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

Informatics in Medicine Unlocked 30 (2022) 100935

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Statistical analysis of COVID-19 infection severity in lung lobes from chest CT

Mehdi Yousefzadeh [a,b,1], Mozhdeh Zolghadri [c,1], Masoud Hasanpour [d], Fatemeh Salimi [e], Ramezan Jafari [f], Mehran Vaziri Bozorg [g], Sara Haseli [h], Abolfazl Mahmoudi Aqeel Abadi [e], Shahrokh Naseri [c], Mohammadreza Ay [d,e], Mohammad-Reza Nazem-Zadeh [d,e,*]

[a] Department of Physics, Shahid Beheshti University, Tehran, Iran
[b] School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran
[c] Department of Medical Physics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[d] Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Tehran, Iran
[e] Medical Physics and Biomedical Engineering Department, Tehran University of Medical Sciences (TUMS), Tehran, Iran
[f] Department of Radiology and Health Research Center Baqiyatallah University of Medical Sciences, Tehran, Iran
[g] Department of Radiology, Kasra Hospital, Tehran, Iran
[h] Chronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Diseases (NRITLD), Shahid Beheshti University of Medical Sciences and Health Services, Tehran, Iran

ARTICLE INFO

Keywords:
Statistical analysis
COVID-19
CT scan
Lung lobes segmentation
Probability density function
Support vector machine

ABSTRACT

Detection of the COVID 19 virus is possible through the reverse transcription-polymerase chain reaction (RT-PCR) kits and computed tomography (CT) images of the lungs. Diagnosis via CT images provides a faster diagnosis than the RT-PCR method does. In addition to low false-negative rate, CT is also used for prognosis in determining the severity of the disease and the proposed treatment method. In this study, we estimated a probability density function (PDF) to examine the infections caused by the virus. We collected 232 chest CT of suspected patients and had them labeled by two radiologists in 6 classes, including a healthy class and 5 classes of different infection severity. To segment the lung lobes, we used a pre-trained U-Net model with an average Dice similarity coefficient (DSC) greater than 0.96. First, we extracted the PDF to grade the infection of each lobe and selected five specific thresholds as feature vectors. We then assigned this feature vector to a support vector machine (SVM) model and made the final prediction of the infection severity. Using the T-Test statistics, we calculated the *p*-value at different pixel thresholds and reported the significant differences in the pixel values. In most cases, the *p*-value was less than 0.05. Our developed model was developed on roughly labeled data without any manual segmentation, which estimated lung infection involvements with the area under the curve (AUC) in the range of [0.64, 0.87]. The introduced model can be used to generate a systematic automated report for individual patients infected by COVID-19.

1. Introduction

SARS-COV-2 infectious virus was identified by the World Health Organization and now is recognized as COVID -19 [1,2]. The current gold standard of its diagnosis is the reverse transcription-polymerase chain reaction (RT-PCR) test, which is a specific molecular diagnosis on respiratory specimens (throat swab/nasopharyngeal swab) [3]. Many studies have reported a higher sensitivity for evaluations based on computed tomography (CT) than RT-PCR test [4]. Although the CT scan of chest is more expensive and less accessible than PCR test for the patients with lung infections, a fast and accurate measurement of the infection severity has a prognostic value and direct influence on

* Corresponding author. Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Tehran, Iran.
E-mail addresses: yousefzadeh.meh@gmail.com (M. Yousefzadeh), mozhdeh.zolghadri80@gmail.com (M. Zolghadri), hasanpour.masoud@gmail.com (M. Hasanpour), salimifateme13@gmail.com (F. Salimi), rezajafari201089@gmail.com (R. Jafari), v_mehran@yahoo.com (M. Vaziri Bozorg), sarahaseli@gmail.com (S. Haseli), a_mahmoudia@alumnus.tums.ac.ir (A. Mahmoudi Aqeel Abadi), naserish@mums.ac.ir (S. Naseri), Mohammadreza_ay@tums.ac.ir (M. Ay), mnazemzadeh@tums.ac.ir (M.-R. Nazem-Zadeh).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.imu.2022.100935
Received 23 February 2022; Received in revised form 28 March 2022; Accepted 29 March 2022
Available online 1 April 2022

patient's treatment, which is not the case by a simple RT-PCR test [5–8].

Univariate and multivariate analyses have evaluated the role of CT in prognosis of the patients with COVID-19. The results showed that infection scores extracted from CT images are highly correlated with laboratory findings and disease severity. This can be useful in speeding up the diagnostic process in symptomatic cases [9]. COVID-19 related infections appear on CT images of the chest with various patterns, including ground glass opacity (GGO), consolidation, crazy paving, etc. [10]. Most lesions are distributed in the peripheral or sub-pleural area. In addition, the lower lobes are more frequently infected than the upper and middle lobes [11].

Segmentation of the lung lobes can provide useful information regarding the location of different infections; thus, a better diagnosis and prognosis can be obtained from the results of the segmented lobes [12–14].

Numerous research in prediction of COVID-19 has been carried out by the means of machine learning and deep learning [15,16]. Using deep feature maps of chest CT images, a system for detection of the infection severity in COVID-19 (For two different classes, severe vs. mild (has been introduced called COVIDC [17]. However, in most studies, the entire lung image was used to predict an infection severity globally, with no local prediction across the individual lung lobes [18].

In this study, we gathered a Chest CT dataset and labeled it with several classes of infection severity in each lobe. Utilizing a workflow of machine learning methods, we aimed to predict the infection severity of each lobe. In summary, the main advantages and novelties of this study are as follows:

- Internal dataset was collected and the severity of lung infection in six classification classes was prepared.
- A U-Net model with trained weight was used for lobe segmentation.
- By segmenting the lung lobes in CT images and using the probability density function (PDF) in five specific intensities, a feature vector was obtained for each lobe with no dependence on the size or number of voxels in each of the five lung lobes.
- The resulting vector was given to a support vector machine model to predict how severely each lung lobe was affected by a lung infection.
- Finally, the area under the curve (AUC) score of each of the severity classes of lung involvement infection was reported.

In the following part of this article, we described the data collection and label assigning. Then, we explained the statistical tools for estimating the PDF and the proposed machine learning model. Subsequently, we presented the results of the PDF estimation and the SVM classification, followed by discussing the results and summarizing the findings. We ultimately addressed the limitations and proposed future works.

2. Methods and materials

In this study, we used 186 CT volumes to train the proposed machine learning prediction model and 46 CT volumes to evaluate it. To segment the lung lobes, a deep learning model was used [19], which is explained in the following section. Next, we estimated a PDF of pixel intensities for each lobe, by analyzing which, we obtain a feature vector for each lung lobe to predict the severity of infection with the help of a SVM model.

2.1. Estimate the severity of infection in each lung lobe

According to the standard systematic report on Radiology Assistant [20] and Rad2share, the percentage of infection was determined by the number of points 0 to 5 (Table 1). According to the experience and expertise in interpreting CT images of the lungs and mastering the sectional anatomy of different areas of the chest, the radiologists visually and approximately estimated the percentage of involvement and the type of infection (especially GGO and consolidation) in each lobe as a standard systematic report.

Table 1
Percentage of infection equivalent to the points of infection.

Number of points involved in infection	Percentage of infection
0 points	0%
1 point	<5%
2 points	5–25%
3 points	25–50%
4 points	50–75%
5 points	75–100%

2.2. Datasets

CT dataset of 232 patients (female: male, 97: 135; average age, 46 years) were obtained from Ghiassi Hospital in Tehran (Alexion 16-slice CT, TOSHIBA Medical Systems). The images were reported by a radiologist and the patient identity was removed from the images to protect the patient's privacy. Also, an official license was obtained from the supervisor and the head of the department to use the hospital data in research. Moreover, an ethics license has been obtained from the ethics committee of the university for the data acquisition and management.

Using a pre-trained U-Net method, we segmented the lung lobes in the CT dataset of 232 vol and 30,157 slices, then the severity of infection was labeled in each lung lobe. Furthermore, a manual segmentation for the lung lobes was performed (only for 35 CT volumes as it is a labor-intensive task) to validate the performance of the pre-trained U-Net method used in segmenting the lung lobes on our data (We did not perform any retraining for this segmentation method as the pre-trained model performed quite well on our data). Finally, 80% of the data (232 CT scans dataset with the severity labels) was used for training, and the remaining 20% was used to evaluate the machine learning prediction model. The number of lobes with different levels of infection is presented in Table 2.

2.3. CT image acquisition

Chest volumes had been obtained in the supine position by the Toshiba Alexion 16-slice CT scanner in low dose exposure condition (MA minimum) with the image of 512 × 512 and the cut thickness of 2 mm. Thick metal objects were removed from the chest zone to prevent metal artifacts in CT images. The collected CT volumes contain an average of 130 slices.

2.4. Lung lobes segmentation

We used a pre-trained model presented in Ref. [19] to segment the lung lobes. This model utilized an R231 U-Net to categorize all the pixels identified inside the lung in one of the five lobes. CIP (Chest Imaging Platform) is an open-source lung segmentation tool based on threshold and morphological operations. The Dice similarity coefficient (DSC) of this network is reported 0.97 ± 0.05 for standard approaches. However, to validate the performance of lobe segmentation on our dataset, we manually segmented the lung lobes in a number of images and evaluated the segmentation results of the model by calculating the DSC.

Table 2
Number of lobes involved in COVID-19 infections in our dataset.

Number of points involved in infection	Number of lobes
0 points	361
1 point	254
2 points	327
3 points	149
4 points	56
5 points	13

For semi-automatic segmentation of the lung lobes, imaging volumes of 35 patients in DICOM format were fed to the 3D Slicer software[2] from the CIP module (Chest Imaging Platform) into the section "Interactive Lobe Segmentation" and the desired images were selected, then "Label Map volume" was created for the image collection. Moreover, in the section "Fisser Volume section", a file was created for the set of imported images. A Gaussian filter was activated for smoothing the image edges. In the option section, the image dimensions and the work speed were customized by a skilled expert. Sagittal images were selected, and the lung lobes were segmented with assigning five different colors to each of the five lung lobes. The segmented images were inspected by two radiologists for accuracy and manually edited by two skilled radiography residents. To edit the segmented images in the 3D Slicer software, the images were entered in the "Segment Editor" and the editing was performed. At the end, the labeled images from the "Save" module were archived in a desired address with the ".nrrd". image format.

2.5. Pre-processing

In the first stage, data preparation was performed to unify the data and incomplete or distorted data were deleted (e.g., the volumes with no breathing confinement). Image pixel intensities were mapped to the range of [−1024, 150]. This was performed by data clipping and windowing functions (i.e., with setting the window level and window width values to 1500 and -600, respectively).

2.6. Statistical analysis and machine learning methods

To manage errors related to the data scale, type, and size, an efficient tool should not depend on the lobe size. To address this, we performed a one-point statistical analysis for generating a PDF of intensities of the lobe infections. If there is a bright area associated with an infection, the PDF of the image pixel intensities gives a great number to reflect a high level of infection severity in the lobe of interest. In addition, for each level of infection involvement (0–5), its PDF of image intensities was compared with the PDF of one incremental unit greater level of infection (1–6), which was also separately performed for five lung lobes altogether. The PDF was calculated for 50 different points with equal distances in the range of [−1024, 150] and the *p*-values for these points in each graph were calculated.

We used a SVM model for the aim of data classification. In the proposed SVM model, we used a nonlinear radial basis function (RBF) kernel. Also, due to the non-uniformity of the data in different classes in the SVM model, we used a weight for each class so that the non-uniformity of the samples does not bias at the discretion of a particular class. Finally, we used the output of the PDF estimated for each lobe at five specific thresholds and constructed a feature vector with a length of five so that the SVM model can detect the severity of infection in each class. Furthermore, the feature vector was standardized in such a way that for each column of these five features, the average value was zero and the standard deviation was normalized to one for different samples. In the SVM model, because the total amount of data in the 5-point class was small and it was not possible to train and evaluate this class adequately, we merged it with the 4-point class. Therefore, the new 4-point class aimed to diagnose lobes with 50–100% involvement of a lobe in the CODID - 19 -related infection.

We used the package "scipy.stats.ttest_ind" in Python 3.7.9 to calculate the *t*-test statistics. In addition, to draw diagrams and results, the package "matplotlib" was used. The "scikit-learn package" was also used to train the SVM model [21].

[2] https://www.slicer.org, version 4.11.0.

3. Result

3.1. Lung lobes segmentation

We fed the image of 232 cases reported by the radiologist to the pre-trained deep learning U-Net model for lobe segmentation. To estimate the severity of infection in each lobe, the images were tagged by two radiology residents. By analyzing the PDF, we obtained a feature vector for each lobe. Finally, by utilizing a SVM model, we predicted the severity of infection in all five lung lobes.

The pixels of each lobe were extracted from 3D CT volumes of the lung [19]. The DSC was calculated for 35 lung CT volumes with the lobes labeled by a semi-automatic method subsequently edited manually by a radiologist (reported in Table 3 for each lobe). The average of all DSC was greater than 0.96. Note that the DSC of the middle lobe was smaller than the other lobes due to the difficulty in distinguishing the middle lobe from others by the radiologist and also the deep learning model. The output of the lobe segmentation model for a sample lung CT image is shown in Fig. 1.

3.2. Statistical analysis results

The PDF of pixels intensities at six different degrees of infection (0–5 points) are plotted in Fig. 2, which manifests a significant difference between them. Fig. 3 shows the *p*-value of T-Test statistics between different categories of lung involvement for all lobes together (The dashed black lines show the thresholds that we used in the SVM model). Fig. 4 suggests a significant difference between PDF of infection intensities for five lung lobes separately, demonstrating that our proposed method can provide adequate differentiation between different configurations of the infections per lobe. As we see in Fig. 4, there was a significant difference between the two groups with the degree of infection 4 and 5, because some of each of the five lobes did not happen to undergo a level 5 infection, but some of the other lobes did. The best threshold values associated with the lowest p-values (i.e., the most significant differences) of each comparison are given in Table 4. In most cases, the confidence intervals for significant differences were more than 95%. Also, with all infection categories in each lobe, the p-value was 0.21 (confidence interval of 67%), but when considering all infection categories and all lobes together, the significant difference became much more apparent. Finally, the difference between the intensities in two right upper and left lower lobes was less significant than the other three lobes.

3.3. SVM model for data classification

We trained the SVM model expressed in the method section and the five thresholds selected with the most significant differences at different intervals of pixel intensity (Fig. 3). The model was trained on 80% of the data and evaluated on 20% of the rest (test set). In addition, due to the small number of instances, the 5-point class was merged with the 4-point one. For the output of the test set, a diagram of the Receiver Operating Characteristic (ROC) curve is depicted in Fig. 5., along with the AUC (of the ROC) values of the final accuracy report in each case. As in Fig. 5, we perhaps see the highest accuracy for the 0-point versus {4,5} -point classes. The lowest results were achieved for the 1-point and the 2-point

Table 3
Average (mean ± standard error) of Dice similarity coefficient (DSC) of the segmentation model for each lung lobe applied on the 35 CT volumes.

Lung Lobes	Right Upper	Right Middle	Right Lower	Left Upper	Left Lower
Lobe Segmentation DSC (mean ± standard error)	0.969 ± 0.036	0.921 ± 0.072	0.971 ± 0.050	0.982 ± 0.015	0.981 ± 0.018

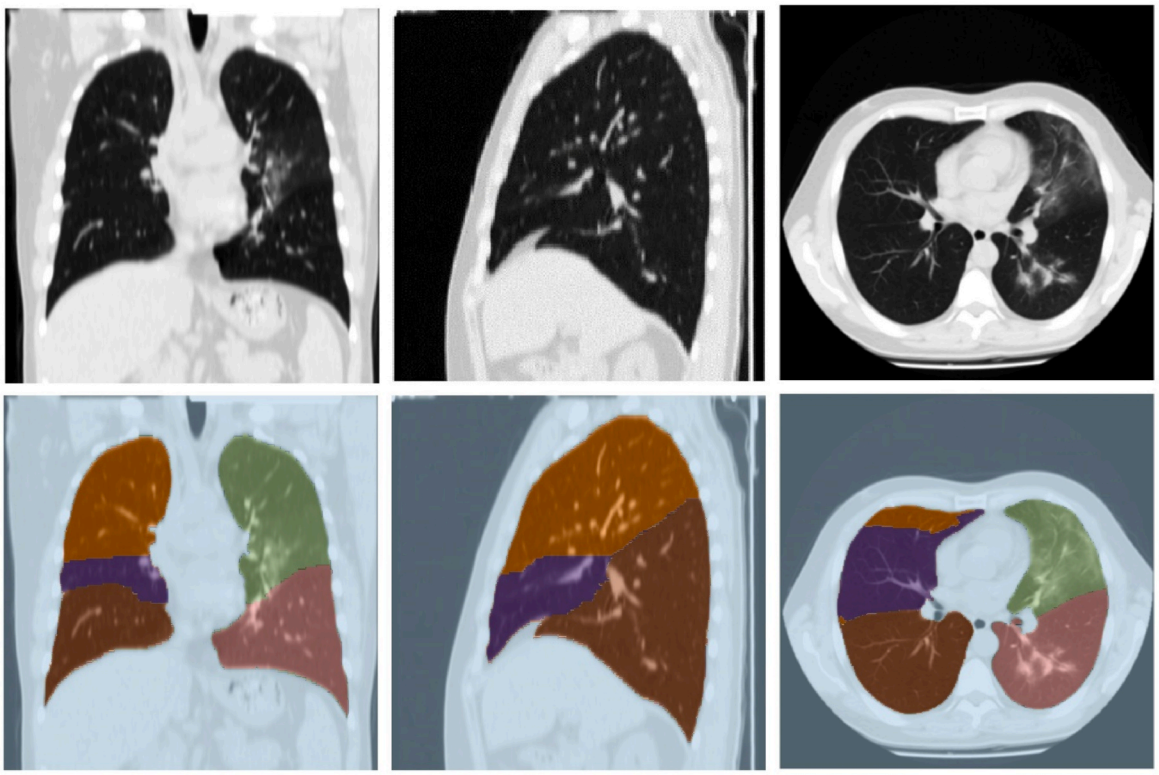

Fig. 1. Lobe segmentation in three different views. The top row shows the CT images and bottom row represents the five segmented lobes.

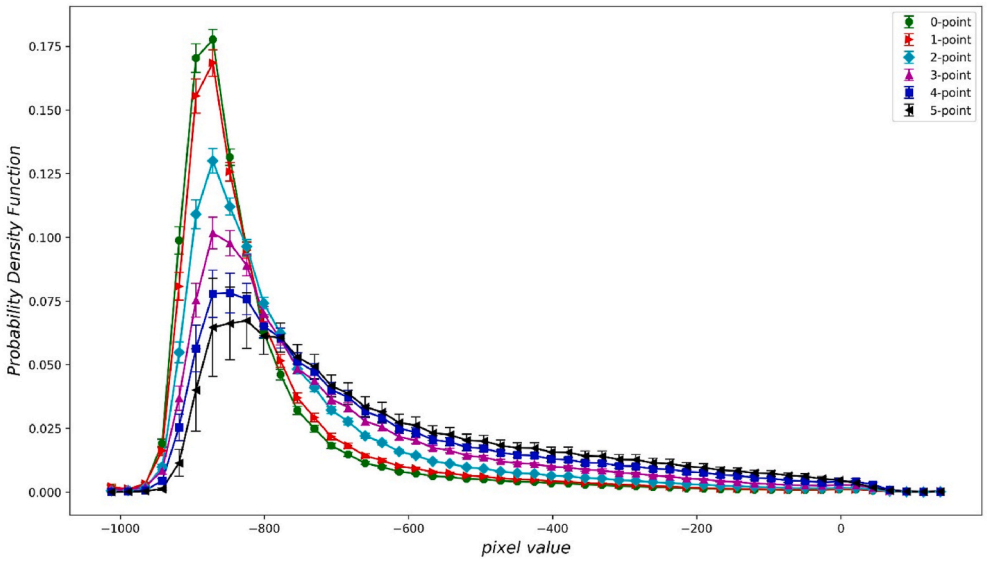

Fig. 2. Probability density function of pixels intensities at different levels of involvement (degrees of infection). The bar at each point represents the standard deviation of the mean for pixels intensities.

classes, where their PDF assimilated the PDF of immediate different levels of infection with a lower probability of separation.

4. Discussion

Due to a high prevalence of the recently epidemic COVID-19 and the occurrence of infectious symptoms on the lung CT images, the use of diagnostic tools based on deep machine learning approaches has an important and prominent role in the field of diagnostic imaging [22]. In this study, we used a deep learning model with U-Net architecture to segment the lung lobes from the source [19], which showed a promising result quite similar to manual labeling.

Most statistical studies to evaluate the severity of infection seek to find a link between CT findings and clinical and laboratory attributes, but due to small available populations, generalization of the results is often limited. In a study in 2020, using, Pearson's correlation test and multivariate analysis curves have yielded a score of relationship between patients' clinical findings, examining tests, and chest CT images. Promisingly, CT images were found capable to predict the mortality from the disease, which could aid specifying the therapeutic approaches

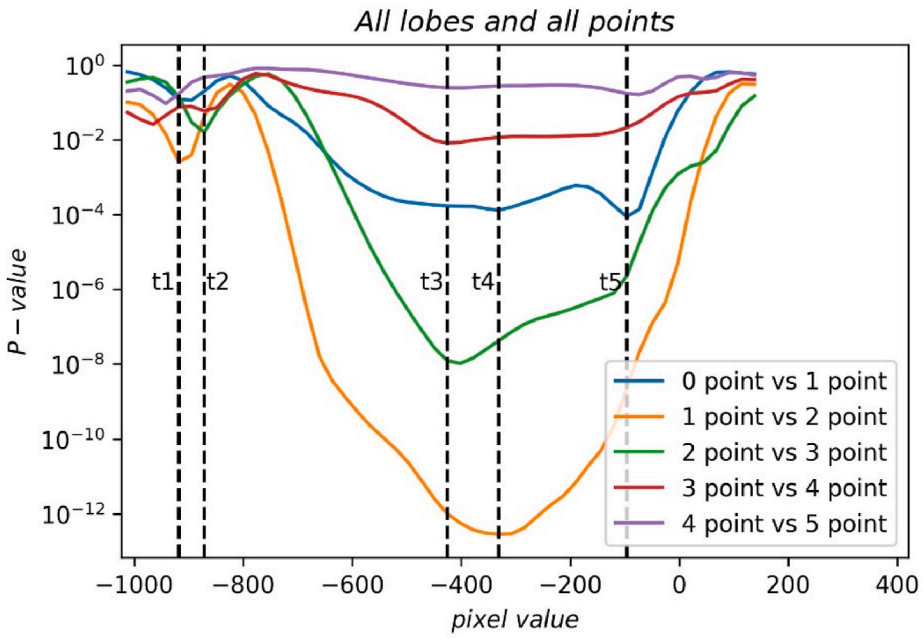

Fig. 3. The *p*-value between different categories of lung involvement for all lobes together. The dashed black lines show the thresholds that we used in the SVM model.

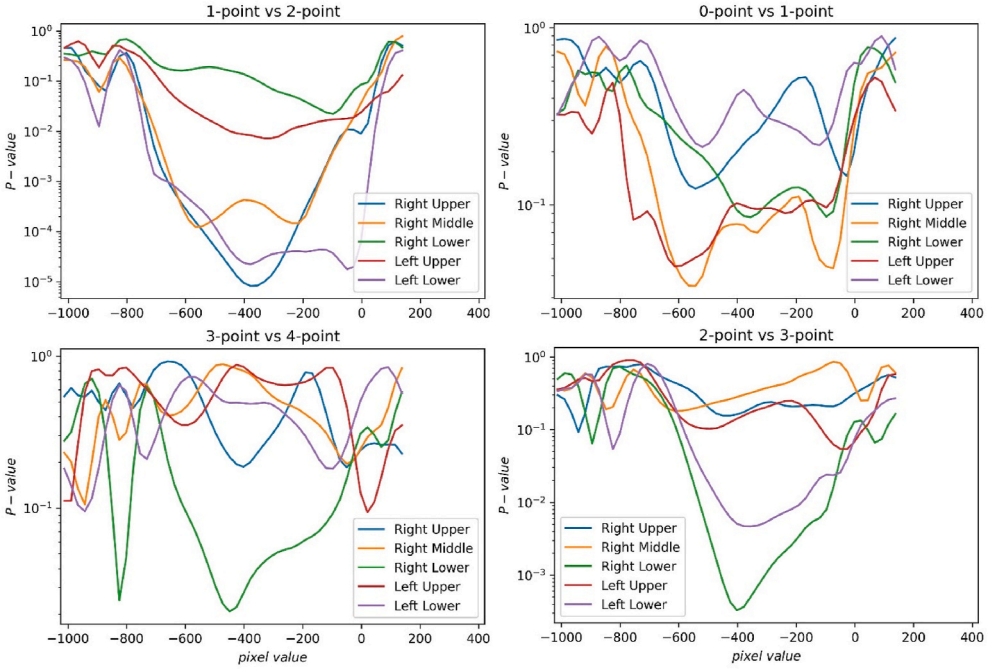

Fig. 4. Significant differences between PDF of infection intensities for different lobes, top left: between two categories with infection degrees of 0 and 1, top right: between two categories with infection degrees of 1 and 2, bottom left: between two categories with infection degrees of 2 and 3, and bottom right: between two categories with infection degrees of 3 and 4.

Table 4
Minimum *p*-values for each degree of infection compared with one level higher for five individual lobes and for all lobes together.

p-value	Right Upper Lobe	Right Middle Lobe	Right Lower Lobe	Left Upper Lobe	Left Lower Lobe	All Lobes
0 vs. 1 point	0.12	**0.03**	0.09	**0.04**	0.21	$\mathbf{10^{-4}}$
1 vs. 2 points	$\mathbf{10^{-5}}$	**0.0001**	**0.0221**	**0.007**	$\mathbf{10^{-5}}$	$\mathbf{10^{-13}}$
2 vs. 3 points	0.092	0.18	**0.0003**	0.054	**0.005**	$\mathbf{10^{-8}}$
3 vs. 4 points	0.185	0.105	**0.021**	0.094	0.095	**0.008**
4 vs. 5 points	-	-	-	-	-	0.093

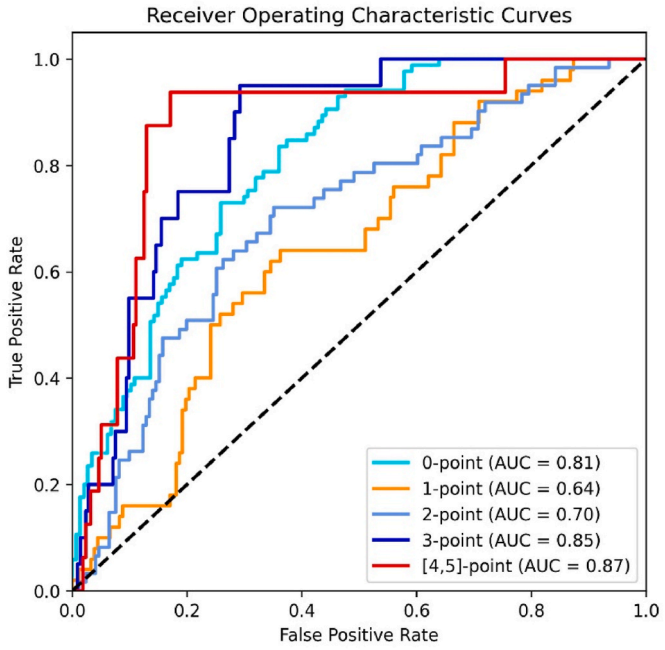

Fig. 5. ROC curve for different five intensity classes. The 4-point and 5-point classes were put in a same category.

for the patients [23]. Previous research has focused on the severity of total lung infection, as well as predicting the severity of disease growth and death from CT images. Hu, Yiqi, et al. showed that Chest CT findings in patients who died of COVID-19 worsened with a moderate positive correlation between CT severity and inflammatory factors associated with leucocytes, neutrophils, and IL-2R [24]. In another work, Huang, Lu, et al. [25] showed that the quantification of lung opacification in COVID-19 measured on Chest CT by using a commercially available deep learning-based tool was significantly different among groups with different clinical severity. In this work, we focused on more details and tried to predict the severity of infection in each lobe without segmentation labeling. Bin Li et al. [26] presented a method for segmentation of ground-glass opacity (GGO) and juxta-vascular nodules that showed weak edges with inhomogeneous intensity characteristic. They used an adaptive local area energy model with a PDF based on a similarity distance and a dynamic multi-feature clustering method. Their proposed method outperformed some previous segmentation methods.

Numerous research has been done in the field of machine learning and deep learning in COVID-19 prediction. Yousefzadeh et al. [15] presented a deep learning framework called AI-CORONA as a radiologist's assistant to diagnose COVID-19. Using an EfficientNetB3-based feature extractor, their proposed method outperforms the state-of-the-art models and experts. In another study, Mukul Singh et al. [16] used a deep learning model with architecture VGG16 (Visual Geometry Group from Oxford) to extract features from CT images. Feature selection was performed using principal component analysis (PCA). For the final classification, four classifiers (i.e., deep convolutional neural network (CNN), Extreme Learning Machine (ELM), Consecutive Online ELM, and the bagging ensemble with SVM) were compared, where the latter achieved the best classification performance (with an accuracy of 95.7%, precision of 95.8%, AUC of 0.958, and an F1 score of 95.3%). In a related study, Shan, Fei, et al. [18] developed an in-depth learning model that first segmented the entire lung infection and predicted the severity of the whole lung infection. The accuracy of predicting the severity of infection in the best case of the methods studied in this work was reported to be 73.4%. In a study, Kelei He and his colleagues presented a multi-task multi-instance deep neural network (called M^2 U-Net) for assessment of COVID-19 by performing segmentation of the lung lobes. Employing a hierarchical multi-instance learning strategy, a lung lobe segmentation has been carried out, including a decoder module with a DSC of 0.785 [27]. Filipe T.et al., used a fully regularized V-Net (FRV- Net), where a 3D CNN was trained end-to-end to segment the five lung lobes. The results of this work were used in computer aided diagnosis (CAD) systems related to pulmonary diseases with an average DSC of 0.93 per lobe [28]. Using deep feature maps of chest CT images, Wajid Arshad Abbasi et al. [17] presented a COVID-19 severity prediction system for two different classes (severe vs. mild) called COVIDC (COVID-19 detection using CT scan). In this approach, they utilized various pre-trained CNN-based models in ImageNet to obtain feature maps and surface learning algorithms such as SVM to obtain the final training model. They reached an overall accuracy of 86% along with a generalization performance with F1 score of 0.94 and an AUC (of the ROC) of 0.98. In a research work, a U-Net neural network was used for the segmentation of five lung lobes with an average DSC greater than 0.96.

The aim of this study was to determine the severity of infection according to the PDF of the intensity of infections, which was studied both in general for all lobes and also separately for each lobe. We used statistical tools based on the PDF, and the features related to the severity of infection were extracted from 232 CT volumes in six classes with different levels of infection. The advantage of this method was its independency on the size of the lobes and its management of errors related to the data scale and size. We observed that the pixel values above −800 in normal lungs were associated with pulmonary arteries. With increasing lung involvement in each lobe, it was observed that the probability of distribution in an interval of [−1024, −800] decreased. Due to the clear areas created by the infection, this distribution increased in an interval of [−800,0]. By calculating the amount of *p-value* in six levels of lung infection, we concluded that there was a significant difference in pixels with a value of −900 due to lesser air in the lung lobes that involved with an infection and the pixel intensities greater than −800 due to infections with different densities (the higher the infection density, the brighter the image in the infected area), would show more significant differences. It should also be noted that a smaller number of samples in each category would lead to a greater standard deviation and weaker power of the method to detect significant differences. For example, it was difficult to compare the two categories of 4 and 5 point-infection levels, due to a small number of samples in our data. Using the PDF statistical tool, we obtained a feature vector for each lobe, which was used as an input of the SVM machine learning model. Different models of machine learning were examined and the SVM method with RBF kernel was found the most differentiating model. We also investigated various statistical tools including the two-point correlation function that were not effective, and for this reason they were not reported here.

This study had some limitations. The main limitation was the manpower error in data labeling which would cause some inaccuracies in analysis. Also, we did not have access to an lung segmentation label to utilize deep learning models, and we had to perform some manual segmentations for a few number of cases, in order to merely evaluate the performance of the used segmentation method on our data. Finally, using deep learning models directly on image data is also associated with many challenges. For example, the size of the lobes was very different, and also deep learning models demand a bigger data to train, in order to prevent an overfitting situation.

Several tasks can be suggested as future works. The first is to collect image segmentation labels to estimate the extent of infection more accurately at each lobe. The next suggestion is to use a wider range of one-point and multi-point statistical tools. Finally, by collecting more data, we can utilize deep learning models instead of combining statistical methods and machine learning approaches.

5. Conclusion

Segmentation of the lung lobes helps to determine the exact location of the infection caused by COVID-19. In this study, we used a pre-trained artificial neural network. Combining the estimated probability density function of the image intensities and a SVM model, we examined the pixels of CT images for the degree of infection at five different levels. The results of this study can be used to generate a systematic automated report for COVID-19.

Data availability statement

The CT dataset has been collected in Ghiassi Hospital, Tehran, Iran. It cannot be shared publicly because of ethical restrictions and sensitive human study participant data. The anonymized (non-personally identifiable) data is, however, available from the Institutional Data Access through the Ethics Committee of Ghiassi Hospital for the researchers who meet the criteria to access the confidential data (contact the corresponding author at mnazemzadeh@tums.ac.ir).

Declaration of competing interest

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

Acknowledgment

This project was partially funded by the Tehran University of Medical Sciences (TUMS), the International Campus (IC) Research Affairs. The CT scan data has been collected in Ghiassi Hospital, Tehran, Iran. It cannot be shared publicly because of ethical restrictions. The data is, however, available from the Institutional Data Access/Ethics Committee of Ghiassi Hospital (contact via mnazemzadeh@tums.ac.ir) for researchers who meet the criteria for access to confidential data. The MDH cohort is also available upon request to the corresponding author's email address.

References

[1] Jiang X, Rayner S, Luo MH. Does SARS-CoV-2 has a longer incubation period than SARS and MERS? J Med Virol 2020 May;92(5):476–8.
[2] Lorusso A, Calistri P, Petrini A, Savini G, Decaro N. Novel coronavirus (SARS-CoV-2) epidemic: a veterinary perspective. Vet Ital 2020 Feb 12;56(1):5–10.
[3] Centers for Disease Control and Prevention. Interim guidelines for collecting, handling, and testing clinical specimens from persons under investigation (PUIs) for coronavirus disease. 2019. COVID-19).
[4] Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, Xia L. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 2020 Aug;296(2):E32–40.
[5] Abbasi-Oshaghi Ebrahim, et al. Diagnosis and treatment of coronavirus disease 2019 (COVID-19): laboratory, PCR, and chest CT imaging findings. Int J Surg 2020; 79:143–53.
[6] Axiaq Ariana, et al. The role of computed tomography scan in the diagnosis of COVID-19 pneumonia. Curr Opin Pulm Med 2021;27(3):163–8.
[7] Fang Yicheng, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 2020;296(2):E115–7.
[8] Khatami Fatemeh, et al. A meta-analysis of accuracy and sensitivity of chest CT and RT-PCR in COVID-19 diagnosis. Sci Rep 2020;10(1):1–12.
[9] Francone M, Iafrate F, Masci GM, Coco S, Cilia F, Manganaro L, Panebianco V, Andreoli C, Colaiacomo MC, Zingaropoli MA, Ciardi MR. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. Eur Radiol 2020 Dec;30(12):6808–17.
[10] Xu YH, Dong J-H, An W-M, et al. Clinical and computed tomographic imaging features of novel coronavirus pneumonia caused by SARS-CoV-2. J Infect 2020;80 (4):394–400.
[11] Pan Y, Guan H, Zhou S, Wang Y, Li Q, Zhu T, Hu Q, Xia L. Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. Eur Radiol 2020 Jun;30(6):3306–9.
[12] Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of COVID-19 in CT images with deep learning. arXiv preprint arXiv:2003.04655 2020 Mar 10.
[13] Jin S, Wang B, Xu H, Luo C, Wei L, Zhao W, et al. AI-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system in four weeks. medRxiv 2020 Jan 1.
[14] Qi X, Jiang Z, Yu Q, Shao C, Zhang H, Yue H, et al. Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. medRxiv 2020 Jan 1.
[15] Yousefzadeh M, Esfahanian P, Movahed SM, Gorgin S, Rahmati D, Abedini A, Nadji SA, Haseli S, Bakhshayesh Karam M, Kiani A, Hoseinyazdi M. ai-corona: radiologist-assistant deep learning framework for COVID-19 diagnosis in chest CT scans. PLoS One 2021 May 7;16(5):e0250952.
[16] Singh M, Bansal S, Ahuja S, Dubey RK, Panigrahi BK, Dey N. Transfer learning–based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data. Med Biol Eng Comput 2021 Apr;59(4):825–39.
[17] Abbasi WA, Abbas SA, Andleeb S, Ul Islam G, Ajaz SA, Arshad K, Khalil S, Anjam A, Ilyas K, Saleem M, Chughtai J. COVIDC: an expert system to diag-nose COVID-19 and predict its severity using chest CT scans: application in ra-diology. Inf Med Unlocked 2021 Jan 1;23:100540.
[18] Shan Fei, et al. Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. Med Phys 2021;48(4):1633–45.
[19] Hofmanninger Johannes, et al. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp 2020;4(1):1–13.
[20] https://radiologyassistant.nl/chest/covid-19/covid19-imaging-findings.
[21] Kramer O. Scikit-learn. InMachine learning for evolution strategies. Cham: Springer; 2016. p. 45–53.
[22] Oulefki A, Agaian S, Trongtirakul T, Laouar AK. Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. Pattern Recogn 2021 Jun 1;114:107747.
[23] Francone M, Iafrate F, Masci GM, Coco S, Cilia F, Manganaro L, Panebianco V, Andreoli C, Colaiacomo MC, Zingaropoli MA, Ciardi MR. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. Eur Radiol 2020 Dec;30(12):6808–17.
[24] Hu Yiqi, et al. Chest CT findings related to mortality of patients with COVID-19: a retrospective case-series study. PLoS One 2020;15(8):e0237302.
[25] Huang Lu, et al. Serial quantitative chest CT assessment of COVID-19: a deep learning approach. Radiology: Cardiothorac Imag 2020;2(2):e200075.
[26] Li B, Chen Q, Peng G, Guo Y, Chen K, Tian L, Ou S, Wang L. Segmentation of pulmonary nodules using adaptive local region energy with probability density function-based similarity distance and multi-features clustering. Biomed Eng Online 2016 Dec;15(1):1–28.
[27] He K, Zhao W, Xie X, Ji W, Liu M, Tang Z, Shi Y, Shi F, Gao Y, Liu J, Zhang J. Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images. Pattern Recogn 2021 May 1;113:107828.
[28] Ferreira FT, Sousa P, Galdran A, Sousa MR, Campilho A. In: End-to-end supervised lung lobe segmentation. In2018 International Joint conference on neural networks (IJCNN). IEEE; 2018. p. 1–8.