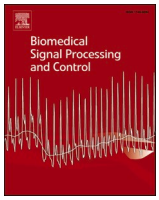
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Biomedical Signal Processing and Control 85 (2023) 104905

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



Semi-supervised COVID-19 volumetric pulmonary lesion estimation on CT images using probabilistic active contour and CNN segmentation

Diomar Enrique Rodriguez-Obregon [a], Aldo Rodrigo Mejia-Rodriguez [a,*], Leopoldo Cendejas-Zaragoza [b,c], Juan Gutiérrez Mejía [d], Edgar Román Arce-Santana [a], Sonia Charleston-Villalobos [e], Tomas Aljama-Corrales [e], Alejandro Gabutti [f], Alejandro Santos-Díaz [b,g,*]

[a] Faculty of Sciences, Universidad Autónoma de San Luis Potosí, S.L.P., Mexico
[b] Tecnologico de Monterrey, School of Engineering and Sciences, Mexico City, Mexico
[c] Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico
[d] Tecnologico de Monterrey, School of Medicine and Health Sciences, Mexico City, Mexico
[e] Universidad Autónoma Metropolitana-Iztapalapa, Mexico City, Mexico
[f] Department of Radiology and Imaging, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico
[g] Tecnologico de Monterrey, School of Medicine and Health Sciences, Monterrey, Mexico

ARTICLE INFO

Keywords:
COVID-19
Computed tomography
Volumetric lesion segmentation
Semi-supervised segmentation
Active contours

ABSTRACT

Purpose: A semi-supervised two-step methodology is proposed to obtain a volumetric estimation of COVID-19-related lesions on Computed Tomography (CT) images.
Methods: First, damaged tissue was segmented from CT images using a probabilistic active contours approach. Second, lung parenchyma was extracted using a previously trained U-Net. Finally, volumetric estimation of COVID-19 lesions was calculated considering the lung parenchyma masks.
Our approach was validated using a publicly available dataset containing 20 CT COVID-19 images previously labeled and manually segmented. Then, it was applied to 295 COVID-19 patients CT scans admitted to an intensive care unit. We compared the lesion estimation between deceased and survived patients for high and low-resolution images.
Results: A comparable median Dice similarity coefficient of 0.66 for the 20 validation images was achieved. For the 295 images dataset, results show a significant difference in lesion percentages between deceased and survived patients, with a p-value of 9.1×10^{-4} in low-resolution and 5.1×10^{-5} in high-resolution images. Furthermore, the difference in lesion percentages between high and low-resolution images was 10 % on average.
Conclusion: The proposed approach could help estimate the lesion size caused by COVID-19 in CT images and may be considered an alternative to getting a volumetric segmentation for this novel disease without the requirement of large amounts of COVID-19 labeled data to train an artificial intelligence algorithm. The low variation between the estimated percentage of lesions in high and low-resolution CT images suggests that the proposed approach is robust, and it may provide valuable information to differentiate between survived and deceased patients.

1. Introduction

In November 2019, an outbreak of the novel coronavirus SARS-CoV-2 was reported in the province of Wuhan, China. In January 2020, the World Health Organization (WHO) declared an international emergency. As of May 2022, the COVID-19 pandemic has caused more than 6 million deaths worldwide [1].

The 2019 coronavirus disease (COVID-19) pandemic has triggered profound changes in healthcare practices and policies, beginning with triage management, diagnostic resources application, and patient destination according to positive test results. At the pandemic's beginning, complete hospital resources were dedicated to COVID-19 patients, so diagnosis became a relevant step in resource management.

Many of these challenges required developing solutions focused on

* Corresponding authors at: Universidad Autónoma de San Luis Potosí, Facultad de Ciencias Campus Pedregal, Av. Chapultepec 1570, Privadas del Pedregal, C.P. 78295, San Luis Potosí, S.L.P., Mexico (A.R. Mejia-Rodriguez), Calle del Puente 222, Ejidos de Huipulco, Tlalpan, CDMX, C.P. 14380, Mexico (A. Santos Díaz).
E-mail addresses: aldo.mejia@uaslp.mx (A.R. Mejia-Rodriguez), alejandro.santos@tec.mx (A. Santos-Díaz).

https://doi.org/10.1016/j.bspc.2023.104905
Received 4 August 2022; Received in revised form 11 March 2023; Accepted 18 March 2023
Available online 22 March 2023

this new reality. For example, the gold-standard test for diagnosing COVID-19, real-time polymerase chain reaction (RT-PCR), was only available in some places. When it was, results may be delayed hours or even days. The test is sensitive to variations and high false negative results can be obtained depending on the protocol used, the technique to obtain the sample, handling of it, the viral load of the patient, the time from the begging of symptoms and arrival to healthcare facilities, and intrinsic patient's health [2].

In another effort, the Dutch Radiological Society standardized a COVID-19 reporting and data system (CO-Rads) aiming to increase the options for early diagnosis, providing information based on non-enhanced Computed Tomography (CT) images where levels of suspicion increased from Co-Rads 1 up to Co-Rads 5 [3]. In parallel, there was significant interest in positive discrimination with CT of higher or lower survival chance to help relatives to make joint decisions with the healthcare team, defining scarce resource allocation and enhancing lung protective measures approach to patients in higher risk even if this would require more extended muscle relaxant use and higher sedative doses well-known risk factors for the need of vasoactive medications.

In addition, hospitals faced the necessity to balance the lack of available experts in reading COVID-19 CT scans, the large number of patients arriving to be assessed, and the interest in decreasing radiation exposure in all patients while still being able to define early the survival probabilities. These conditions exposed the pursuit of automatic tools for patient assessment, many of which are based on AI algorithms.

High-resolution computed tomography (HRCT) imaging has been recognized as an essential tool for diagnosing this novel disease [4,5]. This imaging technique allows the identification of different stages and phenotypes of COVID-19, and it is helpful in accurately evaluating the patient's evolution [6,7]. Furthermore, HRCT is often used to assess the complications in hospitalized patients and evaluate the response to different treatments. HRCT, in combination with AI algorithms, can be used as an automated diagnosis system to assess medical treatments and could be relevant as a tool for radiological research [8].

Most studies on AI related to COVID-19 CT images aim to classify positive and negative cases as automatic diagnosis systems [9,10,11,12,13]. Beyond diagnosis, lesion segmentation is essential in COVID-19 quantifying and correctly assessing the disease. Machine learning algorithms have been applied to CT images of COVID-19 patients to solve the lesion segmentation problem. Wu et al. [14] present a joint classification and segmentation diagnosis system; they used a fine-grained model for segmentation and achieved a Dice similarity coefficient (DSC) value of 0.78 on their dataset built by axial CT slices. On the other hand, authors in [15] proposed a weakly supervised method based on point annotations to segment disease on CT images and tested it on three different datasets, one of them is the COVID-19 CT Lung and Infection Segmentation Dataset (LISD) [16], which is worth to mention since it is used to test our approach for validation. Additionally, Liu et al. [17] proposed another weakly supervised algorithm based on scribble annotations, and they tested it on their dataset [10,11,18]. In [19], a pair of networks called teacher-student framework is trained using Fourier transform augmentation to build a training dataset from lung cancer data without collecting COVID-19 images for training.

Other recent studies, like [20], performed a segmentation task for regions of interest and then use an Artificial Bee Colony optimized Neural Network to classify these regions as COVID-19 or non-COVID-19 regions. They only presented the segmentation performance of two images with a Dice similarity coefficient (DSC) of 0.91 and 0.90 and a Jaccard coefficient of 0.88 and 0.87. Their main contribution was the classification network, where they obtained a classification accuracy of 92.3 %.

In [21], they presented a network with a pixel-wise sparse graph reasoning module which is inserted on a U-net configuration between the encoder and decoder paths. They tested their network on three datasets obtaining DSC values of 0.84, 0.78, and 0.67 on each dataset. It is important to mention that they compared their model with other models on the state of the art, and in most cases, they achieved the best result using the new proposed model with the pixel-wise sparse graph reasoning module.

To deal with the challenging task of obtaining large amounts of annotations for training, Weiyi Xi, et al. proposed in [22] a network training based on dense regression activation maps. These maps were generated by a neural network trained to estimate a per-lobe lesion percentage, then they refined their results using an attention neural network module, achieving a DSC coefficient of 0.7.

All the studies above used convolutional neural network (CNN) architectures to obtain a classification of the images or a segmentation mask. However, it is worth pointing out that such approaches demand large amounts of labeled data to train and evaluate the models. The need for more data to train and test CNN models is grounded on the challenge of generating gold-standard segmentation. Unfortunately, more publicly available datasets are needed to accomplish these tasks, even more for a particular new disease such as COVID-19.

In medical imaging, gold standards are usually drawn manually by expert radiologists. Furthermore, in many cases, multiple radiologists perform the task to account for inter-subject bias. This task is very time-consuming since each HRCT study may hold between 100 and 300 slices, and images must be segmented one slice at a time. Moreover, the CNN training process is time-consuming and computationally expensive, even with access to sufficient labeled data.

Data augmentation [23,24] and transfer learning [25–28] are often used to overcome the lack of data when training a CNN. In the latter strategy, images from other pulmonary diseases can be used to learn similar features on the images and thus segment COVID-19 images. For instance, in [29], images from other lung conditions and COVID data were used simultaneously during training. However, better results are still obtained using only COVID-19 images, even with a minimal training set. Even with the best results, the lesion segmentation does not perform the same as parenchyma segmentation, considering a DSC of 0.65 for COVID-19 lesions in the best case, 0.85 (using only COVID-19 images) and around 0.95 for lung segmentation.

There are other recent attempts to handle these difficulties, such as the one shown in [30], here a label-free segmentation was accomplished by generating synthetic lesion regions that were added to images of healthy subjects. Then, these images were used to train a neural network to differentiate between healthy and synthetic disease images. Three publicly available datasets were used to test the model, including the one we used in this work, achieving a DSC value of 0.69 for this dataset and 0.59 and 0.61 for the others [31].

Consequently, this work is motivated by the need to develop new methodologies to automatically quantify the lung lesion percentage, dealing with the need for datasets to train AI models.

Following this motivation, the aim of this work is to propose a semi-supervised method for lesion quantification on COVID-19 CT images, eliminating the need for a large dataset to train and generalize the results. The proposed methodology removes the need for a training dataset of COVID-19 scans and is also repeatable on other images because it is not a model fitted to a specific dataset. Moreover, we evaluated the helpfulness of the proposed scheme in estimating the outcome of patients from a Mexican third-level hospital with proven COVID-19, and compared the differences in high and low-resolution CT images for COVID-19 lesion segmentation.

The proposed methodology consists of two steps. First, the lesion is segmented using a semi-supervised algorithm based on probabilistic active contours (PACO) [32], which is a multi-region segmentation approach that requires an initial delimitation of regions of interest (ROIs) in order to calculate probability density functions that constraint and guide contours to the border of the multiple ROIs; this algorithm has proved to be a good alternative for the segmentation on medical imaging, however it has yet to be explored for COVID-19 segmentation. Lesion segmentation is made over a full body image (which includes tissue outside the lung). Thus, the raw output of this algorithm may

identify body regions outside the lungs as lesions. For this reason, a second step is implemented to perform a complete lung-parenchyma segmentation. This is done using a U-net pre-trained on a lung parenchyma segmentation dataset. The output mask of the U-net is then used to clean the initial lesion mask. This cleaned version of the lesion and the lung parenchyma is then compared to estimate the lesion concerning the whole lung volume.

2. Methods

Our lesion segmentation approach involves two steps. The first aims to segment the COVID-19-induced lesion, and the second is focused on segmenting the lung parenchyma. This way, we obtained an excellent lung parenchyma segmentation using non-COVID-19 CT images and then used this information to mask the lesion segmentation obtained using our semi-supervised approach. The reason for proceeding as indicated was that experimentally, we noticed that our semi-supervised method performed much better when applied to the images containing the entire body than to parenchyma-only ones. Based on these results, we proposed the methodology described in Fig. 1.

2.1. Study population and CT data acquisition

Data were collected retrospectively from 295 patients (84 female, 211 male, 54 ± 12 years) admitted to the intensive care unit at the National Institute of Medical Sciences and Nutrition Salvador Zubirán (INCMNSZ, Mexico City, Mexico) due to severe pneumonia caused by COVID-19 infection in the period from March to April of 2020. Positive RT-PCR tests confirmed SARS-CoV-2 infection. The institutional ethics committee approved this protocol. Chest CT images were acquired using

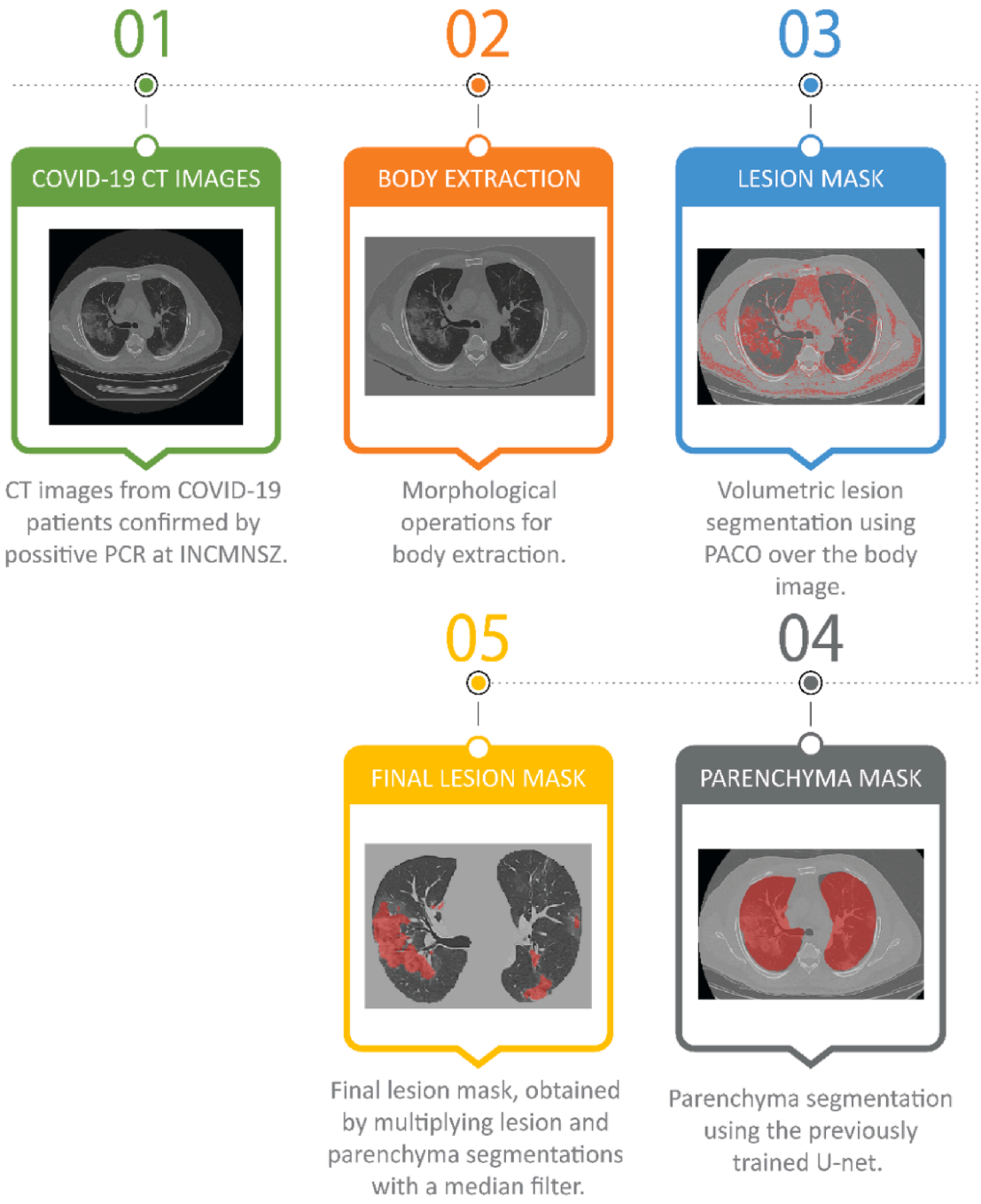


Fig. 1. Flow chart of the proposed methodology.

a GE Revolution EVO Gen 3 (GE Healthcare, Milwaukee, WI) system. The used imaging series consisted of low-resolution (70–101 slices) and high-resolution (117–361 slices) helical acquisitions using a lung window, matrix size of 512x512 pixels, slice thickness of 1.25–3.75 mm, and 1.25 mm spacing between slices. The X-ray tube voltage and current were set to 140 kV and 100–300 mA, respectively.

There is a low and high-resolution image for each patient, giving a total of 590 CT (295 low-resolution and 295 high-resolution volumes). Of those 295 patients, 132 were reported as deceased, and 163 survived.

2.2. Data pre-processing

All the images were pre-processed using a body extraction algorithm. First, a binary threshold was performed, with values between 175HU and 750HU. Then, a 3D morphological closure was applied using a structural element of radius (5,5,5) and a binary fill-hole filter. Finally, the 3D-connected components with a volume of fewer than 2 L were removed. Fig. 2 illustrates these data pre-processing steps.

2.3. Probabilistic active contours (PACO)

For COVID-19 lesion segmentation, the PACO algorithm proposed in [32] was used. This multi-region segmentation algorithm is based on minimizing an energy functional with constraints that guide contours to the border of the multiple regions of interest. For a better understanding, the PACO algorithm consists of the minimization of an energy functional E defined as:

$$E[I] = \sum_{k=1}^{N} \left\{ \lambda_k \int_L -P_k(V(x))H(\phi_k(x)) \prod_{j=1 j\neq k}^{N} (1 - H(\phi_j(x))) dx + \rho_k \int_L |\nabla H(\phi_k(x))| dx \right\} \quad (1)$$

where V is a volume observed over the voxel lattice L with N disjoint regions, the functions $P_k(.)$ measure how much $V(x)$ belongs to one of the N regions, x is a vector defining the voxel positions, and H is the Heaviside function defined by 2

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

where $|\cdot|$ is the L2 norm and ∇ is the gradient function. $\phi_k(\cdot)$ are the functions that define the contours of the objects as their zero-level-set, and if a voxel x belongs to a region R_k, then $\phi_k(x) \rangle 0$.

The first term of 1 corresponds to probability density functions (PDF), and it is used for measuring how much a voxel belongs to each region. This term is multiplied by the product term, which contains the Heaviside function, forcing the voxels not to be included in more than one region. The second term is the classical regularization term used in active contours algorithms; it measures the surface area that delimitates each volumetric region. The constants λ_k and ρ_k are used to control the influence of each term.

A two-step methodology is used to minimize the energy functional in the PACO algorithm. First, the PDFs are estimated using the regions delimited by the ϕ functions. Then, the energy functional E is minimized with respect to ϕ_k. These two steps are repeated until a convergence criterion or when the indicated number of iterations is reached. The

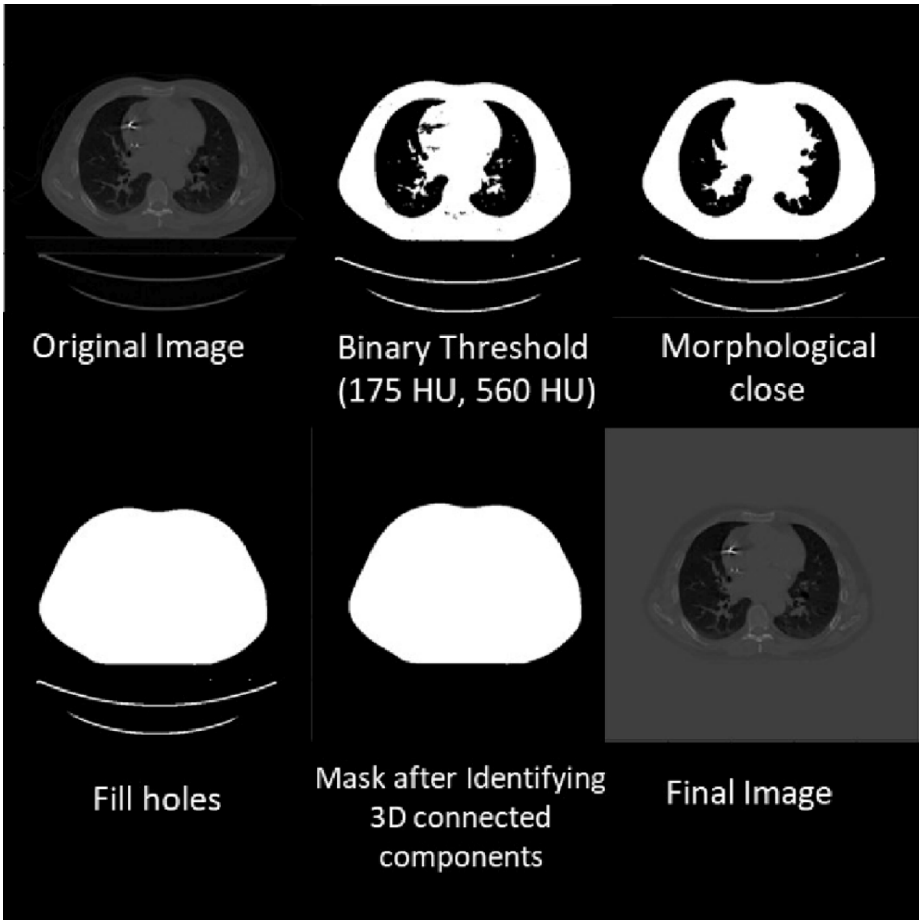


Fig. 2. Body extraction sequence.

reader is referred to [32] for further details on the PACO implementation.

For this work, the algorithm was implemented in MATLAB 2020b, and normal probability distributions were assumed. Also, the λ_k parameters were empirically set to 1 and the ρ_k to 0.5 while the number of iterations was set to 200. The Dirac delta and the Heaviside functions were approximated as in [29] for numerical implementation.

2.4. Initialization

The PACO algorithm requires initial contours to start the minimization process. The initial contours were defined manually, drawing the four initial regions (healthy tissue, lesion, body, and background) in a representative slice from the CT volume. This slice was selected empirically, looking for slices with large lesion areas from the middle section of the lungs. Fig. 3 shows an example of the initial contours for one patient and the initials PDFs from these regions. The background PDF is centered on −2000 for graphic presentation purposes in the figure, but the actual value assigned to the background is −5000. This image shows that the regions of interest have different PDFs.

2.5. Lung parenchyma segmentation

Convolutional neural networks have been widely used for image classification tasks and, more recently, for image segmentation. The latter could be achieved by training a CNN as a pixel-level classifier, as shown in [33], or by using an encoder-decoder architecture similar to the one used in this work and very popular in recent years [34]. This encoder-decoder model is named U-net. It extracts information from the image in the decoder stage to build low-resolution features; then it reconstructs the labeled image in the decoder path; this means it increases the resolution of the coded information to provide an output image with equal resolution as the input and combines the low-resolution with the high-resolution information by concatenating the decoding path with the corresponding encoding counterpart.

In this study, lung parenchyma segmentation was achieved using a modified 3D U-net architecture, like the one presented in [34] (see Fig. 4). The network takes as input a volumetric image of size (256,256,128) and returns a lung parenchyma mask image of the same size.

As Fig. 4 shows, the U-NET architecture was built with a 5-level encoder stage and two 3D convolution blocks on each level, using a kernel size of 3x3 and 8, 16, 32, 64, and 128 convolution filters. Convolutions are followed by a rectified linear unit (ReLu) activation function for each level. After both convolutions have been performed, a batch normalization layer and a Max-pooling operation are applied. The decoder stage is also composed of five levels with an upsampling process each, followed by a transpose 3D convolution layer, a ReLu activation, and a concatenation with the residuals of the same level from the encoder stage. A final convolutional layer is applied using a sigmoid activation function to create the desired volumetric mask output [35].

The network was trained on a publicly available lung segmentation dataset from The Cancer Imaging Archive: Thoracic Volume and Pleural Effusion Segmentations in Diseased Lungs for Benchmark Chest CT Processing Pipelines (PleThora) [36]. This dataset contained 402 CT scans with corresponding parenchyma segmentation masks from subjects with diseased lungs. The images were randomly rearranged in training and validation datasets, including 70 % (n = 281) and 30 % (n = 121) images.

Data augmentation was applied for the training stage using the following methods: random flip of the x or y axes, 3D random rotation among the three axes with a range of [-20, 20] degrees, and 3D translation in the range of [-5, 5] mm for every axis. Image normalization was performed to have the pixel intensity with zero mean and standard deviation of 1. Finally, random Gaussian noise was added with a standard deviation in the [0.0, 0.3] range. The model was trained using a variant learning rate with an initial value of 0.005 that was reduced by a factor of 0.1 when there was no improvement in the validation loss for two consecutive epochs. In addition, the network was trained for 100 epochs using the Adam optimizer. However, it contained a callback function to stop the training when no improvement in the validation set was obtained for five consecutive epochs. The batch size was set to 3, and the objective loss function used was the DICE loss (DL), as shown in eq.3 and commonly used in CNN-based segmentation tasks [37,38]. Finally, a k-fold cross-validation routine (k = 5) was implemented over the PleThora dataset to evaluate the network's overall performance.

$$DL(A,B) = 1 - \frac{2|A \cap B| + 1}{|A| + |B| + 1} \tag{3}$$

where one is added on the numerator and denominator to ensure the stability of the function in edge scenarios, i.e., A = B = ∅, then |A ∩ B| = |A| + |B| = 0.

After the U-NET was trained on the PleThora dataset, it was used to

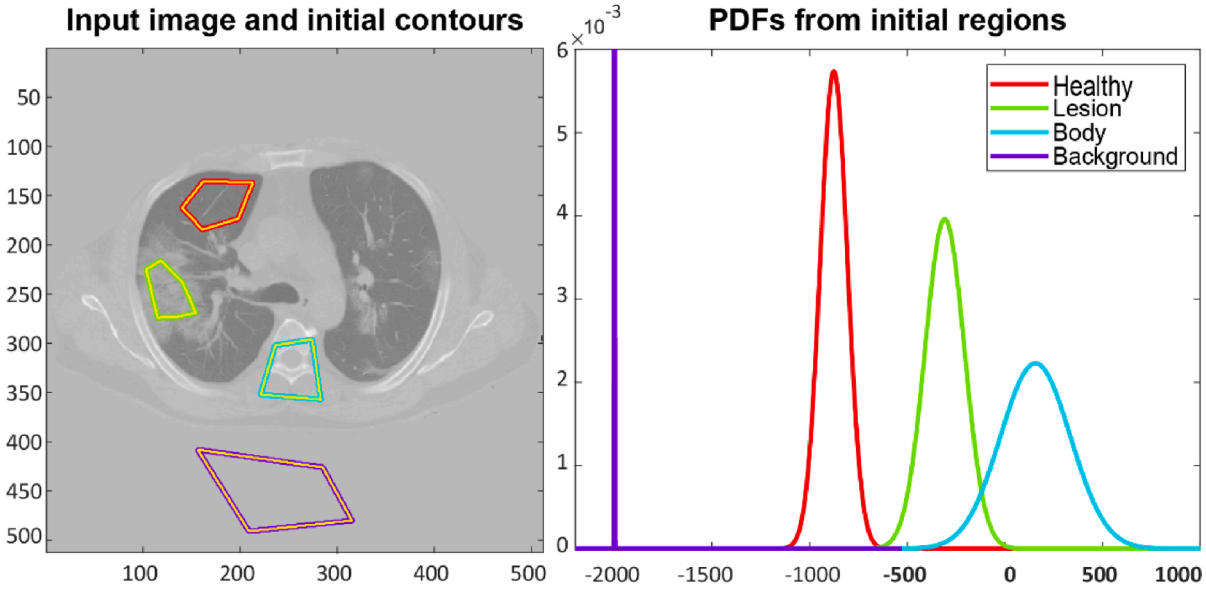


Fig. 3. Initial contours and initial PDFs for the PACO algorithm.

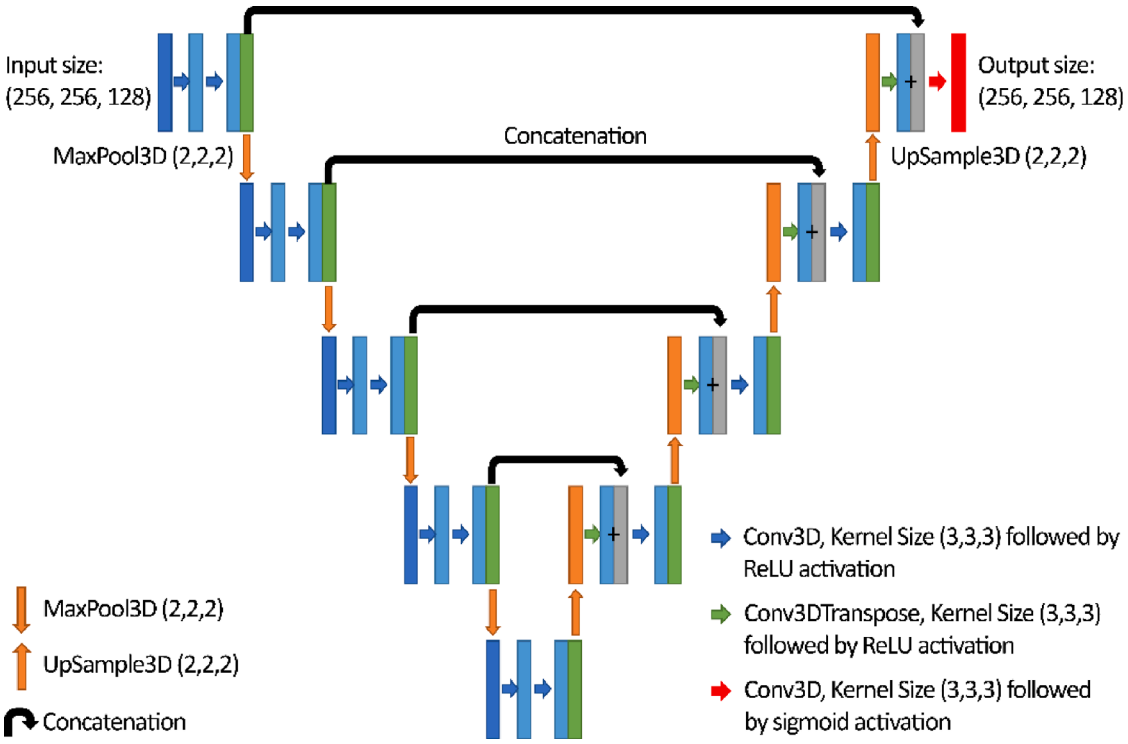


Fig. 4. U-NET architecture.

extract the lung parenchyma from the lesion-segmented CT images of the COVID-19 patients. Additionally, a median filter was applied to the final segmentation masks to eliminate some of the airways and vessels labeled as diseased tissue.

2.6. Lesion segmentation validation

To assess and compare our results with other strategies in the literature, the proposed workflow was applied to the publicly available COVID-19 CT Lung and Infection Segmentation Dataset (LISD) [16]. This dataset contains 20 COVID-19 CT scans manually labeled by three expert radiologists for the right lung, left lung, and infection. Furthermore, the images have different slices per volume and resolution (voxel size).

The 20 CT volumes from this dataset were segmented using the proposed methodology and then the volumetric DSC was computed to evaluate the resultant segmentation [39]. The DSC coefficient measures the overlap between two masks and is defined in eq. (4):

$$DSC(A,B) = \frac{2|A \cap B|}{|A| + |B|} \tag{4}$$

where A and B are binary masks, one is the ground truth, and the other is the segmented mask. The cardinality of a set is defined by $|\cdot|$. This coefficient takes values from 0 to 1, where 0 means no overlap, and 1 is a perfect match between the masks.

2.7. Statistical analysis

The algorithm was then used to perform lung segmentation in our dataset, which contained low (n = 295) and high-resolution CT images (n = 295). Further classification of images was made by identifying patients that survived and did not survive. For each image, the percentage of the lesion was computed by counting each voxel in the parenchyma and lesion mask. Then, the ratio between these two quantities was obtained to estimate the lesion percentage.

A Shapiro-Wilk test was performed to evaluate the normality in the distribution of lesion percentages in both survived and deceased patients for the high-resolution and low-resolution datasets. This test suggested that such data followed a non-normal distribution. Thus, a Flinger-Killeen test was applied for variance homogeneity. A Wilcoxon rank-sum test was used to compare the groups of deceased patients against survived patients in high and low-resolution modalities. A significance level $\alpha = 0.05$ value was considered in all tests.

3. Results

3.1. Evaluation of lung parenchyma segmentation

Fig. 5 shows the result of the lung parenchyma segmentation workflow for a CT image from the PleThora dataset. The upper row images show the ground truth (left) and segmented mask (right) using the U-net architecture, whereas the bottom row depicts the original and cropped parenchyma images, respectively.

Table 1 exhibits the 5-fold cross-validation result for the U-NET using the PleThora dataset. The average DSC obtained was 0.93 and 0.94 for the training and validation datasets. These DSC values reveal that the network performance is optimal for parenchyma segmentation since DSC can take values from 0 to 1, where 1 indicates a perfect match between the segmented mask and the ground truth [40]. The loss value is the reciprocal of the DSC value; for this metric, average values of 0.06 and 0.05 were obtained for training and validation, respectively, where a lower value close to zero means better performance on training.

All CT low and high-resolution images from the 295 patients with COVID-19 were processed using the body extraction workflow and the U-net to segment the parenchyma. An expert observer visually inspected the output masks. An example of the resultant image after applying the lung parenchyma workflow to a CT image from a patient with COVID-19 is presented in Fig. 5B. These results confirmed the optimal performance of the U-NET network, as expected. Similarly, all 20 CT images from LISD were processed using the same workflow.

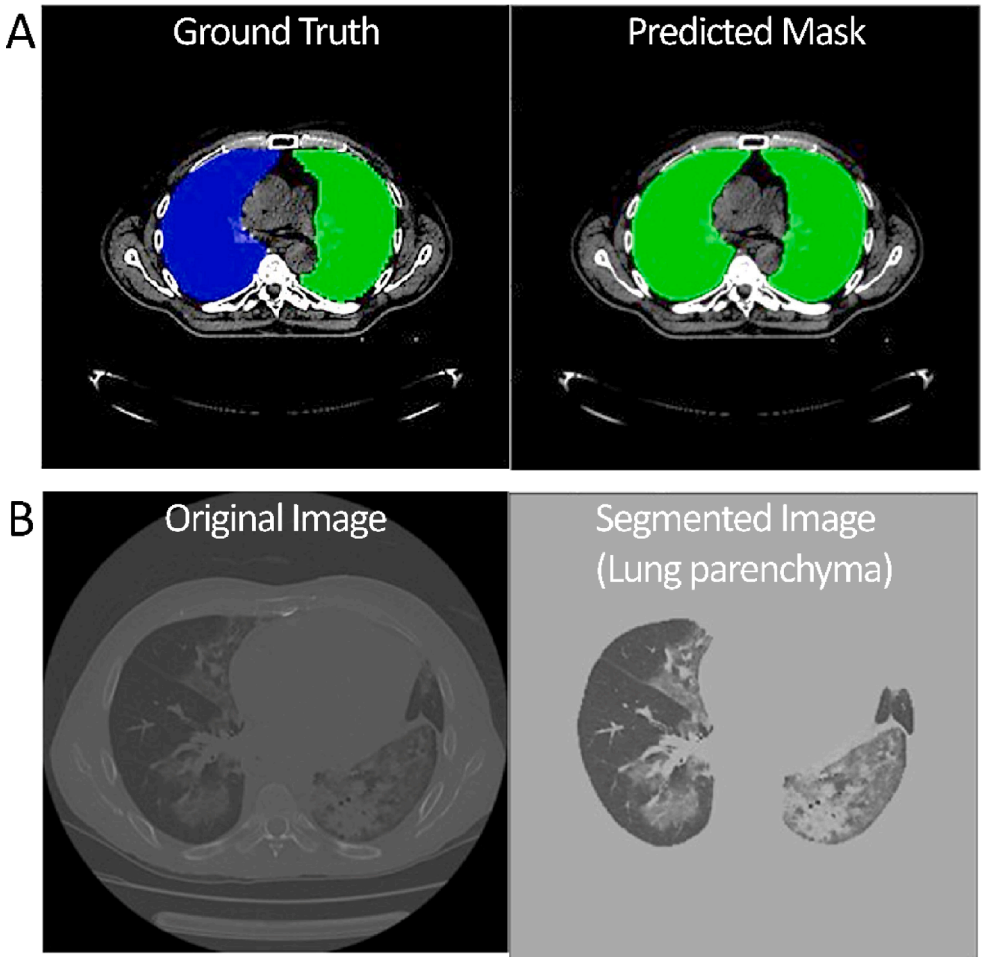


Fig. 5. U-NET parenchyma segmentation result (masks and cropped image).

Table 1
U-NET K-Fold Cross Validation.

Fold	Training Loss	Training DSC	Validation Loss	Validation DSC
1	0.0649	0.9351	0.0688	0.9312
2	0.0680	0.9320	0.0583	0.9417
3	0.0499	0.9501	0.0442	0.9557
4	0.0634	0.9366	0.0608	0.9392
5	0.0578	0.9422	0.0565	0.9434
Average	0.0608	0.9392	0.0577	0.9423

3.2. Evaluation of COVID-19 lesion volume estimation.

Lesion segmentation results for selected subjects within the INCMNSZ dataset are presented in Fig. 6. A visual comparison is made in four patients with CT scans from both high and low resolutions. Similar slices are selected for spatial correspondence. The following images are shown for each patient: (1) The input body image used for PACO, (2) The multiclass PACO output mask, and (3) The lesion overlapping on the body image. It is important to recall that the lesion mask results from multiplying the PACO output segmentation and the parenchyma mask. This figure discloses that segmentation with the proposed methodology correctly locates the disease areas on both CT modalities.

Furthermore, Fig. 6 illustrates the main limitations of the proposed methodology. For both image resolutions, non-disease regions appear on the lesion mask, mainly due to the intensity level, which is the same as diseased areas. These non-lesion regions are vasculature, airways, and in some cases, voxels of the body (out of the parenchyma) that were misclassified by the U-net as parenchyma.

The mean difference in the percentage of the lesion between high and low-resolution images is 10 %; this value could interest clinicians in locations where high-resolution images are difficult to access. In this scenario, an error of 10 % could be acceptable, considering that the algorithm could estimate the percentage of lesions on a low-resolution image and that the relationship between survived and deceased patients remains the same for both resolutions.

3.3. Validation

All 20 volumes from LISD were used for validation, with a mean DSC value of 0.58 and a median of 0.66. The lowest DSC values correspond to the volumes with fewer lesion regions on the ground-truth masks, where the areas delimited by the radiologists are minimal. It is worth mentioning that drawing an initial contour within the affected area is challenging in these cases. For PACO is more difficult to detect these small regions, and even when the region is detected, it possibly locates the lesion in different positions than the ground truth; in these cases, the DSC is very low because there is no overlap between the segmentation mask and the ground truth mask. It is essential to mention that in these cases, the mentioned regions have areas of around one to six pixels.

The obtained DSC value is near the value of 0.7 achieved by [29], and the main difference is that this methodology does not require a ground truth dataset of COVID-19 lesion segmentations; it just needs parenchyma segmentation to train the first network, but parenchyma segmentations are more accessible than COVID-19 related lesions segmentations.

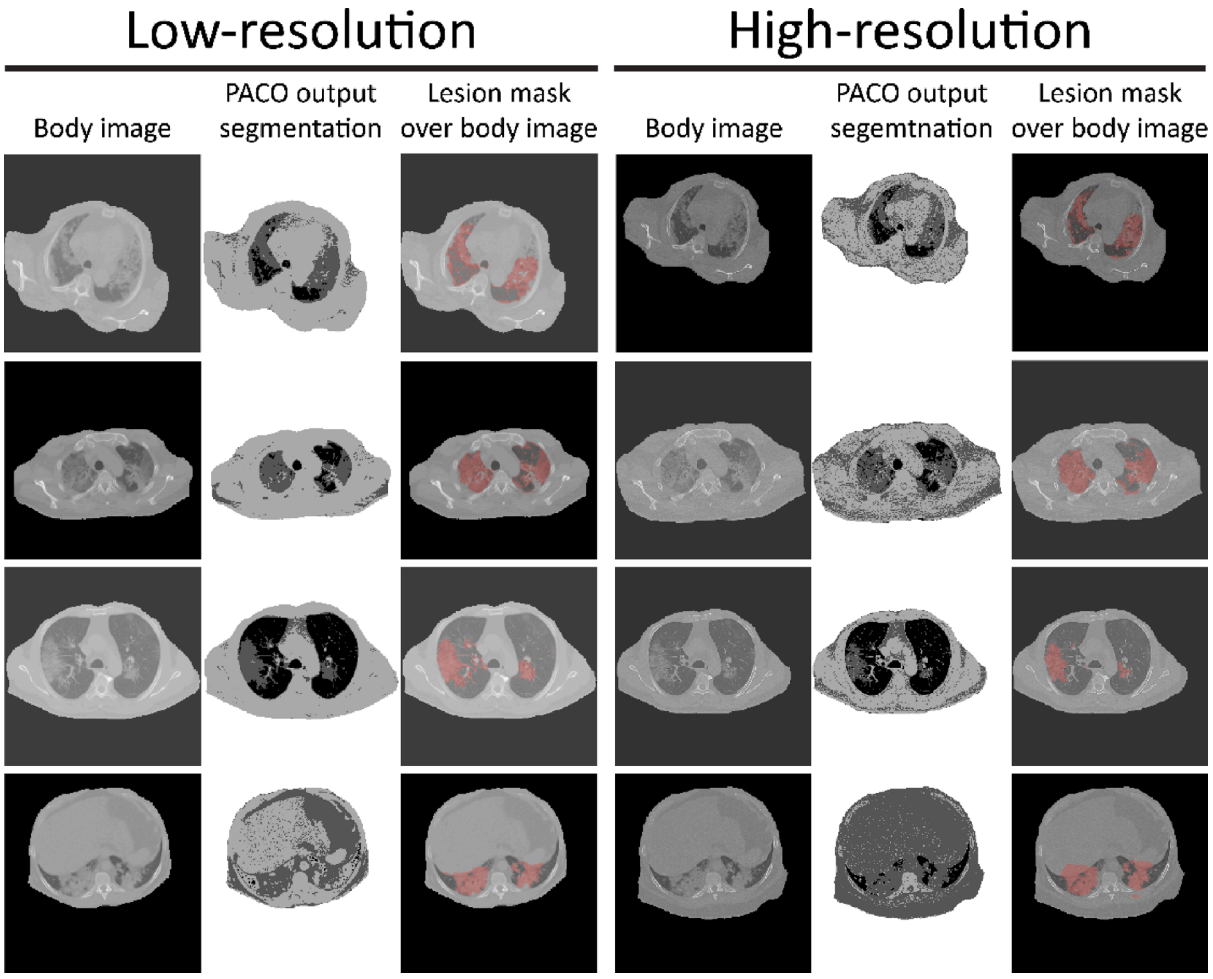


Fig. 6. Examples of COVID-19 lesion segmentation with the PACO algorithm in low and high resolution CT images from four patients.

3.4. Comparison of lesion percentages between survived and deceased patients.

The proposed workflow was applied to low and high-resolution images of the 295 patients and the percentage of lesions per patient was obtained. All 295 patients, were categorized into two groups according to their clinical outcome, survived or deceased. Fig. 7 compares the lesion-related percentages observed for each of these groups.

A Saphiro-Wilk test showed non-normality with a p-value of 2.3×10^{-2} and 2.5×10^{-4} for low and high-resolution groups, respectively. Also, a two-sample F-test for equal variances was computed between survived and deceased patients. The p-value was 0.39 and 0.77 for high and low-resolution groups, indicating different variances in both CT modalities for survived and deceased groups. Wilcoxon rank-sum test revealed different medians for survived and deceased groups in both CT modalities with a p-value of 9.1×10^{-4} in low-resolution CT and 5.1×10^{-5} for high-resolution images.

Fig. 8 shows examples of two patients with an unexpected percentage of the lesion. On the left, the HRCT shows a low percentage of lesions, but the patient died. However, visual inspection of HRCT of patients with this outcome shows that they have more consolidation areas than others. In the first image, dorsal consolidation can be observed; the algorithm cannot detect it because these regions have the same intensities as the rest of the body outside the parenchyma. On the right is the opposite case, where a patient survived, but the HRCT shows a high amount of lesion; the patients with this characteristic have much damage, mainly reticulation, crazy-paving pattern, and ground-glass opacities but almost no consolidation.

4. Discussion

Visual inspection of the final masks confirmed that the proposed pipeline successfully identifies lesions. However, this mask has additional areas that do not belong to lesioned tissue. These areas can be seen in Fig. 6, which correspond to lung airways and vasculature. Also, tissue outside the lung parenchyma can be mislabeled as part of the final mask. These errors are due to mistakes in the U-NET output. It is worth recalling that such an output is used to modify the lesion mask provided by the PACO algorithm. If the lung-parenchyma segmentation has mislabeled regions outside the lung, these regions can also be mislabeled as lesions. However, the 0.589 mean DSC value and the 0.66 median DSC value led us to think that the final lesion mask is still a good approximation compared to manual segmentation (gold standard), with the advantage that it does not require a lot of segmented images to train for lesion detection. The proposed pipeline looks promising, considering the lack of labeled data and the results of other works that used the same LISD dataset. Besides, for this methodology over-fitting is not a problem that is common for training with a small dataset. In addition, our proposed method could be repeatable for new images because no model is fitted to the data. It is worth remembering that the best DSC value obtained with a neural network for the same dataset is 0.7 [29].

The k-fold cross-validation results suggest that the chosen U-NET architecture is optimal for the lung-parenchyma segmentation task. Due to the lack of publicly available data, the training and validation sets contain non-COVID-19 images. This training strategy could affect performance when dealing with COVID-19 images. The next step to improve the proposed methodology would be to visually inspect every

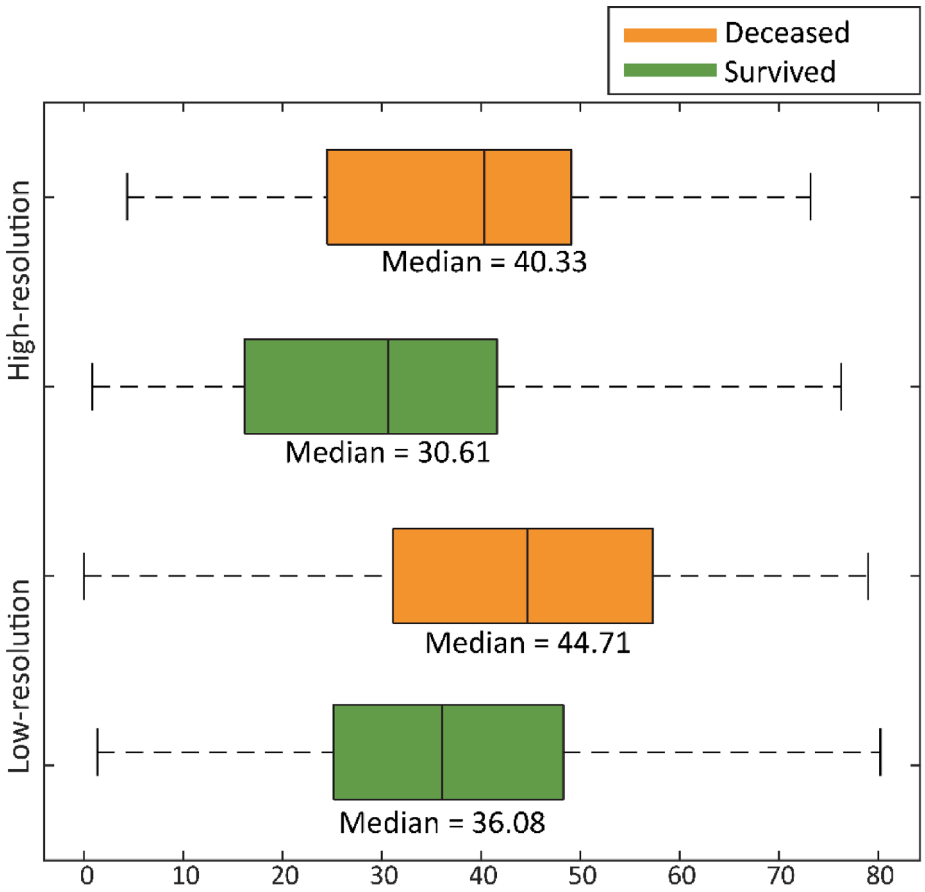


Fig. 7. Box plots for percentages of the lesion in survived and deceased patients.

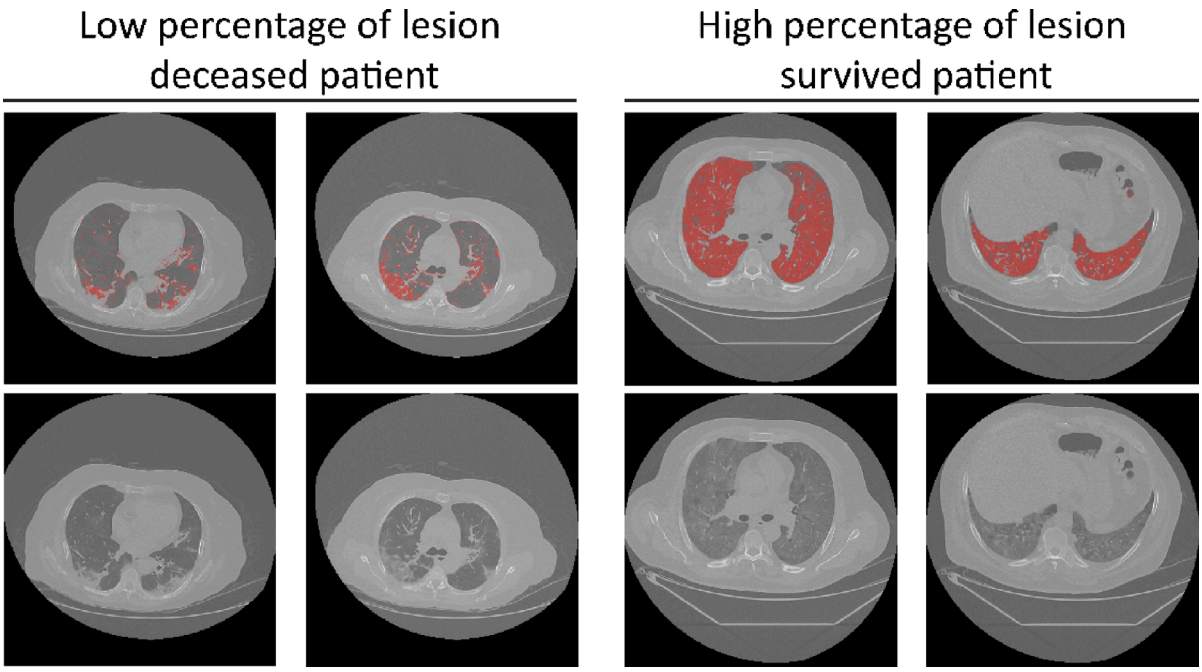


Fig. 8. Lesion mask comparison between survived patients with a high number of estimated lesions and deceased patients with a low number of estimated lesions.

lung parenchyma predicted by the U-NET. Then, a manual correction should be performed. These corrected images can be used as a new training dataset to retrain the network.

Statistical tests confirmed differences between the predicted lesion percentages of survived and deceased patients. Thus, it can be used as an indicator of patient prognosis. It is worth noting that this result remains consistent, even when low-resolution CT images are used instead of high-resolution scans. Additionally, there is a low range of variability

between the lesion percentage estimate in both resolutions (mean difference of 10.15 %). This also suggests that this metric may be used in various CT acquisition protocols at different slice thickness degrees without compromising the lesion estimate.

Despite the difference in the survived and deceased groups, there is a significant overlap between groups. This is mainly because there are patients with a low percentage of lesions who died and patients with high percentages who survived; see Fig. 8.

The deceased patients with low percentages of lesions have more consolidation areas than other patients, who usually present only ground-glass opacities, reticulation, and crazy paving patterns. The PACO algorithm does not detect these consolidation regions because they have the same intensity level as the body; these consolidation regions are then segmented as body and make the estimated percentage of lesions take a lower value than the real. The degree of consolidation could be related to severe disease and the patient's death. However, this miss-classification of the algorithm for consolidation regions could decrease the difference in percentages of lesions for both groups. In contrast, the patients who survived but had many lesions did not have consolidation areas. Again, this could be related to their survival, the treatment, and the patient's reaction to the medicine.

Clinical outcome for COVID-19 patients is multifactorial. Gender (male), age (oldest is worst), comorbidities (hypertension and obesity increase risk), immunological conditions (immunosuppression increases adverse outcomes), hypoxia (early oxygen therapy improves outcomes), and still many others not yet defined. This study found a trend between the amount of lung parenchyma volume involved and the risk of dying, but many other factors were not considered. Adding them to the analysis may improve accuracy. We also found that there is a difference between the type of lesion and not just the amount of area involved. Indeed, the earlier patients present a more extensive involvement of damage to the lungs; the worst prognosis may be expected, and the time frame is essential for conclusions.

5. Conclusions

Our proposed pipeline successfully labeled affected lung tissue in CT images. Exceptions of mislabeled regions include airways, vasculature, and segmentation mistakes from the lung parenchyma. Despite these errors, the workflow allows to estimate the lung lesion percentage.

The core of our methodology uses a semi-supervised probabilistic region contour algorithm to detect the lesion. The proposed algorithm performed similarly to previously proposed neural network architectures on high and low-resolution CT scans. Compared to other approaches based mainly on convolutional neural networks, it does not require high volumes of data for training, thus avoiding common overfitting problems.

At the same time, this work reveals that the percentage of the lesion shows a statistically significant difference between survived and deceased patients. This finding enforces the possibility of using such a measure as a quantitative indicator to evaluate the prognosis of severe COVID-19 patients.

To our knowledge, most of recent publications regarding COVID-19 lesion segmentation focus on AI algorithms. Unfortunately, the lack of essential amounts of labeled data restrains their performance. Thus, a semi-supervised approach such as the one herein may be a successful alternative. Our segmentation results are comparable with those in literature according to the obtained DSC values, without the expense of training an AI algorithm to segment the lesion. In addition, no other work has compared performance between low and high-resolution CT images.

In this work, for PDF estimation, only intensity information was used; however, the PACO algorithm allows the use of multivariate information by transforming images to other domains. Then, the assessment of a multivariate PDF scheme could be explored in future work. In addition, an extended classification evaluation at a pixel level using indices such as accuracy and specificity should be considered. Finally, a comprehensive validation with other publicly available datasets may provide more evidence about the performance of the proposed methodology.

CRediT authorship contribution statement

Diomar Enrique Rodriguez-Obregon: Conceptualization, Methodology, Software, Writing – original draft, Visualization, Validation, Formal analysis. **Aldo Rodrigo Mejia-Rodriguez:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing. **Leopoldo Cendejas-Zaragoza:** Methodology, Software, Visualization, Validation, Formal analysis, Data curation. **Juan Gutiérrez Mejía:** Investigation, Resources. **Edgar Román Arce-Santana:** Software, Writing – review & editing. **Sonia Charleston-Villalobos:** Investigation, Writing – review & editing. **Tomas Aljama-Corrales:** Investigation, Writing – review & editing. **Alejandro Gabutti:** Investigation, Resources. **Alejandro Santos-Díaz:** Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review & editing.

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.

Data availability

Data will be made available on request.

Acknowledgments

The authors acknowledge the Critical Care and Radiology Departments of the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (INCMNSZ) for supporting the present research.

Funding: The Work of D. E. Rodriguez-Obregon was supported by CONACyT through a doctoral Grant [# 787212], and the work of L. Cendejas-Zaragoza was supported by CONACyT through a posdoctoral grant [#647956].

References

[1] WHO, World health organization coronavirus (COVID-19) dashboard (May 2022). URL https://covid19.who.int/.
[2] Y. Li, L. Yao, J. Li, L. Chen, Y. Song, Z. Cai, C. Yang, Stability issues of RT-PCR testing of SARS-COV-2 for hospitalized patients clinically diagnosed with COVID-19, J. Med. Virol. 92 (7) (2020) 903–908.
[3] M. Prokop, W. Van Everdingen, T. van Rees Vellinga, H. Quarles van Ufford, L. Stöger, L. Beenen, B. Geurts, H. Gietema, J. Krdzalic, C. Schaefer-Prokop, et al., Co-rads: a categorical CT assessment scheme for patients suspected of having COVID-19—definition and evaluation, Radiology 296 (2) (2020) E97–E104.
[4] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in china: a report of 1014 cases, Radiology 296 (2) (2020) E32–E40, https://doi.org/10.1148/radiol.2020200642.
[5] R. Chrzan, M. Bociaga-Jasik, A. Bryll, A. Grochowska, T. Popiela, Differences among COVID-19, bronchopneumonia and atypical pneumonia in chest high resolution computed tomography assessed by artificial intelligence technology, J. Personal. Med. 11 (5) (2021) 391, https://doi.org/10.3390/jpm11050391.
[6] L. Gattinoni, D. Chiumello, P. Caironi, M. Busana, F. Romitti, L. Brazzi, L. Camporota, COVID-19 pneumonia: different respiratory treatments for different phenotypes? (2020). doi:doi.org/10.1007/ s00134-020-06033-2.
[7] A.R. Larici, G. Cicchetti, R. Marano, L. Bonomo, M.L. Storto, COVID-19 pneumonia: current evidence of chest imaging features, evolution and prognosis, Chin. J. Acad. Radiol. (2021) 1–12, https://doi.org/10.1007/s42058-021-00068-0.
[8] U. Bhattacharjya, K.K. Sarma, J.P. Medhi, B.K. Choudhury, G., Bar man, Automated diagnosis of COVID-19 using radiological modalities and artificial intelligence functionalities: A retrospective study based on chest HRCT database, Biomed. Signal Process. Control 80 (2023), 104297, https://doi.org/10.1016/j.bspc.2022.104297.
[9] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, C. Zheng, A Weakly-Supervised Framework for COVID-19 Classification and Lesion Localization From

Chest CT, IEEE Trans. Med. Imaging 39 (8) (2020) 2615–2625, https://doi.org/10.1109/TMI.2020.2995965.
[10] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, L. Xia, Serial quantitative chest ct assessment of COVID-19: a deep learning approach, Radiology: Cardiothoracic Imaging 2 (2) (2020), e200075, https://doi.org/10.1148/ryct.2020200075.
[11] H. Yue, Q. Yu, C. Liu, Y. Huang, Z. Jiang, C. Shao, H. Zhang, B. Ma, et. al., Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study, Ann. Transl. Med., 8 (14) (2020), https//doi.org/10.21037/atm-20-3026.
[12] F. Al-Areqi, M.Z. Konyar, Effectiveness evaluation of different feature extraction methods for classification of covid-19 from computed tomography images: A high accuracy classification study, Biomed. Signal Process. Control 76 (2022), 103662.
[13] K. Gupta, V. Bajaj, Deep learning models-based ct-scan image classification for automated screening of covid-19, Biomedical Signal Processing and, Control 80 (2023), 104268.
[14] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, R.-G. Zhang, M.-M. Cheng, Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation, IEEE Trans. Image Process. 30 (2021) 3113–3126.
[15] I. Laradji, P. Rodriguez, O. Manas, K. Lensink, M. Law, L. Kurzman, W. Parker, D. Vazquez, D. Nowrouzezahrai, A weakly supervised consistency-based learning method for covid-19 segmentation in ct images, in: In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 2453–2462.
[16] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng, W. Hao, M. Sen, Y. Xiaoyu, N. Ziwei, L. Chen, T. Lu, Z. Yuntao, Z. Qiongjie, D. Guoqiang, H. Jian, COVID19 ct lung and infection segmentation dataset (apr 2020). URL https://doi.org/10.5281/zenodo.3757476.
[17] X. Liu, Q. Yuan, Y. Gao, K. He, S. Wang, X. Tang, J. Tang, D. Shen, Weakly supervised segmentation of COVID19 infection with scribble annotation on ct images, Pattern Recogn. 122 (2022), 108341.
[18] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, Y. Shi, Abnormal lung quantification in chest ct images of COVID-19 patients with deep learning and its application to severity prediction, Med. Phys. 48 (4) (2021) 1633–1645.
[19] H. Chen, Y. Jiang, H. Ko, M. Loew, A teacher–student framework with fourier transform augmentation for covid-19 infection segmentation in ct images, Biomed. Signal Process. Control 79 (2023), 104250.
[20] S. Punitha, T. Stephan, R. Kannan, M. Mahmud, M. S., Kaiser, S. B., Belhaouari,, Detecting COVID-19 From Lung Computed Tomography Images: A Swarm optimized Artificial Neural Network Approach. IEEE, Access 11 (2023) 12378–12393.
[21] H. Jia, H. Tang, G. Ma, W. Cai, H. Huang, L. Zhan, Y. Xia, A convolutional neural network with pixel-wise sparse graph reasoning for COVID-19 lesion segmentation in CT images, Comput. Biol. Med. (2023), 106698.
[22] W. Xie, C. Jacobs, J.P. Charbonnier, B. van Ginneken, Dense regression activation maps for lesion segmentation in CT scans of COVID-19 patients, Med. Image Anal. (2023), 102771.
[23] M. Elgendi, M. U. Nasir, Q. Tang, D. Smith, J.-P. Grenier, C. Batte, B. Spieler, W. D. Leslie, C. Menon, R. R. Fletcher, et al., The effectiveness of image augmentation in deep learning networks for detecting COVID-19: A geometric transformation perspective, Frontiers in Medicine 8. Front. Med. (Lausanne) (2021), https//doi.org/10.3389/fmed.2021.629134. PMID: 33732718.
[24] U. Asghar, M. Arif, K. Ejaz, D. Vicoveanu, D. Izdrui, O. Geman, An improved COVID-19 detection using gan-based data augmentation and novel qunet-based classification, Biomed Res. Int. (2022).
[25] M.M. Taresh, N. Zhu, T.A.A. Ali, A.S. Hameed, M.L. Mutar, Transfer learning to detect COVID-19 automatically from x-ray images using convolutional neural networks, Int. J. Biomed. Imaging (2021).
[26] S. Albahli, W. Albattah, Deep transfer learning for COVID-19 prediction: case study for limited data problems, Current medical imaging 17 (8) (2021) 973.
[27] I. Lahsaini, M.E.H. Daho, M.A. Chikh, Deep transfer learning based classification model for COVID-19 using chest ct-scans, Pattern Recogn. Lett. 152 (2021) 122–128.
[28] S. Ahuja, B.K. Panigrahi, N. Dey, V. Rajinikanth, T.K. Gandhi, Deep transfer learning-based automated detection of COVID-19 from lung ct scan slices, Appl. Intell. 51 (1) (2021) 571–585.
[29] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, Q. Zhu, G. Dong, J. He, Z. He, et al., Toward data-efficient learning: A benchmark for COVID-19 ct lung and infection segmentation, Med. Phys. 48 (3) (2021) 1197–1210, https://doi.org/10.1002/mp.14676.
[30] Q. Yao, L. Xiao, P. Liu, S.K. Zhou, Label-free segmentation of covid-19 lesions in lung ct, IEEE Trans. Med. Imaging 40 (10) (2021) 2808–2819.
[31] W. Sun, X. Feng, J. Liu, H. Ma, Weakly supervised segmentation of covid-19 infection with local lesion coherence on ct images, Biomedical Signal Processing and Control 79 (2023), 104099.
[32] E.R. Arce-Santana, A.R. Mejia-Rodriguez, E. Martinez-Peña, A. Alba, M. Mendez, E. Scalco, A. Mastropietro, G. Rizzo, A new probabilistic active contour region-based method for multiclass medical image segmentation, Med. Biol. Eng. Comput. 57 (3) (2019) 565–576, https://doi.org/10.1007/s11517-018-1896-y.
[33] D. Ciresan, A. Giusti, L. Gambardella, J. Schmidhuber, Deep neural networks segment neuronal membranes in electron microscopy images, Advances in neural information processing systems 25 (2012).
[34] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241. doi:10.48550/arXiv.1505.04597.
[35] L. Cendejas-Zaragoza, D.E. Rodriguez-Obregon, A.R. Mejia-Rodriguez, E.R. Arce-Santana, A. Santos-Diaz, COVID-19 volumetric pulmonary lesion estimation on ct images using a u-net and probabilistic active contour segmentation, Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (2021) 3850–3853, https://doi.org/10.1109/EMBC46164.2021.9629532.
[36] K.J. Kiser, S. Ahmed, S. Stieb, A.S.R. Mohamed, H. Elhalawani, P.Y.S. Park, N. S. Doyle, B.J. Wang, A. Barman, Z. Li, W.J. Zheng, C.D. Fuller, L. Giancardo, Plethora: Pleural effusion and thoracic cavity segmentations in diseased lungs for benchmarking chest ct processing pipelines, Med. Phys. 47 (11) (2020) 5941–5952, https://doi.org/10.1002/mp.14424.
[37] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, in: Deep learning in medical image analysis and multimodal learning for clinical decision support, Springer, 2017, pp. 240–248.
[38] S. Jadon, A survey of loss functions for semantic segmentation, in: in: 2020 IEEE Conference on Computational intelligence in Bioinformatics and Computational Biology (CIBCB), 2020, pp. 1–7.
[39] L.R. Dice, Measures of the amount of ecologic association between species, Ecology 26 (3) (1945) 297–302. http://www.jstor.org/stable/1932409.
[40] S. Jadon, A survey of loss functions for semantic segmentation, Bioinformatics and Computational Biology (CIBCB) IEEE (2020) 1–7. https://doi.org/10.1109/CIBCB48159.2020.9277638.