

Detection and Classification of COVID-19 by Lungs Computed Tomography Scan Image Processing using Intelligence Algorithm

Abstract

The latest World Health Organization statistics show that the number of people living with COVID-19 disease is now more than 42 million worldwide. Some diagnosis methods include detecting and observing clinical symptoms associated with the disease (fever, dry cough, shortness of breath, sore throat, and muscle fatigue). Some other methods, such as computed tomography (CT)-scan imaging from the lungs, are the more accurate diagnostic methods. In this study, we examine the types of abnormal COVID-19 can cause in the lungs of infected subjects and detect and classify this disease. In this paper, we used data from the lung's CT-scan images from the 79 participants. To do this, in this article, for processing CT-scan images of the lungs to diagnose and classification of the COVID-19 disease in men and women of different ages, for rapid diagnosis and high accuracy of this disease by the automatic classification algorithm is used. The final results showed that the proposed method could base on different categories (gender, age categories, and type of damage caused by COVID-19) with high detection and classification accuracy. The algorithm presented in this article has accurately identified the data of healthy subjects and patients with coronavirus.

Keywords: Classification, computerized tomography, COVID-19, detection, medical image processing

Submitted: 24-Jul-2020

Revised: 20-Sep-2020

Accepted: 25-Nov-2020

Published: 20-Oct-2021

Introduction

Coronavirus is a disease caused by a virus called COVID-19. The condition is now a pandemic (or spread worldwide) and has killed many people in different countries. Corona often affects the respiratory system, and the causative agent of the disease, the COVID-19 virus, is shared between humans and animals. A new generation of coronaviruses has spread to a region of China that has claimed thousands of lives so far. The latest version of coronaviruses spread to human specimens in December 2019 in Wuhan, China. Coronavirus is the target of the human respiratory system, so there are receptors on the surface of our respiratory system's cells that the virus uses to enter the sections in this area. In general, the coronavirus presence can be hazardous if a person has a chronic illness or has inflamed lungs and respiratory diseases such as asthma, or if a person has a weakened

immune system. Significant symptoms of the virus in humans include fever, cough, rapid breathing, shortness of breath, sore throat, and runny nose. Other symptoms of coronavirus include chills, body aches, headaches, diarrhea, nausea, and vomiting. According to previous research, compared to middle-aged and older people with COVID-19, it has been found that children or young people with moderate symptoms are more easily treated, and these properties are easily proven.^[1] According to a joint mission by the World Health Organization (WHO) and the China Truth Finder, the death toll from 5.8% in Wuhan to 0.7% in other parts of China. Most fatalities occur in elderly patients or those with underlying disease (cardiovascular, diabetes, chronic lung disease, high blood pressure, and cancer). In the process of diagnosing people with coronavirus, computed tomography (CT)-scan imaging of the chest has higher accuracy and sensitivity than diagnostic kits. CT scans of the chest are a standard imaging tool

**Naser Safdarian¹,
Nader Jafarnia
Dabanloo²**

¹Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, ²Engineering Research Center in Medicine and Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran

Address for correspondence:
Prof. Nader Jafarnia Dabanloo,
Department of Biomedical
Engineering, Faculty of Medical
Sciences and Technology,
Science and Research Branch,
Islamic Azad University, Tehran,
Iran.
E-mail: jafarnia@srbiau.ac.ir

Access this article online

Website: www.jmssjournal.net

DOI: 10.4103/jmss.JMSS_55_20

Quick Response Code:



How to cite this article: Safdarian N, Dabanloo NJ. Detection and classification of COVID-19 by lungs computed tomography scan image processing using intelligence algorithm. J Med Sign Sens 2021;11:274-84.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

for diagnosing pneumonia. In general, white spots on the lung image indicate the coronavirus's involvement, which can be accurately detected by imaging. In CT scans of the chest, the rate of quality images of high lung tissue is high, and a radiologist can quickly determine the extent of lung involvement with the disease. CT scans of the chest identify standard radiological features in patients with COVID-19. These features include ground-glass opacity (GGO), multifocal patchy consolidation, and interstitial changes with the peripheral distribution. The method of diagnosing COVID-19 through radiology has been studied in some previous studies. In one study that the researchers used from the 41 patients with COVID-19, all of the patients had abnormal pneumonia in chest CT scan images.^[2] In another study, abnormal chest CT scan images showed that all samples had significant multifocal patchy ground-glass opacities in the lungs' peripheral parts.^[3]

This study's primary purpose is to examine the abnormal COVID-19 types that can cause in the lungs of infected subjects and detect and classify this disease using extracted features from CT-scan images. Therefore, in this article, we used the intelligent algorithms for processing lungs CT scan images to diagnose and classify COVID-19 disease in different ages men and women, for rapid diagnosis and high accuracy of this disease by automatic classification algorithms.

Literature of previous researches

In March 2020, a review article on coronavirus-2019 (COVID-19) was published by Harapan *et al.*^[4] In that article, the researchers analyzed various studies in the field of the disease. They used the available information to summarize knowledge about the existing epidemic. The study looked at pathogenic and immune responses, epidemiology, diagnosis, treatment, disease management, and control and preventive strategies. In that paper, researchers discussed various diagnostic methods, risk factors, the types of immune damage caused by the disease, and clinical findings.^[4]

In March 2020, Chen *et al.* Used high-resolution CT (HRCT) imaging techniques to diagnose COVID-19 in patients of different age groups.^[5] In that study, HRCT data related to 98 patients (52 men; 46 women) in the age range of 4–88 years (with an average age of 43 years) divided into the four groups. These data were collected from COVID-19 patients at Hangzhou City Hospital, Zhejiang Province, in China. Furthermore, some data obtained from patients related to the days between the disease initial symptoms and HRCT testing, inside four different groups studied. These groups consist of gender and clinical signs (including fever, muscle fatigue, cough dryness, difficulty breathing or dyspnea, profuse sputum, sore throat, and direct or indirect contact with people in Wuhan, China). The laboratory analyzes data (including CRP, SSA tests, white blood cell counts, neutral particle

counts, and the number of lymphocytes – for patients with COVID-19) also used to assess patients' health status.^[5] The analysis performed items such as the rate of bilateral lung involvement, a comparison between the distribution of lesions in different lung lobes, and how lesions were distributed in the lungs at different ages. These data compared the distribution of HRCT density distribution for lesions in different age groups, finally reaching the size distribution of lesions observed in HRCT images in other age groups, review, and analysis by the authors of that article. The results of this study showed that the signs of COVID-19 disease in HRCT data of people with small patches, GGO, and consolidations in 98 patients confirmed the presence of this disease. They showed that patients in the 45–59-year-old age range, aged 60 and older, had lesions in both lungs, lung lobes, and lung content, and had a more significant number of lesions than patients in the following 18 years old age range.^[5]

The diagnostic value of CT scans of the chest compared to the reverse transcription-polymerase chain reaction (RT-PCR) test has been studied.^[6] The study was conducted at Tongji Hospital in Wuhan, China, for 2 months from January 6 to February 6, 2020. The results of this test were published in the Journal of the American Radiological Association (RSNA) on February 26, 2020.^[6] The study looked at images of 1014 CT scans of patients with corona and RT-PCR test results, ranging in age from about 51 years (51 ± 15). These patients underwent both RT-PCR tests from the end of the throat and CT scan of the chest. In this study, several RT-PCR tests were performed to compare their test results with the initial test results after 3 days. Repeated tests were analyzed after 3 days for analysis along with CT scans of the chest. Of the 1014 patients suspected of having corona in that study, 601 patients (59%) reported positive RT-PCR, and 888 out of 1014 patients (88%) had positive corona results on CT scans of the chest. The CT scan test's accuracy and sensitivity in the diagnosis of COVID-19 based on RT-PCR positive findings were about 97%. Based on the RT-PCR test results as a corona diagnostic reference, chest CT scan imaging sensitivity for COVID-19 reported being about 97% (equivalent to 580 patients out of 601).^[6] The results showed that patients with negative RT-PCR had COVID-19 in 48% of cases (equivalent to 147 patients out of 308), although CT scan images confirmed this.^[6]

In April 2020, Dai *et al.* conducted a program to extract the characteristics of CT scan images of the chest and the clinical features of patients with COVID-19 disease in Jiangsu, China.^[7] The study involved 234 people with COVID-19, and their chest CT scans showed symptoms such as GGO, consolidation, fibrosis, and air trapping. They found that more than 90% of people with the condition had a history of epidemiology, with a high fever and cough incidence. The study of CT scan scans of the chest also showed that bilateral abnormalities were seen in the lung

lobes and that most people were involved in the lower and peripheral parts of the lungs.^[7]

Among the methods of diagnosis of COVID-19 disease through the processing of CT scan images of the lungs, some engineering studies have done globally that the most important of them mentioned below.

In July 2020, Minaee *et al.* presented a method based on applying deep learning models to diagnose COVID-19 disease by processing radiographic images of the chest on 5000 X-ray images of the chest presence of COVID-19 disease by a radiologist was determined.^[8] These images were processed on a subset of 2000 radiograms to train four convulsive neural networks called ResNet18, ResNet50, SqueezeNet, and DensNet-121 to diagnose COVID-19 disease. The networks tested on the remaining 3000 images with a sensitivity rate of 98% (± 3) and a specificity rate of about 90%. The researchers also presented a method for generating heatmaps of the lung areas potentially infected with the coronavirus. They showed that most of the identified locations also approved by radiologists.^[8]

In July 2020, Hassantabar *et al.* presented a three-model approach based on Deep Neural Network learning on fractal features of images and convolutional neural networks (CNNs) using X-ray lung's images. They explained the diagnosis and classification of patients with COVID-19 using lung image processing techniques.^[9] In that research method, the researchers presented three algorithms called classification using extracted features, classification without extracted features, and segmentation method to display Ground Truth from input images of subjects with disease COVID-19 and subjects who have not this disease. In the process of segmenting, a CNN architecture to find damaged tissue in lung images also presented. The final results showed that the proposed method identified almost all affected areas with an accuracy of more than 83.84% and could also monitor and control the growth of the damaged regions of patients' lungs.^[9]

In May 2020, Khan *et al.* proposed a method called CoroNet, a deep convolution neural network model for the automatic diagnosis of COVID-19 disease through X-ray images of the lungs.^[10] The model proposed is based on the Xception architecture, which is already trained on the ImageNet dataset. The prepared datasets related to COVID-19 and other X-ray images of the lungs from two different databases in the format of public access are end-to-end trained. The database used in that study consisted of two existing X-ray image databases, in which approximately 290 COVID-19 chest radiographs were available. Classification of 4 classes (including COVID-19, pneumonia viral, Pneumonia bacterial, and normal subjects), as well as classification of three categories (including COVID-19, Pneumonia, and normal) and binary classification has been presented in the research of Khan *et al.*^[10]

In April 2020, Ucar and Korkmaz proposed a diagnosing method for COVID-19 disease based on artificial intelligence.^[11] A deep SqueezeNet with Bayesian optimization (deep Bayes-SqueezeNet) was designed and presented in that study. They used two public databases, including CT scans of lungs, to quickly diagnose their disease. One was the COVID chest X-ray dataset (containing 5949 data from 2839 chest radiographs, including 1583 normal cases, 4290 pneumonia cases, and 76 patients with Quid-19), and the other was the Kaggle chest X-ray pneumonia dataset.^[11]

In June 2020, Peng and Nagata experimentally reviewed nonlinearity and overfitting in machine learning using COVID-19 data.^[12] In the proposed model, the support vector regression method used to predict more than 150,000 patients with COVID-19 in 12 most affected countries (Brazil, France, Germany, India, Iran, Italy, Peru, Russia, Spain, Turkey, United Kingdom, and the USA). Different nonlinear structures using different kernel functions and performance sensitivity analysis, the proposed model presented. The model with the highest degree of nonlinearity (Gaussian kernel) is an example of overfitting in a machine learning model. This model's results can help researchers and professionals in future to better understand the trade-offs in the construction of the model and their subsequent effects on the generalization performance of the model.^[12]

Materials and Methods

Introduce the database

In this study, we used the CT scan images data from the 79 participants (with an average age of 50.3 ± 16.0 ; with a minimum age of 19 years and a maximum age of 91 years) that 49 of them were men and 30 were women. That included 13 healthy participants and 66 subjects with coronavirus. These images related to scans of the subject's chests in various slides, in which images of the lungs of subjects under test were visible. Furthermore, the radiologist's reports have been attached to the images, through which the subjects can carefully consider the desired labels. The radiologist's final report on the subject's status with COVID-19 disease is considered the gold standard of this research for labeling the subject's under test data. All images recorded using 120 kV and 20 mA (for capturing the entire chest image data in Scout Chest mode), and 120 kV and 150 mA (for the Axial Spiral image view data to capture images of different upper body slides for Imaging of the lungs). According to the international protocols, this research has the ethics code approved by the Medical Ethics Committee for ethical considerations.

Lungs computed tomography-scan image processing

The images data were called in MATLAB software (R2018a version) to examine the subjects' CT scan images. Each subject's images included an axial view of the chest and

various viscera slides inside the upper torso (which consists of the right and left lungs). After importing all the subject's images, all of them preprocessed. If there are some image noises, the noises are eliminated by applying a median filter with 3×3 size. The following parts of the images included the subjects' right and left lungs extracted from the clean images obtained from the previous stage using the segmentation method. In this way, using the image segmenter toolbox in MATLAB software, the original images entered the segmentation partition environment in MATLAB. We were applying the appropriate threshold on the images at this stage to extract the lungs from the images of the subjects by changing the contrast of the images. The boundaries of the lungs are extracted by making the changes to the threshold on the image. Then, to more accurately identify the lungs' boundaries, we removed the other areas using clear borders and then used the fill holes option to fill in the lung images' excess cavities. Using the morphological method of abrasion, we proceeded to reveal the lungs' boundaries more accurately. The operation performed using the Erode Mask. Then, through the lungs' mask extracted from the previous step, we obtained the original lungs as carefully from the authentic images. Now, after the final and accurate extraction of the right and left lungs of the subjects, in the final stage, some necessary and essential features were extracted that make a proper distinction between healthy subjects and subjects with coronavirus disease. The first relevant feature was called lung volume. The pulmonary volume feature was extracted using the "region props" function in MATLAB software to extract volumetric pulmonary regions. We extracted the voxels spaces required to perform these calculations from the information in the appendix to each image in the DICOM format (Gathered from the file metadata previously). Finally, the total lung capacity for each lung was extracted and stored as the first feature. The second and third features extracted from the right and left lungs' images were the features of the right lung area and the left lung area, respectively. All the features extracted from the images of the right and left lungs of all the subjects were recorded and stored in the matrix to perform the final classification analysis of coronavirus disease and differentiate the patient data from healthy individuals. How to do this process is explained in future sections.

Statistical analysis on data extracted from database

In this section, we first analyze the descriptive statistics (by SPSS, IBM SPSS Version 20.0) related to the data used in this paper. According to the results of the new age groups for participants, we suggested four age groups in this study, including Group A (range of 19–25 years; 3 subjects), Group B (range of 26–40 years; 24 subjects), Group C (range of 41–60 years; 31 subjects) and Group D (range over 60 years; 21 subjects). In this way, at first, descriptive statistics analyses were performed on the data. These analyses performed on age, gender, and class (healthy, COVID-19) about all subjects, and the results were reported as frequency

and percentage. Furthermore, the subjects were divided into four different age groups and then analyzed. Subsequently, statistical analysis of the normality of the features extracted from CT scan images of the lungs, including lung volume distribution, the area of the right lung area, and the left lung area, was performed. These tests were performed using Kolmogorov–Smirnov and Shapiro–Wilk tests, and their results were reported. Besides, *t*-independent statistical tests were performed for variable lung volume characteristics and two independent male and female groups to investigate the observed differences in male and female data. Subsequently, an independent *t*-test was performed for variable lung volume features and two separate groups of healthy individuals and patients with coronavirus to investigate the statistically significant differences between the two groups. In the next step, Mann–Whitney and Kolmogorov–Smirnov statistical tests were performed for the variable area of the right lung area, and two independent groups of healthy individuals and patients with coronavirus. To investigate the differences between four different age groups studied in this article, Kruskal–Wallis statistical tests for the features that called the areas in the right and left lungs (as separate tests), and four independent groups about the age of the participants in these tests also performed. Furthermore, we performed Mann–Whitney and Kolmogorov–Smirnov statistical tests for the left lung area and two independent groups of healthy individuals and patients with COVID-19.

According to the radiologist's reports about the presence or absence of lesions in different parts of the lungs, we analyzed descriptive statistics on the conflict's various lung lobes in subjects with coronavirus in the SPSS software. Descriptive statistics were also analyzed for lung degeneration in subjects with coronavirus (resolve condition and highly suggestive). The results of all analyses expressed in this section are described separately in the statistical analysis results section, along with the output results of each analysis in SPSS software.

Classification of COVID-19

In this section, we described the types of coronavirus classification methods. In this way, these classifications are done in three different modes. The first mode is that the first category of features, such as the lung volume, the right lung area, and the left lung area, is extracted from the CT scan images of all subjects applied to the classification methods, distinguishing between the healthy and patients with COVID-19. The second mode is that the features include the presence or absence of lesions in the lower part of the right and left lungs of all subjects to examine and differentiate between healthy individuals and patients with coronavirus. Finally, the third mode is that the features consist of the lungs volume, the right and left lung areas, the presence or absence of lesions in all lobes of the two lungs in all subjects, and classification studies performed between healthy individuals and subjects COVID-19. This paper used

the SVM classification method, the Random Forest (RF) method, and the KNN method. In this way, each category of features introduced above applied to these classification algorithms as feature matrices. The last column of these features matrix was healthy subjects (label 1) or patients with COVID-19 (label 2). In this paper, we used the 4-Fold Cross-Validation and 10-Fold Cross-Validation methods. The results of each introduced method were demonstrated in the results section classifiers, with their final results' tables.

Results

First, we present the results of the introduced preprocessing and processing of the CT-scan images of the subject's lungs in the MATLAB software. Figure 1 shows the results of these steps of the method presented in this article.

Results of descriptive statistical analysis

Results of normalization statistical analysis of the lung image's extracted features

Table 1 shows the statistical tests' results in the statistical methods section to check the normality of lung volume

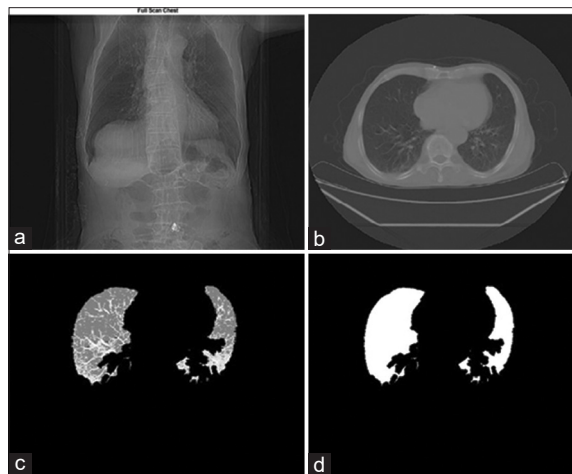


Figure 1: An 80-year-old male with coronavirus disease, subject number 18; (a) Front view of the upper body as scout Chest; (b) Slice number 16 (out of a total of 48 slides obtained for subject) related to the person's lungs; As is clear from the lower lobe of the right and left lungs, there are signs of environmental consolidation; The radiologist's diagnosis of this person as a person with COVID-19 that he was highly suggestive; (c) Extraction of a person's lungs by applying thresholding methods, segmentation, and morphological methods, and finally extracting the lungs' exact boundaries from other tissues in the computed tomography-scan image of this sample person; (d) Binary image of lung extraction; the lower lobes of both right and left lungs, and of course, a small portion of the left lung upper lobe of this subject with COVID-19, have been damaged

distribution, right lung area, and left lung area as extracted features from the lungs images.

This table shows Kolmogorov–Smirnov and Shapiro–Wilk tests performed to test the normality of the three general features extracted from the images. The df column indicates the degree of freedom of the data. The sig column suggests information on the significance of the test, and the statistic column represents the values of the statistical index of the relevant test. According to the introduced test results, the *P* value for the “lung volume distribution” feature is higher than 0.05, which indicates that the test is not significant, and these data are normal. Therefore, we should test the parametric tests (such as *t*-test or ANOVA test) to analyze this feature. According to the above test results, the *P* value for the “right lung area” feature is <0.05, which indicates that the test is significant and that the data are not normal. Therefore, nonparametric tests (such as Mann–Whitney and Kolmogorov–Smirnov) should be used to analyze this feature's data. According to the above test results, the *P* value for the “left lung area” in the Kolmogorov–Smirnov test is more than 0.05, which indicates that the test is not significant for the statistical population. However, this value is <0.05 in the Shapiro–Wilk test, which indicates that the test is significant for the samples under study, which shows that the samples' data are not normal. Therefore, nonparametric tests should be used to analyze the data of this feature.

Results of independent *t*-test, for lung volume feature, and two independent male and female groups

The results of Table 2 show that the variable of the lung volume feature, for both independent male and female groups, has a significant difference ($F = 8.0$; $P = 0.006$), and the variances of them are unequal ($t = 1.7$). It should be noted that the first column of Table 2 represents the results of Levene's Test for Equality of Variances, meaning that *F* represents the statistical index, and sig represents the significance level. Thus, if significance <0.05 is obtained, the results of the second row of the table (equal variances not assumed) are confirmed. In other words, the results of the *t*-test statistical index given in the third column of this table indicate the outcome of the *t*-test. Therefore, in these results, because significance <0.05, the second row's results are confirmed, in which case the results of the *t*-test columns for equality of means which include mean difference and difference and also 95% confidence interval of the difference in the upper and lower band from 95% are reported. These

Table 1: Results of the normality test for all extracted features from the lungs images

	Tests of normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Significance	Statistic	df	Significance
Lungs volume	0.056	79	0.200*	0.990	79	0.781
Area right lung	0.124	79	0.004	0.968	79	0.042
Area left lung	0.094	79	0.079	0.961	79	0.015

*This is a lower bound of the true significance; ^aLilliefors significance correction

results mean that the scattered data of male and female individuals in the studied samples differ. This finding can be considered a difference in lung volume between the two independent groups mentioned. Furthermore, *P* values related to the mean equality test show that the test results are not significant. Therefore, the means are almost equal for the two independent groups of men and women.

Results of independent t-test for lung volume feature and two independent groups of healthy subjects and patients with COVID-19

The results of Table 3 show that the lung volume feature variable, for two independent groups of healthy individuals and patients with COVID-19, has a significant difference ($F = 1.2$; $P = 0.274$), and the variances are equal ($t = 1.9$). Furthermore, sig values related to the mean equality test show that these test results are not significant. Therefore, the means are almost equal for the two independent groups of healthy individuals and patients with the coronavirus. The results and interpretation of the different sections of Table 3 are similar to what is said about Table 2, except that in Table 3, because significance >0.05 , the output of this table's first row is equal variances assumed to be considered. These results indicate that lung volume data are similar in comparing healthy individuals and patients with COVID-19 disease (despite almost identical dispersions). This finding is also evident in the classification results. Although the lack of significant differences in the mean lung volume of healthy subjects and the subjects with COVID-19, the diagnosis of healthy individuals and patients with coronavirus performed with high accuracy.

Statistical tests results for the right lung area as feature and two independent groups of healthy subjects and patients with COVID-19

First, the Mann–Whitney test results, which examined the right lung area as a feature, were presented in two

independent groups of healthy subjects and patients with coronavirus. The results of this test are presented in Table 4.

It should be noted that in the right column of Table 4, the values of the Mann–Whitney test, Wilcoxon, *Z* statistic, and *P* value are reported, respectively.

Then, we presented the results of the Kolmogorov–Smirnov test in Table 5.

Furthermore, in the right column of Table 5 is the statistical index *Z* in the Kolmogorov–Smirnov test, along with its *P* value.

The results of Table 4 show that the right lung area feature, for two independent groups of healthy subjects and patients with COVID-19, had a significant difference in the Mann–Whitney test ($Z = -3.425$; $P = 0.001$). Furthermore, these results had a significant difference in the Kolmogorov–Smirnov test ($Z = 1.897$; $P = 0.001$) in Table 5. The *P* values associated with these tests show that the results of these tests are significant. Therefore, the right lung area values for healthy individuals and patients with corona as two independent groups are almost distinct. These results suggest that the right lung area variable is not the same in comparing healthy individuals and patients with coronavirus. This finding is also evident in the classification results, due to the significant difference in the right lung area of healthy and subjects with COVID-19, the algorithm presented with high accuracy to identify healthy people and patients with coronavirus.

Statistical test results for the right lung area feature and four independent age groups of all subjects

In Table 6, we presented the Kruskal–Wallis test results related to studying the right lung area feature in four independent age groups of the participants in this article.

Table 2: Results of t-test for checking the lung volume feature, for both independent male and female groups

Lungs volume	Levene's test for equality of variances		t-test for equality of means						
	F	Significance	t	df	Significance (two-tailed)	Mean difference	SE difference	95% CI of the difference	
								Lower	Upper
Equal variances assumed	8.004	0.006	1.528	77	0.131	0.23871	0.15625	-0.07242	0.54983
Equal variances not assumed			1.727	76.793	0.088	0.23871	0.13821	-0.03651	0.51393

Independent samples test. CI – Confidence interval; SE – Standard error

Table 3: Results of t-test for checking the lung volume feature, for two independent groups of healthy individuals and patients with COVID-19

Lungs volume	Levene's test for equality of variances		t-test for equality of means						
	F	Significance	t	df	Significance (two-tailed)	Mean difference	SE difference	95% CI of the difference	
								Lower	Upper
Equal variances assumed	1.213	0.274	1.945	77	0.055	0.39425	0.20267	-0.00931	0.79782
Equal variances not assumed			2.263	20.205	0.035	0.39425	0.17421	0.03109	0.75741

Independent samples test. CI – Confidence interval; SE – Standard error

It should be noted that, in the first part of Table 6, we have a table for the four groups performed on the age of subjects (the first group has three people; the second group has 24 people; the third group has 31 people, and the fourth group has 21), and the second table shows the results of the Kruskal–Wallis test.

The results of Table 6 show that the right lung area, for four independent age groups, had a significant difference in the Kruskal–Wallis test ($\chi^2 = 10.7$; $P = 0.013$). The P value associated with this test shows that these test results are significant; therefore, the right lung area feature values for the four age groups of the subjects are separate. These results show that the right lung area data are not the same in comparing subjects in four different age groups.

Statistical tests results for left lung area as feature and two independent groups of healthy subjects and patients with COVID-19

First, the Mann–Whitney test results presented for the left lung area feature in two independent groups of healthy subjects and patients with coronavirus shown in Table 7.

It should be noted that, in the right column of Table 7, the values of the Mann–Whitney test, Wilcoxon, Z statistic, and P value are reported, respectively. Then, the results of the Kolmogorov–Smirnov test are presented in Table 8.

Furthermore, in the right column of Table 8 is the statistical index Z in the Kolmogorov–Smirnov test, along with its P value. The results of Table 8 show that the left lung area, for two independent groups of healthy subjects and patients with coronavirus, had a significant difference in the Mann–Whitney test ($Z = -3.5$; $P = 0.000$). Furthermore, the results had a significant difference in the Kolmogorov–Smirnov test ($Z = 1.8$; $P = 0.002$). The P values associated with these tests show that the results of these tests are significant. Therefore, the left lung area values for two independent groups of healthy subjects and patients with COVID-19 are almost separate. These results suggest that the left lung area’s value is not the same in comparing healthy individuals and patients with coronavirus. This finding can be seen in the classification results, in which due to the significant difference in the left lung area feature of healthy and patient subjects, the algorithm presented with high accuracy to identify healthy people and patients with COVID-19.

Statistical tests results for the left lung area as feature and four independent age groups of all subjects

In the following, we presented the Kruskal–Wallis test results in Table 9, related to the study of the left lung area values, and raised in four independent age groups of the participants.

It should be noted that Table 9 also shows the results of the Kruskal–Wallis test, which also includes the Chi-square, the degree of freedom, and the P value. The results of Table 9

Table 4: Results of the Mann-Whitney test on the right lung area feature about the two independent groups of healthy subjects and patients with coronavirus

Test statistics ^a	
	Area right lung
Mann-Whitney U	170.000
Wilcoxon W	2381.000
Z	-3.425
Asymptotic significance (two-tailed)	0.001

^aGrouping variable: Class

Table 5: Results of the Kolmogorov-Smirnov test on the right lung area feature about the two independent groups of healthy subjects and patients with coronavirus

Test statistics ^a	
	Area right lung
Most extreme differences	
Absolute	0.576
Positive	0.576
Negative	-0.045
Kolmogorov-Smirnov Z	1.897
Asymptotic significance (two-tailed)	0.001

^aGrouping variable: Class

Table 6: Results of the Kruskal-Wallis test on the right lung area feature in four independent age groups of the participants

	Ranks		
	Age new	n	Mean rank
Area right lung	1.00	3	52.67
	2.00	24	50.88
	3.00	31	37.06
	4.00	21	30.10
	Total	79	
Test statistics ^{a,b}			
	Area right lung		
χ^2	10.722		
df	3		
Asymptotic significance	0.013		

^aKruskal-Wallis test; ^bGrouping variable: Age new

Table 7: Results of the Mann-Whitney test on the left lung area feature in two independent groups of healthy subjects and patients with coronavirus

Test statistics ^a	
	Area left lung
Mann-Whitney U	162.000
Wilcoxon W	2373.000
Z	-3.530
Asymptotic significance (two-tailed)	0.000

^aGrouping variable: Class

show that the left lung area feature, for four independent age groups of subjects, had a significant difference in the Kruskal–Wallis test ($\chi^2 = 13.6$; $P = 0.003$). The P value

Table 8: Results of the Kolmogorov-Smirnov test on the left lung area feature in two independent groups of healthy subjects and patients with coronavirus

Test statistics ^a	
	Area left lung
Most extreme differences	
Absolute	0.558
Positive	0.558
Negative	-0.061
Kolmogorov-Smirnov Z	1.840
Asymptotic significance (two-tailed)	0.002

^aGrouping variable: Class

Table 9: Results of the Kruskal-Wallis test for the values of the left lung area in four independent age groups of participants

Test statistics ^{a,b}	
	Area left lung
χ^2	13.609
df	3
Asymptotic significance	0.003

^aKruskal-Wallis test; ^bGrouping variable: Age new

associated with this test indicates that the results of this test are significant, and therefore the values of the left lung area for the four age groups of all subjects are separate. These results suggest that the left lung area values are not the same in comparing subjects in four different age groups.

Descriptive statistics on the involvement of different lung lobes in subjects with COVID-19

We obtained the results of descriptive statistics on the involvement of different lung lobes in subjects with coronavirus. The right lung is divided into three segments: upper, middle, and lower lobes and the left lung comprises two upper and lower lobes. Therefore, by examining the output of the proposed model in this article, following the radiologist's report, each lobe's involvement in the right and the left lung are visible. If that particular lobe is involved, label one and otherwise (i.e., no responsibility of the lung tissue in that specific lobe), the label given a zero label. In the right lung upper lobe, 30.4% of the subjects (out of 100% of the total, 79 people studied) had lesions in this part of the right lung. In the middle lobe of the right lung and the right lung lower lobe, 62% of the subjects had an injury with lesions in these parts. Furthermore, in the left lung upper lobe and the left lung lower lobe, 63.3% and 73.4% of the subjects had some lesions in these parts. Based on these results, it is clear that the left lung lower lobe has the highest number of lesions in the lungs and followed by the left lung upper lobe, the middle lobe, and the lower lobe of the right lung. Finally, the right lung upper lobe had lower values of lesions.

Furthermore, among the 66 patients with COVID-19 among those studied in this paper, few patients had tissue

destruction in the right lung's upper lobe. Patients number 1 and 2 were also affected in all parts of the lungs (except the right lung upper lobe), and thus all lungs parts of these patients suffered from destruction and lung lesions. It should be noted that these patients, due to the involvement of four out of five divisions lungs, were two patients who died by a coronavirus. However, in all subjects with COVID-19, the lower lobe's left lung was involved in tissue destruction by this virus. Of course, the lower lobe's right lung is involved in all patients (except in a few patients) and also in the middle lobe's right lung in all patients (except in a small number of patients) suffered from involvement and tissue damage in this area of the lungs. In the same two patients, Number 10 and 12, the left lung's upper lobe is not damaged, but it was destroyed in other patients. After two patients 1 and 2 (who died), patient number 15 also had a shallow lung volume (compared to healthy specimens) and the destruction of four parts of the five lobes of the right and left lungs. Due to the deterioration of the patient's clinical condition, stricter care and treatment guidelines should be provided.

Descriptive statistics on the type of lung destruction in subjects with COVID-19

By examining the model output proposed in this paper, and according to the radiologist's report, different types of lung destruction depending on the peripheral destruction or fragmentation of lung tissue damage (or interrupted of injury) in subjects with COVID-19, these destructions are defined as follows: Ground-Glass Circumferential, Consolidation Circumferential, Ground Glass Patchy, and Consolidation Patchy. This lung destruction rate can be divided into two different states, "Resolve Condition" and "Highly Suggestive." Suppose any of the above conditions are present in the report of the radiologist. In that case, label 1, and otherwise (i.e., the absence of any above situations) label zero is given to the section under study. In 51% of subjects with coronavirus, the lung injury condition seen as "Highly Suggestive," and the others (i.e., 15%) related to the "Resolve Condition" injury. The highest number of injuries was reported as "Consolidation Circumferential" with 45% of patients with corona disease, and the lowest quantity of lung injuries was reported as "Consolidation Patchy" (with 7% of subjects with the COVID-19). In general, peripheral damage to the lungs is much higher than interrupting damage to lung tissue.

Classification of the COVID-19

In this section, we described the final results of the different classifiers on the introduced features. These results are expressed as the mean and standard deviation of the Sensitivity, Specificity, and Accuracy parameters for using each classification method. Table 10 shows the best results from the data classification of healthy subjects and the subjects with coronavirus.

Table 10: Classification results of the subjects as healthy and with coronavirus-19, for all extracted features from the subjects (with 4-fold and 10-fold cross validation)

Classification results of the subjects as healthy and with COVID-19, for lung volume, right lung area and left lung area as three extracted features from the subjects				
Classifier	Cross validation	Sensitivity	Specificity	Accuracy
SVM (RBF kernel)	4-fold	90±0.15	100±0	98.42±0.02
KNN		66.67±2.9 e-14	68.75±0	68.42±8.6 e-14
RF		66.67±2.9 e-14	62.63±0.008	63.26±0.007
SVM (RBF kernel)	10-fold	95±0.22	100±0	99.38±0.03
KNN		0±0	100±0	87.5±0
RF		0±0	100±0	87.5±0
Classification results of subjects as healthy and COVID-19, for the presence or absence of lesions in the lower lobe of the right and left lungs				
Classifier	Cross validation	Sensitivity	Specificity	Accuracy
SVM (RBF kernel)	4-fold	100±0	93.75±0	95±0
KNN		100±0	93.75±0	95±0
RF		100±0	93.75±0	95±0
SVM (RBF kernel)	10-fold	100±0	100±0	100±0
KNN		100±0	100±0	100±0
RF		100±0	100±0	100±0
Classification results of healthy and patients COVID-19 subjects, for extracted features such as lung volume, right lung area, left lung area, presence or absence of lesions in all lobes of the right and left lungs				
Classifier	Cross validation	Sensitivity	Specificity	Accuracy
SVM (RBF kernel)	4-fold	86.67±0.23	100±0	97.89±0.04
KNN		100±0	100±0	100±0
RF		100±0	100±0	100±0
SVM (RBF kernel)	10-fold	95±0.16	100±0	98.89±0.35
KNN		100±0	100±0	100±0
RF		100±0	100±0	100±0

RF – Random forest; COVID-19 – Coronavirus 2019

Discussion

On December 31, 2019, the Chinese Ministry of Health reported to the WHO that several new pneumonia cases had been observed in Wuhan, China, in the province of Hubei in central China.^[4] The disease has spread rapidly to China and all world's parts, with the latest WHO statistics showing that the number of people with COVID-19 has reached more than 42 million worldwide. The number of deaths has risen, which now it is increasing more than 1 million people.^[13] According to these statistics, the United States of America is one of the countries with the highest number of cases of the disease, followed by India, Brazil, and Russian, with the highest number of cases, until October 24, 2020.^[13] According to WHO reports, the total number of patients observed in 213 countries or regions.^[13] While most countries worldwide have put precautionary measures, the importance of early diagnosis of those with the disease can be pointed out. Therefore, this paper presents a method with very high accuracy for the diagnosis and classification of COVID-19 disease.

The final results of this article determined that the proposed method could be very accurately based on different categories (gender, age groups, type of injury caused by COVID-19, etc.) to diagnose and classify this disease.

In examining the extent of lung tissue involvement, and subjects classified into different groups in this research, data on lung tissue destruction by coronavirus can be discussed. Thus, according to the images of different lung slides in the subjects, the right lung tissue was divided into three parts, the upper lobe, the middle lobe, and the lower lobe. Moreover, the left lung tissue was divided into two parts, the upper lobe and the lower lobe, and we examined these lobes.

In general and considering the position of lung involvement in subjects and affecting lung tissue in patients with coronavirus, it was found that subjects with lower lung damage and destruction of lung tissue were more than the middle and upper lobes. This could be a compelling reason to confirm previous studies^[7] regarding the further involvement of the lower part of the lungs in subjects with COVID-19. According to statistics test, the destruction of lung tissue in the left lung's lower lobe and the left lung's upper lobe obtained in this article, respectively, the coronavirus appears to be more likely to destroy left lung tissue than the right lung in patients with the coronavirus. These findings could be due to the smaller volume of the left lung than the right lung, which could be due to the placement of the virus in the left lung and the faster destruction of this part of the lung. In both the left and

Table 11: Compression the best results of proposed method with best results of previous research

Researcher	Method	Database	Best results (%)		
			Sensitivity	Specificity	Accuracy
Minaee <i>et al.</i> ^[8]	CNN: ResNet18 model, ResNet50 model, SqueezeNet model (with best results), DensNet-121	Lung X-ray images	98±2.7	92.9±0.9	-
Hassantabar <i>et al.</i> ^[9]	DNN	Lung X-ray images	86.01	82.39	82.39
	CNN	Lung X-ray images	96.1	99.71	93.2
Khan, <i>et al.</i> ^[10]	CoroNet: Deep CNN (4-class CoroNet)	Lung X-ray images	89.92	96.4	89.6
	CoroNet: Deep CNN (3-class CoroNet)	Lung X-ray images	96.9	97.5	95
	CoroNet: Deep CNN (binary CoroNet)	Lung X-ray images	99.3	98.6	99
Ucar and Korkmaz ^[11]	Bayes-SqueezeNet (3-class)	Lung X-ray images	98.3	99.1	98.3
Bharati, <i>et al.</i> ^[14]	Combining VGG, data augmentation and STN with CNN or VGG Data STN with CNN (VDSNet)	Lung X-ray images	63	-	73
Proposed method	Statistical analysis	Lung X-ray images	95±0.22	100±0	99.38±0.03
	Intelligence algorithm (SVM, KNN, RBF) Classification results of the subjects as healthy and with COVID-19, for three extracted features from the data				
	Statistical analysis		100±0	100±0	100±0
	Intelligence algorithm (SVM, KNN, RBF) Classification results of the subjects as healthy and with COVID-19, for presence or absence of lesions				
Proposed method	Statistical analysis		100±0	100±0	100±0
	Intelligence algorithm (SVM, KNN, RBF) Classification results of the subjects as healthy and with COVID-19, for all extracted features				

CNN – Convolutional neural network; DNN – Deep neural network; COVID-19 – Coronavirus 2019

right lungs, lung tissue destruction in the lower lobes of the lungs was higher than in other lung lobes. However, these findings are consistent with previous research on the greater susceptibility of the lower parts of the lungs to coronavirus. In addition to the above, good results were obtained in this paper on the nonuniformity of the right lung area's variable data area in the comparison between healthy subjects and patients with coronavirus. These desirable results are visible both in the statistical analysis results and in the classification result. Due to the significant difference in the right lung area of healthy and patients with COVID-19, it is clear that the algorithm presented in this article has been able to accurately identify the data of healthy subjects and patients with coronavirus. Furthermore, in the results of statistical analysis on the subject's data in four different age groups, it was found that the values of the right lung area and also the values of the left lung area per four different age groups are distinct from each other, which, of course, confirms the findings presented in.^[5] It should note that the method presented in this article can speed up the process of diagnosing coronavirus disease and also hasten the start of necessary treatments for these patients to prevent the death of people with this disease. Table 11 indicates the compression of the best results of the proposed method in this article with previous research results. The algorithm presented in this paper has been able to diagnose and classify the coronavirus's damage to the lungs. Compared to other previous methods, we see that this method has shown better and more appropriate performance. Besides,

the likelihood of surviving patients can depend on the strength of the immune system. This idea can only be hypothesized in this paper due to the unavailability of information related to the subjects' immune system.

Compliance with ethical standards

According to the international protocols, this research has the ethics code approved by the Medical Ethics Committee for ethical considerations. This article contains studies with human participants performed by the authors. All ethical considerations have been applied, and the Medical Ethics Committee approved this study in the Islamic Azad University, Science, and Research Branch, Tehran, Iran.

Financial support and sponsorship

None.

Conflicts of interest

There are no conflicts of interest.

References

- Chen ZM, Fu JF, Shu Q, Chen YH, Hua CZ, Li FB, *et al.* Diagnosis and treatment recommendations for pediatric respiratory infection caused by the 2019 novel coronavirus. *World J Pediatr* 2020;16:240-6.
- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395:497-506.
- Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, *et al.* A familial cluster of pneumonia associated with the 2019 novel

- coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet* 2020;395:514-23.
4. Harapan H, Itoh N, Yufika A, Winardi W, Keam S, Te H, *et al.* Coronavirus disease 2019 (COVID-19): A literature review. *J Infect Public Health* 2020;13:667-73.
 5. Chen Z, Fan H, Cai J, Li Y, Wu B, Hou Y, *et al.* High-resolution computed tomography manifestations of COVID-19 infections in patients of different ages. *Eur J Radiol* 2020;126:108972.
 6. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, *et al.* Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. *Radiology* 2020;296:32-40.
 7. Dai H, Zhang X, Xia J, Zhang T, Shang Y, Huang R, *et al.* High-resolution chest CT features and clinical characteristics of patients infected with COVID-19 in Jiangsu, China. *Int J Infect Dis* 2020;95:106-12.
 8. Minaee S, Kafieh R, Sonka M, Yazdani S, Jamalipour Soufi G. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Med Image Anal* 2020;65:101794.
 9. Hassantabar S, Ahmadi M, Sharifi A. Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches. *Chaos Solitons Fractals* 2020;140:110170.
 10. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Comput Methods Programs Biomed* 2020;196:105581.
 11. Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based Diagnostic of the coronavirus disease 2019 (COVID-19) from X-ray images. *Med Hypotheses* 2020;140:109761.
 12. Peng Y, Nagata MH. An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. *Chaos Solitons Fractals* 2020;139:110055.
 13. Organization W.H. Coronavirus Disease 2019 (COVID-19): Situation Report, 72. Organization W.H; 2020.
 14. Bharati S, Podder P, Mondal MR. Hybrid deep learning for detecting lung diseases from X-ray images. *Inform Med Unlocked* 2020;20:100391.