COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models

Mohamed Loey [a,b,*], Seyedali Mirjalili [c,d]

[a] *Department of Computer Science, Faculty of Computers and Artificial Intelligence, Benha University, Benha, 13518, Egypt*
[b] *Information Technology Program, New Cairo Technological University, New Cairo, Egypt*
[c] *Center for Artificial Intelligence Research and Optimization, Torrens University Australia, Fortitude Valley, Brisbane, QLD, 4006, Australia*
[d] *Yonsei Frontier Lab, Yonsei University, Seoul, South Korea*

ARTICLE INFO

Keywords:
COVID-19
Cough sound
Deep learning
Scalogram images
Sound classification

ABSTRACT

Deep Learning shows promising performance in diverse fields and has become an emerging technology in Artificial Intelligence. Recent visual recognition is based on the ranking of photographs and the finding of artefacts in those images. The aim of this research is to classify the different cough sounds of COVID-19 artefacts in the signals of altered real-life environments. The introduced model takes into consideration two major steps. The first step is the transformation phase from sound to image that is optimized by the scalogram technique. The second step involves feature extraction and classification based on six deep transfer models (GoogleNet, ResNet18, ResNet50, ResNet101, MobileNetv2, and NasNetmobile). The dataset used contains 1457 (755 of COVID-19 and 702 of healthy) wave cough sounds. Although our recognition model performs the best, its accuracy only reaches 94.9% based on SGDM optimizer. The accuracy is promising enough for a wide set of labeled cough data to test the potential for generalization. The outcomes show that ResNet18 is the most stable model to classify the cough sounds from a limited dataset with a sensitivity of 94.44% and a specificity of 95.37%. Finally, a comparison of the research with a similar analysis is made. It is observed that the proposed model is more reliable and accurate than any current models. Cough research precision is promising enough to test the ability for extrapolation and generalization.

1. Introduction

The coronavirus disease (COVID-19) continues to spread globally with over 230 million confirmed cases and 4.7 million deaths worldwide as of October 2021 [1]. The coronavirus classes are subgroups of the (α, β, δ, γ) coronavirus [2]. The COVID-19 outbreak was caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) in 2019 [3]. The origin of the β community of coronaviruses has been confirmed to be SARS-CoV-2 [2]. By 2020, the exponential dissemination of the disease pushed the World Health Organization to declare COVID-19 as a pandemic. SARS-CoV-2 can spread in a number of forms, especially in polluted and overcrowded environments [4]. Governments and health-care institutions have introduced new policies to cope with over-crowding by implementing infection control systems [5].

The COVID-19 outbreak has triggered a drastic increase in international scientific collaboration. Deep Learning (DL) and Machine Learning have constructive impacts in the battle against Coronavirus [6,

7]. DL involves using a dataset to help detect and avoid deadly illnesses. The healthcare industry requires support from emerging technologies, such as Artificial Intelligence (AI) [8], Internet of Things (IoT) [9], and big data [10], to deter the emergence of new Coronavirus diseases [4]. DL is best used to diagnose the virus and to correctly predict its propagation. The primary signs of COVID-19 include cough and exhaustion. Coughing is the consequence of several other illnesses and the effects of each infection differ; hence, diseases of the lung may impair the acoustics of cough [11].

The main objective of this research is to identify the observable characteristics of the COVID-19 cough, as demonstrated in Fig. 1. The proposed model, when provided with a coughing tone, would identify the sound as either sick or stable. To control the COVID-19 pandemic, wide-scale research is an imperative in areas such as X-ray, CT scan image of chest [2,12], and medical face mask detection [5]. Developing nations are experiencing a shortage of health care professionals and devices such as personal protective equipment. This could cause

significant problems in the emerging countries when compared to the industrialized countries, especially if the disease continues to spread at a rate similar to that of the West. It is important to develop effective means of early detection and diagnosis to reduce the death rates. Cough may play an important role in initial diagnostic tests [13–16].

All the sound classification methods employ machine learning and DL. Machine learning classifiers include support vector machine [17] and decision tree [18], while DL classifiers include the Convolutional Neural Network (CNN) models (AlexNet [19], VGGNet [20], GoogleNet [21], ResNet [22]). The CNN family is designed for high speed and performance in image classification. The main contributions of this research are as follows:

1) A novel DL algorithm that can identify COVID-19 based on a subset of tone.
2) The proposed model increases sound detection efficiency by introducing scalogram technique to convert sound to image.
3) Six models for DL training are implemented to get optimal efficiency.

The present study is organized as follows. Section 2 of this paper reviews the extant literature. Section 3 explains the main features of the dataset. Section 4 displays the proposed COVID-19 cough sound model in an instructive manner. Section 5 reports the findings of the tests, and Section 6 presents the conclusions and potentials for future research.

2. Related Works

This section presents a survey of the latest literature on the diagnosis of COVID-19 using cough sounds. Subsequently, the latest assessment of DL toward cough sound scan analysis is addressed. This section compiles existing knowledge on the use of machine learning and DL in sound classification. The sound classification levels may be subdivided into a pre-processing stage, an extraction stage, and a classification stage. Most of the sound detection studies focus on sound construction and sound recognition based on traditional machine learning techniques [23–25]. The present research focuses on the classification and recognition of cough sounds produced by those with the COVID-19 virus. Schuller et al. [26] implemented a DL method to classify raw breathing and coughing of COVID-19 patients based on CNN. They adjusted the CNN method that utilizes breathing and coughing audio to recognize if a patient is infected with COVID-19 or if they are healthy. The proposed approach is almost greater than the conventional baseline. While the CNN model achieved an accuracy of 80.7%, with the current data available, a DL model is able to achieve the best performance.

In [27], Bansal et al. proposed a CNN model for COVID-19 audio recognition based on mel-frequency cepstral coefficients (MFCC). Two

learning dependent technique using the VGG 16 architecture. The outcome of the introduced model, with a high-quality results approach, produced 70.58% test accuracy with 81% sensitivity. In Ref. [28], the authors proposed a model to distinguish the COVID-19 sounds from the several types of non-COVID-19 sounds. They used 1838 cough and 3597 non-cough sounds that were categorized into 50 classes for training and testing (70 COVID-19 and 247 healthy sounds). The study showed that the overall accuracy of DL-based multiclass classifier was 92.64%.

Other studies before the COVID-19 pandemic, such as [29], introduced a transfer learning model to classify cough sound events. Neural network models are constructed from two stages, pretraining and fine-tuning, and then the decoded details are collected by a Hidden Markov Model (HMM). In this research, three cough HMMs and one non-cough HMM are added to the proposed model. The tests were performed on a dataset obtained from 22 patients who suffered from multiple respiratory disorders. Their proposed method shows that the qualified deep model can reach a high degree of precision of 90%. Hee et al. [30] proposed a machine learning classifier for asthmatic and healthy children. The dataset included 1192 cough samples from asthmatic children and 1240 cough samples from healthy children. Features such as MFCC were derived from the audio. The learned machine learning algorithm was generated with a Gaussian Mixture Model–Universal Context Model. The study showed that the overall sensitivity and specificity of machine learning classifier were 82.81% and 84.76%, respectively. In Ref. [31], Amrulloh et al. introduced a classification model for pneumonia and asthma. Their approach quantified the sound features by MFCC, Shannon entropy, and non-Gaussian, and these characteristics were found to form the basis for artificial neural network classifiers. The proposed approach reached 89% sensitivity and 100% precision. The findings demonstrate how our approach could be used to discriminate between pneumonia and asthma in open areas.

Most of the above research studies used mathematical analyses and machine learning to accurately identify COVID-19 infection. Fewer studies were found to utilize transfer learning and CNN of cough sound datasets for the variables of coronavirus patients and stable patients. Therefore, further studies are needed on DL with streamlined efficiency metrics. As per the literature review presented here, it is advised to use cough sounds for the diagnosis of COVID-19. The new paradigms tend to be quicker and more successful in combatting the COVID-19 pandemic.

3. Dataset characteristics

A wide range of difficult data collection methods is expected in COVID-19 patients. A database of respiratory sounds documented during acute COVID-19 infection is presented in Ref. [32]. The aim of the Coswara project is to establish a diagnostic method for COVID-19
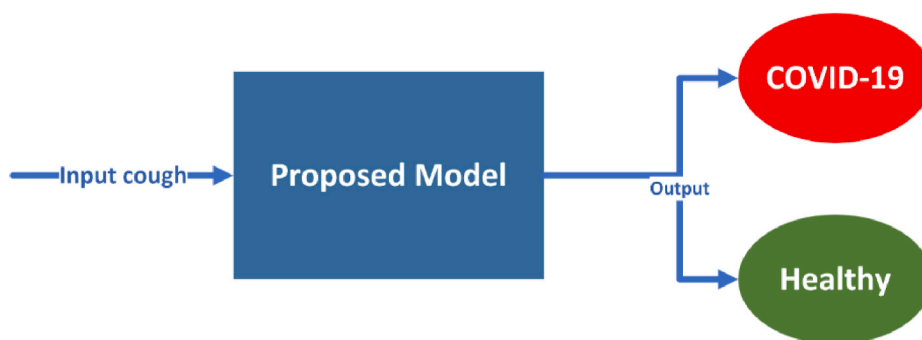

Fig. 1. The proposed COVID-19 cough sound classification model.

methods were discussed to identify audio sounds in this study. The first process uses the spectrogram as an input to the MFCC algorithm. Second, the collection of image processing pipeline is assessed via the transfer

pertaining to respiratory, cough, and speech sounds [33]. The dataset contained accessible deep_cough recordings from 92 COVID-19 positive patients and 1079 stable tones. The Sarcos[1] (SARS COVID-19 South Africa) dataset is extremely limited with only 8 COVID-19 cases and 13 healthy cases. Both Coswara and Sarcos are imbalanced because positive subjects are overshadowed by nonpositive subjects.

This study conducted its experiments based on CoughDataset. The CoughDataset dataset is based on Coughvid.[2] The dataset is organized into five classes (COVID-19, Healthy, Lower, Upper, and Obstructive) and includes 3325 sound files with 16 KHz, mono, 1 s duration. The proposed DL model was trained and tested on COVID-19 (755 sound files) and healthy (702 sound files) classes, as shown in Fig. 2.

In order to render the proposed model in a public dataset, the model was built to provide cough classification. This cough classifier is used by the diagnostic engine to assess whether or not a sound is related to COVID-19. To evaluate the classifier, we used the CoughDataset data and the COVID-19 and non-COVID-19 sounds from the dataset. CoughDataset is a rich archive with data from an incredibly broad range of COVID-19 patients. This set comprises 3325 sounds divided into five main groups (755 of COVID, 702 of healthy, 1032 of lower, 186 of obstructive, and 650 of upper). The current method used 1457 cough sounds for preparation and research.

4. The proposed model

The architecture diagram of the proposed DL cough classification model is shown in Fig. 3. The introduced model includes two main components: The first component is the feature extraction, and it transforms sound to image based on scalogram, while the second component is the feature extraction and classification model based on the DL models (GoogleNet, ResNet18, ResNet50, ResNet101, MobileNetv2, and NasNetmobile). GoogleNet, ResNet, MobileNet, and NasNet are amongst the most widely used DL transfer learning models [34–36]. The proposed model used DL models for feature extraction and classification in the training, validation, and testing stages.

4.1. Scalogram

A scalogram is the real value of the Continuous Wavelet Transform (CWT) coefficients of a wave [37]. This study adopts the scalogram approach in two measures. First, the 1-D electrocardiogram (ECG) signals are preprocessed for noise reduction. Second, 2-D scalograms utilizing CWT are used with the preprocessed signals. The ECG transforms the signal from the time to the frequency domain based on CWT, as shown in Fig. 4. Low frequency and high frequency noise are removed by convolution using an average filter. The CWT utilizes internal products to calculate the resemblance of a wave and an examination function like the Fourier transform. The CWT of a function $f(s)$ at a scale $(x > 0)$ is calculated using equation (1). $\vartheta(s)$ is a continuous function in both the time domain and the frequency domain called the father signal. x is the continuously varying values of the scale parameter, and y is the position parameter. The outcome of the CWT coefficients is a matrix filled with wavelet located by scale and position. The goal of the father signal is to provide the generation root feature of the children signals. ECG is calculated by scale parameter and father signal in CWT [38,39].

$$CWT(x, y) = \frac{1}{|x|^{0.5}} \int_{-\infty}^{\infty} f(s)\, \vartheta\left(\frac{s-y}{x}\right) ds \tag{1}$$

4.2. DL models

There are several efficient pre-train CNN that have the potential of passing learning. However, they require training and analysis of the dataset at their input layer. Several combinations and methods are applied to build the networks. In 2014, new transfer learning CNN model was proposed by C. Szegedy et al. [21] at Google. GoogleNet object classification deep network is composed of feature extraction network and classification network, as shown in Fig. 5. It contains 22 convolutional layers [34,35]. GoogLeNet has inception layers, each of which conducts a particular method of convolution and then concatenates the filters together for the next layer [40]. Nine inception modules are stacked vertically in sequence. At the global average pooling layer, the ends of the inception modules are linked to the global average. Many visual functions have also significantly gained from the deep models. A tendency to sharpen increasingly difficult issues and improve scoring precision has been observed over the years. The preparation of neural network becomes complicated and often degrades as the precision begins to saturate.

Residual Network is also referred to as ResNet [22]. Residual Learning is structured to resolve many of the problems associated with DL. Deep ResNet smartly attempts to master several low-, mid-, and high-level features. The individual network is trained to retrieve minor fragments of knowledge. The idea of 'residual' may be understood as throwing away the functionality acquired throughout the previous layer. ResNet was inspired by VGGNet architecture. ResNet has several models, such as ResNet18/50/101. ResNet18 has 18 convolution layers with 3×3 filter, as shown in Fig. 6. Furthermore, ResNet50 and ResNet101 have 50 layers and 101 layers, respectively, and each block has 3 convolution layers with 3 filters (1×1, 3×3, and 1×1). Each 2-layer block is replaced with this 3-layer bottleneck block, as illustrated in Fig. 7. Data Augmentation is a methodology that can be extended to diversity databases to enhance the preparation of recognition. During the preparation, the photos were not changed at all. Data convergence enhances the efficiency of the classification [34,41,42].

MobileNetV2 and NasNetMobile are DL models for mobile devices. The design of MobileNetv2 comprises a total of 155 layers and 164 links [43,44]. MobileNetv2 is inspired by mobile architecture based on depth-wise separable convolutions. The Global Average Pooling [45] layer was applied, which greatly decreases forward error estimation failure. The training weights were standardized using a batch normalization layer [46]. NasNet is a modular CNN that consists of simple building blocks that have been optimized using reinforcement learning [47]. A cell is made up of just a few operations, and these are replicated time and again due to the required size of the network. Mobile edition (NasNetMobile) comprises of 12 partitions of 5.3 million values and 564 million multiply-accumulates (MACs).

A time-domain signal is converted into a frequency-domain signal using a scalogram, and the signal is analyzed on multiresolution. Nevertheless, the mechanism retains the morphological difficulty of signal processing. This indicates that current basic classifiers-based machine learning can be poor at predicting complex signals. We performed a picture to DL of CNN, which exhibits optimum efficiency for the recognition of visual morphology. No attempt has been made to equate the output of DL models with the feedback of the 2-D matrix using wavelet transform. Thus, the present study concentrated on designing the most representative DL models (GoogleNet, ResNet18, ResNet50, ResNet101, MobileNetv2, and NasNetmobile) that are most commonly used for image classification. The use of scalogram to depict signal characteristics and its capacity to distinguish biometrically are the novelties of this study. As described in Section 5, the proposed model was tested with input signal in the form of picture, as is appreciated for DL models.
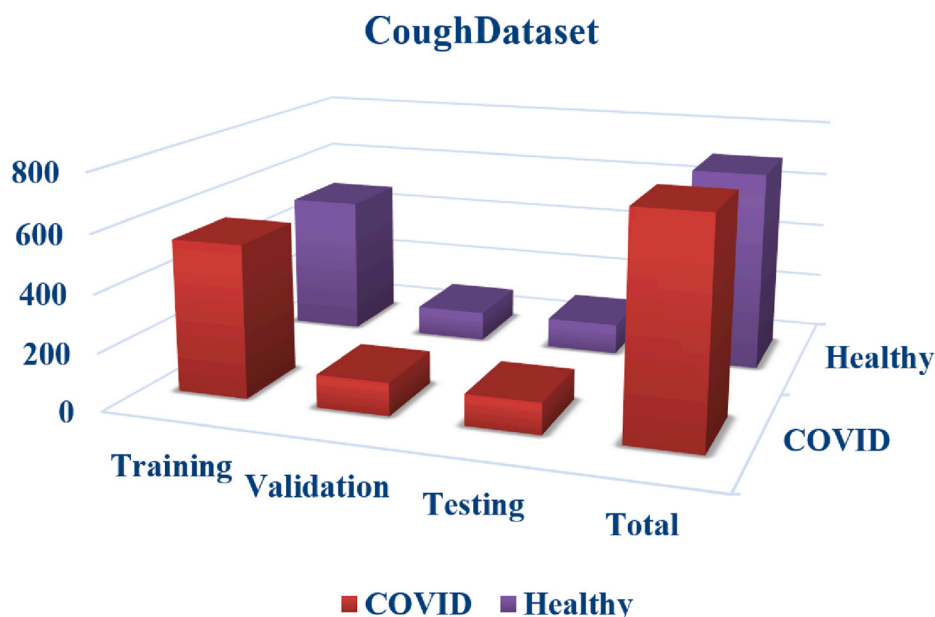
---

[1] https://coughtest.online.
[2] https://coughvid.epfl.ch.

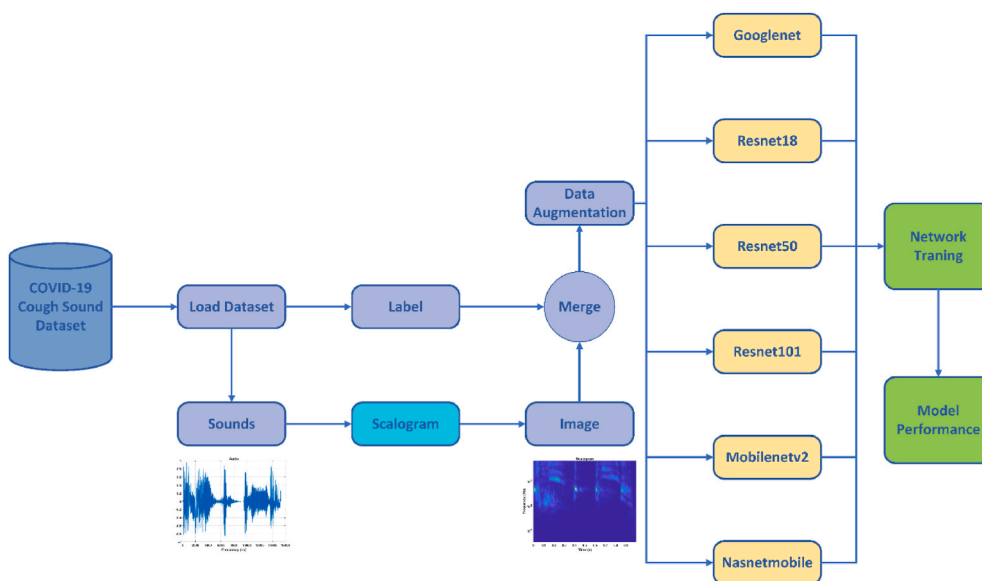**Fig. 2.** The dataset statistics used in the proposed model.

Fig. 3. The proposed cough sound classification model.

5. Experimental results

The introduced DL model is performed in the transfer mode with the suggested initial training setup (batch norm decay = 0.5, batch norm epsilon = e^{-3}, dropout = 0.5, weight decay = e^{-3}). The learning rate of 0.01, with batch size = 8, was automatically reduced until it reached e^{-5}. This decreased the preparation period without lowering the efficiency levels. The DL models were trained for 20 h on a single NVidia 2070 RTX with the CUDA and Deep Neural Network library (CuDNN) in Tensorflow and MATLAB.

The dataset is split into 70% training images, 15% validation images, and 15% testing images. Both labeled and evaluation data were used in our experiment. The validation accuracy is a classification score for checking the learning method during the process. It enables the identification of overfitting as a potential trigger. If the accuracy of assessment and training is different, it signifies that overfitting has already

happened. The test's consistency depends on how the data is learned from training. The split ratio depends on the volume of the dataset. To ensure the greatest degree of model efficiency, an effective balance must be struck between training and testing. Furthermore, a straightforward response is not possible concerning the process or parameter takes one over the top. It was observed that the model output increased as more samples were used [48]. Stochastic Gradient Descent with momentum (SGDM) [49] was chosen as the optimizer technique in the current study to improve detector performance. The setup of the DL models is shown in Table 1. Table 1 displays the findings from each DL transfer models with initial learning rate at 0.01 and the number of epochs at 20. The batch size was set to 8 and early-stopping was allowed if an accuracy change was not seen. The best optimizer strategy was found to be SGDM, which aims to change the weight parameters.

To prevent over-fitting issues with the DL net, we employed the dropout approach [50]. Loss function $L(x,y)$ was used as the teaching
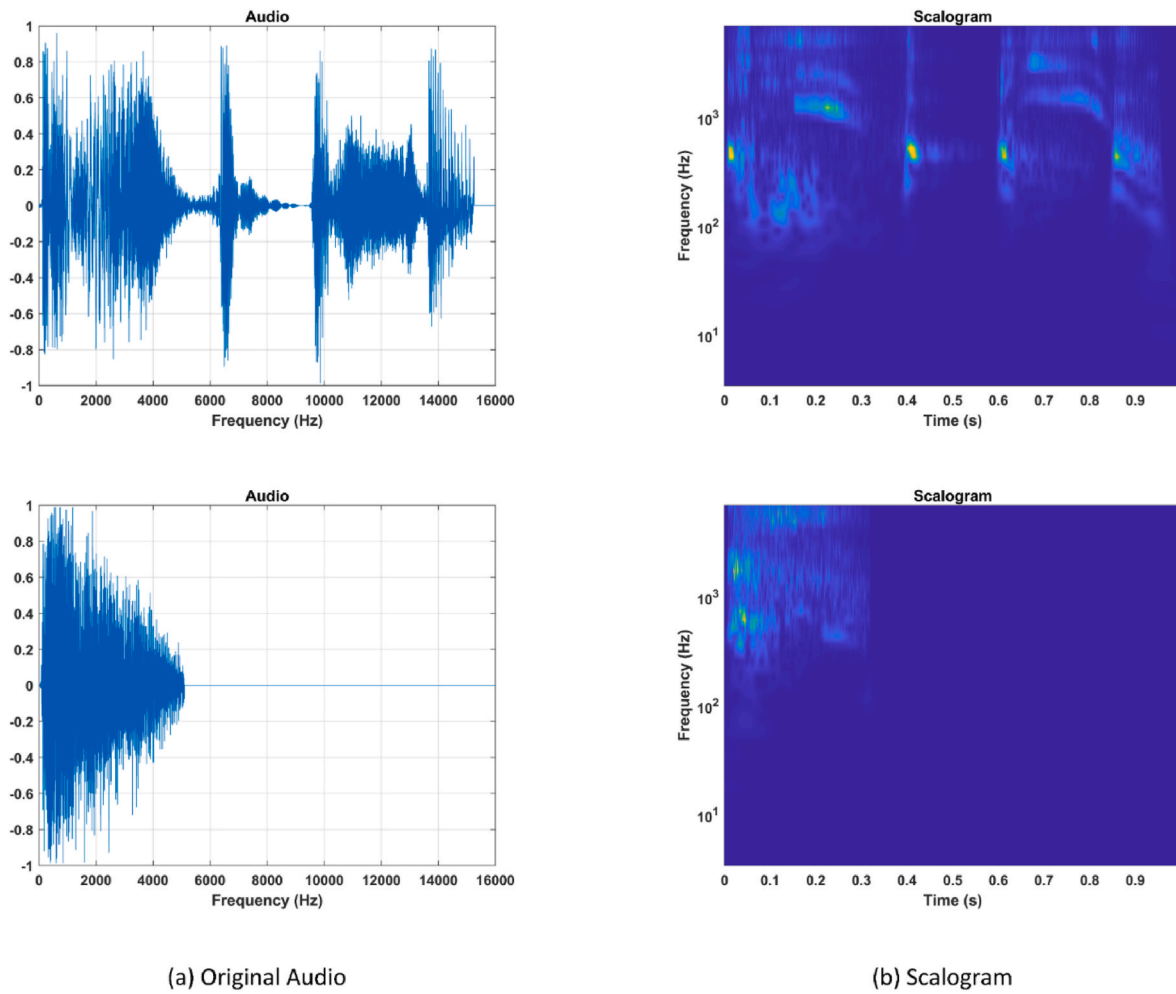
(a) Original Audio (b) Scalogram

Fig. 4. Scalogram of an electrocardiogram of COVID-19 cough sounds.
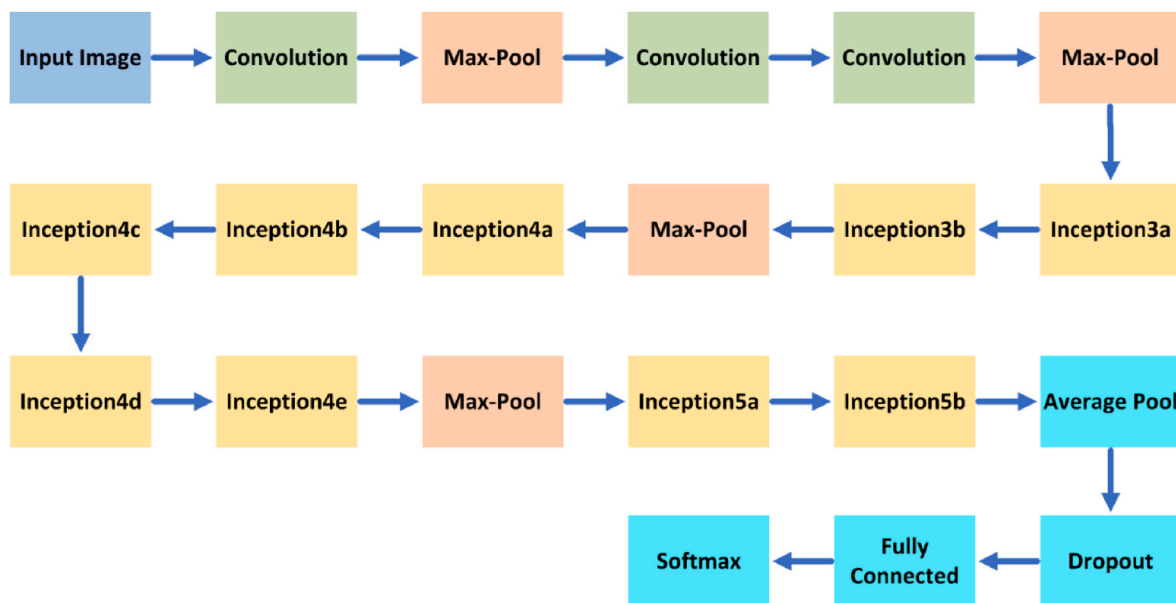
Fig. 5. GoogleNet architecture.

Input Image

↓

Convolution 7x7 , 64/2

Max Pooling

↓

Convolution 3x3 , 64
Convolution 3x3 , 64

↓

Convolution 3x3 , 64
Convolution 3x3 , 64

↓

Convolution 3x3 , 128
Convolution 3x3 , 128

↓

Convolution 3x3 , 128
Convolution 3x3 , 128

↓

Convolution 3x3 , 256
Convolution 3x3 , 256

↓

Convolution 3x3 , 256
Convolution 3x3 , 256

↓

Convolution 3x3 , 512
Convolution 3x3 , 512

↓

Convolution 3x3 , 512
Convolution 3x3 , 512

Average Pool

↓

Fully Connected

Fig. 6. ResNet18 architecture.

Input Image

↓

Convolution 7x7 , 64/2

Max Pooling

↓

Convolution 1x1 , 64
Convolution 3x3 , 64
Convolution 1x1 , 256

↓

Convolution 1x1 , 128
Convolution 3x3 , 128
Convolution 1x1 , 512

↓

Convolution 1x1 , 256
Convolution 3x3 , 256
Convolution 1x1 , 1024

↓

Convolution 1x1 , 512
Convolution 3x3 , 512
Convolution 1x1 , 2048

Average Pool

↓

Fully Connected

| Resnet50 | 3 | 4 | 6 | 3 |
| Resnet101 | 3 | 4 | 23 | 3 |

Fig. 7. Resnet50 and Resnet101 architecture. The table below shows the number of each block.

Table 1
DL models setup.

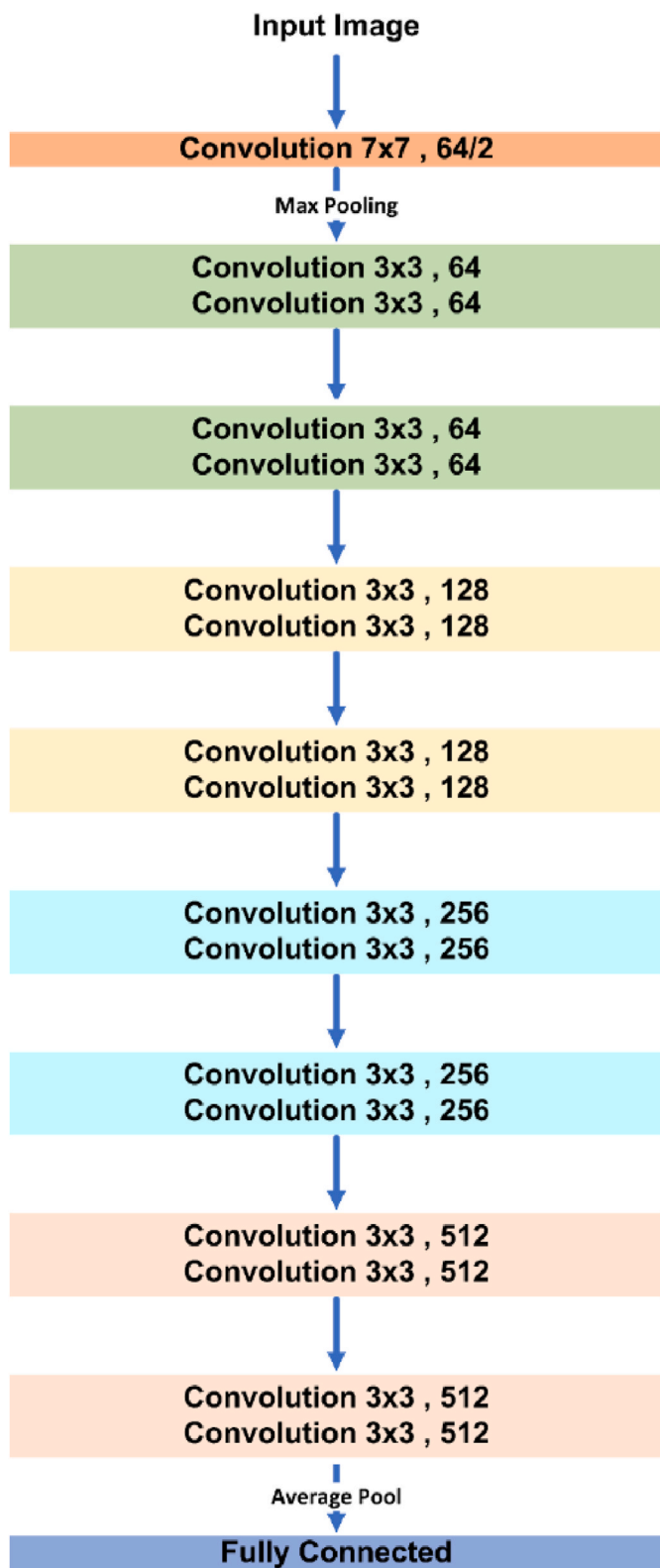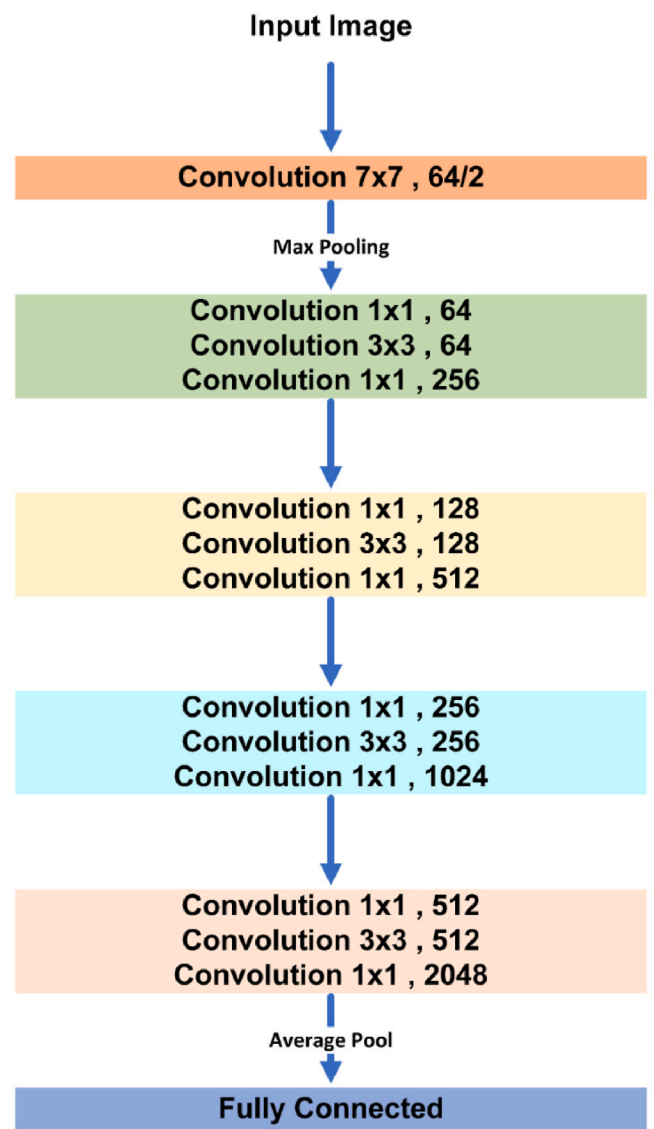| Model	Layers	Batch Size	Epoch	Learning Rate	Optimizer
GoogleNet	22	8	20	0.01	SGDM
ResNet18	18				
ResNet50	50				
ResNet101	101				
MobileNetv2	53				
NasNetMobile	cells				

criteria, which is described as the total of binary and box loss functions, as illustrated in equation (2)

$$L(x,t) = L_{cl}(x_c) + \delta[b > 0]L_{re}(k, k^*) \tag{2}$$

where the bounding boxes of k and k^* are denoted by $[k_a, k_b, k_w, k_h]$, w is the width and h is the height of the box, and x_c represents the expected

score class c. $\delta[b > 0]$ represents the boxes of non-background at 0. The regression loss involves the bounding box *and*the classification loss L_{cl} represented in equation (3):

$$L_{cl}(x_{c^*}) = -\log(x_{c^*}) \tag{3}$$

The regression loss L_{re} is calculated as given in equations (4) and (5):

$$L_{re}(m, m^*) = \sum_{j \in (a,b,w,h)} V_{L1}\left(m_j - m_j^*\right) \tag{4}$$

where,

$$V_{L1}(q) = \begin{cases} \frac{1}{2}*q^2, & if \ |q| < 0 \\ |q| - \frac{1}{2}, & otherwise \end{cases} \tag{5}$$

5.1. Performance evaluation

Testing can show a good outcome that proves the reliability of the DL models. The confusion matrix is a technique to calculate the research statistical performance. The six statistical measurements include accuracy, sensitivity, specificity, precision, F1 score, and the Matthews Correlation Coefficient (MCC). The confusion matrices for the two groups (COVID-19 and Healthy) are shown in Fig. 8, and Fig. 9. For a measurement close to the truth, accuracy = $(N_{TP}+N_{TN})/((N_{TP}+N_{FP})+(N_{TN}+N_{FN}))$ where N_{TP} is the number of correctly labeled, N_{FN} is the number of mislabeled, N_{TN} is the number of correctly labeled instances of the rest of the classes, and N_{FP} is the number of mislabeled instances of the rest of the classes. For the GoogleNet model, the confusion matrix of the test is shown in Fig. 8(a) and the overall accuracy is 90.7%. The performance of the three ResNet models (ResNet18/50/101) are illustrated in Fig. 8(b,c,d) and the overall accuracy is 94.9%, 90.7%, and 91.7% respectively. Resnet18 achieved the highest accuracy due to the
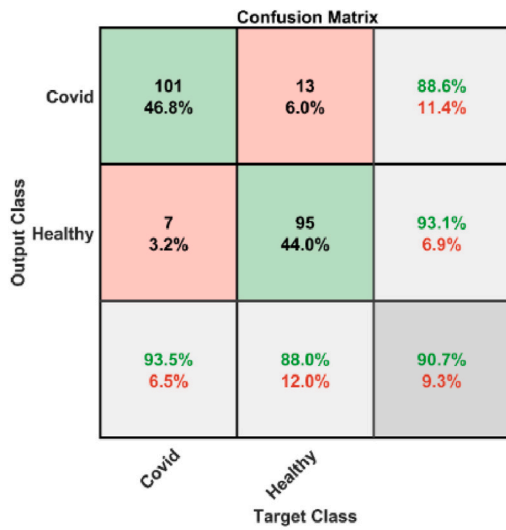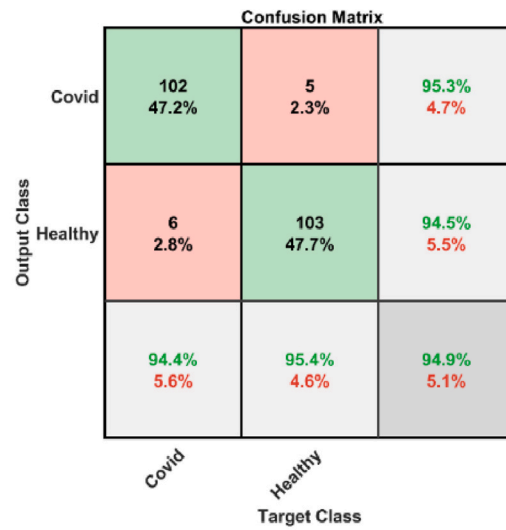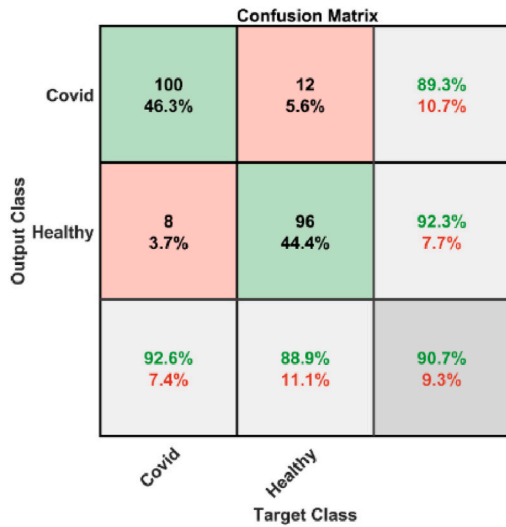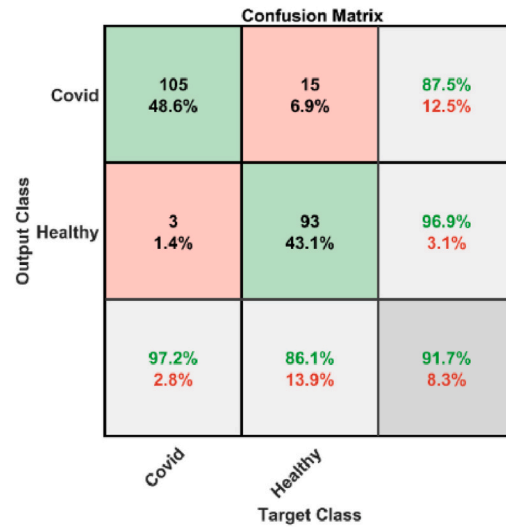

Fig. 8. Confusion matrix of GoogleNet, ResNet18, ResNet50, and ResNet101.
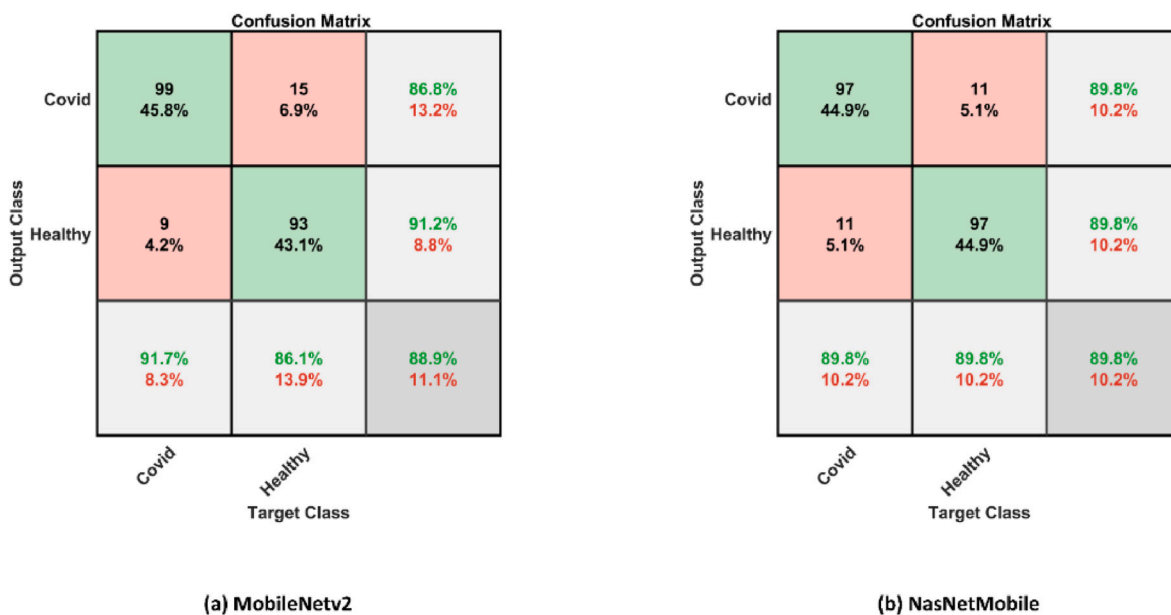
Fig. 9. Confusion matrix of MobileNetv2 and NasNetMobile.
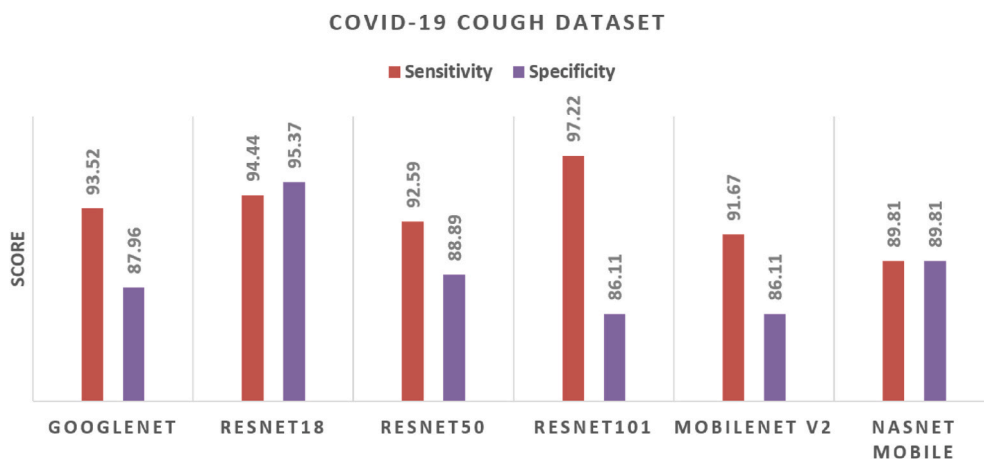

Fig. 10. Sensitivity and specificity for all DL models.
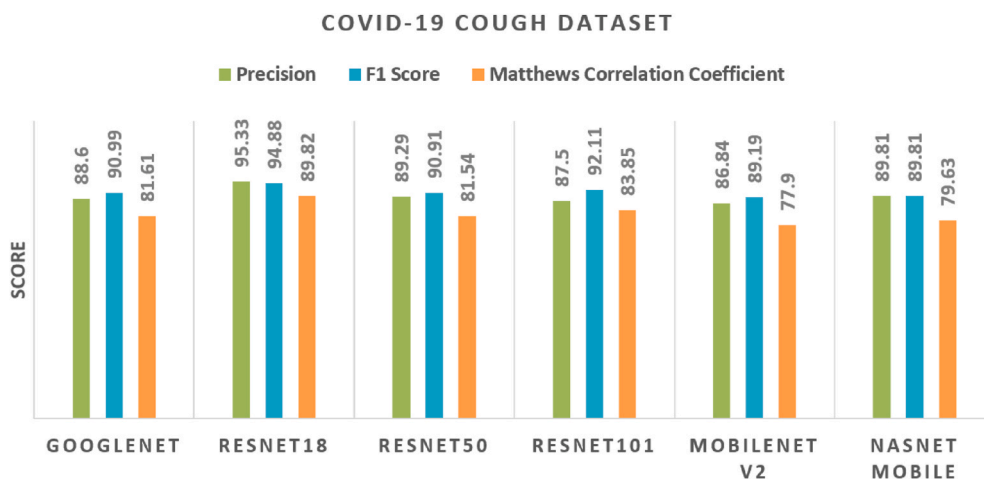

Fig. 11. Precision, F1 score, and the Matthews Correlation Coefficient for all the DL models.

small dataset. In Fig. 9, the DL mobile models show that the overall accuracy is 88.9% and 89.9% for MobileNetv2 and NasNetMobile respectively.

The accuracy of the DL models' predictions was measured quantitatively. The two widely employed classification efficiency indicators are sensitivity and precision. For the measurements, $Sensitivity = N_{TP} / ((N_{TP} + N_{FN}))/((N_{TP}+N_{FN}))$ and $Specificity = N_{TP}/((N_{FP}+N_{TN}))$. Fig. 10 introduces the sensitivity and specificity for the six DL models. The highest sensitivity of 97.22% is achieved by ResNet101, which refers to the test's ability to recognize the cough sounds of patients with COVID-19. A test with the high specificity of 95.37% for ResNet18 determines that the test is able to recognize patients who do not have COVID-19.

For measurements, the $Precision = N_{TP} /(N_{TP} + N_{FP})$, $F1\ score = 2* N_{TP} / (2*N_{TP} + N_{FP} + N_{FN}))$ and $MCC = N_{TP}*N_{TN} - N_{FP}* N_{FN} / \sqrt{(N_{TP} + N_{FP})*(N_{TP} + N_{FN})*(N_{TN} + N_{FP})*(N_{TN} + N_{FN})}$. Fig. 11 introduces the precision and F1 score and MCC for the six DL models. The highest precision of 95.33% is achieved by ResNet18, which suggests that this model returns more pertinent outcomes than the others. A test with a high F1 score of 94.88% for ResNet18 determines the efficiency of the DL model. Finally, the MCC indicates that the more accurate statistical rate provided a strong score in each of the four uncertainty matrix groups. The best MCC is for ResNet18 of 89.82%. Further data collection is required to determine the ability of the deep transfer models to be much more reliable. Despite its positive precision rates, the proposed research needs to be repeated on a wider scale, since it may also be applied for other medical applications.

Table 2
Performance comparison of different methods in term of Accuracy.

Reference	Methodology	Dataset	Result
[26]	CNN	1427	80.7%
[27]	CNN	871	70.5%
[28]	CNN	317	92.6%
Proposed	**DL transfer model**	**1457**	**94.9%**

5.2. Comparative analyses and discussions

The outcomes of the introduced model concerning the use of DL models in CoughDataset with scalogram images of COVID-19 patients and healthy individuals is shown in Fig. 12. Fig. 12 illustrates how our proposed model can classify data with high accuracy. Furthermore, the use of scalogram with DL models to depict signal characteristics and its capacity to distinguish biometrically are the novelties of the present study. Most of the related studies focus on the classification of cough sounds based on machine learning. Performance comparison of different methods in term of Accuracy (AC) is illustrate in Table 2. A related study presented in Refs. [26,27] used a small dataset, which included the real COVID-19 cough sound dataset. Much of the above study is based on the classification of cough and non-cough tones. By analyzing the performance of DL transfer models in handling COVID-19 cough sounds with the SGDM, we find that the performance measurement of all the DL models increases sharply in the case of cough signals with strong frequency. Although our recognition model performs the best, its accuracy only reaches 94.9% based on SGDM optimizer, the accuracy of the training data, and the attempt to check the labelling of the data. However, any flaw in the marking of the data that managed to slip past our scrutiny is likely to affect the recorded results. This effect is more
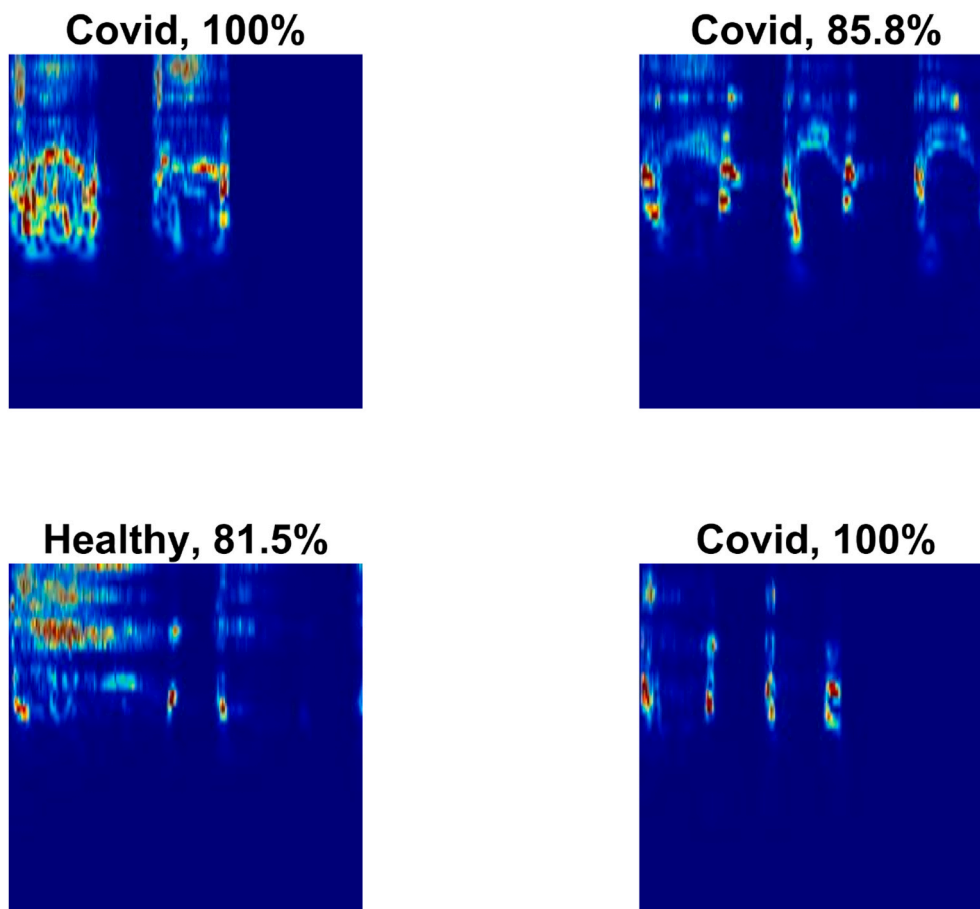

Fig. 12. Representative classifier outcomes of DL models.

pronounced where the data amounts to be comparatively small.

The observations of the present study and those from the other separate studies listed in the Related Works section, indicate that distinct latent features of cough sounds can be used for effective DL diagnosis of various respiratory disorders. The cough sound may be used as a preliminary diagnostic instrument since it differentiates healthy coughs from the COVID-19 coughs. We investigate the possibility of using the scalogram of sound as feedback to DL models to determine which model performs the best at categorizing medical images to sound. ResNet and GoogleNet are found to have high accuracy in this research and are called deep versions of DL transfer models. NasNetMobile and MobileNetv2 have high accuracy for mobile versions. The tests are conducted on one dataset for measurement which contains audio wave files. ResNet18 was 4.17%, 4.17%, and 3.24% more accurate than GoogleNet, ResNet50, and ResNet101 respectively. The NasNetMobile had 0.92% higher accuracy than MobileNetv2. The studies use the scenario in order to determine the efficiency and consistency of the current classification model. The findings reveal that the maximum classification accuracy is obtained by the ResNet18 model on the cough sound from the checked COVID-19 dataset. The experimental findings for the DL classifier demonstrate how it fits the cough sounds of COVID-19 patients better than other CNN classifiers. Therefore, it can prove more successful in diagnosis by sparing the doctors the intense workload involved with the initial sounding of the COVID-19 cough.

6. Conclusion and future works

The present study has introduced novel DL models for cough sound classification by focusing on tone, which would help curb the spread of COVID-19. The proposed model integrates two main components. The first one involved transforming sound waves to a picture, which was implemented using scalogram. The second component is the generation of universal features accompanied by additional classification using deep transfer models (GoogleNet, ResNet18, ResNet50, ResNet101, MobileNetv2, and NasNetmobile). The CoughDataset used includes 1457 wave cough sounds of 755 patients and 702 healthy individuals. While our recognition model is the most reliable, its accuracy is still at 94.9%. Cough research precision is promising enough to test this model for extrapolation and generalization. The results demonstrated that ResNet18 is the model most effective at classifying the cough sound relative to the other models evaluated on a smaller dataset. This study has been compared with others studies on COVID-19 cough sounds, and it was found to be more predictive and important than all of the established classifiers.

The outcomes of the current study signal important contributions for future studies in machine learning and DL. Our results can be compared with another common type of time frequency representation named spectrogram. Despite its positive precision rates, the proposed research needs to be repeated on a wider scale in order to be employed in other medical applications.

Funding

Declaration of competing interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] WHO Coronavirus, COVID-19) Dashboard. https://covid19.who.int. Oct. 01, 2021.
[2] M. Loey, F. Smarandache, 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, Art. no. 4, Apr, Symmetry 12 (4) (2020), https://doi.org/10.3390/sym12040651.

[3] N.E.M. Khalifa, F. Smarandache, G. Manogaran, M. Loey, A Study of the Neutrosophic Set Significance on Deep Transfer Learning Models: an Experimental Case on a Limited COVID-19 Chest X-Ray Dataset, Jan, '' Cogn Comput, 2021, https://doi.org/10.1007/s12559-020-09802-9.
[4] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Jan, Measurement 167 (2021) 108288, https://doi.org/10.1016/j.measurement.2020.108288.
[5] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, Fighting against COVID-19: a novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection, Feb, Sustainable Cities and Society 65 (2021) 102600, https://doi.org/10.1016/j.scs.2020.102600.
[6] S. Bhattacharya, et al., Deep learning and medical image processing for coronavirus (COVID-19) pandemic: a survey, Sustainable Cities and Society 65 (Feb. 2021) 102589, https://doi.org/10.1016/j.scs.2020.102589.
[7] A. Sufian, A. Ghosh, A.S. Sadiq, F. Smarandache, A survey on deep transfer learning to edge computing for mitigating the COVID-19 pandemic, J. Syst. Architect. 108 (Sep. 2020) 101830, https://doi.org/10.1016/j.sysarc.2020.101830.
[8] Y. Ma, Z. Wang, H. Yang, L. Yang, Artificial intelligence applications in the development of autonomous vehicles: a survey, IEEE/CAA Journal of Automatica Sinica 7 (2) (Mar. 2020) 315–329, https://doi.org/10.1109/JAS.2020.1003021.
[9] M. Kavre, A. Gadekar, Y. Gadhade, Internet of Things (IoT): a survey, Dec, in: 2019 IEEE Pune Section International Conference (PuneCon), 2019, pp. 1–6, https://doi.org/10.1109/PuneCon46936.2019.9105831.
[10] P. Saranya, P. Asha, "Survey on Big Data Analytics in Health Care," in 2019 International Conference On Smart Systems And Inventive Technology (ICSSIT), Nov, 2019, pp. 46–51, https://doi.org/10.1109/ICSSIT46314.2019.8987882.
[11] M. Pahar, M. Klopper, R. Warren, T. Niesler, COVID-19 Cough Classification Using Machine Learning and Global Smartphone Recordings, arXiv:2012.01926 [cs, eess], Dec, 2020. Accessed: Jan. 16, 2021. [Online]. Available: http://arxiv.org/abs/2012.01926.
[12] M. Loey, G. Manogaran, N.E.M. Khalifa, A Deep Transfer Learning Model with Classical Data Augmentation and CGAN to Detect COVID-19 from Chest CT Radiography Digital Images, Oct, '' Neural Comput & Applic, 2020, https://doi.org/10.1007/s00521-020-05437-x.
[13] Y.A. Amrulloh, D.A.R. Wati, F. Pratiwi, R. Triasih, A novel method for wet/dry cough classification in pediatric population, May, in: 2016 IEEE Region 10 Symposium (TENSYMP), 2016, pp. 125–129, https://doi.org/10.1109/TENCONSpring.2016.7519390.
[14] E. Nemati, M.M. Rahman, V. Nathan, K. Vatanparvar, J. Kuang, A comprehensive approach for classification of the cough type*, Jul, in: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society, EMBC), 2020, pp. 208–212, https://doi.org/10.1109/EMBC44109.2020.9175345.
[15] J. Schröder, J. Anemuller, S. Goetze, "Classification of human cough signals using spectro-temporal Gabor filterbank features, Mar, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 6455–6459, https://doi.org/10.1109/ICASSP.2016.7472920.
[16] S. Ranjani, V. Santhiya, A. Jayapreetha, A real time cough monitor for classification of various pulmonary diseases, in: 2012 Third International Conference On Emerging Applications Of Information Technology, Nov, 2012, pp. 102–105, https://doi.org/10.1109/EAIT.2012.6407872.
[17] V. Bhateja, A. Taquee, D.K. Sharma, Pre-processing and classification of cough sounds in noisy environment using SVM, in: 2019 4th International Conference on Information Systems and Computer Networks (ISCON), Nov. 2019, pp. 822–826, https://doi.org/10.1109/ISCON47742.2019.9036277.
[18] W. Gao, W. Bao, X. Zhou, Analysis of cough detection index based on decision tree and support vector machine, Jan, J. Combin. Optim. 37 (1) (2019) 375–384, https://doi.org/10.1007/s10878-017-0236-8.
[19] A. J A. in neural information processing systems Krizhevsky I. Sutskever, G. E. Hinton, A. %J A. in neural information processing systems Krizhevsky, "Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. (2012) 1097–1105.
[20] S. Liu, W. Deng, Very deep convolutional neural network based image classification using small training sample size, in: 2015 3rd IAPR Asian Conference On Pattern Recognition, ACPR), 2015, pp. 730–734, https://doi.org/10.1109/ACPR.2015.7486599.
[21] C. Szegedy, et al., Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition, CVPR), 2015, pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594.
[22] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
[23] R.X.A. Pramono, S.A. Imtiaz, E. Rodriguez-Villegas, A cough-based algorithm for automatic diagnosis of pertussis, PLoS One 11 (9) (Sep. 2016), e0162128, https://doi.org/10.1371/journal.pone.0162128.
[24] N. Rochmawati, et al., Covid symptom severity using decision tree, in: 2020 Third International Conference On Vocational Education And Electrical Engineering (ICVEE), Oct. 2020, pp. 1–5, https://doi.org/10.1109/ICVEE50212.2020.9243246.
[25] M. Soliński, M. Łepek, Ł. Kołtowski, Automatic cough detection based on airflow signals for portable spirometry system, Informatics in Medicine Unlocked 18 (Jan. 2020) 100313, https://doi.org/10.1016/j.imu.2020.100313.
[26] B.W. Schuller, H. Coppock, 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.

[27] V. Bansal, G. Pahwa, N. Kannan, Cough Classification for COVID-19 based on audio mfcc features using Convolutional Neural Networks, Oct, in: 2020 IEEE International Conference On Computing, Power And Communication Technologies (GUCON), 2020, pp. 604–608, https://doi.org/10.1109/ GUCON48875.2020.9231094.

[28] A. Imran, et al., AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app, Informatics in Medicine Unlocked 20 (2020) 100378, https://doi.org/10.1016/j.imu.2020.100378. Jan.

[29] J.-M. Liu, M. You, Z. Wang, G.-Z. Li, X. Xu, Z. Qiu, Cough event classification by pretrained deep neural network, Nov, BMC Med. Inf. Decis. Making 15 (4) (2015) S2, https://doi.org/10.1186/1472-6947-15-S4-S2.

[30] H.I. Hee, et al., Development of machine learning for asthmatic and healthy voluntary cough sounds: a proof of concept study, Art. no. 14, Jan, Appl. Sci. 9 (14) (2019), https://doi.org/10.3390/app9142833.

[31] Y. Amrulloh, U. Abeyratne, V. Swarnkar, R. Triasih, "Cough Sound Analysis for Pneumonia and Asthma Classification in Pediatric Population," in *Modelling And Simulation 2015 6th International Conference On Intelligent Systems*, Feb. 2015, pp. 127–131, https://doi.org/10.1109/ISMS.2015.41.

[32] M. Cohen-McFarlane, R. Goubran, F. Knoefel, Novel coronavirus cough database: NoCoCoDa, IEEE Access 8 (2020) 154087–154094, https://doi.org/10.1109/ ACCESS.2020.3018028.

[33] N. Sharma, et al., Coswara – A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis, *arXiv:2005.10548 [cs, eess]*, Aug. 2020. Accessed: Jan. 16, 2021. [Online]. Available: http://arxiv.org/abs/2005.10548.

[34] M. Loey, A. ElSawy, M. Afify, Deep learning in plant diseases detection for agricultural crops: a survey, Int. J. Serv. Sci. Manag. Eng. Technol. (2020) (IJSSMET), Apr. 01, www.igi-global.com/article/deep-learning-in-plant-diseases-detection-for-agricultural-crops/248499. accessed Apr. 11, 2020.

[35] M. Loey, M.R. Naman, H.H. Zayed, A survey on blood image diseases detection using deep learning, Int. J. Serv. Sci. Manag. Eng. Technol. (2020). Jul. 01, www. igi-global.com/article/a-survey-on-blood-image-diseases-detection-using-d eep-learning/256653. Jun. 17, 2020.

[36] A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends, Jan. 28, 2021), https://ieeexplore.ieee.org/document/8351898.

[37] Y.-H. Byeon, S.-B. Pan, K.-C. Kwak, Intelligent deep models based on scalograms of electrocardiogram signals for biometrics, Art. no. 4, Jan, Sensors 19 (4) (2019), https://doi.org/10.3390/s19040935.

[38] T. Li, M. Zhou, ECG classification using wavelet packet entropy and random forests, Art. no. 8, Aug, Entropy 18 (8) (2016), https://doi.org/10.3390/ e18080285.

[39] H. Khorrami, M. Moavenian, A comparative study of DWT, CWT and DCT transformations in ECG arrhythmias classification, Aug, Expert Syst. Appl. 37 (8) (2010) 5751–5757, https://doi.org/10.1016/j.eswa.2010.02.033.

[40] N. Sharma, V. Jain, A. Mishra, An analysis of convolutional neural networks for image classification, Jan, Procedia Computer Science 132 (2018) 377–384, https:// doi.org/10.1016/j.procs.2018.05.198.

[41] M. Loey, M. Naman, H. Zayed, Deep transfer learning in diagnosing leukemia in blood cells, Art. no. 2, Jun, Computers 9 (2) (2020), https://doi.org/10.3390/ computers9020029.

[42] M.L.R. AbdElNabi, M. Wajeeh Jasim, H.M. El-Bakry, M. Hamed N. Taha, N.E. M. Khalifa, Breast and colon cancer classification from gene expression profiles using data mining techniques, Art. no. 3, Mar, Symmetry 12 (3) (2020), https:// doi.org/10.3390/sym12030408.

[43] H. Yasar, M. Ceylan, A novel comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning, and deep learning methods, Oct, Multimed. Tool. Appl. (2020), https://doi.org/10.1007/s11042-020-09894-3.

[44] I.D. Apostolopoulos, S.I. Aznaouridis, M.A. Tzani, Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases, Jun, J. Med. Biol. Eng. 40 (3) (2020) 462–469, https://doi.org/10.1007/s40846-020-00529-4.

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

[46] S. Santurkar, D. Tsipras, A. Ilyas, A. Mądry, How does batch normalization help optimization?, Dec, in: Proceedings Of the 32nd International Conference On Neural Information Processing Systems Red Hook, NY, USA, 2018, pp. 2488–2498.

[47] B. Zoph, V. Vasudevan, J. Shlens, 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.

[48] Y. Xu, 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, Jul, J. Anal. Test. 2 (3) (2018) 249–262, https://doi.org/10.1007/s41664-018-0068-2.

[49] I. Sutskever, J. Martens, G. Dahl, G. Hinton, On the importance of initialization and momentum in deep learning, Atlanta, GA, USA, in: Proceedings of the 30th International Conference on International Conference on Machine Learning vol. 28, 2013. III-1139-III–1147.

[50] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. 15 (56) (2014) 1929–1958.