety of illnesses. Recent study has started

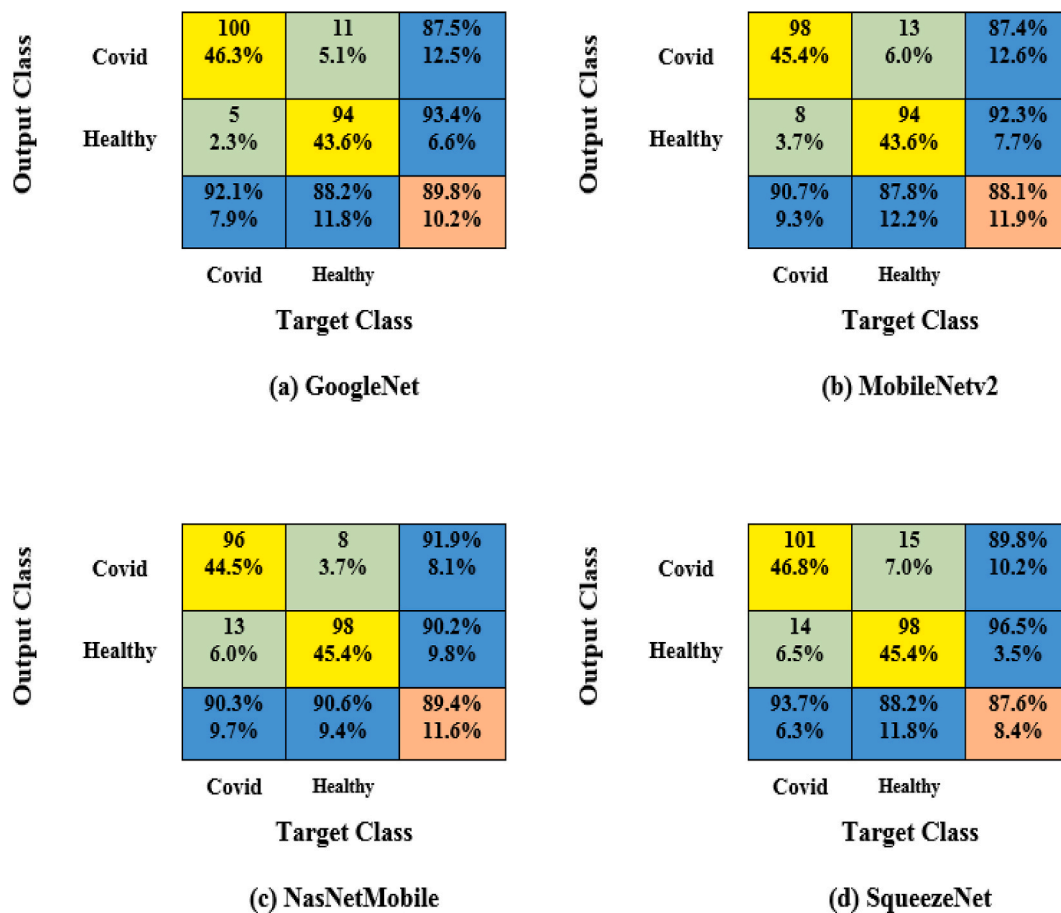


Fig. 6. Shows confusion matrix of GoogleNet, MobileNetv2, NasNetMobile, and SqueezeNet.

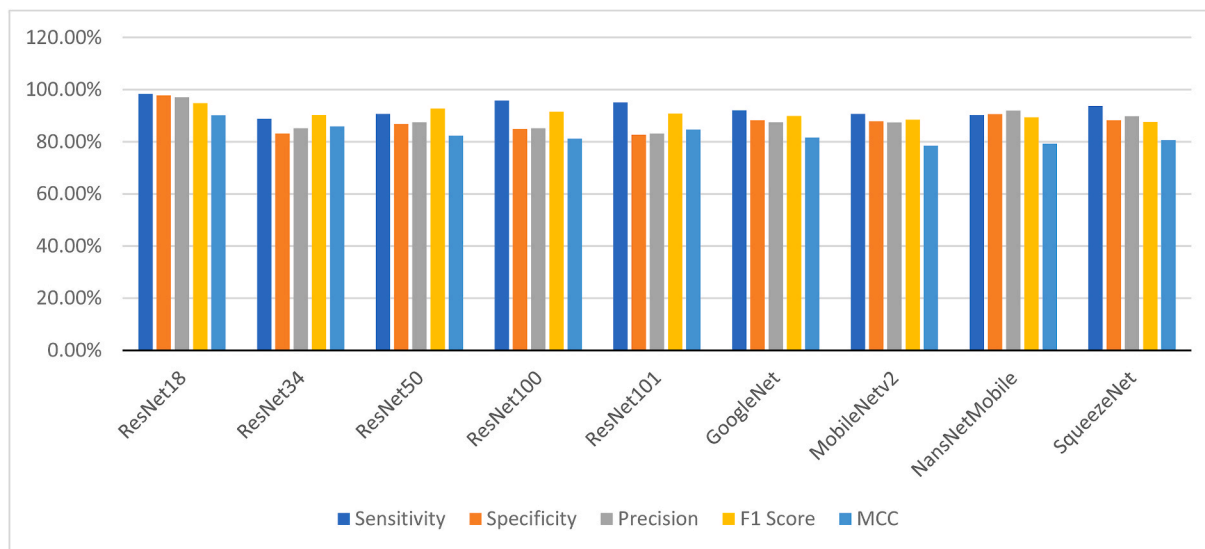


Fig. 7. Shows Sensitivity, Specificity, Precision, F1 Score, and MCC for all Deep Learning models.

Table 1
Deep learning Models setup.

	Batch Size	No. of Epochs	Epoch Patience	Stopping Criteria	Learning rate	Optimizer
Parameters	8	22	11	11	0.02	SGDM

to look at how breathing sounds (e.g., cough, breathing, and voice) captured in hospital by equipment from COVID-19 positive individuals vary from healthy people’s sounds. So, Eq. (1) is applied in this research to convert the sampled cough and breath sounds into Mel scale m for data pre-processing. Also, a Chebyshev I filter is used to reduce low and high-frequency noise. This is so that the proposed model can accurately diagnose the disease. In addition, ResNet18 has a high specificity of 97.8%, indicating that it can detect individuals who do not have COVID-19. ResNet18 has the highest sensitivity of 98.3%, this means that the test’s ability to identify the cough and breath sounds of COVID-19 patients. This clearly indicates that the proposed model’s Mel-scale spectrogram image will be able to distinguish COVID-19 from other diseases and normal lung states.

ResNet18 has the best precision of 97.1%; this means that ResNet18 model produces more relevant results than the others. The efficiency of the Deep Learning model is determined by a test with a high F1 score of 99.17% for ResNet18. Moreover, the MCC shows that the more reliable statistical rate obtained a good score in each of the 4 uncertainty matrix groupings. ResNet18 has the highest MCC of 90.6%. Additional data gathering is necessary to establish the deep transfer models’ capacity to be considerably more accurate. In spite of its high precision rates, the suggested study should be replicated on a larger scale, because of it might be used to other medical apps.

4. Discussion

Fig. 8 represents the results of the proposed model for the implementation of DL models in breath Dataset with Mel-scale spectrogram images of COVID-19 patients and healthy persons. Fig. 8 shows how our suggested model may correctly identify data with excellent accuracy. Moreover, the proposed study’s originality include the use of Mel-scale spectrograms with DL models to show signal properties and its ability to

Table 2

In terms of Accuracy, sensitivity, and specificity, a comparison of several methodologies is made.

Reference	Layers	Dataset	Accuracy	Sensitivity	Specificity
[31]	CNN	1427	80.7%	74.9%	76.1%
[32]	CNN	871	70.5%	81%	90%
[33]	CNN	317	92.6%	89.1%	96.7%
[37]	CNN	1457	94.9%	94.4%	95.4%
[38]	CNN	723	96.5%	96.4%	95.5%
[39]	CNN	439	85.4%	87.1%	87.3%
[40]	CNN	1949	96.8%	94.4%	97.7%
Current study	Deep Learning transfer model	1600	99.2%	98.3%	97.8%

identify biometrically. The majority of related research focuses on the categorization of cough and breath sounds using machine learning. Table 2 shows a performance comparison of several approaches in terms of Accuracy (AC). Also, Table 2 shows a comparable research that utilized a small dataset, that includes the actual COVID-19 cough sound collection. Most of the previous research is focused on the distinction between cough and non-cough tones. By determining the efficiency of deep learning transfer models in treating COVID-19 cough and breath tones with the SGDM, we discover that the performance assessment of several DL models improves obviously in the state of cough signals with high frequency. Despite our classification model finalises the superior, its accuracy is only 99.2% rely on the SGDM optimizer, the quality of the training data, as well as the try to validate the data labelling. Therefore, any problem in the data marking that escaped our examination is likely to have an impact on the recorded findings. This impact is more noticeable when the amount of data is very little.

The results of this study, as well as those of the other different researches mentioned in the Related Works section, suggest that discrete hidden properties of cough and breath tones can also be employed for

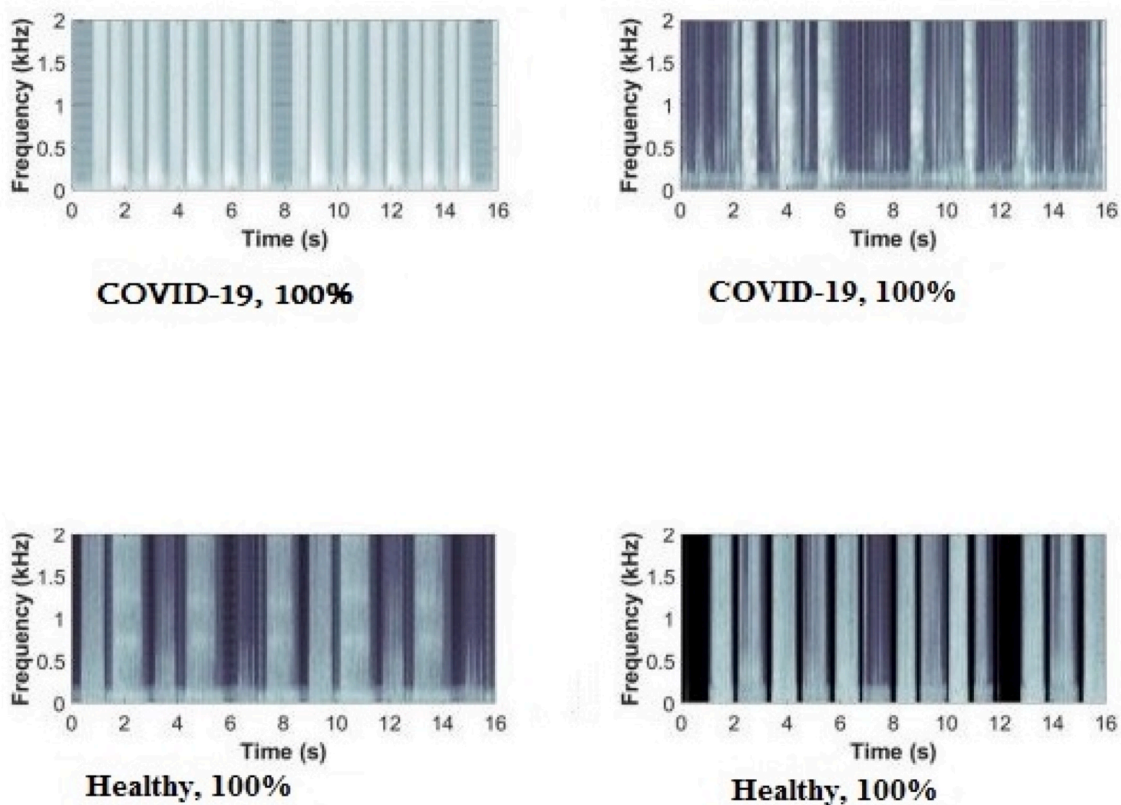


Fig. 8. Shows samples of shallow breathing noises and the spectrograms that are included in them.

successful deep learning detection of several respiratory illnesses. The cough and breath tones can be utilized as an initial diagnostic test tool since they distinguish healthy coughs and breaths from COVID-19 coughs and breaths. Our study used a Mel-scale spectrogram of tone as feedback to deep learning transfer models in order to identify which model is better at classifying medical images to audio.

In this paper, ResNet18/100/101 as well as SqueezeNet are shown to have peak accuracy and are referred to as deep types of deep learning transfer models. For mobile versions, MobileNetv2 has a better accuracy. The experiments are run on a single dataset for evaluation, which includes sound wave files. ResNet18 significantly outperforms GoogLeNet, SqueezeNet, and ResNet/50/100/101. The NasNetMobile was more accurate than MobileNetv2. This scenario is used in the experiments to evaluate the efficiency as well as consistency of the present classification model. The results show that the ResNet18 model achieves the highest classification accuracy on the cough and breath sounds from the validated COVID-19 dataset. The testing results for the deep learning classifier show that it matches the cough and breath tones of COVID-19 sufferers very well than other CNN classifiers. As a result, it may be more successful in diagnosing by relieving clinicians of the significant workload associated with the initial sounding of the COVID-19 breath and cough.

5. Conclusion

The current research has developed novel Deep Learning models for breathing and coughing sound categorization that concentrate on sound, which could assist with COVID-19 transmission control. Two primary components are combined in the suggested model. The first included employing a Mel-scale spectrogram to convert sound waves into an image. The development of universal characteristics, as well as extra classifying using deep transfer models, is the second component (GoogLeNet, ResNet18, ResNet34, ResNet50, ResNet100, ResNet101, MobileNetv2, SqueezeNet and NasNetmobile). Over 1600 individuals (1185 male and 415 female) from all across the world contributed data to the dataset (mostly the Indian population). Despite the fact that our recognition model has become the most reliable, its performance stays at 99.2%. Cough and breath studies efficiency is good enough to examine this model for estimations and generalization. ResNet18 was shown to be the most efficient in categorizing cough and breathing sounds when compared to another models evaluated on a smaller dataset. When compared to earlier COVID-19 cough sound research, this research was shown to be higher accurate and significant than all of the other current classifiers.

The outcomes of the proposed study lead to important recommendations for upcoming machine learning and deep learning studies. Our findings may be compared to a scalogram, which is another frequent sort of time-frequency representation. Considering its excellent overall accuracy, the proposed study requires wider replication before it is employed in additional healthcare applications. This study paves the door for the implementation of deep learning in COVID-19 diagnoses by demonstrating that it is a quick, time-efficient, in addition to the reduced technology that does not violate social separating requirements throughout pandemics such as COVID-19.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval

The submitted work is original and not has been submitted to another journal for simultaneous consideration. The manuscript is not published elsewhere in any form or language.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Funding

This research received no external funding.

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.

Acknowledgments

This research received no external funding

References

- [1] Chang C, et al. Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage* 2010;50(1):81–98. .
- [2] Aydın S, et al. Emotion recognition with eigen features of frequency band activities embedded in induced brain oscillations mediated by affective pictures. *Int J Neural Syst* 2016;26:1650013. 03.
- [3] Lachaux J, et al. Estimating the time-course of coherence between single-trial brain signals: an introduction to wavelet coherence. *Neurophys Clin Clin Neurophys* 2002;32(3):157–74. .
- [4] Klein A, et al. Conventional and wavelet coherence applied to sensory-evoked electrical brain activity. *IEEE Trans Biomed Eng* 2006;53(2):266–72. .
- [5] Aydın S, et al. Cortical correlations in wavelet domain for estimation of emotional dysfunctions. *Neural Comput Appl* 2018;30(4):1085–94 [].
- [6] Aydın S. Cross-validated adaboost classification of emotion regulation strategies identified by spectral coherence in resting-state. *Neuroinformatics* 2021:1–13. .
- [7] Kılıç B, et al. Classification of contrasting discrete emotional states indicated by EEG based Graph theoretical network measures. *Neuroinformatics* 2022:1–15. .
- [8] Rosati R, et al. A novel deep ordinal classification approach for aesthetic quality control classification. *Neural Comput Appl* 2022:1–15. .
- [9] WHO Coronavirus. COVID-19Weekly epidemiological update [Online], Available: <https://covid19.who.int>; Dec. 30, 2021.
- [10] Loey M, et al. Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning. *Symmetry* 2020;12(4):651.
- [11] Mohammad-Rahimi H, et al. Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review. *Front Cardiovasc Med* 2021;8:638011. .
- [12] Özkaya U, et al. Coronavirus (COVID-19) classification using deep features fusion and ranking technique. In: *Big data analytics and artificial intelligence against COVID-19. Innovation Vision and Approach*; 2020. p. 281–95. .
- [13] Wang X, et al. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Trans Med Imag* 2020;39(8):2615–25. .
- [14] Majid A, et al. COVID19 classification using CT images via ensembles of deep learning models. *Comput Mater Continua (CMC)* 2021:319–37 [].
- [15] Khan MA, et al. Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine. *Comput Electr Eng* 2021; 90:106960 [].
- [16] Akram T, et al. A novel framework for rapid diagnosis of COVID-19 on computed tomography scans. *Pattern Anal Appl* 2021;24(3):951–64. .
- [17] Zhao W, et al. Deep learning for COVID-19 detection based on CT images. *Sci Rep* 2021;11(1):1–12. .
- [18] Khan MA, et al. COVID-19 case recognition from chest CT images by deep learning, entropy-controlled firefly optimization, and parallel feature fusion. *Sensors* 2021; 21(21):7286. .
- [19] Shui-Hua W, et al. Deep rank-based average pooling network for COVID-19 recognition. *Comput Mater Continua (CMC)* 2022:2797–813 [].
- [20] Zhang YD, et al. Pseudo zernike moment and deep stacked sparse autoencoder for COVID-19 diagnosis. *Cmc Comput Mater Continua* 2021:3145–62 [].
- [21] Kaushik H, et al. Screening of COVID-19 patients using deep learning and IoT framework. *Cmc Comput Mater Continua* 2021:3459–75 [].
- [22] Bhateja V, et al. Pre-processing and classification of cough sounds in noisy environment using SVM. In: *2019 4th international conference on information systems and computer networks (ISCON)*; 2019. p. 822–6. .
- [23] Gao W, et al. Analysis of cough detection index based on decision tree and support vector machine. *J Combin Optim* 2019;37(1):375–84 [].
- [24] Krizhevsky A, et al. Imagenet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst* 2012;25.

- [25] Liu S, et al. Very deep convolutional neural network based image classification using small training sample size. In: 2015 3rd IAPR Asian conference on pattern recognition (ACPR); 2015. p. 730–4.
- [26] Szegedy C, et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 1–9. .
- [27] He K, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–7781. .
- [28] Pramono RXA, et al. A cough-based algorithm for automatic diagnosis of pertussis. *PLoS One* 2016;11(9):e0162128.
- [29] Rochmawati N, et al. Covid symptom severity using decision tree. In: 2020 Third international conference on vocational education and electrical engineering (ICVEE); 2020. p. 1–5. .
- [30] Soliński M, et al. Automatic cough detection based on airflow signals for portable spirometry system. *Inform Med Unlocked* 2020;18:100313. .
- [31] Nessiem MA, et al. Detecting COVID-19 from breathing and coughing sounds using deep neural networks. In: 2021 IEEE 34th international symposium on computer-based medical systems (CBMS); 2021. p. 183–8.
- [32] Bansal V, et al. Cough Classification for COVID-19 based on audio mfcc features using Convolutional Neural Networks. In: 2020 IEEE international conference on computing, power and communication technologies (GUCON); 2020. p. 604–8. .
- [33] Imran A, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlocked* 2020;20:100378. .
- [34] Liu JM, et al. Cough event classification by pretrained deep neural network. *BMC Med Inf Decis Making* 2015;15(4):1–10. .
- [35] Hee HI, et al. Development of machine learning for asthmatic and healthy voluntary cough sounds: a proof of concept study. *Appl Sci* 2019;9(14):2833.
- [36] Amrulloh Y, et al. Cough sound analysis for pneumonia and asthma classification in pediatric population. In: 2015 6th international conference on intelligent systems, modelling and simulation; 2015. p. 127–31.
- [37] Loey M, et al. COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models. *Comput Biol Med* 2021;139:105020. .
- [38] Rahman T, et al. QUCoughScope: an intelligent application to detect COVID-19 patients using cough and breath sounds. *Diagnostics* 2022;12(4):920.
- [39] Liu S, et al. COVID-19 detection with a novel multi-type deep fusion method using breathing and coughing information. In: 2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC); 2021. p. 1840–3. .
- [40] Mehrotra R, et al. Diagnosis of hypercritical chronic pulmonary disorders using dense convolutional network through chest radiography. *Multimed Tool Appl* 2022;81(6):7625–49. .
- [41] Alkhodari M, et al. Detection of COVID-19 in smartphone-based breathing recordings: a pre-screening deep learning tool. *PLoS One* 2022;17(1):e0262448.
- [42] Shi Y, et al. Theory and application of audio-based assessment of cough. *J. Sensor*. 2018:1–7. <https://doi.org/10.1155/2018/9845321>.
- [43] Addison PS. Wavelet transforms and the ECG: a review. *Physiol Meas* 2005;26(5):R155.
- [44] Yasar H, et al. A novel comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning, and deep learning methods. *Multimed Tool Appl* 2021;80(4):5423–47. .
- [45] Talha M, et al. Transferable learning architecture for scalable visual quality inspection. In: 2022 2nd international conference on artificial intelligence (ICAI); 2022. p. 26–32.
- [46] Xu Y, et al. Human activity recognition and embedded application based on convolutional neural network. *J Artif Intell Technol* 2021;1(1):51–60. .
- [47] Kasim NABM, et al. Celebrity face recognition using deep learning. *Indo J. Elect Eng Comput Sci* 2018;12(2):476–81.
- [48] Srivastava N, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014;15(1):1929–58. .
- [49] Xu Y, et al. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J Anal Test* 2018;2(3):249–62. .
- [50] Sutskever I, et al. On the importance of initialization and momentum in deep learning. In: International conference on machine learning; 2013. p. 1139–47. .