Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 2807 – 2823
© TÜBİTAK
doi:10.3906/elk-2105-64

A transfer learning-based deep learning approach for automated COVID-19 diagnosis with audio data

Devrim AKGÜN[1,*], **Abdullah Talha KABAKUŞ**[2], **Zehra KARAPINAR ŞENTÜRK**[2], **Arafat ŞENTÜRK**[2], **Enver KÜÇÜKKÜLAHLI**[2]

[1]Department of Software Engineering, Faculty of Computer and Information Sciences, Sakarya University, Sakarya, Turkey

[2]Department of Computer Engineering, Faculty of Engineering, Düzce University, Düzce, Turkey

Received: 09.05.2021 • **Accepted/Published Online:** 09.08.2021 • **Final Version:** 04.10.2021

Abstract: The COVID-19 pandemic has caused millions of deaths and changed daily life globally. Countries have declared a half or full lockdown to prevent the spread of COVID-19. According to medical doctors, as many people as possible should be tested to identify their status, and corresponding actions then should be taken for COVID-19 positive cases. Despite the clear necessity of these medical tests, many countries are still struggling to acquire them. This fact clearly indicates the necessity of a large-scale, cheap, fast, and accurate alternative prescreening tool that can be used for the diagnosis of COVID-19 while waiting for the medical tests. To this end, a novel end-to-end transfer learning-based deep learning approach that uses only a given cough sound for the diagnosis of COVID-19 was proposed in this study. The proposed models employed various pretrained deep neural networks, namely, VGG19, ResNet50V2, DenseNet121, and MobileNet, via the transfer learning technique. Then, these models were evaluated on a gold standard dataset, namely, Cambridge data. According to the experimental result, the proposed model, which employed the MobileNet via the transfer learning technique, provided the best accuracy, 86.42%, and outperformed the state-of-the-art. Thus, the proposed model has the potential to provide automated COVID-19 diagnosis in an easily applicable and fast yet accurate way.

Key words: COVID-19, diagnostics, audio analysis, transfer learning, convolutional neural network, deep neural network

1. Introduction

The COVID-19 pandemic outbreak on a global scale has caused millions of deaths and caused millions of people to stay in intensive care units. It has also caused some bad social effects, such as the loss of a large number of people's jobs due to national or local lockdowns that made people's lives even more difficult. Hence, a quick diagnosis of COVID-19 is crucial for people to take the proper precautions before the effects of the infection get worse. Therefore, some researchers have started to study machine learning-based detection for finding alternative approaches or improved models for the quick detection of patients so that they can be isolated immediately to (i) prevent the spread of infection, (ii) and start treatment before it gets worse. As of the 9th of March 2021, the World Health Organization (WHO) reported that the COVID-19 pandemic had caused about 2.6 million deaths and more than 116 million confirmed cases on a global scale.[1] These huge numbers clearly

*Correspondence: dakgun@sakarya.edu.tr

[1]World Health Organization (WHO) (2021). WHO Coronavirus (COVID-19) Dashboard [online]. Website https://covid19.who.int [accessed XX March 2021].

demonstrate the alarming threat level of the COVID-19 pandemic, and its effect still goes on at the time of the writing of this paper.

Scientists from many disciplines, including but not limited to medical doctors, pharmacologists, economists, sociologists, and data scientists, have then started studying the COVID-19 pandemic from their perspectives. Ongoing COVID-19 studies using machine learning techniques involve datasets that are usually categorized into medical images, textual, and speech data. Medical images usually involve X-ray images and computed tomography (CT) scans which require special equipment to obtain. Textual data are usually classified into COVID-19 case reports, social media posts, mobility data, national provider identifier (NPI) data, and scholarly article collections. Speech data include cough and breath sounds from healthy people and COVID-19 patients. One of the known practical methods for the diagnosis of respiratory illnesses is the analysis of respiratory sounds [1, 2]. In practice, electronic stethoscopes provide a good means for discovering various illnesses for clinicians [3]. This fact motivated researchers to develop automated tools for analyzing sounds to make quick diagnoses by exploring the differences in the audio records of patients and healthy people, such as coughs, voice, and breathing. This provides a cheap and fast alternative for COVID-19 infection detection, which meets the needs of many low-income people and also country-wise as it was reported that the test of the whole population of the US would take $\$8.6B$ alone, assuming a test would cost.$\$23$ [2]

Therefore, the tests can only be applied to a limited number of people daily. For example, the daily diagnostic testing capacity in the US as of July 2020 was reported to be fluctuating between $520,000$ and $823,000$ despite that some experts forecasted the need for five million tests per day during this time [4]. In order to measure the efficiency of the proposed model, we used a gold standard open-source COVID-19 dataset, namely, Cambridge data [5], that was proposed by the researchers of the Cambridge University and was composed of 4352 unique users whose sound (cough, and breath) records were collected through the developed web and mobile apps. To this end, various CNN models based on the VGG19, ResNet50V2, DenseNet121, and MobileNet, have been proposed. The proposed models were evaluated on the Cambridge data to reveal their efficiencies. The main contributions of the proposed approach are as follows:

- A very rare machine learning-based disease diagnosis has been performed. We used audio data in order to detect COVID-19 rather than using the medical images, which is quite prevalent in the literature.
- Various pretrained deep neural network architectures were employed for the problem and their efficiencies were evaluated.
- 1D voice signals were converted into 2D mel-spectrogram images as the input of the proposed deep neural network models.
- We proposed an end-to-end supervised decision support system for the COVID-19 pandemic without requiring advanced preprocessing steps such as feature extraction, feature selection, ROI detection, etc.
- The drawbacks of medical imaging-based diagnosis methods, such as the cost of imaging, a longer time, its harmful effects, the unavailability of devices, and the infection risk during imaging, were eliminated.
- The diagnosis of COVID-19 in an easier, faster, and cheaper yet accurate way was provided with an acceptable classification accuracy as high as 88.04%.

[2]Advisory Board (2021). Why your coronavirus test could cost 23 or $2,315$ [online]. Website https://www.advisory.com/daily-briefing/2020/06/17/covid-test-cost [accessed XX March 2021]

- We shed light on the learned lessons through the conducted experiments on the proposed deep neural networks as a contribution to the body of the knowledge of the field.

The rest of the paper is structured as follows: Section 2 describes the related study. The material and method are presented in Section 3. The experimental result and discussion are described in Section 4. The final section, Section 5, concludes the paper with future direction. Performance of the proposed model compared with the state-of-art studies.

2. Related study

In this section, we will briefly explain the scientific papers on COVID-19 diagnosis. We grouped the studies based on the way the diagnosis is made. In the literature, the diagnosis of COVID-19 is mostly made through the evaluation of the medical images or very recently through the cough and breath sound.

2.1. COVID-19 diagnosis via chest images

Chen et al. [5] proposed a momentum contrastive learning method for COVID-19 diagnosis through chest CT images. Their deep learning (DL) approach requires few samples on the contrary to the existing DL methods for this problem. They used a contrastive learning method to train an encoder that is capable of capturing features of public massive lung datasets. 88% classification accuracy was achieved by the proposed method. Another contrastive learning approach was proposed by Li et al. [6]. A contrastive multitask convolutional neural network model was adapted to the problem to distinguish COVID patients from other pneumonia patients and healthy controls. They benefited from contrastive learning in order to obtain transformation-invariant features of CT and X-ray images. They proved that this feature-learning task improves the generalization ability of CNN. Another CT image-based DL approach for the diagnosis of COVID-19 was proposed by Wu et al. [7]. They developed a DL model with weekly supervised active learning. A 2D U-Net and a 3D residual network were designed for lung segmentation and diagnosis through CT scans. An accuracy of 95% was achieved for the classification of COVID-19. Javor et al. [8] showed the accurate prediction ability of deep learning on chest CT images for COVID-19 diagnosis. The performance of their ResNet50 deep model was compared by the performances of two radiologists, and their deep learning approach was shown to be better than human readers with 95.6% AUC. Different from the unsupervised DL approach for the problem, Wu et al. [9] followed the supervised methodology. The texture features of CT scans were extracted by experienced radiologists. Their supervised approach achieves 82% accuracy for individual classification via random forest classifier.

There are also a considerable number of papers on the X-ray image-based automated decision support systems for COVID-19 diagnosis. The most recent and successful studies are summarized as follows.

Shorfuzzaman and Hossain [10] adapted Siamese neural network structure with contrastive loss and n-shot learning for the problem. They proposed a fine-tuned and pretrained CNN encoder to capture features of chest X-ray images. Their model achieved 95.6% accuracy with few training images. Vaid et al. [11] used CNN to detect structural abnormalities and so to categorize the disease with X-ray scans. Their VGG19-based model with three fully connected layers achieved 96.3% accuracy. Apart from these, Jin et al. [12] developed a hybrid ensemble model to differentiate COVID-19 and common viral pneumonia using chest X-ray images. Their model includes feature extraction, feature selection, and classification processes. The proposed model achieved 98.64% accuracy with relief as feature extractor, trial and error approach as feature selector, and support vector machines (SVMs) as a classifier.

In addition to all these, there is also some research that supports the classification process of COVID-19 diagnosis. Colombi et al. [13] compared the performances of CT and lung ultrasound for COVID-19 diagnosis in populations with different disease prevalence levels. CT was shown to have better classification performance than lung ultrasound in both high prevalence and moderate prevalence groups. Canayaz [14] emphasized the effect of feature extraction on the classification performance after the X-ray images are preprocessed by contrast enhancement technique. He supported the classification process with metaheuristic algorithms. Pretrained neural network structures were used for feature extraction, and then, the best features were determined using binary particle swarm optimization and gray wolf optimization methods. The author achieved more than 99% accuracy via SVM with these features.

As seen from the abovementioned very recent literature, COVID-19 can be diagnosed very accurately using chest CT or X-ray images. Deep learning has been frequently used in that kind of COVID-19 classification approach. Although very high classification accuracies were achieved by image-based diagnosis methods, there are also some difficulties, which are listed as follows: First, many image-based diagnosis approaches [7, 15–18] require high computational load for image segmentation. Besides, training with high-resolution CT or X-ray images takes a long time. Beyond all these, obtaining the image can be difficult for some patients because of some reasons such as unavailability or scantiness of proper devices in clinics (especially the case in underdeveloped countries), cost, concerns about the harmful effects of rays, or claustrophobia. Therefore, the researchers must tend to novel approaches that eliminate the mentioned issues.

2.2. COVID-19 diagnosis via cough or speech sound

Speech-based diagnosis of COVID-19 is a recent idea. It aims to eliminate the disadvantages of image-based classification using human voices. In addition, the possibility of infection of this insidious and contagious disease can be eliminated by reducing the contact rate using an individual's speech data.

There are several audio datasets to aid in the diagnosis of COVID-19. Shuja et al. [19] evaluated speech data in three point of views for COVID-19 diagnosis. They explained that cough sounds, breathing rate, and speech stress detection techniques could help COVID-19 detection with machine learning methods. Imran et al. [20] used cough samples to develop a composite model including traditional multiclass machine learning classifiers, binary deep learning classifiers, and multiclass deep learning classifiers based on the fact that cough is one of the major indicators of COVID-19. The challenge of cough-based COVID-19 diagnosis is that cough is a very prevalent symptom in many diseases. The authors, therefore, investigated the distinct pathomorphological features of the respiratory system with and without COVID-19. Their rigorous approach first detects cough, then classifies cough as COVID-19, Bronchitis, Pertussis, or normal. The detection accuracy of COVID-19 cases was approximately 90%.

Similarly, Brown et al. [21] also worked with cough and breath to classify healthy and sick people. They created a crowdsource audio dataset. Voice samples were acquired from 7000 donors, and 235 of them were diagnosed as COVID-19 patients. The researchers developed a web and mobile application to gather voice data from all over the world. Their approach uses both handcrafted features and features from transfer learning. In addition to the features like duration, onset, tempo, period, MFCC (mel-frequency cepstral coefficients), they also used VGGish, a CNN proposed for audio classification. These features were used in logistic regression (LR), gradient boosted trees, and support vector machines classifiers. They achieved an AUC of 80% with LR for the COVID-19 positive cases from non-COVID-19 cases. Sharma et al. [22] also established a dataset of respiratory sounds through a web application. They also provided a web and mobile application to be used as a diagnosis tool. People record their voices through the application, and this tool provides the probability of COVID-19 for the given voice sample. They used a 28D feature vector including temporal and spectral acoustic features for the random forest classifier. Their test accuracy was 66.74%.

Unlike others, Han et al. [23] investigated only the data obtained from the COVID-19 positive patients. They analyzed speech recordings of COVID-19 patients and developed an audio-based model to predict the health state (severity, sleep quality, fatigue, and anxiety) of the patients automatically. They established two feature sets and used SVM for classification with an average accuracy of 69%.

The above sections summarize the studies on COVID-19 diagnosis. As we emphasized previously, image-based detection methods have several drawbacks. Although very high classification accuracies were achieved in DL-based approaches, the proposed models came to a solution with rigorous processing. Most of them require a segmentation step. Also, there are other reasons such as unavailability of proper imaging devices, concerns about the damages of rays, cost of imaging, claustrophobia for CT scan, infection probability during imaging, etc. that make image-based diagnosis less useful. These reasons tend the researchers to develop easier, cheaper, and safer methods in terms of infection risk. The audio-based related works that have been proposed for the problem are summarized in Table 1. Once the reliable audio-based method is developed, all aforementioned drawbacks and concerns can vanish.

Although there are several papers on the diagnosis of COVID-19 using audio data, either their dataset is not publicly available in order to compare and evaluate the performance of the proposed methods with others, like the study by Imran et al. [20], or the models suffer from sufficient accuracy like [21–23]. In this paper, we aimed to cope with these issues. We used an open audio dataset of Brown et al. [21] that we previously called Cambridge data and provided an easier training process without compromising the classification accuracy.

In this study, we proposed a MobileNet-based deep neural network to differentiate COVID-19 cases from the non-COVID-19 cases. When we reviewed the related studies in the literature in terms of utilizing transfer learning, we have noticed that there are not enough studies that adapt several popular pretrained networks to the COVID-19 diagnosis problem. However, the power of these networks, which were trained by millions of images of thousands of classes, cannot be ignored, and many studies in different domains in the literature are already aware of this power [11, 12, 14, 24, 25]. However, for the diagnosis of COVID-19, it is not utilized properly. Although there exist some papers, which use pretrained networks such as ResNet, InceptionV3, SqueezeNet, etc., they are proposed for image-based diagnosis. To the best of our knowledge, there is no paper that benefits from popular pretrained networks for the audio-based automatic diagnosis of COVID-19. Therefore, we intended to fill this gap and used different pretrained models for COVID-19 diagnosis, including VGG19, ResNet50V2, DenseNet121, and MobileNet.

3. Material and method

In this section, the utilized dataset for the proposed study, the architecture of the proposed model, and how the model was optimized and trained were described.

3.1. Data preparation

Cambridge data [5] is a large-scale crowdsourced dataset of respiratory sounds collected for the diagnosis of COVID-19. This dataset consists of 4352 unique samples collected from 2261 users over a web and an Android application. Therefore, the measurement instrument is the microphone of the corresponding device (e.g., computer, smartphone, and tablet). According to Brown et al. [21], the Cambridge data was crowdsourced

Table 1. A comparison of the related work that was proposed for audio-based COVID-19 diagnosis.

Paper	Dataset	Purpose	Method	Metrics
Imran et al. [20]	Self-established and not donated dataset	COVID-19 detection	CNN-based cough detection followed by a combination of DTL-MC, CML-MC(SVM), and DTL-BC for COVID-19 detection. Features of the cough detector were transferred to COVID-19 detector.	77.3%(TPR)* 83.8%(TNR)
Brown et al. [21]	Cambridge Data [21]	COVID-19 detection	Feature extraction and then Logistic regression classifier	82% (AUC)
Sharma et al. [22]	Coswara[22]	COVID-19 detection	Feature extraction and then Random forest classifier	66.74% (accuracy)
Han et al. [23]	Self-established and not donated dataset	To categorize the health state of patients from four aspects: the severity of illness, sleep quality, fatigue, and anxiety.	Feature extraction and then SVM	69% (accuracy)
Proposed model	Cambridge Data [21]	COVID-19 detection	Transfer learning-based deep neural network that utilizes the extracted mel-spectrograms of given cough sounds	86.42% (accuracy)

DTL-MC: deep transfer learning-based multiclass classifier,
CML-MC: classical machine learning-based multiclass classifier,
DTL-BC: deep transfer learning-based binary classifier,
SVM: support vector machines,
AUC: area under the ROC curve,
*TPR for Imran et al. [20] refers to the probability of detecting a COVID-19 positive case as COVID-19 positive, and TNR refers to the probability of detecting a COVID-19 negative case as COVID-19 negative

from a diverse range of countries including but not limited to Greece, the United Kingdom, Italy, Germany, Iran, India, and Bangladesh. The dataset was crowdsourced through the developed web, and Android app. Unlike some other COVID-19 audio datasets, the Cambridge data contains only cough and breath sounds; voices (e.g., background noise and conversations between the patients and medical doctors) were not included. The creators of this dataset designed it to be ready to use for classification tasks, and the one that utilized with our study, namely, task 1, consists of 460 cough sounds from 141 COVID-19 positive users (users who have declared they tested positive), and 319 COVID-19 negative users (users who have not declared a positive test for COVID-19, have a clean medical history, have never smoked, and have no symptoms) for distinguishing users for the COVID-19. The durations of the cough recordings vary from 0.48 s to 24.96 s. The mean and standard deviation of the durations of the cough recordings were calculated as 5.77 s and 2.38 s, respectively. The histogram of durations of the cough recordings is presented in Figure 1. In order to analyse audio files using two-dimensional convolutional neural networks, mel-spectrogram images were obtained for each audio files. A mel-spectrogram provides a visual representation of a given sound's frequency, amplitude, and time characteristics [26] that represents useful information from nonstationary signals [27]. Therefore, mel-spectrograms were generated from these cough sounds thanks to an open-source Python audio and music analysis library, namely, Librosa [28] , which provides various functions for information retrieval from the signals. Figure 2 shows the steps to compute mel-spectogram images corresponding to an input audio file. Mel-frequency scale produces a model similar to human frequency perception [28] and used frequently in audio applications. The first step is to split the input signal into small frames determined with the window size parameter. Fourier transform of the sequential frames, which may overlap each other depending on the hop length, are used to compute the mel-spectrogram. In the experiments, an FFT window length of 2048, a hop length, which is the number of samples between consecutive frames of 512, and the window type of "hann" was employed during the generation of mel-spectrograms. Since the sampling rates of the audio files vary between 8000 Hz and 48,000 Hz, a sampling rate of 22,050, which is a standard value for most audio tasks [5], was fixed for all files. The obtained mel-spectrograms were exported as RGB images to replicate the success of CNNs on the tasks based on Music Information Retrieval (MIR) [29] for the COVID-19 detection task. The main motivation behind this idea was that a proposed CNN might efficiently utilize the color scales of mel-spectrograms that indicate the amplitude of the frequency [27]. Thereafter, these mel-spectrogram images were given to the proposed model as the input to the model after reshaping.

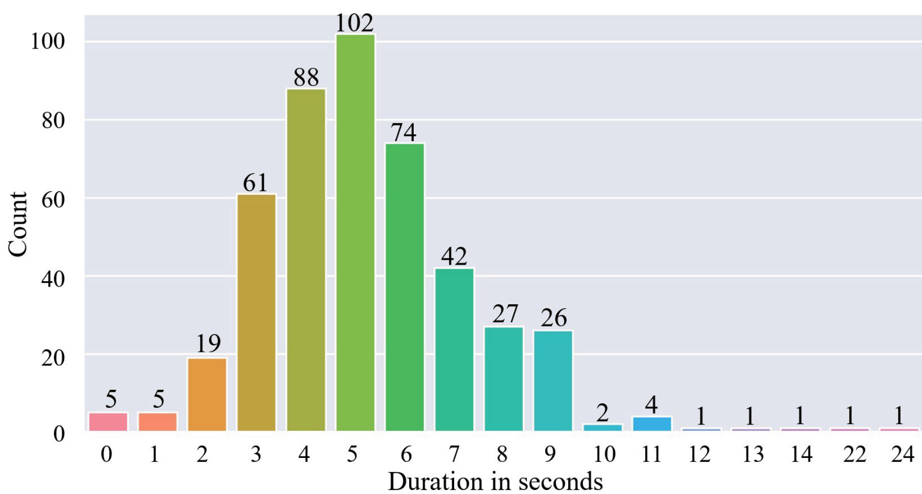

Figure 1. The histogram of durations of the cough recordings.

3.2. Proposed model

The proposed model was implemented using Keras [30], an open-source wrapper library written in Python for various deep neural network backends, namely, TensorFlow [31], Microsoft Cognitive Toolkit, and Theano. TensorFlow, a widely-used machine learning platform developed by Google, was employed as the back end of the proposed model. The proposed model employed the transfer learning technique, which is a technique of employing a model, that was proposed for a task (T_1), for another but related task (T_2). Given the massive resources needed to train deep neural networks, transfer learning is a useful technique when the training dataset

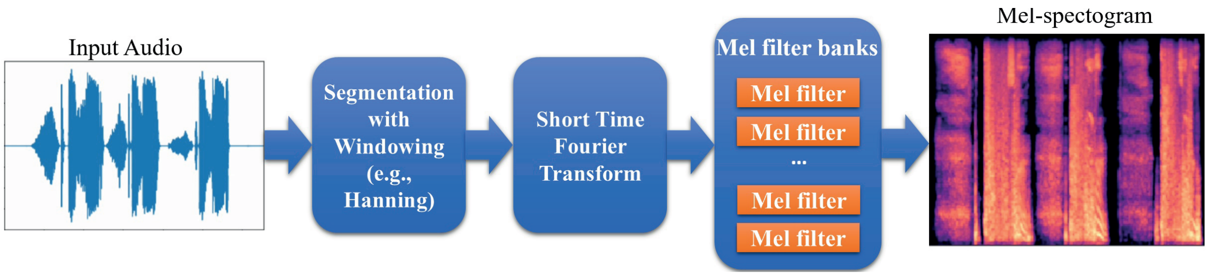

Figure 2. The steps to compute mel-spectogram images.

is not large enough to train models that provide high accuracy, which was the case for the proposed study. Since the aims of these tasks are different, the layers of the transferred model responsible for classification are excluded from the task T_2. An overview of the workflow of the proposed transfer learning-based COVID-19 diagnosis model is presented in Figure 3.

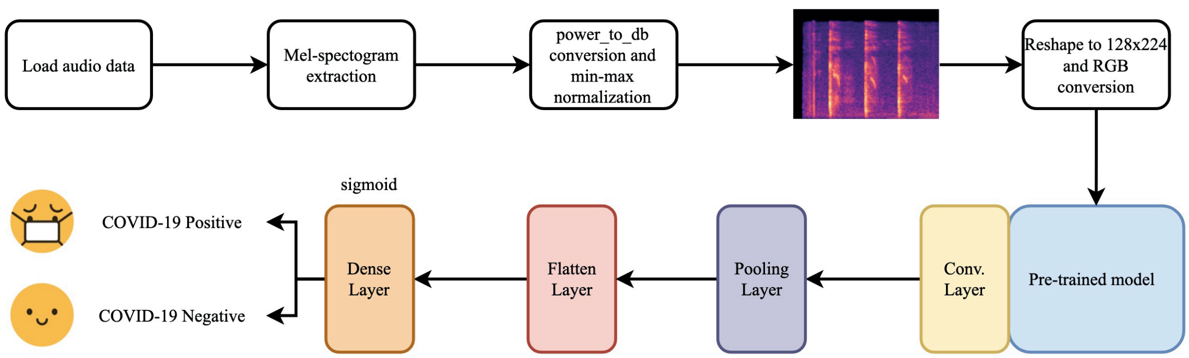

Figure 3. An overview of the workflow of the proposed transfer learning-based COVID-19 diagnosis model.

The proposed models employed the VGG19 [32], ResNet50V2 [33], DenseNet121 [34], and MobileNet [35] pretrained models through the transfer learning technique. VGG19 is a CNN model that consists of 26 layers and was proposed for the ImageNet Challenge 2015. ResNet50V2 is the improved version of the original ResNet50 model because of the modification that was made in the propagation of the connections between blocks [36]. ResNet50V2 was proposed in 2016 and consisted of 192 layers. DenseNet121 and MobileNet were both proposed in 2017 and consist of 429 and 92 layers, respectively. Since the aforementioned pretrained models are already provided by Keras, they were employed as standalone models for the sake of comparison. A comparison of the employed pretrained models is given in Table 2 in terms of the number of layers they consist of, the number of parameters they have, and the file sizes.

Although these transfer learning models were mainly developed for computer vision tasks, they can be used for problems where features can be represented by two-dimensional arrays like images. In this case, the inputs are mel-spectrogram images which are obtained from the mel-spectrogram transform of audio files. For each model, the dense layers, which are located at the top of the models and are responsible for the ImageNet classification task, were excluded from the transfer learning-based models. Since convolutional layers are trained to extract general features from the images, these can be very useful for extracting the features for another problem where the size of the available data set is usually small, as in this case. For the current models, the last

Table 2. A comparison of the employed pretrained models within the proposed models.

Model	Number of layers	Number of parameters	Size
VGG19	26	$146,667,240$	549 MB
ResNet50V2	192	$25,613,800$	98 MB
DenseNet121	429	$8,062,504$	33 MB
MobileNet	92	$4,253,864$	16 MB

convolutional layers right before dense layers were also included in the training to increase the success of the model. The reason behind this design preference is that the convolutional layers, which are close to the dense layer, select more specific features. The dense layers are solely responsible for the classification of the given samples into two classes, namely, (i) COVID-19 positive and, (ii) COVID-19 negative. The number of units of the dense layers was set empirically like the other hyperparameters. Since the aim of the proposed model is classifying the given samples into two classes (which is also known as the binary classification task), the binary cross-entropy was employed as the loss function of the proposed model. For the same reason, the activation function of the last dense layer was fixed to the sigmoid. When it comes to other dense layers, ReLU (rectified linear unit), which is considered the most employed activation function for deep neural networks [38, 39], was employed as the activation function to introduce nonlinearity to the employed model. Another advantage of ReLU is that it is easy to compute as the output equals the input if the input is nonnegative; otherwise, it equals 0. This ability can alleviate the gradient vanishing and exploding problems that usually occur with the sigmoid or tanh activation functions [40]. Various optimization algorithms, namely, Adam (adaptive moment estimation), SGD (stochastic gradient descent), and RMSprop were employed during the optimization of the model. The RMSprop [42], which divides the learning rate for weight by a running average of the magnitudes of recent gradients for that weight, was employed as the optimization algorithm of the proposed model since it provided the best classification performance. An overview of the architectures of the proposed models is presented in Figure 4.

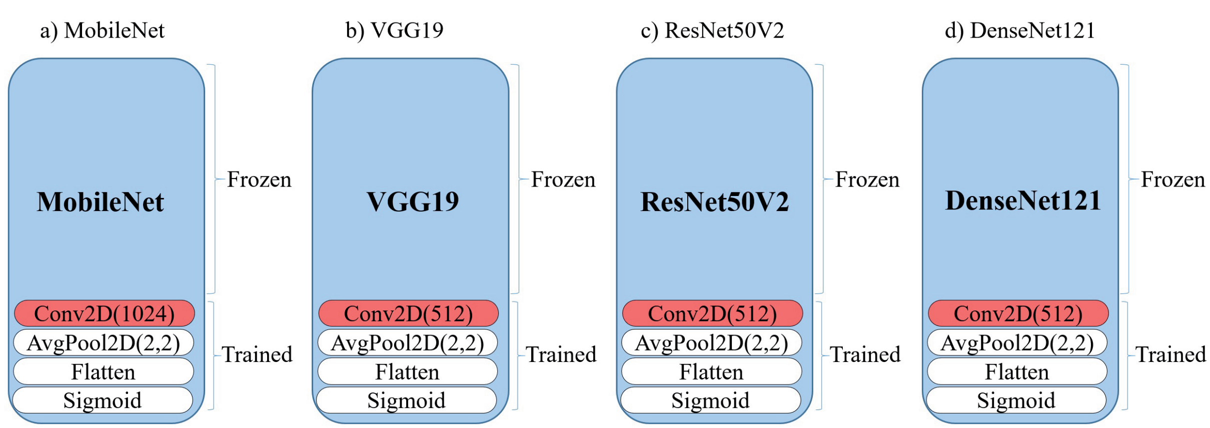

Figure 4. An overview of the architectures of the proposed models.

3.3. Model optimization

Hyperparameters of deep neural networks are set empirically, and they greatly affect the learning process. Therefore, a wide range of values should be experimented with in order to reveal the most optimized model in terms of provided classification performance. To this end, a wide range of values, which are listed in Table 3, were experimented with during the optimization of the proposed model.

Table 3. The hyperparameters and their values that were experimented with during the model optimization.

Hyperparameter	Experimented values
Activation function	ReLU, sigmoid
Loss function	Binary cross-entropy
Optimization algorithm	Adam, SGD, RMSprop, Nadam
Learning rate	5×10^{-4}, 1×10^{-3}, 5×10^{-3}, 1×10^{-2}
Batch size	4, 8, 16, 32
Dropout rate	0.2, 0.4

3.4. Model training

Evaluations for the models were realized using k-fold validation, which is useful in the case of a small dataset. For this purpose, the number of folds (the k value) was set to 5. Thanks to this technique, samples in each fold were guaranteed to be different. This was realized with the Stratified k-fold function of the widely-used scikit-learn [43] library, which is the implementation of the stratified k-fold technique and preserves the same number of samples for each class. Stratified k-fold technique helps to avoid overfitting [44]. Hence, this technique is being deployed to further improve the prediction accuracy of classifiers [45]. Unlike the related work, the proposed model had not been trained for a fixed number of epochs. Instead of this approach, the early stopping callback was employed, which is responsible for stopping the training when the monitored criterion had not got better for a predefined number of epochs (aka patience) in order to prevent overfitting [46]. To the best of our knowledge, the early stopping callback has not been employed in any of the related work. The monitored criteria and patience of the employed early stopping callback were defined as the accuracy obtained for the validation set (aka validation accuracy) and 30, respectively. The training had continued for 32 to 84 epochs until it was stopped by the employed early stopping callback since the validation accuracy had not been decreased for 30 epochs.

4. Experimental result

The proposed transfer learning-based models were trained and evaluated on the Cambridge data dataset to reveal their efficiencies in terms of the diagnosis of COVID-19. Hyperparameter tuning is a critical task [48] for deep neural networks to find out the hyperparameters that yield the best accuracy for the employed network. Therefore, hyperparameter tuning was applied to the proposed models via the grid search technique. According to the experimental result, the proposed model based on the MobileNet obtained the best accuracy, an accuracy of 86.42%, when the batch size and learning rate were set to 4 and 10^{-4}, respectively. Following this model, the proposed model based on the DenseNet121 obtained an accuracy of 85.33% when the batch size and learning rate were set to 16 and 5×10^{-3}. Using the same values for the batch size and learning rate, the proposed model based on the VGG19 obtained a similar accuracy, an accuracy of 85.11%. The best accuracy of the proposed model based on ResNet50V2 was calculated as 82.06% when the batch size and learning rate were set to 8 and 10^{-4}, respectively. All accuracies obtained for the proposed models on the test set with respect to the employed values for hyperparameters are listed in Table 4. Also, the obtained TPR and FPR values for each proposed model when the 5-fold validation was employed are plotted in Figure 5. Inference time is a measurement that reveals the time duration required for a model to evaluate a given single sample. As the elapsed inference times of the proposed models are plotted in Figure 6, the proposed models, rated from fastest to slowest, were obtained as follows: VGG19, MobileNet, ResNet50V2, and DenseNet121.

Table 4. The obtained accuracy values for the proposed models on the test set with respect to the employed values for hyperparameters.

Batch size	Learning rate	MobileNet	VGG19	DenseNet121	ResNet50V2
4	1×10^{-2}	0.8380	0.8424	0.8270	0.7899
	5×10^{-3}	0.8468	0.8358	0.8292	0.7855
	1×10^{-3}	0.8533	0.8380	0.8358	0.8052
	1×10^{-4}	**0.8642**	0.8490	0.8359	0.8139
8	1×10^{-2}	0.8468	0.8424	0.8269	0.7921
	5×10^{-3}	0.8489	0.8381	0.8292	0.8030
	1×10^{-3}	0.8577	0.8402	0.8337	0.8118
	1×10^{-4}	0.8468	0.8468	0.8117	**0.8206**
16	1×10^{-2}	0.8380	0.8446	0.8380	0.7877
	5×10^{-3}	0.8402	**0.8511**	**0.8533**	0.8118
	1×10^{-3}	0.8358	0.8425	0.8402	0.8096
	1×10^{-4}	0.8358	0.8468	0.8094	0.8096
32	1×10^{-2}	0.8292	0.8249	0.8315	0.7812
	5×10^{-3}	0.8292	0.8446	0.8314	0.7790
	1×10^{-3}	0.8467	0.8314	0.8424	0.8009
	1×10^{-4}	0.8402	0.8381	0.7701	0.8162

5. Discussion

Preliminary diagnosis of COVID-19 via the proposed deep transfer learning-based end-to-end method is not substituting it into the hospitals or medical associations, ignoring the values of doctors. The proposed method is just a decision support system to allow nonprofessional people to perform their own COVID-19 test with the purpose of decreasing the infection risk of this viral disseminative disease that influences the world. Trying to get the test done at the hospital will most probably cause new COVID-19 cases if the candidate person carries Coronavirus. Waiting at home after the result of the self-test is seen as positive until the professional medical team arrives will prevent new positive cases. Therefore, it will provide self-isolation and social distancing through self cough test and immediate test result. One of the gains of it can be automated Corona map production through the proper web interfaces. For example, once the proposed method is integrated into the application of the Turkish Ministry of Health, not only the cases determined by the hospitals but also the cases who performed self-tests in their homes can be shown in the Corona map for the locations, and this will avoid possible undesirable contacts for healthy people. Besides, the proposed approach can be used in case of the

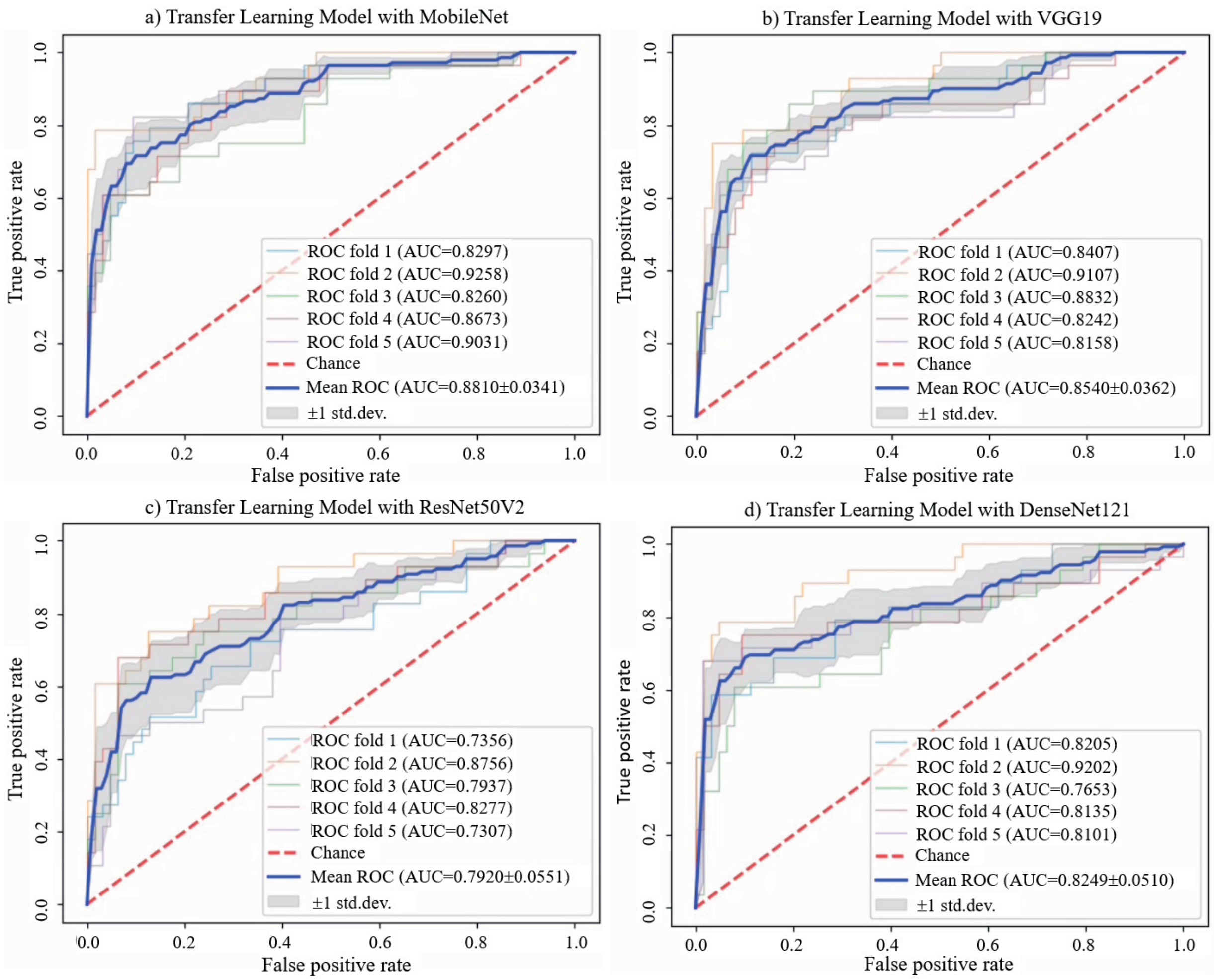

Figure 5. The obtained TPR and FPR values for the proposed transfer learning models when the 5-fold validation was employed.

unavailability of medical test facilities since it just needs the cough sounds of the patients.

Recent ML-based studies related to COVID-19 diagnosis mostly use X-ray or CT scan [49–52]. Although they provide high classification accuracies, they oblige people to go to hospitals for screening. However, this approach is undesirable at least for two reasons: infection risk and shortage of medical facilities. On the contrary, our approach does not require screening. There are some factors, that affect the performance of the proposed transfer learning-based COVID-19 diagnosis method, which are listed as follows:

1. The number of cough samples is one of the critical factors for the generalization capability of the networks. In the present study, this improved with transfer learning-based models.

2. The quality of the samples is desired to be high so that better discriminative features can be extracted for decisions with high accuracies. The existence of irrelevant and redundant data in voice samples like human voices or other environmental noise will decrease the quality of the input features. Cropping irrelevant and relevant data from the dataset may improve the performance and reduce the training time.

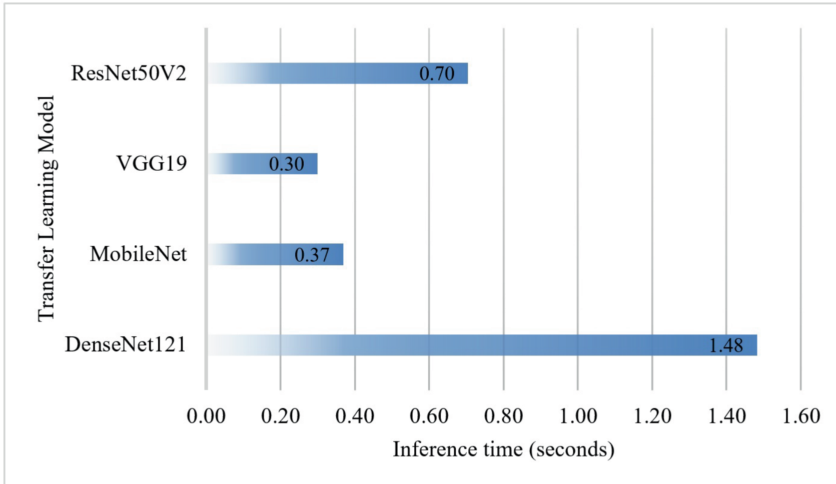

Figure 6. The elapsed inference times for the proposed transfer learning models.

3. The variations in the sample sizes/durations of the sound samples are another factor that decreases the quality. It is compensated with zero paddings for the samples below the determined size and resizing for the samples above the determined size.

The experimental result implies that the best values for the batch size and learning rate depend on the proposed model. Hence, they cannot be generalized. Another conclusion that was reached in the light of the experimental result is that a deeper model does not guarantee better performance as the best performed model was based on the MobileNet, which consists of fewer layers (92) than both the ResNet50V2 (192) and DenseNet121 (429). A comparison of inference times given in Figure 6 shows that the best model that provides the quickest response, which is about 0.30 s is the model based on VGG19. It is followed with the models based on MobileNet and ResNet50V2 which provide the computation result in about 0.37 s and 0.70 s, respectively. With 1.38 s of inference time, the model based on DenseNet121 has the longest inference time among the developed models due to the number of layers.

6. Conclusion

In this study, a novel transfer learning-based deep neural network model was proposed for the diagnosis of the ongoing COVID-19 pandemic. The proposed model was intentionally designed to be a large-scale, cheap, fast, and accurate prescreening tool that can be utilized by as many people as possible who have been still waiting for the COVID-19 medical tests. Therefore, the only input of the proposed model is the given cough sound, which is easily obtainable thanks to the integrated microphones in many daily-use devices (e.g., smartphones, tablets, and notebooks). The proposed models employed various pretrained deep neural networks, namely, VGG19, ResNet50V2, DenseNet121, and MobileNet, via the transfer learning technique. Then, these models were evaluated on a gold standard dataset, namely, Cambridge data, which consists of a total of 9986 samples from 6613 unique users. According to the experimental result, the best accuracy, 89.1%, was obtained when the proposed model employed the MobileNet via the transfer learning technique. The experimental result confirms that the analysis of cough sounds via deep neural networks is an efficient method for the COVID-19 diagnosis. The proposed model outperformed the state-of-the-art and thus is promising enough to encourage us to continue further development. As future work, more pretrained models can be employed via the transfer learning technique. Also, the authors want to evaluate the proposed model on other COVID-19 audio datasets. Finally, more audio-based features can be utilized to try to improve the performance of the proposed model in terms of the correct classification of given samples.

Acknowledgments

The authors are grateful to Cecilia Mascolo from Cambridge University, UK, for providing the audio dataset faithfully and fast. Also, we thank all the people who contributed to the data collection process, either through voice donation or through data processing, to provide opportunities for better diagnostic methods.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Srivastava A, Jain S, Miranda R, Patil S, Pandya S et al. Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease. PeerJ Computer Science 2021; 1-22. doi: 10.7717/peerj-cs.369

[2] Andrès E, Gass R, Charloux A, Brandt C, Hentzler A. Respiratory sound analysis in the era of evidence-based medicine and the world of medicine 2.0. Journal of Medicine and Life 2018; 11 (2): 89-106. doi: 10.5772/48402

[3] Huang M, Liu H, Pi X, Ao Y, Wang Z. Computer-aided diagnosis and new electronic stethoscope. Zhongguo Yi Liao Qi Xie Za Zhi 2017; 41 (3): 161-165.

[4] Laguarta J, Hueto F, Subirana B. COVID-19 artificial intelligence diagnosis using only cough recordings. IEEE Open Journal of Medical and Biological Engineering 2020; 1: 275-281. doi: 10.1109/ojemb.2020.3026928

[5] Chen X, Yao L, Zhou T, Dong J, Zhang Y. Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. Pattern Recognit. 2021; 113:107826. doi: 10.1016/j.patcog.2021.107826

[6] Jimping L, Zhao G, Tao Y, Zhai P, Chen H et al. Multi-task contrastive learning for automatic CT and X-ray diagnosis of COVID-19. Pattern Recognition 2021; 114: 107848. doi: 10.1016/j.patcog.2021.107848

[7] Wu X, Chen C, Zhong M, Wang J, Shi J. COVID-AL: the diagnosis of COVID-19 with deep active learning. Medical Image Analysis 2021; 68: 101913. doi: 10.1016/j.media.2020.101913

[8] Javor D, Kaplan H, Kaplan A, Puchner SB, Krestan C et al. Deep learning analysis provides accurate COVID-19 diagnosis on chest computed tomography. Europen Journal of Radiology 2020; 133: 109402. doi: 10.1016/j.ejrad.2020.109402

[9] Wu Z, Li L, Jin R, Liang L, Hu Z et al. Texture feature-based machine learning classifier could assist in the diagnosis of COVID-19. Europen Journal of Radiology 2021; 137: 109602. doi: 10.1016/j.ejrad.2021.109602

[10] Shorfuzzaman M, Hossain MS. MetaCOVID: A Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients. Pattern Recognition 2020; 113: 107700. doi: 10.1016/j.patcog.2020.107700

[11] Vaid S, Kalantar R, Bhandari M. Deep learning COVID-19 detection bias: accuracy through artificial intelligence. International Orthopaedics 2020; 44 (8): 1539-1542. doi: 10.1007/s00264-020-04609-7

[12] Jin W, Dong S, Dong C, Ye X. A hybrid ensemble model for differential diagnosis between COVID-19 and common viral pneumonia by chest X-ray radiograph. Computers in Biology and Medicine 2021; 131: 104252. doi: 10.1016/j.compbiomed.2021.104252

[13] Colombi D, Petrini M, Maffi G, Villani GD, Bodini FC et al. Comparison of admission chest computed tomography and lung ultrasound performance for diagnosis of COVID-19 pneumonia in populations with different disease prevalence. Europen Journal of Radiology 2020; 133: 109344. doi: 10.1016/j.ejrad.2020.109344

[14] Canayaz M. MH-COVIDNet: Diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images. Biomedical Signal Processing and Control 2021; 64: 102257. doi: 10.1016/j.bspc.2020.102257

[15] Yang D, Xu Z, Li W, Myronenko A, Roth HR et al. Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. Medical Image Analysis 2021; 70: 101992. doi: 10.1016/j.media.2021.101992

[16] Gao K, Su J, Jiang Z, Zeng L, Feng Z et al. Dual-branch combination network (DCN): towards accurate diagnosis and lesion segmentation of COVID-19 using CT images. Medical Image Analysis 2021; 67: 101836. doi: 10.1016/j.media.2020.101836

[17] Vidal PL, Moura J, Novo J, Ortega M. Multi-stage transfer learning for lung segmentation using portable X-ray devices for patients with COVID-19. Expert Systems Applications 2021; 173: 114677. doi: 10.1016/j.eswa.2021.114677

[18] Abdel-Basset M, Chang V, Hawash H, Chakrabortty RK, Ryan M. FSS-2019-nCov: a deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. Knowledge-Based Systems 2021; 212: 106647. doi: 10.1016/j.knosys.2020.106647

[19] Shuja J, Alanazi E, Alasmary W, Alashaikh A. COVID-19 open source data sets: a comprehensive survey. Applied Intelligence 2021; 51: 1296-1325. doi: 10.1007/s10489-020-01862-6

[20] Imran A, Posokhava I, Qureshi HN, Masood U, Riaz MS et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Informatics in Medicine Unlocked 2020; 20: 100378. doi: 10.1016/j.imu.2020.100378

[21] Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A et al. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2020; 11: 3474-3484. doi: 10.1145/3394486.3412865

[22] Sharma N, Krishnan P, Kumar R, Ramoji S, Chetupalli S et al. Coswara – a database of breathing, cough, and voice sounds for COVID-19 diagnosis. Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, vol. 2020-October. pp. 4811–4815. May 2020; Accessed: Mar. 04, 2021. [Online]. Available: http://arxiv.org/abs/2005.10548.

[23] Han J, Qian K, Song M, Yang Z, Ren Z et al. An early study on intelligent analysis of speech under COVID-19: severity, sleep quality, fatigue, and anxiety. Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH. vol. 2020-October. pp. 4946–4950. Apr. 2020; Accessed: Mar. 04, 2021. [Online]. Available: http://arxiv.org/abs/2005.00096.

[24] Wen L, Gao L, Li X. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. IEEE Transactions on Systems, Man, and Cybernetics: Systems 2019; 49 (1): 136-144. doi: 10.1109/TSMC.2017.2754287

[25] Hu F, Xia GS, Hu J, Zhang L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing 2015; 7 (11): 14680-14707. doi: 10.3390/rs71114680

[26] Khanna H, Gaunt SLL, McCallum DA. Digital spectrographic cross-correlation: tests of sensitivity. Bioacoustics 1997; 7: 209-234. doi: 10.1080/09524622.1997.9753332

[27] Verstraete D, Ferrada A, Droguett EL, Meruane V, Modarres M. Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings. Shock and Vibration 2017; 2017: 1-17. doi: 10.1155/2017/5067651

[28] McFee B, Raffel C, Liang D, Ellis DP, McVicar M et al. librosa: audio and music signal analysis in Python. In: Proceedings of the 14th Python in Science Conference (SciPy 2015); City, Country; 2015. pp. 18-25. doi: 10.25080/majora-7b98e3ed-003

[29] Dörfler M, Bammer R, Grill T. Inside the Spectrogram: Convolutional Neural Networks in Audio Processing. In: Proceedings of the 2017 12th International Conference on Sampling Theory and Applications (SampTA 2017); City, Country; 2017. pp. 152-155. doi: 10.1109/SAMPTA.2017.8024472

[30] Chollet F. Deep Learning with Python. New York, NY, USA: Simon and Schuster, 2017.

[31] Abadi M, Barham P, Chen J, Chen Z, Davis A et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2016); City, Country; 2016. pp. 265-283.

[32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015); City, Country; 2015. pp. 1–14.

[33] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Leibe B, Matas J, Sebe N, Welling M (editors). Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, Vol. 9908. Cham, Switzerland: Springer, 2016.

[34] Huang G, Liu Z, Van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017); City, Country; 2017. doi: 10.1109/CVPR.2017.243

[35] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv 2017; 1704.04861: 1-9.

[36] Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Informatics in Medicine Unlocked 2020; 19: 1-9. doi: 10.1016/j.imu.2020.100360

[37] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 2014; 15 (1): 1929-1958.

[38] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015; 521: 436-444. doi: 10.1038/nmeth.3707

[39] Ramachandran P, Zoph B, Le QV. Searching for activation functions. arXiv preprint 2017; 1710.05941: 1-13.

[40] Hu Z, Li Y, Yang Z. Improving convolutional neural network using pseudo derivative ReLU. In: Proceedings of the 2018 5th International Conference on Systems and Informatics (ICSAI 2018); City, Country; 2018. pp. 283-287. doi: 10.1109/ICSAI.2018.8599372

[41] Ng HW, Nguyen VD, Vonikakis V, Winkler S. Deep learning for emotion recognition on small datasets using transfer learning. In: Proceedings of the 2015 ACM International Conference on Multimodal Interaction (ICMI 2015); City, Country; 2015. pp. 443–449. doi: 10.1145/2818346.2830593

[42] Hinton G, Srivastava N, Swersky K. Neural Networks for Machine Learning. 2012.

[43] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research 2011; 12: 2825-2830.

[44] Zhang Y, Wu L. Bankruptcy prediction by genetic ant colony algorithm. Advanced Materials Research 2011; 186: 459-463. doi: 10.4028/www.scientific.net/AMR.186.459.

[45] Zeng Y, Jiang K, Chen J. Automatic seismic salt interpretation with deep convolutional neural networks. In: Proceedings of the 2019 3rd International Conference on Information System and Data Mining (ICISDM 2019); City, Country; 2019. pp. 16-20. doi: 10.1145/3325917.3325926

[46] Ying X. An overview of overfitting and its solutions. In: Proceedings of the International Conference on Computer Information Science and Application Technology (CISAT 2018); City, Country; 2018. pp. 1–6. doi: 10.1088/1742-6596/1168/2/022022

[47] Gopalakrishnan K, Khaitan SK, Choudhary A, Agrawal A. Deep convolutional neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. Construction and Building Materials 2017; 157: 322-330. doi: 10.1016/j.conbuildmat.2017.09.110

[48] Kong W, Dong ZY, Luo F, Meng K, Zhang W et al. Effect of automatic hyperparameter tuning for residential load forecasting via deep learning. In: Australasian Universities Power Engineering Conference (AUPEC 2017); City, Country; 2017. pp. 1-6. doi: 10.1109/AUPEC.2017.8282478

[49] Chaudhary PK, Pachori RB. FBSED based automatic diagnosis of COVID-19 using X-ray and CT images. Computers in Biology and Medicine 2021; 134: 104454. doi: 10.1016/j.compbiomed.2021.104454

[50] Saygılı A. A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods. Applied Soft Computing 2021; 105: 107323. doi: 10.1016/j.asoc.2021.107323

[51] Li X, Zhai M, Sun J. DDCNNC: dilated and depthwise separable convolutional neural Network for diagnosis COVID-19 via chest X-ray images. International Journal of Cognitive Computing in Engineering 2021; 2: 71-82. doi: 10.1016/j.ijcce.2021.04.001

[52] Jia G, Lam HK, Xu Y. Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method. Computers in Biology and Medicine 2021; 134: 104425. doi: 10.1016/j.compbiomed.2021.104425