A New Model to Detect COVID-19 Coughing and Breathing Sound Symptoms Classification from CQT and Mel Spectrogram Image Representation using Deep Learning

Mohammed Aly[1]*
Department of Artificial Intelligence
Faculty of Computers and Artificial Intelligence,
Egyptian Russian University, Badr City, 11829, Egypt

Nouf Saeed Alotaibi[2]
Department of Computer Science
College of Science, Shaqra University
Shaqra City 11961, Saudi Arabia

Abstract*—Deep Learning is a relatively new Artificial Intelligence technique that has shown to be extremely effective in a variety of fields. Image categorization and also the identification of artefacts in images are being employed in visual recognition. The goal of this study is to recognize COVID-19 artefacts like cough and also breath noises in signals from real-world situations. The suggested strategy considers two major steps. The first step is a signal-to-image translation that is aided by the Constant-Q Transform (CQT) and a Mel-scale spectrogram method. Next, nine deep transfer models (GoogleNet, ResNet18/34/50/100/101, SqueezeNet, MobileNetv2, and NasNetmobile) are used to extract and also categorise features. The digital audio signal will be represented by the recorded voice. The CQT will transform a time-domain audio input to a frequency-domain signal. To produce a spectrogram, the frequency will really be converted to a log scale as well as the colour dimension will be converted to decibels. To construct a Mel spectrogram, the spectrogram will indeed be translated onto a Mel scale. The dataset contains information from over 1,600 people from all over the world (1185 men as well as 415 women). The suggested DL model takes as input the CQT as well as Mel-scale spectrograms derived from the breathing and coughing tones of patients diagnosed using the coswara-combined dataset. With the better classification performance employing cough sound CQT and a Mel-spectrogram image, the current proposal outperformed the other nine CNN networks. For patients diagnosed, the accuracy, sensitivity, as well as specificity were 98.9%, 97.3%, and 98.1%, respectively. The Resnet18 is the most reliable network for symptomatic patients using cough and breath sounds. When applied to the Coswara dataset, we discovered that the suggested model's accuracy (98.7%) outperforms the state-of-the-art models (85.6%, 72.9%, 87.1%, and 91.4%) according to the SGDM optimizer. Finally, the research is compared to a comparable investigation. The suggested model is more stable and reliable than any present model. Cough and breathing research precision are good enough just to test extrapolation as well as generalization abilities. As a result, sufferers at their headquarters may utilise this novel method as a main screening tool to try and identify COVID-19 by prioritising patients' RT-PCR testing and decreasing the chance of disease transmission.

Keywords*—*COVID-19; median filter; deep learning; Mel-scale spectrogram; sound classification; constant-Q Transform

I. Introduction

COVID-19 is an unique SARS disease which first surfaced in 2019 and has since spread over the world, producing a worldwide pandemic [1]. In accordance with the World Health Organization's (WHO) April 2021 report [2] there really are currently over 150 million documented illnesses including over 3 million deaths. Moreover, across over 32.5 million new cases and 500,000 fatalities, the USA has the highest overall number of illnesses as well as deaths.

These large numbers have placed a strain on numerous healthcare systems, particularly given the virus's propensity to cause more genetic variants and spread faster among people.

Recent studies have now employed relatively new artificial intelligence (AI) algorithms to recognise and categorise COVID-19 in CT and X-ray images [3]. Several studies (including CT scans as raw data) used machine and also DL techniques to distinguish among healthy and infected subjects with a discriminating accuracy of much more than 95% [4-9]. The capability of different classifiers including such support vector machines (SVM) and also convolutional neural networks (CNN) to identify COVID-19 in CT images with few which was before stages is the great contribution of these investigations. Additionally, some publications have used DL with supplementary feature fusion approaches as well as entropy-controlled enhancement to detect COVID-19 in CT images [10-13].

In light of the above, this research proposes a thorough deep learning technique for COVID-19 identification using coughing as well as breathing signals (**Fig. 1**). The suggested method might be used as a quick, low-cost, as well as readily distributed COVID-19 pre-screening tool, particularly in locations where the virus has spread rapidly. Despite the fact that the current gold new standard for detecting viral infection, RT-PCR, seems to have a good success rate, it has several drawbacks, such as high costs for equipment and chemical agents, the require for expert medical and nursing staff for tests, breaches of social separation, as well as the long time it needed to achieve outcomes (2-3 days).

*Corresponding Author.

As a result, the construction of a DL model removes the majority of these constraints, resulting in stronger resurrection in the medical and financial domains of many countries.

All of the techniques to sound classification utilise machine learning (ML) and DL. ML classification methods include SVM [14] and decision trees [15], while DL classifiers include CNN models (AlexNet [16], VGGNet [17], GoogleNet [18], ResNet [19]). CNN image classification models are built for speed as well as efficiency. The below are the study's major contributions:

- A novel DL strategy for recognizing COVID-19 from a set of tones.
- The proposed model enhances sound recognition effectiveness by employing a Mel-scale spectrogram and CQT method to convert sound into image.
- Nine DL training models are employed to achieve optimal efficiency.

The present study is structured as follows: Section II of this study examines the existing literature. Section III highlights the major properties of the dataset. Section IV includes a presentation of the proposed COVID-19 cough as well as breath tones model. Section V displays the test findings, whereas Section VI gives the paper's conclusions.

II. Related Work

The rest of the published studies mentioned here used ML and statistical analytics to detect COVID-19 disease. There has been less research that applies CNN and transfer learning on coughing signals datasets to determine the features of normal as well as coronavirus patients. Additional studies on DL with simpler efficiency assessments are thus needed. In this research, a novel model using DL, CQT, and a Mel-scale spectrogram was developed to detect COVID-19. According to the research study here, it is suggested that cough sounds be used to diagnose COVID-19. In fighting the COVID-19 epidemic, the advances are much more efficient and quicker.

This section investigates the much more available research on COVID-19 diagnosis that use coughing signals. As a result, the most recent evaluation of DL for coughing signal scan processing is addressed. This section includes details about the use of ML and also DL in sound detection. The stages of signal categorization can be classified into three phases: pre-processing, extraction, and classification. The core of tone detection study is concentrated on sound generation as well as recognition using classic machine learning approaches [20–22]. This paper concentrated on categorizing and identifying breathing and coughing sounds caused by COVID-19 virus infected individuals. Schuller et al. [23] used CNN to create a deep learning strategy to identify raw breathing as well as coughing in COVID-19 patients. Researchers improved the CNN method, which employs breathing as well as coughing sounds to test whether a person has COVID-19 or is fit. The suggested model is about twice as effective as the typical starting point. The CNN model achieved an overall score of approximately 81%, indicating that a DL model can deliver the best results with the available data.

A CNN model for COVID-19 is shown audio categorization proposed via frequency cepstral coefficients (MFCC) in [24]. The VGG 16 architecture is used in two learning strategies. The provided model achieved an overall of nearly 71 percentage as well as a sensitivity of 81 percentage using a high-quality outcomes method. The authors of Ref. [25] established a methodology for distinguishing COVID-19 and healthy sounds. For training and evaluation they employed 1838 coughs and 3597 other signals divided into 50 groups. In accordance with the study, the DL-based multi-class classifier scored about 92 percent, overall total accuracy. Prior to the COVID-19 pandemic, other research, like [26], to identify cough sound occurrences, a transfer learning technique applied. The NN models are developed in two stages: pre training & fine-tuning, after which the decoded data is collected by a Hidden Markov Model (HMM). In this work, three cough HMMs and one non-cough HMM are included to the proposed model. The experiments were carried out using a dataset generated from twenty two people suffering from various respiratory illnesses. This approach demonstrates that the qualifying deep model can now achieve a 90% precision level. M. aly et al. [27] proposed a classification model to identify COVID-19 in their investigation. The offered dataset contains 1600 wave coughing as well as breathing tones. To convert signal to image, the Mel-scale spectrogram method was utilised. Based on the data, the recommended model's overall accuracy, sensitivity, as well as specificity reached 99.2%, 98.3%, and 97.8%, respectively.

In [28], the author proposed a classification model for pneumonia and asthma. Their approach used MFCC, Shannon entropy, as well as non-Gaussian distributions to quantify signal parameters, and all these attributes have been determined to be the basis for artificial neural network classifiers. The suggested technique has 89 percent sensitivity as well as 100 percent accuracy. The results demonstrate how this technique may be applied to distinguish between pneumonia as well as asthma in public environments. According to [29], the goal of this study is to characterize the unique coughing sounds tones of COVID-19 artefacts in signals from various real-life scenarios. The model provided here tends to take two crucial stages into consideration. Converting the signal to image is the first step, which is improved using the scalogram approach. The second step is feature extraction as well as classification. The dataset utilized comprises 1457 wave coughing tones (755 from COVID-19 as well as 702 from healthy). The machine learning classifier's overall sensitivity and specificity were approximately 94% and 95 %, respectively.

An obvious and common problem with most previous COVID-19 research is that it uses a small dataset. Big data is preferred to little data because the higher the sample size, the more exact your estimations will be. Small data has a few advantages. For example, tiny data makes visualization, examination, as well as knowing what is going on in the data much simpler than enormous data. Furthermore, the innovation of this study is in the development of a DL model based on CQT as well as Mel-scale spectrogram-based breath and cough recordings into this DL model, which performs much better than conventional respiratory auscultation devices. Where, electronic stethoscopes are preferred because they are more accessible to a larger population. This is essential for obtaining medical data regarding COVID-19 patients in a responsible way, while keeping isolated behavior amongst persons. Moreover, this research examined patients from India, whose COVID-19 has a unique genetic variant likely of eluding the immune system as well as most available immunizations. As a result, it focuses attention on the ability of artificial intelligence algorithms to detect this viral illness in persons with this unique variant, even those who are asymptomatic.

AI design does not need a large amount of memory. This is a good strategy for the future expansion of telehealth and smartphone applications for COVID-19 (or other pandemics) that can offer real-time information and efficient and quick exchanges between patients and healthcare professionals. As a result, as a COVID-19 pre-screening tool, this enables for better and quicker isolation as well as contact tracing than presently existing methodologies.

III. Dataset Characteristics

The dataset for this investigation was obtained from a project aimed at creating an available dataset for pulmonary sounds of normal and unwell patients, which also included participants with COVID-19, according to coswara [30]. Ever since, it has gathered information from over 1,600 people from all around the world (Male: 1185, Female: 415; mostly Indian population). Crowdsourcing was used to gather breathing, coughing, and speech sound using an interactive online app tailored for smartphone devices [31]. All voices were recorded with a smartphone microphone and recorded at a frequency of 48 kHz. All audio samples (in.WAV file) were chosen at random to employ a web interface that enables many writers to review each audio file while also improving labelling performance and accuracy. There are now 120 COVID-19 instances in the database, representing a one-to-ten ratio as compared to normal (control) patients. To produce a balanced dataset, all COVID-19 participants' data was assessed, and the exact number of assessments was assigned randomly from the control participants' data. Furthermore, just two types of breath sounds, shallow and deep, were captured from each patient and used for subsequent research.

The proposed model was developed to classify breathing as well as coughing in order to offer it in a public dataset. This is used by the diagnostic engine. Classifiers for breathing as well as coughing are applied to determine if a sound is connected with COVID-19. We utilized data from the breathing dataset in addition to the COVID-19 and healthy tones dataset to assess the classifier.

IV. Proposed Model

Fig. 1 presents the suggested COVID-19 cough and breath sounds classification model. The architecture design of the proposed Deep Learning cough-breathing classification model is shown in Fig. 2. The presented DL cough classification model needs pre-processing, feature extraction, and classification. The suggested model consists of two major phases. The first phase is feature extraction, which transforms sound to picture using CQT and a Mel-spectrogram, and the second step is feature extraction and classification model. Deep Learning models such as GoogleNet, ResNet18, ResNet34, ResNet50, ResNet100, ResNet101, Mobile-Netv2, NasNetmobile, and SqueezeNet are used in feature extraction and classification. GoogleNet, ResNet, Mobile-Netv, NasNetmobile, and SqueezeNet are the most extensively used Deep Learning transfer learning models. Deep Learning models were employed in the suggested model's learning, validating, and assessing processes for feature extraction and classification.

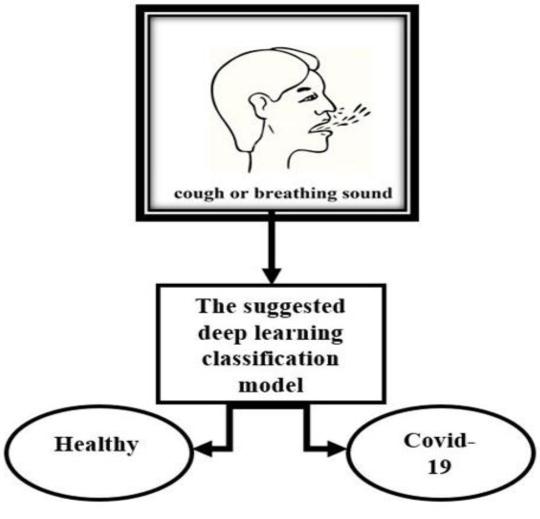


Fig. 1. The Suggested Deep Learning Classification Model.

A. Constant-Q Transform (CQT) and Mel-Scale Spectrogram

Human ears do not register variations across all frequency ranges equally. As frequency increases, it becomes increasingly difficult for individuals to distinguish between separate frequencies. The sound wave will be represented digitally by the voice recording. The CQT transforms a time-domain audio input to a frequency-domain signal [42]. To generate a spectrogram, the frequency will be converted to a log scale, and the amplitude will indeed be converted to decibels (db). The spectrogram will just be translated onto a Mel scale to generate a Mel spectrogram. In order to accurately imitate human ear behaviour using DL models, we employed the Mel scale to quantify frequencies. Each equivalent length among frequencies on the Mel- scale sounds equally distinct to human ears. To transform frequency from Hertz (f) to Mel (m), the Mel-scale utilizes the following equation:

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

A Mel-scale spectrogram is a spectrogram with frequencies estimated in Mel. A Mel-scale spectrogram is a short-time Fourier transform (STFT) value [32]. The CQT and a mel-scale spectrogram are employed in two ways in this work. To decrease noise, the 1-D electrocardiogram (ECG) data will be first standardized. Second, the preprocessed signals are presented to a 2-D mel-scale spectrogram using Continuous Wavelet Transform (CWT). As illustrated in

Fig. 3, the ECG employs CWT to transform the signal from time domain to frequency domain. Convolution using only a median filter utilised to decrease low and high-frequency noise. Small amplitude features of the ECG that are of physiological or clinical importance are generally obscured by noise and interference. Because noise's bandwidth overlaps that of desired signals, basic filtering is insufficient to enhance the signal-to-noise ratio (SNR). The CWT typically uses findings to determine the resemblance of a wave to an evaluation function such as the Fourier transform (FT). The (CWT) is a time-frequency analysis method that differs from the more common (STFT) in that it enables for unlimited high-frequency signal feature localisation in time.

The CWT will accomplish this by using a variable window width proportional to the observer scale—flexibility that will provide for the separation of high-frequency features. The CWT is distinct from the STFT in that it is not limited to using sinusoidal analysing 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.

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

The coefficients of CWT coefficients provide a matrix filled with located and scaled wavelets. The father signal's goal is to provide the generation fundamental characteristic of the child signals. Cough tones and breath sounds were separated from the dataset. The COVID-19 and normal groups' symptomatic breathing and coughing sounds were compared over time to see whether there had been any differences (Fig. 3).

B. Deep Learning Models

Many successful pre-train CNNs are capable of passing learning. Furthermore, they require dataset preparation and analysis at the input layer. A multitude of procedures and combinations are used to build the networks. MobileNetV2 and NasNetMobile are 2 DL models for smart phones. MobileNetv2's design has 155 layers as well as 164 connections [33, 34]. Separable convolutions are employed in mobile design, are utilized in MobileNetv2. NasNetMobile's mobile edition is divided into twelve sections. NasNet is a flexible CNN composed of fundamental construction components enhanced using recurrent neural networks [35]. A cell is made up of only a few actions that are frequently duplicated due to the network's required size. The layer Global Average Pooling [36] was used, which significantly minimises forwarding error prediction failure.

SqueezeNet is a small network designed to provide a more compacted alternative to AlexNet [37]. It has nearly 50 times less parameters than AlexNet yet performs three times quicker. SqueezeNet's core ideas are as follows:

- 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.

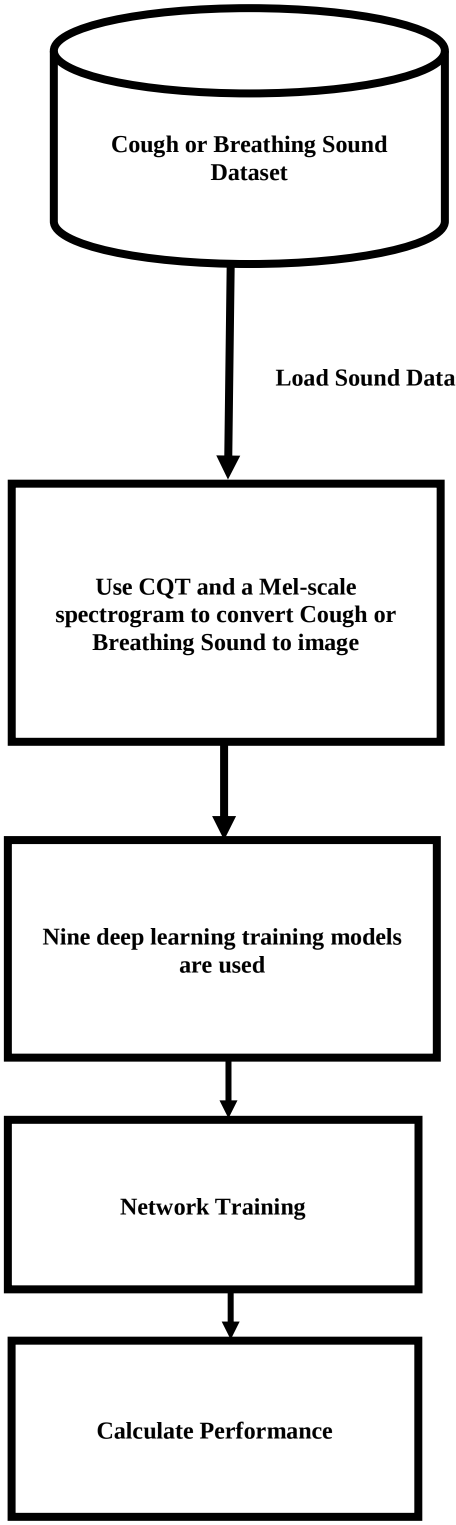
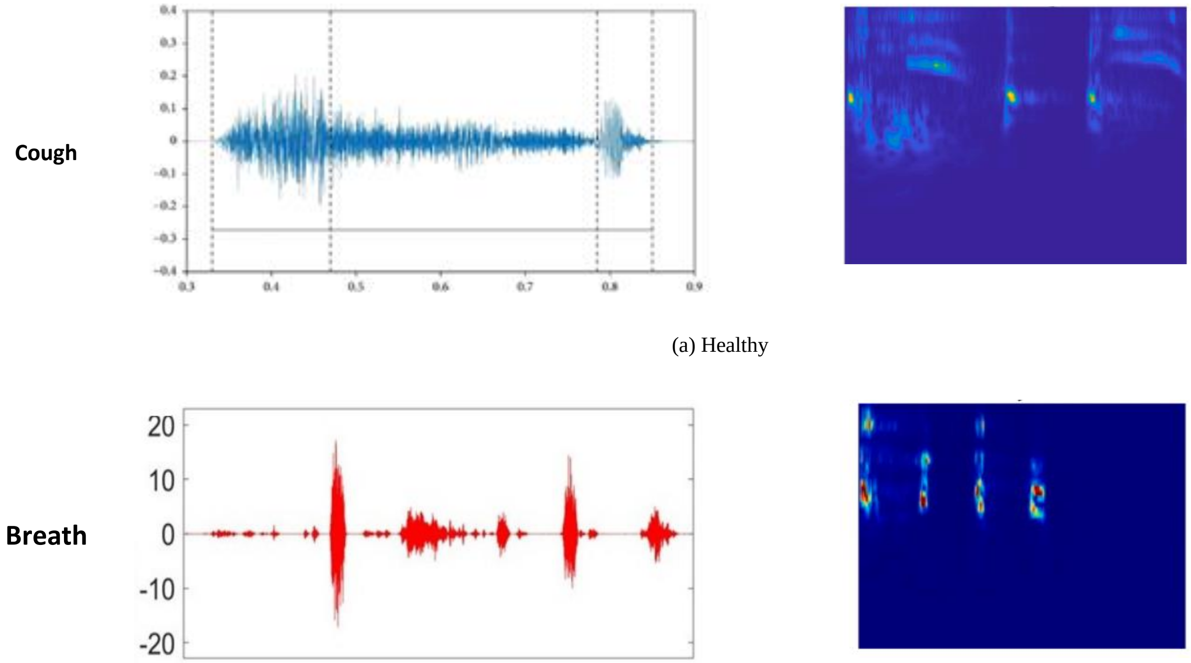

Cough or Breathing Sound Dataset
Load Sound Dataset
Use CQT and a Mel-scale spectrogram to convert Cough or Breathing Sound to image
Nine deep learning training models are used
Network Training
Calculate Performance

Fig. 2. The Presented COVID-19 Coughing and Breathing DL Classification Model.

The SqueezeNet design contains 15 layers, with 5 distinct layers and 2 convolution layers.



(a) Healthy

(b) COVID-19

Fig. 3. A Mel-Scale Spectrogram and Constant-Q Transform of an Electrocardiogram of Covid-19 and Healthy Cough and Breath Sounds.

The Residual Network is a well-known deep learning model (ResNet). The development of these Residual blocks lessened the difficulty of training very deep networks, and the ResNet model is founded on them. ResNet provides several models, like 18/34/50/101/152. ResNet18 has 18 convolutional layers and a 33 filter. The ResNet-34 design entailed placing shortcut links onto a plain network in order to convert it into its residual network counterpart. In this scenario, the plain network was impacted by VGG neural networks (VGG-16, VGG-19) with a 33 filter in the convolutional networks. The ResNet-34 design contains 34 convolution layers. Regardless of the fact that the Resnet50 architecture is centered on the preceding generation, it differs in one significant way. Large Residual Networks, like as ResNet101 (101 layers) as well as ResNet152 (152 layers), are constructed utilising extra 3-layer blocks. Also when network depth is raised, the 152-layer ResNet had significantly decreased complexity.

GoogLeNet is constructed on numerous extremely small convolutions to significantly reduce the amount of parameters. The GoogLeNet architecture contains 22 layers, although the parameters have been reduced from 60 million (AlexNet) to 4 million. GoogLeNet has nine inception modules to investigate clustering and network inside a network. During the inception modules, the module range is computed, and the entirely connected layers are deleted. Pooling parameters in the inception modules decreases the number of parameters. In furthermore, a shadow network and also an auxiliary classifier were used to enhance the findings [38].

A CQT is used to transform a time-domain signal to a frequency-domain signal, that is then evaluated with many resolutions. The use of a Mel-scale spectrogram to display signal characteristics, in addition to its ability to distinguish biometrically, distinguishes this paper. In light of this, the signal processing system maintains its morphological difficulties. This implies that ML based on basic classifiers may be unsuccessful in identifying complex signals. We sent an image through CNN's DL, which showed to be the most effective in detecting visual morphology. The DL model output has not been equal. As a result, the current study intended to create the most representative DL models for image categorization (GoogleNet, ResNet18/34/50/100/101, MobileNetv2, SqueezeNet, and NasNetmobile).

V. Experimental Results

The provided DL model is performed in transfer mode using the suggested basic training setup (batch norm epsilon= e^{-3}, weight decay= $\boldsymbol{e^{-3}}$, and batch norm decay= 0.5, and dropout= 0.5). The batch size= $\mathbf{8}$, as well as the learning rate $= \mathbf{0.02}$, which was lowered till it reached $\boldsymbol{e^{-5}}$ automatically. The Deep Learning models are tested for 20 hours on a DELL PC with a 2.4 GHz Intel Core (TM) i7-M520 CPU, MATLAB R2016 64-bit, and 16 GB RAM running Windows 10 as well as tensorflow's Deep Neural Network library (CuDNN).

The dataset was divided into three parts: 80% for training, 10% for validation, and 10% for testing. We used both labelled and assessment data in our investigation. Validation accuracy is a classification score that is used to assess the learning technique as it proceeds. The size of the dataset determines the split ratio. To ensure the maximum level of model efficiency, an appropriate balance between training and testing must be attained. Furthermore, there is no immediate reaction to the process or parameter pushes one over the brink. The results of each DL transfer model are shown in **Table I**, with an initial learning rate of 0.02 and 22 epochs. The batch size was set to eight, and early ending was permitted if there was no change in accuracy. It was revealed that by using more samples, the model output improved [39]. Stochastic Gradient Descent with Momentum (SGDM) [40] was the optimizer technique used in this study to improve detector performance. To prevent over-fitting issues with the Deep Learning net, we adopted the dropout approach [41]. As indicated in eq. (3), the teaching criteria were the loss function $L(x,t)$, which is defined as the total of binary plus box loss functions. Also, Eq. (5) and (6) are used to calculate the regression loss L_{re}:

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

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

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

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

Where,

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

A. Examination of Performance

Testing can yield a positive result, demonstrating the Deep Learning models' dependability. The confusion matrix is a statistical performance calculation approach used in study. Among the six statistical metrics are accuracy, sensitivity, specificity, precision, the F1 score, and the Matthews Correlation Coefficient (MCC). **Fig. 4** and **Fig. 5** show the confusion matrices for the two categories (COVID-19 as well as Healthy). Eq. 7 was used to get as close to the truth as feasible

$$accuracy = [N_{TP} + N_{TN}]/[(N_{TP} + N_{FP}) + (N_{TN} + N_{FN})] \qquad (7)$$

Where, N_{TP}, N_{FN}, N_{TN}, and N_{FP} are No. of correctly labeled, mislabeled, clearly labelled instances of the remaining classes and incorrectly labelled instances of the remaining classes respectively. The efficiency of the five ResNet models (ResNet 18/34/50/100/101) is shown in Fig. 4(a, b, c, d, e), and the overall accuracy is 98.9%, 91.4%, 93.1%, 92.9%, and 90.1%. The confusion matrix of the test for the GoogleNet model is given in Fig. 5(a), as well as the accuracy rate is 89.9%. Fig. 5(b, c, d) depicts the performance of MobileNetv2, NasNetMobile, and SqueezeNet, with accuracy rate of 89.2%, 88.9%, and 86.9%, respectively. Because of the tiny dataset, Resnet18 obtained the maximum accuracy. The accuracy of DL models' predictions was quantitatively tested. Both sensitivity and accuracy are widely utilized classification efficiency metrics. Eq. 8 and 9 are applied to calculate Sensitivity and Precision. Fig. 6 depicts the sensitivity and specificity of the nine Deep Learning models.

ResNet101 has a sensitivity of 95.6% when it relates to distinguishing COVID-19 persons' breathing sounds. ResNet18 does indeed have 98.1% specificity, meaning that it can detect people who do not have COVID-19. Precision, F1 score, and MCC are calculated using Eq. 10, 11, and 12. **Fig. 6** depicts the accuracy, F1 score, as well as MCC for the nine Deep Learning models. ResNet18 does have the highest accuracy of 96.4%, indicating that it generates more relevant findings than the other models. In **Fig. 6**, the DL model's efficiency is assessed by a test with a high F1 score of 95.9% for ResNet18. Finally, the MCC demonstrates that the more statistically reliable rate performed well in all four categories of the uncertainty matrix. ResNet18 has the highest MCC of 91.8 percent.

$$Sensitivity = N_{TP}/(N_{TP} + N_{FN}) \qquad (8)$$

$$Specifcity = \frac{N_{TN}}{(N_{TN}+N_{FP})} \qquad (9)$$

$$Precision = N_{TP}/(N_{TP} + N_{FP}) \qquad (10)$$

$$F1\ score = 2 \times N_{TP}/(2 \times N_{TP} + N_{FP} + N_{FN}) \qquad (11)$$

$$MCC = \frac{N_{TN}\times N_{TP}-N_{FP}\times N_{FN}}{\left(\sqrt{(N_{TP}+N_{FP})\times(N_{TP}+N_{FN})\times(N_{TN}+N_{FP})\times(N_{TN}+N_{FN})}\right)} \qquad (12)$$

B. Examination of Performance Discussions and Comparative Analyses

In Fig. 6, the outcomes of the suggested method for applying deep learning DL models in the breathing dataset implementing CQT as well as Mel-scale spectrogram images of COVID-19 illness and healthy are shown. Fig. 7 shows how our proposed approach can effectively recognise data. The present study's innovations include the employment of CQT and a Mel-scale spectrogram with deep learning models to characterise signal characteristics and biometric identification capabilities. The core of related research focuses on classifying breathing and coughing signals through ML. Table II analyses the performance of several techniques in terms of accuracy. The authors in [23, 24] employed a small dataset that included the actual COVID-19 coughing sound sample in a comparable investigation. Much of the prior study has been on distinguishing between coughing and non-coughing tones. We noticed this when analysing the effectiveness of Deep Learning transfers methods in detecting COVID-19 cough sounds using the SGDM, when cough signals occur often, the efficiency of all Deep Learning techniques improves significantly. While having the greatest performance, our detection model's efficiency is just 98.9% relying on the SGDM optimizer, the learning data correctness and the effort to analyse the labelled data. Every inaccuracy in data recording that evaded our notice, on the other hand, is most likely to influence the reported outcomes. Table III demonstrates the accuracy of the proposed suggested model when implemented to the Coswara dataset. The state-of-the-art models are labelled in the first left column. According to Table III, the proposed technique achieves high accuracy compared to the other models. The findings of this study, as well as those of another study mentioned in the related works section, indicate that specific latent properties of coughing sounds may well be successfully exploited for DL identification of a variety of respiratory issues. As it differentiates between normal and COVID-19 coughing, the coughing can be used as a preliminary diagnostic technique. We study the use of a Mel-scale spectrogram of tone as a return to Deep learning to see if the model is greater than effective at identifying medical images to tone. ResNet as well as GoogleNet were proved to have great accuracy in this work despite being recognised as deep variants of DL transfer models. For mobile versions, NasnetMobile as well as mobilenetv2 provide great precision. For assessment, the tests are done on a separate dataset that consists of audio wave files. Resnet18 was much more effective than GoogleNet, while resnet34/50/100/101 was much more effective than GoogleNet. NasnetMobile outperformed Mobilenetv2m and Squeezenet in terms of accuracy. The case is used in the experiments to assess the existing classification model's efficiency and consistency. According to the results, the resnet18 model has the greatest classification accuracy on cough as well as breath signals from the confirmed COVID-19 dataset. The DL classification outperforms conventional CNN classifications in matching coughing and breathing sounds of COVID-19 sufferers. As an outcome, it could really aid in diagnosis by alleviating clinicians of the stress connected from the first sound of the COVID-19 cough as well as breath.

TABLE I. DEEP LEARNING MODELS SETUP

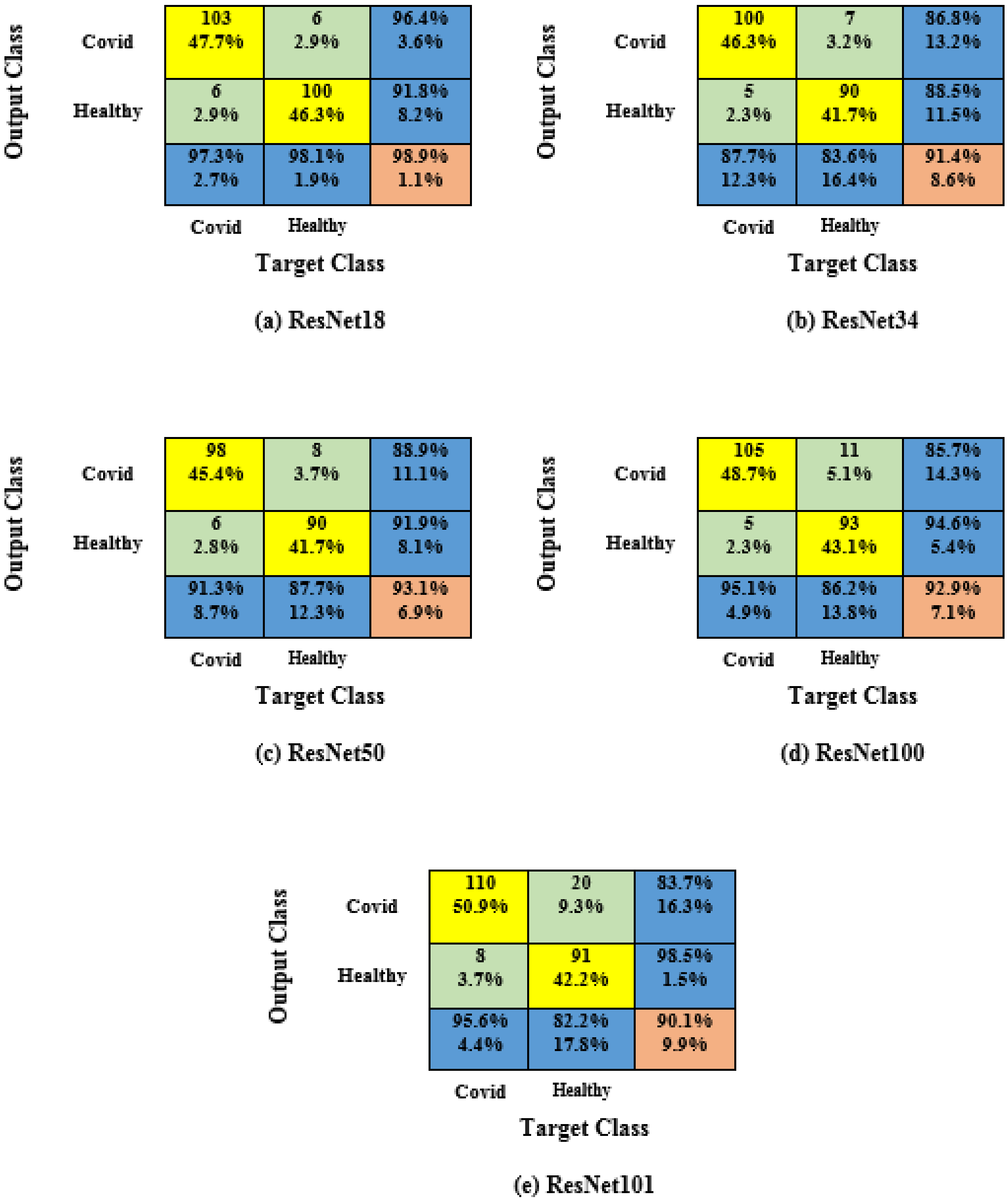
Deep Learning models	**LAYERS**	**Batch Size**	**Epoch**	**Learning rate**	**Optimizer**
Googlenet	20	8	22	0.02	SGDM
ResNet18	18				
ResNet34	34				
ResNet50	50				
ResNet100	100				
ResNet101	101				
MobileNetv2	53				
NasNetMobile	cells				
SqueezeNet	15				

TABLE II. IN TERMS OF ACCURACY, A COMPARISON OF SEVERAL METHODOLOGIES IS MADE

Reference	**LAYERS**	**Dataset**	**Result**
[23]	CNN	1427	80.7%
[24]	CNN	871	70.5%
[25]	CNN	317	92.6%
[29]	CNN	1457	94.9%
Current study	**Deep Learning transfer model**	**1850**	**98.9%**

TABLE III. IN TERMS OF ACCURACY, DISPLAYS OUTCOMES FOR VARIOUS MODELS FOR CLASSIFICATION ON COSWARA DATASET

Reference	**LAYERS**	**Result**
[23]	CNN	85.6%
[24]	CNN	72.9%
[25]	CNN	87.1%
[29]	CNN	91.4%
Current study	**Deep Learning transfer model**	**98.7%**



Output Class
Target Class

Output Class \ Target Class	Covid	Healthy	
Covid	103 47.7%	6 2.9%	96.4% 3.6%
Healthy	6 2.9%	100 46.3%	91.8% 8.2%
	97.3% 2.7%	98.1% 1.9%	98.9% 1.1%

(a) ResNet18

Output Class \ Target Class	Covid	Healthy	
Covid	100 46.3%	7 3.2%	86.8% 13.2%
Healthy	5 2.3%	90 41.7%	88.5% 11.5%
	87.7% 12.3%	83.6% 16.4%	91.4% 8.6%

(b) ResNet34

Output Class \ Target Class	Covid	Healthy	
Covid	98 45.4%	8 3.7%	88.9% 11.1%
Healthy	6 2.8%	90 41.7%	91.9% 8.1%
	91.3% 8.7%	87.7% 12.3%	93.1% 6.9%

(c) ResNet50

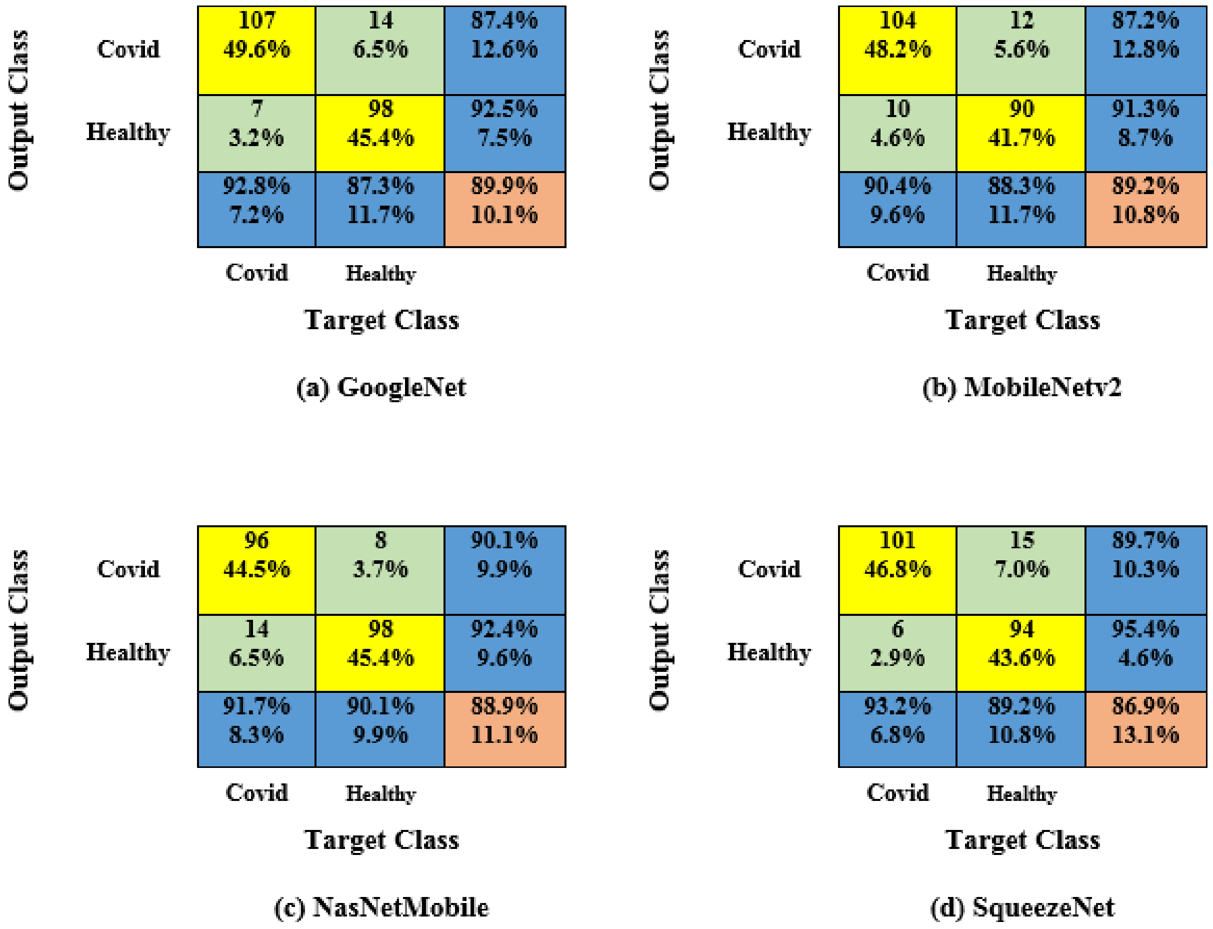
Output Class \ Target Class	Covid	Healthy	
Covid	105 48.7%	11 5.1%	85.7% 14.3%
Healthy	5 2.3%	93 43.1%	94.6% 5.4%
	95.1% 4.9%	86.2% 13.8%	92.9% 7.1%

(d) ResNet100

Output Class \ Target Class	Covid	Healthy	
Covid	110 50.9%	20 9.3%	83.7% 16.3%
Healthy	8 3.7%	91 42.2%	98.5% 1.5%
	95.6% 4.4%	82.2% 17.8%	90.1% 9.9%

(e) ResNet101

Fig. 4. Shows Confusion Matrix of ResNet18/34/50/100/100.



Output Class

	Covid	Healthy	
Covid	107 49.6%	14 6.5%	87.4% 12.6%
Healthy	7 3.2%	98 45.4%	92.5% 7.5%
	92.8% 7.2%	87.3% 11.7%	89.9% 10.1%

Target Class

(a) GoogleNet

Output Class

	Covid	Healthy	
Covid	104 48.2%	12 5.6%	87.2% 12.8%
Healthy	10 4.6%	90 41.7%	91.3% 8.7%
	90.4% 9.6%	88.3% 11.7%	89.2% 10.8%

Target Class

(b) MobileNetv2

Output Class

	Covid	Healthy	
Covid	96 44.5%	8 3.7%	90.1% 9.9%
Healthy	14 6.5%	98 45.4%	92.4% 9.6%
	91.7% 8.3%	90.1% 9.9%	88.9% 11.1%

Target Class

(c) NasNetMobile

Output Class

	Covid	Healthy	
Covid	101 46.8%	15 7.0%	89.7% 10.3%
Healthy	6 2.9%	94 43.6%	95.4% 4.6%
	93.2% 6.8%	89.2% 10.8%	86.9% 13.1%

Target Class

(d) SqueezeNet

Fig. 5. Shows Confusion Matrix of GoogleNet, MobileNetv2, NasNetMobile, and SqueezeNet.

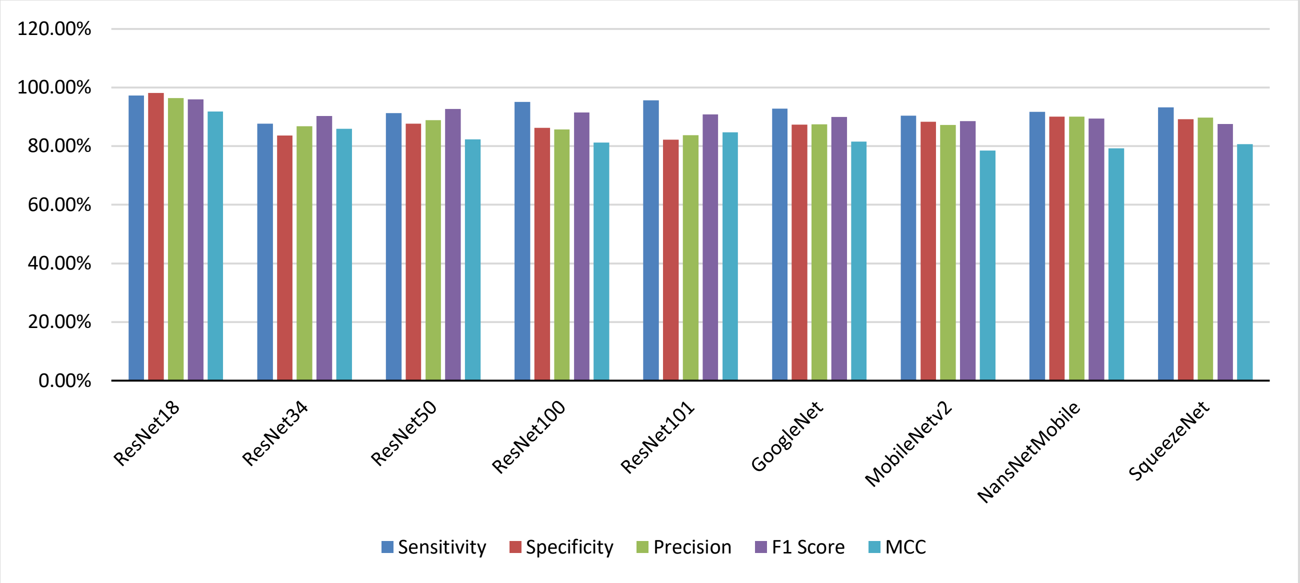
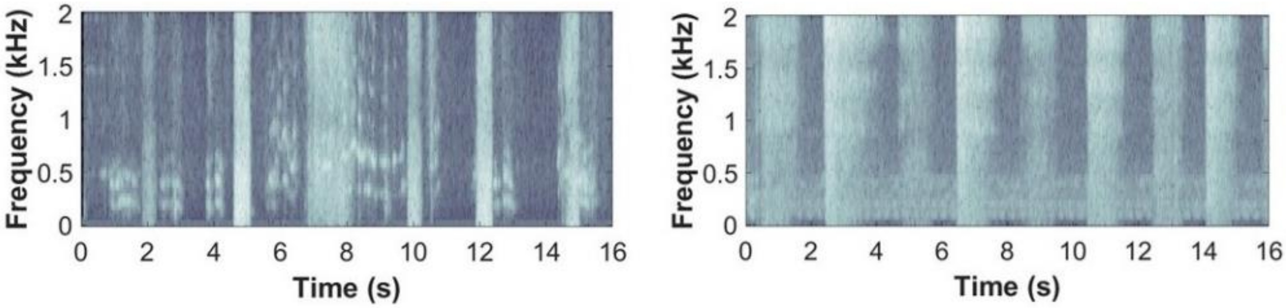


Fig. 6. Shows Sensitivity, Specificity, Precision, F1 Score, and MCC for All Deep Learning Models.



COVID-19, 100%

Healthy, 100%

Fig. 7. Shows Samples of Shallow Breathing Tones with their CQT and Mel- Scale Spectrograms.

VI. Conclusion

The current study created innovative DL models for breath and cough sound classification that focus on sound and might aid in COVID-19 transmission controls. The proposed model combines two key components. The initial method was using a CQT as well as a Mel-scale spectrogram to transform sound waves into images. The second component is the construction of universal features as well as extra classification utilising deep transfer models (GoogleNet, ResNet18, ResNet34, ResNet50, ResNet100, ResNet101, MobileNetv2, SqueezeNet and NasNetmobile). Around 1,600 people from all over the world (1185 men and 415 women) supplied data to the collection (mostly the Indian population). With the better classification performance employing coughing sound CQT and a Mel-spectrogram image, the current proposal outperformed the other nine CNN networks. For symptomatic patients, the accuracy, sensitivity, and specificity were 98.9%, 97.3%, and 98.1%, respectively. The Resnet18 network is the most dependable for symptomatic patients who use coughing as well as breathing tones. When applied to the Coswara dataset, we discovered that the suggested model's accuracy (98.7%) outperforms the state-of-the-art models (85.6%, 72.9%, 87.1%, and 91.4%) based on the SGDM optimizer. The suggested study's findings contribute to key suggestions for future ML and DL research. Our results can be comparable to a scalogram, another common type of time-frequency representation. Given its high overall accuracy, the suggested study will require more replication before it may be used in other healthcare applications. This work opens the door for the use of DL in COVID-19 diagnosis by proving that it is a rapid, time-efficient, and low-tech solution that does not violate social separation criteria in pandemics like COVID-19.

Acknowledgment

This research received no external funding.

References

[1] WHO Coronavirus, COVID-19, Dec. 30, 2021. [Online]. Available: https://covid19.who.int.

[2] M. Loey, F. Smarandache, and N.E.M. Khalifa, “Within the lack of chest COVID-19 X-ray dataset: a novel detection model based on GAN and deep transfer learning,” Symmetry, vol. 12, no. 4, Apr. 2020.

[3] Mohammad-Rahimi H, Nadimi M, Ghalyanchi-Langeroudi A, Taheri M, and Ghafouri-Fard S, “ Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review,” Frontiers in cardiovascular medicine, Vol. 8, Mar. 2021.

[4] Ozkaya U, Ozturk Ş, Barstugan M,“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,” Springer, p. 281–295, 2020.

[5] Wang X, Deng X, Fu Q, Zhou Q, Feng J, and Ma H,“A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT,” IEEE transactions on medical imaging, vol. 39, no. 8, pp. 2615–2625, 2020.

[6] Majid A, Khan MA, Nam Y, Tariq U, Roy S, and Mostafa RR,“ COVID19 classification using CT images via ensembles of deep learning models,” Computers, Materials and Continua, p. 319–337, 2021, [Online]. Available: https:// doi.org/10.32604/cmc.2021.016816.

[7] Khan MA, Kadry S, Zhang YD, AkramT, SharifM, and Rehman A,“ Prediction of COVID-19-pneumonia based on selected deep features and one class kernel extreme learning machine,” Computers & Electri- cal Engineering, vol. 90, 2021.

[8] AkramT, Attique M, Gul S, Shahzad A, Altaf M, and Naqvi SSR,“ A novel framework for rapid diagnosis of COVID-19 on computed tomography scans,” Pattern analysis and applications, pp. 1–14, 2021.

[9] ZhaoW, JiangW, and Qiu X,“ Deep learning for COVID-19 detection based on CT images,” Scientific Reports, vol.11, no. 1, pp.1-12, 2021.

[10] Khan MA, AlhaisoniM, Tariq U, Hussain N, Majid A, and Damas ˇevičius R,“COVID-19 Case Recognition from Chest CT Images by Deep Learning, Entropy-Controlled Firefly Optimization, and Parallel
Feature Fusion,” Sensors, vol. 21, no. 21, 2021.

[11] Shui-HuaW, Khan MA, Govindaraj V, Fernandes SL, Zhu Z, and Yu-Dong Z,“Deep rank-based average pooling network for COVID-19 recognition,” Computers,Materials, & Continua, p. 2797–2813, 2022.

[12] Zhang YD, Khan MA, Zhu Z,Wang SH,“ Pseudo zernike moment and deep stacked sparse autoencoder for COVID-19 diagnosis,” Cmc-Computers Materials & Continua, p. 3145–3162, 2021.

[13] Kaushik H, Singh D, Tiwari S, Kaur M, Jeong CW, and Nam Y,“ Screening of COVID-19 patients using deep learning and IoT framework,” Cmc-Computers Materials & Continua., pp. 3459–3475, 2021.

[14] V. Bhateja, A. Taquee, and D.K. Sharma,“Pre-processing and classifcation of cough sounds in noisy environment using SVM,” in: 2019 4th International Conference on Information Systems and Computer Networks (ISCON), pp. 822–826, Nov. 2019.

[15] W. Gao, W. Bao, and X. Zhou,“Analysis of cough detection index based on decision tree and support vector machine,” J. Combin. Optim., vol. 37, no. 1, pp. 375–384, Jan. 2019.

[16] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton,“Imagenet classification with deep convolutional neural networks,” Advances in neural information processing systems, vol. 25, 2012.

[17] S. Liu, and W. Deng,“Very deep convolutional neural network based image classification using small training sample size,” in: 3rd IAPR Asian Conference On Pattern Recognition, pp. 730–734, 2015.

[18] C. Szegedy,“ Going deeper with convolutions,” in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1–9, 2015.

[19] K. He, X. Zhang, S. Ren, and J. Sun,“Deep residual learning for image recognition,” in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[20] R.X.A. Pramono, S.A. Imtiaz, E. Rodriguez and Villegas,“A cough-based algorithm for automatic diagnosis of pertussis,” PLoS One, vol. 11, no. 9, Sep. 2016.

[21] N. Rochmawati,“ Covid symptom severity using decision tree,” in: Third International Conference On Vocational Education And Electrical Engineering (ICVEE), pp. 1 5, Oct. 2020.

[22] M. Soli´ nski, M. Łepek, and Ł. Kołtowski,“ Automatic cough detection based on airflow signals for portable spirometry system,”Informatics in Medicine Unlocked, Jan. 2020.

[23] B.W. Schuller, H. Coppock, and A. Gaskell,“Detecting COVID-19 from Breathing and Coughing Sounds Using Deep Neural Networks,” arXiv:2012.14553 [cs, eess], Dec. 2020, Accessed: Jan. 16, 2021, [Online]. Available: http://arxiv.org/abs/2012 .14553.

[24] V. Bansal, G. Pahwa, and N. Kannan,“Cough Classifcation for COVID-19 based on audio mfcc features using Convolutional Neural Networks,” in: IEEE International Conference On Computing, Power And Communication Technologies (GUCON), pp. 604–608Oct. 2020.

[25] Imran,“AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app,” Informatics in Medicine Unlocked ,2020.

[26] J.-M. Liu, M. You, Z. Wang, G.-Z. Li, X. Xu, and Z. Qiu,“Cough event classifcation by pretrained deep neural network,” BMC Med. Inf. Decis. Making, vol. 15, no. 4, Nov. 2015.

[27] M. Aly, and N. Alotaibi, “A novel deep learning model to detect COVID-19 based on wavelet features extracted from Mel-scale spectrogram of patients’ cough and breathing sounds, ” Informatics in Medicine Unlocked, Vol. 32, 2022.

[28] Y. Amrulloh, U. Abeyratne, V. Swarnkar, and R. Triasih, “Cough Sound Analysis for Pneumonia and Asthma Classifcation in Pediatric Population,”in Modelling And Simulation 2015 6th International Conference On Intelligent Systems, pp. 127–131, Feb. 2015.

[29] M. Loey and S. Mirjalili,“COVID-19 cough sound symptoms classifcation from scalogram image representation using deep learning models,” Computers in Biology and Medicine, vol. 139, Nov. 2021.

[30] Sharma N, Krishnan P, Kumar R, Ramoji S, Chetupalli SR, and Ghosh PK,“ Coswara–A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis,” arXiv, 2020.

[31] Indian institute of science,“Project Coswara,” [Online]. Available: https://coswara.iisc.ac.in/team.

[32] Paul S Addison,“Wavelet transforms and the ECG: a review,” Physiological Measurement, Nov. 2005.

[33] H. Yasar, and M. Ceylan,“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., Oct. 2020.

[34] I.D. Apostolopoulos, S.I. Aznaouridis, and M.A. Tzani,“Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases,”J. Med. Biol. Eng., vol. 40, no. 3, pp. 462–469, Jun. 2020.

[35] B. Zoph, V. Vasudevan, J. Shlens, and Q.V. Le,“Learning Transferable Architectures for Scalable Image Recognition, ” arXiv:1707.07012 [cs, stat], Apr. 2018, Accessed: Jan. 17, 2021, [Online]. Available: http://arxiv.org/abs/1707.07012.

[36] C. Szegedy, S. Ioffe, V. Vanhoucke, and A.A. Alemi,“Inception-v4, inception-ResNet and the impact of residual connections on learning, ” in: Proceedings Of the Thirty-First AAAI Conference On Artifcial Intelligence, San Francisco, California, USA, , pp. 4278 -4284, Feb. 2017, pp. 4278 -4284.

[37] V.K. Pothos, D. Kastaniotis, I. Theodorakopoulos and N. Fragoulis,“ A fast, embedded implementation of a Convolutional Neural Network for Image Recognition,” Technical Report, Aug. 2016, [Online]. Available: https://www.researchgate.net/publication/306003694_A_fast_embedded_implementation_of_a_Convolutional_Neural_Network_for_Image_Recognition.

[38] Nur Ateqah, Nur Hidayah, Zaidah Ibrahim, and Nur Nabilah,“ Celebrity Face Recognition using Deep Learning,” Indonesian Journal of Electrical Engineering and Computer Science, vol. 12, no. 2, pp. 476-481, Nov. 2018.

[39] Y. Xu, and R. Goodacre,“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, vol. 2, no. 3, pp. 249-262, Jul. 2018.

[40] Sutskever, J. Martens, G. Dahl, and G. Hinton,“ On the importance of initialization and momentum in deep learning,” in: Proceedings of the 30th International Conference on International Conference on Machine Learning, vol. 28, 2013.

[41] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov,“Dropout: a simple way to prevent neural networks from overftting,” J. Mach. Learn. Res., vol. 15, no. 56, pp. 129-1958, 2014.

[42] K. Khoria, et.al, “On significance of constant-Q transform for pop noise detection,” Computer Speech & Language, Vol. 77, 2023.