



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Original article

Value and prognostic impact of a deep learning segmentation model of COVID-19 lung lesions on low-dose chest CT



Axel Bartoli, MD^{a,b,1,*}, Joris Fournel^{a,b,1}, Arnaud Maurin, MD^a, Baptiste Marchi, MD^a, Paul Habert, MD^{a,d,e}, Maxime Castelli, MD^a, Jean-Yves Gaubert, MD^{a,d,e}, Sebastien Cortaredona, MD^{f,g}, Jean-Christophe Lagier, MD, PhD^{f,h}, Matthieu Million, MD, PhD^{f,h}, Didier Raoult, MD, PhD^{f,h}, Badih Ghattas, MCU^c, Alexis Jacquier, MD, PhD^{a,b}

^a Department of Radiology, Hôpital de la Timone Adultes, AP-HM. 264, rue Saint-Pierre, 13385 Marseille Cedex 05, France

^b CRMBM - UMR CNRS 7339, Medical Faculty, Aix-Marseille University, 27, Boulevard Jean Moulin, 13385 Marseille Cedex 05, France

^c I2M - UMR CNRS 7373, Aix-Marseille University, CNRS, Centrale Marseille, 13453 Marseille, France

^d LIEE, Medical Faculty, Aix-Marseille University, 27, Boulevard Jean Moulin, 13385 Marseille Cedex 05, France

^e CERIMED, Medical Faculty, Aix-Marseille University, 27, Boulevard Jean Moulin, 13385 Marseille Cedex 05, France

^f Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005, Marseille, France

^g IRD, VITROME, Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005, Marseille, France

^h IRD, MEPHI, Institut Hospitalo-Universitaire Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005, Marseille, France

ARTICLE INFO

Article History:

Received 4 November 2021

Accepted 9 March 2022

Available online 22 March 2022

Keywords:

COVID-19

Artificial intelligence

Multidetector computed tomography

Deep learning

Diagnostic imaging

ABSTRACT

Objectives: 1) To develop a deep learning (DL) pipeline allowing quantification of COVID-19 pulmonary lesions on low-dose computed tomography (LDCT). 2) To assess the prognostic value of DL-driven lesion quantification.

Methods: This monocentric retrospective study included training and test datasets taken from 144 and 30 patients, respectively. The reference was the manual segmentation of 3 labels: normal lung, ground-glass opacity (GGO) and consolidation (Cons). Model performance was evaluated with technical metrics, disease volume and extent. Intra- and interobserver agreement were recorded. The prognostic value of DL-driven disease extent was assessed in 1621 distinct patients using C-statistics. The end point was a combined outcome defined as death, hospitalization > 10 days, intensive care unit hospitalization or oxygen therapy.

Results: The Dice coefficients for lesion (GGO+Cons) segmentations were 0.75 ± 0.08 , exceeding the values for human interobserver (0.70 ± 0.08 ; 0.70 ± 0.10) and intraobserver measures (0.72 ± 0.09). DL-driven lesion quantification had a stronger correlation with the reference than inter- or intraobserver measures. After stepwise selection and adjustment for clinical characteristics, quantification significantly increased the prognostic accuracy of the model (0.82 vs. 0.90 ; $p < 0.0001$).

Conclusions: A DL-driven model can provide reproducible and accurate segmentation of COVID-19 lesions on LDCT. Automatic lesion quantification has independent prognostic value for the identification of high-risk patients.

© 2022 The Authors. Published by Elsevier Masson SAS on behalf of Société française de radiologie. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Abbreviations: ACE, angiotensin-converting enzyme; BMI, body mass index; Cons, consolidation; CNN, convolutional neural network; COVID-19, coronavirus disease 2019; CT-SS, chest tomography severity score; DL, deep learning; DSC, Dice similarity coefficient; GGO, ground-glass opacity; ICU, intensive care unit; LDCT, low-dose computed tomography; MAE, mean absolute error; MVSF, mean volume similarity fraction; ROC, receiver operating characteristic

Institution from which the work originated: Department of Radiology, Hôpital de la Timone Adultes, AP-HM. 264, rue Saint-Pierre 13385 Marseille Cedex 05, France.

* Corresponding author at: 10, rue de Vauvenargues, 13007, Marseille, France.

E-mail address: axel.bartoli@ap-hm.fr (A. Bartoli).

¹ Authors participate equally to the work.

1. Introduction

In December 2019, an outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spread worldwide from Asia to Europe [1,2]. SARS-CoV-2 is responsible for coronavirus disease 2019 (COVID-19). It was declared a worldwide pandemic by the World Health Organization on March 11th 2020. One of the main risks is the congestion of the health care system due to an unusually rapid inflow of patients, especially in the intensive care unit (ICU). Thus, there is a need for precise patient selection and risk stratification to focus on severe cases [3]. This stratification is based on clinical criteria, viral

load on reverse-transcription polymerase chain reaction (RT-PCR) and pulmonary lesions on chest CT.

Low-dose computed tomography (LDCT) is more effective than chest X-ray for depicting ground-glass opacity (GGO) and consolidation (Cons), with a lower dose of radiation than conventional chest CT [4–7]. Some investigators have shown that a semi quantitative clinical score reflecting the extent of lesions might be useful for patient risk stratification [8,9]. Nevertheless, the computation of semi quantitative scores remains a time-consuming process that is prone to intra- and interobserver variability. Hence, there is a need for a fast, reproducible and fully automated COVID-19 lung lesion segmentation method that can be applied to a large cohort as a predictive risk stratification tool in disease management and prediction.

Deep learning (DL) techniques, especially convolutional neural networks (CNNs), have shown promising results in the automation of medical imaging measures [10]. In thoracic imaging, these techniques have shown excellent performance in nodule detection, lesion segmentation and disease classification [11,12].

The main purpose of this study was to develop and evaluate a complete DL pipeline that allows a fully automated segmentation of COVID-19 pulmonary lesions on LDCT and the computation of lesion volume and extent. Our secondary purpose was to investigate whether automatic lesion quantification was associated with adverse events among COVID-19 patients.

2. Materials and methods

2.1. Study design

This single-center retrospective study was conducted from March 3rd to July 2nd, 2020, and approved by the local Institutional Review Board (N°: 2020-0012, RGPD/Ap-Hm: 2020-48). Training, validation and test datasets including LDCT from 124, 20 and 30 patients, respectively, were included to build a pipeline based on CNNs adapted to assess automatic segmentation and quantification of COVID-19 lesions on LDCT as well as computation of lesion volume and extent. A flow diagram of the procedure is shown in Fig. 1. Then, we evaluated the predictive value of deep learning (DL)-driven quantification of lung lesions on adverse event occurrence in a dataset of 1621 patients, excluding data from the training, validation and test datasets. Among those 1621 patients, 983 have been previously reported [13,14]. The authors did not receive any financial or material support from any industrial company in the execution of this study.

2.2. Population and data

2.2.1. Population

All patients were enrolled from a single center (La TIMONE Hospital – Assistance Publique Hôpitaux de Marseille (APHM)). All patients who presented between March 30th and June 2nd 2020 with a confirmed COVID-19 infection using SARS-CoV RNA detection from a nasopharyngeal swab sample [15,13] and were eligible for unenhanced LDCT were retrospectively included. LDCT was performed on all patients who were over 55 years old or had risk factors for adverse outcomes for COVID-19, such as hypertension, diabetes, obesity (BMI>30), dyspnea or abnormal lung auscultation. The exclusion criteria were refusal to participate in the protocol and an age below 18 years.

2.2.2. Clinical data

The following clinical parameters were recorded by infectiologists (M.M. and J-C.L., with 25 and 20 years of experience, respectively) the same day as the LDCT: age, sex, date of the first symptoms, temperature, heart rate, systolic and diastolic blood pressures, respiratory rate, oxygen saturation, cough, rhinorrhea, dyspnea, diarrhea, myalgia, and lung auscultation abnormalities. Medical history was

recorded: heart disease, tobacco use, chronic obstructive pulmonary disease, asthma, diabetes, obesity, sleep apnea syndrome, oncological status and immunosuppression status. The time between the first symptoms and the LDCT was recorded. Patient follow-up lasted 10 days for patients with no adverse events, and the follow-up period was extended to cover the in-hospital stay for patients who required hospitalization. The primary endpoint of the second objective was a combined outcome consisting of either a need for oxygen therapy, a need for transfer to the ICU, hospitalization ≥ 10 days and/or death.

2.2.3. Radiological data

All patients underwent unenhanced, deep-inspiration LDCT on the same system (Revolution EVO – GE Healthcare, WI, USA) with parameters detailed in Appendix A. To develop our pipeline, we used a training dataset composed of 124 LDCT examinations (68767 CT slices) and a validation dataset of 20 LDCT examinations (6317 CT slices) from consecutive patients in clinical care. To obtain a training dataset including all types of lesions and with a homogeneous repartition of lesion extent and severity, the chest tomography severity score (CT-SS) developed by Yang et al. was used on the whole cohort [16]. This score, ranging from 0 to 40, has been validated as a semi-quantitative clinical method to quantify the extent and severity of lung abnormalities in COVID-19. All CT-SS images were evaluated by two experienced chest radiologists (J-Y.G. and P.H., with 25 and 7 years of experience, respectively). Patients for the training and validation datasets were chosen depending on their CT-SS, resulting in 13/144 (10.5%) and 2/20 (10%) severe patients (CT-SS >19.5) and 111/144 (89.5%) and 18/20 (80%) mild patients (CT-SS < 19.5). The test dataset was composed of 30 consecutive patients (15587 CT slices) from clinical care and did not overlap with the training dataset nor the validation datasets.

2.2.4. Manual segmentation

Manual image segmentation was undertaken for the combined training, validation and test datasets by a single observer (Observer 1 (O1), A.B., with 5 years of experience). For each patient, all images from the lung window LDCT were anonymized. Images were imported in DICOM format into the validated post processing software 3D Slicer (<https://www.slicer.org>, 2014) [17]. Manual segmentation of the lung window CT was applied to the entire lung volume, including all slices, using thresholding, painting and erasing methods to obtain the segmentation masks of three distinct labels: GGO, Cons, and normal pixels within the lungs (LungN). GGO and Cons were distinguished using a threshold based on the attenuation values in HU compared to that of the pulmonary artery [18]. Distal vascular and bronchial trees were not extracted from the labels. The non-segmented part of the image was classified under a fourth label: background (BG). After being validated by one experienced chest radiologist (J-Y.G.), the obtained segmentation masks were considered the ground truth, especially for GGO and Cons. Clinical parameters were obtained from the ground-truth segmentations as follows: lung volume (cm³) was the sum of the LungN, GGO, and Cons labels. The GGO and Cons volumes (cm³) were extracted from the respective labels. The GGO and Cons extents (%) were the ratios of the GGO and Cons volumes, respectively, to the total lung volume. Lesion extent (%) was the sum of the GGO and Cons extents. The user interaction time was recorded for all manual segmentations. All ground-truth manual segmentations and extracted clinical measures were labeled O1a.

2.3. Network architecture

Our pipeline was composed of three 2D slice-based CNN models and aimed to produce automated segmentation of GGO and Cons on LDCT images with corresponding measures in terms of volume (cm³) and extent (%). All automated segmentations and extracted measures

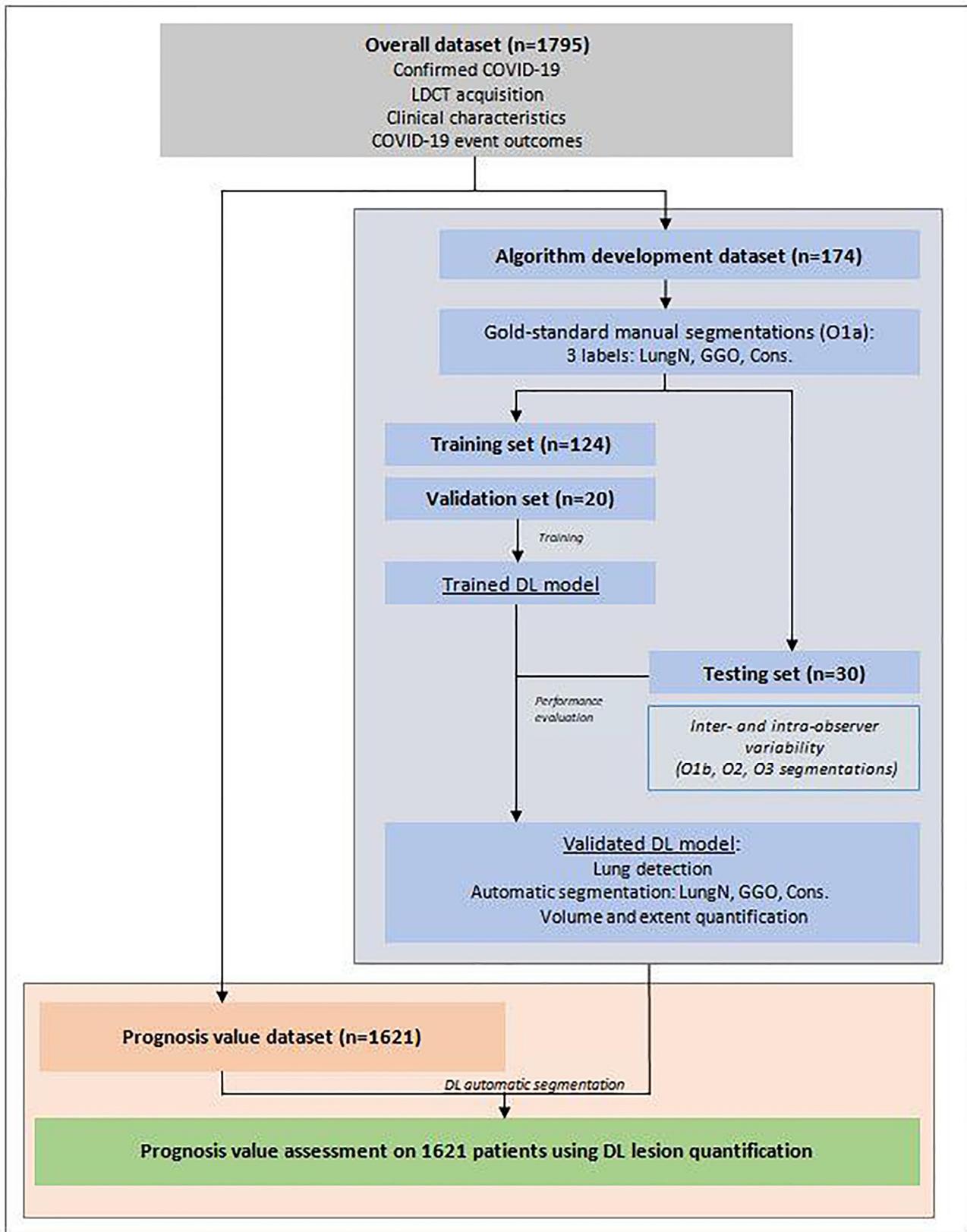


Fig. 1. Overview of the study design and data flow.

Note — LDCT: low-dose computed tomography; O1a: ground-truth manual segmentation by Observer 1. DL: deep learning; CT-SS: chest tomography severity score; LungN: normal lung; GGO: ground-glass opacity; Cons.: consolidation; O1b: intraobserver manual segmentation by Observer 1; O2: manual segmentation by Observer 2; O3: manual segmentation by Observer 3.

assess the predictive performance of the DL-driven automatic lesion extent quantification on the prognostic value dataset, we performed multivariate logistic regression on the following outcome: “transfer to ICU and/or death and/or hospitalization ≥ 10 days and/or oxygen therapy”. We randomly divided the prognosis value dataset ($n=1621$) into a training subset (70% of the initial sample size, $n=1135$) and a validation subset (30% of the initial sample size, $n=486$). Model parameters were estimated on the training dataset, and prognosis performance was assessed on the validation dataset. A reference model (A) was first tested and adjusted for the following covariates: age, sex, comorbidities (cancer, diabetes, coronary artery disease, hypertension, chronic respiratory diseases, obesity), and time from symptom onset to scan date. Next, we tested a second model (B) where we added CT-SS as an independent variable and a third model (C) where we added the automatic lesion extent quantification obtained by DL-driven segmentation. Second-order interaction terms between the scores and the covariates were tested in Models B and C. We used likelihood ratio tests for comparing models. To estimate the models’ ability to discriminate individuals, we computed the C-statistic on the validation dataset [22]. The optimal cutoff value for the automatic lesion extent quantification was selected based on the Youden index to maximize accuracy (sensitivity + specificity – 1).

A two-sided α of less than 0.05 was considered statistically significant. All analyses were carried out using SAS 9.4 statistical software (SAS Institute, Cary, NC).

3. Results

A total of 1785 patients were included, and the clinical characteristics, CT-SS and pulmonary lesion distributions of the training, validation, test and prognostic value datasets are shown in Table 1. Pulmonary lesions evaluated on LDCT from the training, validation and test datasets were extracted from the manual ground-truth segmentations (O1a). Those from the prognosis value dataset were derived from the model segmentation. An example of the automated segmentation results is shown in Fig. 3. The overall test dataset of LDCT scans had a median mean dose–length product of 38.75 ± 39.9 mGy.cm.

3.1. Segmentation evaluation

The results for the DSC and clinical parameters between the automatic and manual segmentations, as well as a comparison to the inter- and intraobserver performances, are shown in Table 2. The correlations between automatic and manual measures of lesion extent are presented in Table 2 and Fig. 4.

3.1.1. Segmentation accuracy

The DSC was 0.75 ± 0.08 for the overall lesion segmentations, 0.71 ± 0.10 for GGO segmentation, and 0.64 ± 0.09 for Cons segmentations. The MVSF results are presented in Appendix C.

The MAE was 70.3 ± 65.8 cm³ for the GGO volume, 29.5 ± 35.9 cm³ for the Cons volume and 71.4 ± 72.6 cm³ for the lesion volume. The biases were -18.3 ± 95.4 cm³ for the GGO volume, 14.4 ± 44.4 cm³ for the Cons volume and -3.9 ± 102.6 cm³ for the lesion volume, and none of these biases was found to be significant. In terms of disease extent, the MAE was $2.2 \pm 2.1\%$ for the GGO extent, $1.0 \pm 1.3\%$ for the Cons extent and $2.1 \pm 2.4\%$ for the lesion extent. The biases were not significant for the lesion extent quantification ($-0.1\% \pm 3.2$; $p = 0.59$). Disease extent measures were highly correlated with ground truth, with a lesion extent correlation of 0.947 ($p < 0.001$).

Concerning segmentation efficiency, the mean interaction time was significantly different between manual and automated segmentation: 14.74 ± 2.9 min versus 19 seconds ($p < 0.001$) for each patient.

3.1.2. Reproducibility

For lesion segmentation, the DSC was higher for the Auto vs. O1a evaluation (0.75 ± 0.08) than for the interobserver (O1a vs. O2: 0.70 ± 0.08 ; O1a vs. O3: 0.70 ± 0.08) or intraobserver agreement (0.72 ± 0.09). It was identical for the GGO and Cons segmentations. The automated lesion volume measures had an MAE of 71.4 ± 72.6 cm³; the interobserver MAEs were as follows: O1a vs. O2: 105.1 ± 102.6 cm³; O1a vs. O3: 122.8 ± 105.4 cm³. The intraobserver MAE was 117.0 ± 82.7 cm³. The correlation with the ground truth was higher for the automated measures of lesion volume (0.94) than for the interobserver (O1a vs. O2: 0.88; O1a vs. O3: 0.87) or intraobserver (0.91) measures. For lesion extent, the MAE was lower for the Auto vs. O1a evaluation ($2.1 \pm 2.4\%$) than for the interobserver (O1a vs. O2: $3.1 \pm 2.9\%$; O1a vs. O3: $3.9 \pm 3.7\%$) or intraobserver evaluations ($3.5 \pm 2.7\%$). The lesion extent correlation r was 0.947 for automated measures versus 0.909, 0.872 and 0.920 for the inter- and intraobserver measures. There were statistically significant biases in lesion extent for the O1a vs. O3 interobserver and intraobserver measures. Bland–Altman plots are presented in Appendix D.

3.2. Prognostic value

There were 227 patients (14%) in the prognostic value dataset who presented with the combined outcome (Table 3). After adjustment for baseline clinical characteristics, the global scores were significantly associated with outcome occurrence (“transfer to ICU and/or death and/or hospitalization ≥ 10 days and/or oxygen therapy”.) and the addition of GGO or Cons did not modify the prognostic prediction for either the human or automatic radiological score. The adjusted odds ratios were 3.02 (95% CI: 2.44; 3.73) for the CT-SS and 3.86 (95% CI: 2.96; 5.05) for automatic quantification. The C-statistic was 0.82 (0.79–0.88) in Model A excluding all radiological scores, 0.89 (0.95–0.93) in Model B including CT-SS and 0.90 (0.86–0.94) in Model C including DL-driven quantification. The differences between Models A and B and between Models A and C were statistically significant (likelihood ratio tests: $p < 0.001$). ROC curve analysis for lesion extents DL-driven quantification is shown on Appendix E.

4. Discussion

The main finding of the study was that the proposed automatic quantification pipeline provides an accurate and reproducible segmentation of GGOs and consolidations in COVID-19 infection. With respect to the human ground-truth segmentation, the variability of the model was lower than the inter- or intraobserver variability. The presented model was computationally efficient, requiring less than 20 seconds for complete DL-driven segmentation. Its accuracy was similar regardless of the extent of the lesions. Furthermore, the presented data showed that the automatic quantification of lesion extent provides a strong prognostic marker of adverse events during COVID-19 infection.

During the COVID-19 pandemic, diagnostic imaging has multiple roles, including diagnosis, prognosis, and follow-up [23]. One potential method to obtain a precise evaluation of disease-related lesions and prognosis is to quantify the extent of the lesions. This study proposes a distinct segmentation of different COVID-19 lesions, differentiating GGO from consolidation. Most previously published works have focused on automated algorithms that help distinguish COVID-19 infection from other pulmonary infections [24,25]. One of the main strengths of the present paper was the use of LDCT as input data. COVID-19 patients might undergo multiple CT examinations for diagnosis, follow-up and evaluation of complications of SARS-CoV-2 infection. At times when LDCT is encouraged in pneumonia diagnosis, automated algorithms should be adapted to these technical modifications [26].

Table 1
Characteristics of the different datasets.

| Characteristics of the different datasets | | | | | |
|--|------------------------------|-------------------------------|----------------------------|--------------------------------------|-------------|
| | Training dataset (n= 124) | Validation dataset (n= 20) | Testing Dataset (n= 30) | Prognosis value dataset (n= 1621) | |
| Sex, | | | | | |
| Men, n (%) | 63 (50.8) | 15 (75.0) | 14 (46.7) | 777 (47.9) | |
| Age | | | | | |
| Age 18–44 years, n (%) | 49 (39.5) | 6 (30.0) | 4 (13.3) | 583 (36.0) | |
| Age 45–64 years, n (%) | 58 (46.7) | 11 (55.0) | 17 (56.7) | 759 (46.8) | |
| Age >64 years, n (%) | 17 (13.7) | 3 (15.0) | 9 (30.0) | 279 (17.2) | |
| Time between symptom onset and LDCT | | | | | |
| ≤7 days/asymptomatic, n (%) | | 61 (49.2) | 10 (50.0) | 17 (56.7) | 1057 (65.2) |
| ≥7 days, n (%) | | 63 (50.8) | 10 (50.0) | 13 (43.3) | 564 (34.8) |
| Comorbidities | | | | | |
| Hypertension, n (%) | 17 (13.97) | 3 (15.0) | 10 (33.4) | 333 (20.5) | |
| Diabetes mellitus, n (%) | 10 (8.1) | 1 (5.0) | 8 (26.7) | 192 (11.8) | |
| Cancer, n (%) | 4 (2.8) | 0 (0.0) | 5 (16.7) | 75 (4.6) | |
| Respiratory diseases, n (%) | 19 (15.3) | 2 (10.0) | 1 (3.3) | 193 (11.9) | |
| Cardiac diseases, n (%) | 7 (5.6) | 1 (5.0) | 3 (10.0) | 135 (8.3) | |
| Obesity (BMI ≥30, n (%)) | 9 (7.3) | 2 (10.0) | 5 (16.7) | 233 (14.4) | |
| Medication | | | | | |
| Beta blockers, n (%) | | 6 (4.8) | 1 (5.0) | 3 (10.0) | 95 (5.9) |
| HMG-CoA reductase inhibitors, n (%) | | 3 (2.4) | 0 (0.0) | 5 (16.7) | 99 (6.1) |
| Dihydropyridine derivatives, n (%) | | 3 (2.4) | 0 (0.0) | 6 (20.0) | 87 (5.4) |
| Angiotensin II receptor blockers, n (%) | | 6 (4.8) | 1 (5.0) | 7 (23.3) | 104 (6.4) |
| ACE inhibitors, n (%) | | 5 (4.0) | 1 (5.0) | 2 (6.7) | 33 (2.0) |
| Symptoms | | | | | |
| Cough n (%) | 63 (50.8) | 12 (60.0) | 16 (53.3) | 658 (40.6) | |
| Rhinitis, n (%) | 23 (18.5) | 5 (25.0) | 4 (13.3) | 293 (18.1) | |
| Fever ≥ 38°C, n (%) | 31 (25.0) | 10 (50.0) | 9 (30.0) | 272 (16.8) | |
| Anosmia, n (%) | 36 (29.0) | 4 (20.0) | 6 (20.0) | 339 (20.9) | |
| Ageusia, n (%) | 29 (23.4) | 4 (20.0) | 6 (20.0) | 324 (20.0) | |
| Dyspnea, n (%) | 27 (21.7) | 3 (15.0) | 9 (30.0) | 326 (20.1) | |
| Thoracic pain, n (%) | 12 (9.7) | 4 (20.0) | 1 (3.3) | 212 (13.1) | |
| CT-SS | | | | | |
| Mean score/40 (standard deviation) | 8.6 (±8.1) | 7.5 (±9.1) | 18.2 (±6.5) | 6.0 (±7.2) | |
| Severe forms (CT—SS > 19.5) | 13 (10.5) | 2 (10.0) | 9 (30) | 127 (7.8) | |
| Pulmonary lesions | | | | | |
| Overall, n (%) | 104 (83.9) | 16 (80.0) | 30 (100.0) | 1408 (86.9) | |
| GGO, n (%) | 103 (83.1) | 16 (80.0) | 30 (100.0) | 1406 (86.7) | |
| Cons, n (%) | 83 (66.9) | 4 (20.0) | 29 (96.7) | 908 (56.0) | |
| GGO & Cons, n (%) | 82 (66.1) | 4 (20.0) | 29 (96.7) | 906 (55.9) | |
| Clinical outcomes | | | | | |
| Oxygen therapy (Oxy), n (%) | | 20 (19.1) | 5 (25.0) | 10 (33.3) | 180 (11.1) |
| ICU, n (%) | | 7 (5.6) | 2 (10.0) | 2 (6.7) | 43 (2.7) |
| Death, n (%) | | 1 (0.7) | 0 (0.0) | 2 (6.7) | 20 (1.2) |
| Hospitalization ≥10 days (Hospit10days), n (%) | | 13 (10.5) | 6 (30.0) | 7 (23.3) | 129 (8.0) |
| ICU/Death, n (%) | | 7 (5.6) | 2 (10.0) | 3 (10.0) | 57 (3.5) |
| ICU/Death/Hospit10days, n (%) | | 14 (11.3) | 6 (30.0) | 7 (23.3) | 150 (9.3) |
| ICU/Death/Hospit10days/Oxy, n (%) | | 21 (16.9) | 6 (30.0) | 12 (40.0) | 227 (14) |

Note — LDCT: low-dose computed tomography; BMI: body mass index; ACE: angiotensin-converting enzyme; CT-SS: chest tomography severity score; GGO: ground-glass opacity; Cons.: consolidation; ICU: intensive care unit.

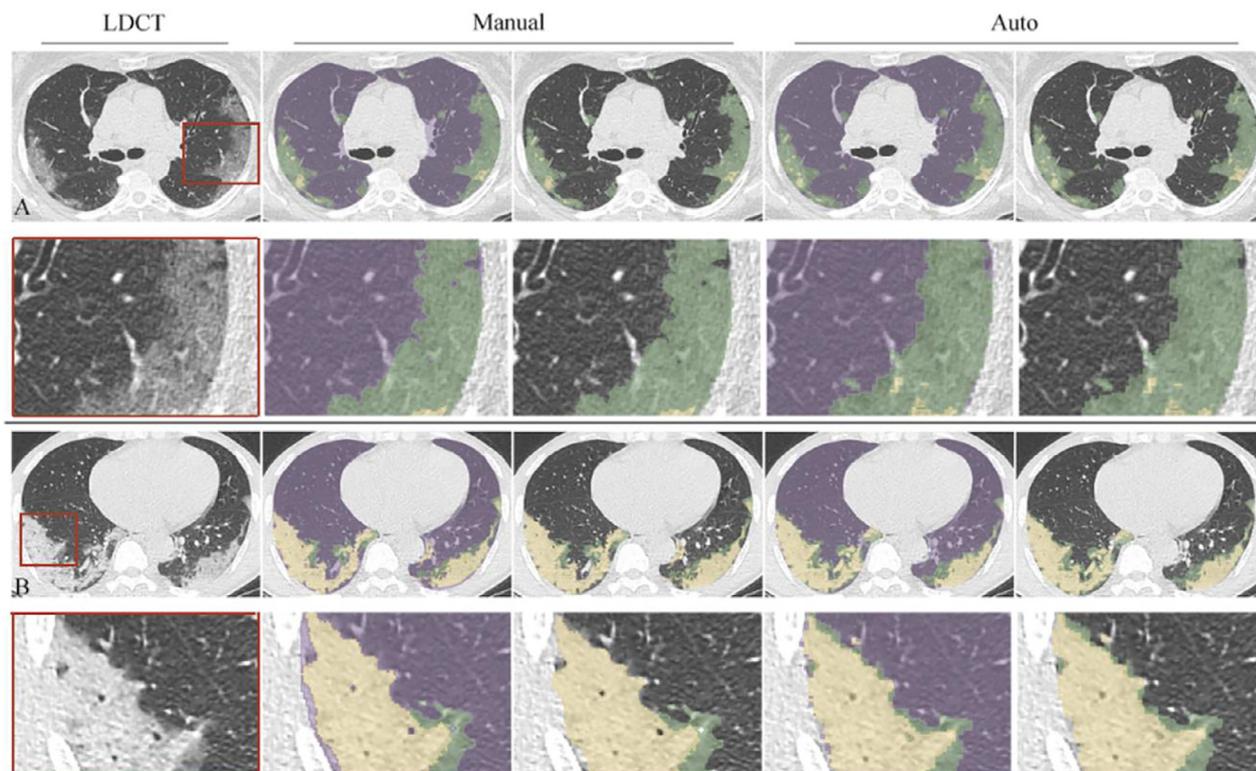


Fig. 3. Examples of the obtained automatic segmentations (Auto) compared to the corresponding LDCT images and manual reference segmentations (Manual).

Note — Normal lung (purple), consolidation (yellow), ground-glass opacity (green).

A. Example 1: Mid-thoracic carina level. The second row shows higher-magnification views of the areas in the red rectangles.

B. Example 2: Inferior mediastinum level. The second row shows higher-magnification views of the areas in the red rectangles.

The training dataset had substantial variability in pulmonary lesion extent and disease severity (from 0% to 36%). One of the main strengths of our study was that manual segmentation was conducted on all LDCT images in the training, validation and test datasets. Contrary to many segmentation models, the algorithm and obtained results were tested on all images in the test dataset (which was numbered at 15587 images for the 30 patients) rather than selected slices.

The literature has seen a wide number of CNN-based methodologies for automatic segmentation of lung abnormality on CT scan. Works may be divided in three categories: those that base the training on CT scans fully annotated by experts [21,27], those that make use of weak/noisy labels to lower the annotation load [20,28,29] and those using transfer learning to transfer knowledge from non-COVID19 lesions [30]). Regarding the network architectures, 2D CNNs [20,21,27] and 3D CNNs [27,28,30] are both represented. Some researchers focus on the detail of the architectures and advocate for additional modules, such as attention blocks [21,29]. Despite of the vast number of papers proposing new architectures and modules, 8 out of the 10 finalists in the COVID-19 Lung CT Lesion Segmentation Challenge chose an UNet architecture as we propose [27]

In 2020, Belfiore et al. highlighted the need to quantify the percentage of ventilated lung parenchyma as distinguished from the affected lung parenchyma [31]. Here, we propose a segmentation tool that differentiates normal from affected lungs (GGO and Cons). Cons DSC, volume and extent measures were always lower than the GGO measures. Interestingly, this demonstrates the difficulty of producing Cons segmentations. This could be due to the anatomic presentation of COVID-19 consolidations, which mostly have a sub pleural distribution and affect the lower segments [9]. Hence, consolidations are sometimes in continuity with the sub pleural fat and the

chest wall, which can lead to segmentation failure. Consolidations were the only measure whose correlation was lower for the automated measure (*Auto* vs. *O1a*) than for the interhuman measure (*O1a* vs. *O3*). This finding suggests that our model might fail partially in cases of peripheral and lower lobe consolidation. Liu et al. proposed CT quantification of pneumonia lesions to predict the progression of severe disease and distinguished three labels: consolidation, semi-consolidation and ground glass [32]. They used a simple threshold to differentiate these labels. The authors obtained a DSC of 0.82 for COVID-19 pneumonia but did not publish the algorithmic details, biases or correlations. Chassagnon et al. presented a COVID-19 segmentation algorithm with a mean lesion DSC between automated and manual segmentation of 0.69 [25]. For GGO lesion segmentation, the DSC of 0.71 ± 0.10 in the present study was below that of Jung et al. for the automated segmentation of GGO (0.78 ± 0.07) [33]. This discordance can probably be explained by the difference in morphological patterns between parenchymal lesions and nodules.

Among all tested factors, age remained the best predictor of clinical outcome. However, the C-statistic was significantly improved when DL-driven quantification was added for the combined outcome, which confirms the benefit of adding the radiological score to evaluate the prognosis. DL-driven quantification was not superior to the CT-SS in predicting the occurrence of clinical outcomes but did not require any human input. Concerning gender, 'men' is no longer a risk factor after adjustment on CTSS and automatic CT scores. This was due to a significant difference in CT scores between men and women. The same statistical reason can explain hypertension results. The present code is protected (IDDN.FR.001.220003.000.S.C.2020.000.31235) and can be shared upon the signing of a collaboration agreement.

Table 2
Model segmentation performances in comparison to human reproducibility on the test dataset ($n=30$).

| Model segmentation performance in comparison to human reproducibility | | | | |
|---|----------------------|-----------------------|-----------------------|-----------------------|
| | <i>Auto vs. O1a</i> | <i>O1a vs. O2</i> | <i>O1a vs. O3</i> | <i>O1a vs. O1b</i> |
| Technical metrics: Dice similarity coefficient | | | | |
| LungN | 0.99 (± 0.01) | x | x | x |
| GGO | 0.71 (± 0.10) | 0.64 (± 0.15) | 0.63 (± 0.15) | 0.62 (± 0.14) |
| Cons. | 0.64 (± 0.09) | 0.54 (± 0.19) | 0.64 (± 0.12) | 0.57 (± 0.17) |
| Lesion | 0.75 (± 0.08) | 0.70 (± 0.08) | 0.70 (± 0.10) | 0.72 (± 0.09) |
| Clinical parameters: volume and extent | | | | |
| Volume | | | | |
| GGO | | | | |
| MAE (cm ³) | 70.3 (± 65.8) | 140.3 (± 126.2) | 117.3 (± 106.5) | 100.9 (± 73.4) |
| Bias (cm ³) | -18.3 (± 95.4) | 21.2 (± 189.3) | 68.3 (± 144.1) | -30.6 (± 122.3) |
| <i>p</i> | 0.29 | 0.80 | 0.02 | 0.20 |
| Corr. | 0.940 | 0.757 | 0.857 | 0.898 |
| Cons. | | | | |
| MAE (cm ³) | 29.5 (± 35.9) | 62.6 (± 81.8) | 29.3 (± 31.7) | 68.1 (± 57.6) |
| Bias (cm ³) | 14.4 (± 44.4) | -24.9 (± 100.5) | -2.9 (± 43.4) | -36.2 (± 82.2) |
| <i>p</i> | 0.39 | 0.49 | 0.89 | 0.03 |
| Corr. | 0.902 | 0.792 | 0.937 | 0.733 |
| Lesion | | | | |
| MAE (cm ³) | 71.4 (± 72.6) | 105.1 (± 102.6) | 122.8 (± 105.4) | 117.0 (± 82.7) |
| Bias (cm ³) | -3.9 (± 102.6) | -3.8 (± 148.1) | 65.4 (± 149.3) | -66.8 (± 128.0) |
| <i>p</i> | 0.88 | 0.96 | 0.03 | 0.01 |
| Corr. | 0.941 | 0.880 | 0.873 | 0.910 |
| Extent | | | | |
| GGO | | | | |
| MAE (%) | 2.2 (± 2.1) | 4.3 (± 3.8) | 3.8 (± 4.0) | 3.3 (± 2.7) |
| Bias (%) | -0.6 (± 3.0) | 0.8 (± 5.7) | 2.3 (± 5.0) | -1.3 (± 4.1) |
| <i>p</i> | 0.25 | 0.70 | 0.03 | 0.13 |
| Corr. | 0.940 | 0.766 | 0.822 | 0.884 |
| Cons. | | | | |
| MAE (%) | 1.0 (± 1.3) | 2.1 (± 2.8) | 1.0 (± 1.1) | 2.2 (± 1.9) |
| Bias (%) | 0.5 (± 1.6) | -0.7 (± 3.5) | -0.1 (± 1.5) | -0.7 (± 2.8) |
| <i>p</i> | 0.56 | 0.61 | 0.83 | 0.08 |
| Corr. | 0.882 | 0.754 | 0.926 | 0.673 |
| Lesion | | | | |
| MAE (%) | 2.1 (± 2.4) | 3.1 (± 2.9) | 3.9 (± 3.7) | 3.5 (± 2.7) |
| Bias (%) | -0.1 (± 3.2) | 0.0 (± 4.3) | 2.2 (± 4.9) | -2.0 (± 4.0) |
| <i>p</i> | 0.59 | 0.96 | 0.03 | <0.01 |
| Corr. | 0.947 | 0.909 | 0.872 | 0.920 |

Note — DSC: Dice similarity coefficient; MVSF: mean value similarity function; LungN: normal lung; GGO: ground-glass opacity; Cons.: consolidation. Lesion: GGO + Cons. The means \pm standard deviations of the metrics are reported.

MAE: mean absolute error. The means and standard deviations (in parentheses) of the absolute differences are reported. Bold characters represent significant results.

Our study has some limitations. All CT images were acquired on the same CT scanner in one clinical center. Additionally, the presented algorithm cannot provide a segmentation of the distal vascular and bronchial trees. A future goal of our work should be to include arterial and bronchial segmentation in our algorithm for even more precise lesion segmentation.

5. Conclusion

A complete DL-driven pipeline for LDCT, which allows minimum radiation exposure, was developed to segment GGOs and consolidation due to COVID-19 lung involvement. The algorithm produces automatic lesion volume and extent measures that can be directly

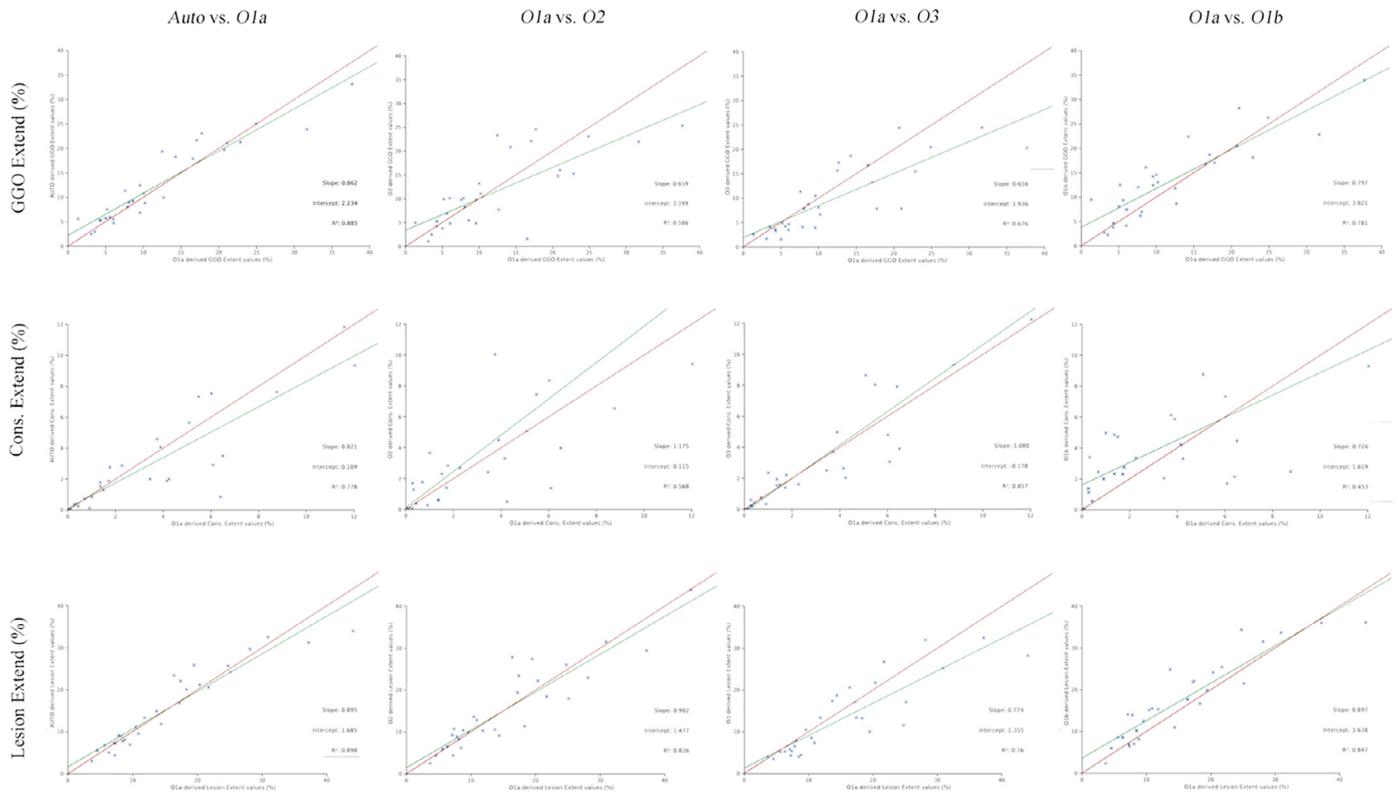


Fig. 4. Correlation of disease extent measures between automatic and manual segmentations on the test dataset ($n=30$). Note — Evaluation of Auto vs. O1a (Column 1), O1a vs. O2 (Column 2), O1a vs. O3 (Column 3), of O1a vs. O1b (Column 4). The green line is the fitted regression line. The red line is the identity line. GGO: ground-glass opacity; Cons: consolidation.

provided to physicians. DL-driven segmentation was more reproducible than human measures, achieving lower biases and mean absolute error than human inter- and intraobserver comparisons of lesion

volume and extent. Lung involvement as quantified by our DL-driven pipeline was significantly associated with the occurrence of adverse events. This framework should be tested on multicenter datasets to evaluate disease severity at the time of the first LDCT evaluation.

Table 3
Multivariate logistic regressions for the primary endpoint.

| | Multivariate logistic regressions for the primary endpoint Death/ICU/Hospit>10d/Oxy (n=227, 14%) | | |
|---|---|-----------------------------------|-----------------------------------|
| | Model A OR 95% CI ^a | Model B OR 95% CI ^a | Model C OR 95% CI ^a |
| Sex (ref. women) | | | |
| Men | 1.47[1.01;2.15] | 1.17[0.77;1.77] | 1.00[0.64;1.55] |
| Age (ref. 18-44 years) | | | |
| 45-64 years | 4.29[2.22;8.30] | 2.30[1.14;4.61] | 2.35[1.16;4.77] |
| >64 years | 15.34[7.60;30.97] | 8.76[4.18;18.34] | 6.85[3.21;14.61] |
| Hypertension | 1.65[1.09;2.50] | 1.44[0.90;2.29] | 1.94[1.19;3.16] |
| Diabetes mellitus | 1.27[0.78;2.07] | 1.06[0.61;1.84] | 0.97[0.54;1.73] |
| Cancer | 2.20[1.09;4.43] | 2.21[1.04;4.67] | 2.34[1.06;5.20] |
| Respiratory diseases | 0.81[0.46;1.43] | 0.80[0.41;1.56] | 0.99[0.52;1.88] |
| Cardiac diseases | 1.89[1.10;3.24] | 1.99[1.08;3.66] | 1.52[0.79;2.92] |
| Time between symptoms/LDCT (ref. ≤ 7 days) | | | |
| > 7 days | 1.08[0.73;1.59] | 0.71[0.46;1.10] | 0.86[0.54;1.37] |
| CT-SS | | 3.02[2.44;3.73] | |
| Automatic lesion extent | | | 3.86[2.96;5.05] |
| C statistic (95% CI)^b | 0.82 [0.79-0.88] | 0.89 [0.85-0.93] | 0.90 [0.86-0.94] |

Note — a: Adjusted odds ratios with 95% confidence intervals. b: The C-statistic is a measure of goodness of fit for binary outcomes in a logistic regression model. It is equal to the area under the receiver operating characteristic (ROC) curve and ranges from 0.5 to 1. Models were based on the training set of the prognostic value dataset ($n=1135$), and the C-statistic was estimated on the validation set ($n=486$) of the prognostic value dataset. All scores were standardized (mean=0, standard deviation=1) prior to the analysis. LDCT: low-dose computed tomography; ICU: intensive care unit.

Author contributions

See Authors statement

Institutional review board statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of ASSISTANCE-PUBLIQUE DES HOPITAUX DE MARSEILLE (AP-HM) (N°: 2020-0012, RGPD/Ap-Hm: 2020-48).

Data availability statement

None.

Funding

The authors state that this work has not received any funding.

Guarantor

The scientific guarantor of this publication is Pr. Alexis Jacquier.

Statistics and biometry

One of the authors has significant statistical expertise.

Informed consent

Only if the study is on human subjects:

Written informed consent was waived by the Institutional Review Board.

Ethical approval

Institutional Review Board approval was obtained.(N°:2020-0012, RGP/DAp-Hm: 2020-48)

Participations

Conception and design of study: Axel Bartoli, Jean-Yves GAUBERT, Jean-Christophe LAGIER, Badih GHATTAS, Alexis JACQUIER.

Acquisition of data: Axel Bartoli, Arnaud MAURIN, Baptiste MARCHI, Paul HABERT, Maxime CASTELLI, Jean-Christophe LAGIER.

Analysis and/or interpretation of data: Sebastien CORTAREDONA, Axel BARTOLI

Drafting the manuscript: Alexis JACQUIER, Matthieu MILLION, Didier RAOULT

Revising the manuscript critically for important intellectual content: Alexis JACQUIER, Didier RAOULT Jean-Yves GAUBERT, Jean-Christophe LAGIER, Matthieu MILLION, Badih GHATTAS.

Approval of the version of the manuscript to be published: Axel BARTOLI, Joris FOURNEL, Arnaud MAURIN, Baptiste MARCHI, Paul HABERT, Maxime CASTELLI, Jean-Yves GAUBERT, Sebastien CORTAREDONA, Jean-Christophe LAGIER, Matthieu MILLION, Didier RAOULT, Badih GHATTAS, Alexis JACQUIER.

Declaration of Competing Interest

The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

CRedit authorship contribution statement

Axel Bartoli: Conceptualization, Data curation, Investigation, Software, Supervision, Writing – original draft. **Joris Fournel:** Formal analysis, Investigation, Software. **Arnaud Maurin:** Data curation. **Baptiste Marchi:** Data curation. **Paul Habert:** Resources. **Maxime Castelli:** Data curation. **Jean-Yves Gaubert:** Methodology. **Sebastien Cortaredona:** Formal analysis. **Jean-Christophe Lagier:** Methodology, Resources. **Matthieu Million:** Writing – review & editing. **Didier Raoult:** Writing – review & editing. **Badih Ghattas:** Methodology, Supervision, Validation, Visualization, Writing – review & editing. **Alexis Jacquier:** Conceptualization, Funding acquisition, Methodology, Project administration, Validation, Writing – review & editing.

Acknowledgements

Authors would like to thank all the paramedical staff of our department who are managing the COVID-19 crisis with professionalism and effectiveness.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.redii.2022.100003](https://doi.org/10.1016/j.redii.2022.100003).

References

- [1] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020;382:727–33. doi: [10.1056/NEJMoa2001017](https://doi.org/10.1056/NEJMoa2001017).
- [2] Munster VJ, Koopmans M, van Doremalen N, van Riel D, de Wit E. A novel coronavirus emerging in China - key questions for impact assessment. *N Engl J Med* 2020;382:692–4. doi: [10.1056/NEJMp2000929](https://doi.org/10.1056/NEJMp2000929).
- [3] Zhou T-T, Wei F-X. Primary stratification and identification of suspected Corona virus disease 2019 (COVID-19) from clinical perspective by a simple scoring proposal. *Mil Med Res* 2020;7:16. doi: [10.1186/s40779-020-00246-8](https://doi.org/10.1186/s40779-020-00246-8).
- [4] Chung M, Bernheim A, Mei X, Zhang N, Huang M, Zeng X, et al. CT imaging features of 2019 novel coronavirus (2019-nCoV). *Radiology* 2020;295:202–7. doi: [10.1148/radiol.2020200230](https://doi.org/10.1148/radiol.2020200230).
- [5] Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, et al. Sensitivity of Chest CT for COVID-19: comparison to RT-PCR. *Radiology* 2020;200432. doi: [10.1148/radiol.2020200432](https://doi.org/10.1148/radiol.2020200432).
- [6] Jalaber C, Lapotre T, Morcet-Delattre T, Ribet F, Jouneau S, Lederlin M. Chest CT in COVID-19 pneumonia: a review of current knowledge. *Diagn Interv Imaging* 2020;101:431–7. doi: [10.1016/j.diii.2020.06.001](https://doi.org/10.1016/j.diii.2020.06.001).
- [7] Hani C, Trieu NH, Saab I, Dangeard S, Bennani S, Chassagnon G, et al. COVID-19 pneumonia: a review of typical CT findings and differential diagnosis. *Diagn Interv Imaging* 2020;101:263–8. doi: [10.1016/j.diii.2020.03.014](https://doi.org/10.1016/j.diii.2020.03.014).
- [8] Li K, Wu J, Wu F, Guo D, Chen L, Fang Z, et al. The Clinical and chest CT features associated with severe and critical COVID-19 pneumonia. *Invest Radiol* 2020;55:327–31. doi: [10.1097/RLI.0000000000000672](https://doi.org/10.1097/RLI.0000000000000672).
- [9] Yuan M, Yin W, Tao Z, Tan W, Hu Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. *PLoS ONE* 2020;15:e0230548. doi: [10.1371/journal.pone.0230548](https://doi.org/10.1371/journal.pone.0230548).
- [10] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciampi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal* 2017;42:60–88. doi: [10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005).
- [11] Maldonado F, Moua T, Rajagopalan S, Karwoski RA, Raghunath S, Decker PA, et al. Automated quantification of radiological patterns predicts survival in idiopathic pulmonary fibrosis. *Eur Respir J* 2014;43:204–12. doi: [10.1183/09031936.00071812](https://doi.org/10.1183/09031936.00071812).
- [12] Zhang G, Jiang S, Yang Z, Gong L, Ma X, Zhou Z, et al. Automatic nodule detection for lung cancer in CT images: a review. *Comput Biol Med* 2018;103:287–300. doi: [10.1016/j.combiomed.2018.10.033](https://doi.org/10.1016/j.combiomed.2018.10.033).
- [13] Million M, Lagier J-C, Gautret P, Colson P, Fournier P-E, Amrane S, et al. Full-length title: early treatment of COVID-19 patients with hydroxychloroquine and azithromycin: a retrospective analysis of 1061 cases in Marseille, France. *Travel Med Infect Dis* 2020;101738. doi: [10.1016/j.tmaid.2020.101738](https://doi.org/10.1016/j.tmaid.2020.101738).
- [14] Gautret P, Lagier J-C, Parola P, Hoang VT, Meddeb L, Sevestre J, et al. Clinical and microbiological effect of a combination of hydroxychloroquine and azithromycin in 80 COVID-19 patients with at least a six-day follow up: a pilot observational study. *Travel Med Infect Dis* 2020;34:101663. doi: [10.1016/j.tmaid.2020.101663](https://doi.org/10.1016/j.tmaid.2020.101663).
- [15] Amrane S, Tissot-Dupont H, Doudier B, Eldin C, Hocquart M, Mailhe M, et al. Rapid viral diagnosis and ambulatory management of suspected COVID-19 cases presenting at the infectious diseases referral hospital in Marseille, France, - January 31st to March 1st, 2020: a respiratory virus snapshot. *Travel Med Infect Dis* 2020;101632. doi: [10.1016/j.tmaid.2020.101632](https://doi.org/10.1016/j.tmaid.2020.101632).
- [16] Yang R, Li X, Liu H, Zhen Y, Zhang X, Xiong Q, et al. Chest CT severity score: an imaging tool for assessing severe COVID-19. *Radiol Cardiothorac Imaging* 2020;2:e200047. doi: [10.1148/ryct.2020200047](https://doi.org/10.1148/ryct.2020200047).
- [17] Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 2012;30:1323–41. doi: [10.1016/j.mri.2012.05.001](https://doi.org/10.1016/j.mri.2012.05.001).
- [18] Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner society: glossary of terms for thoracic imaging. *Radiology* 2008;246:697–722. doi: [10.1148/radiol.2462070712](https://doi.org/10.1148/radiol.2462070712).
- [19] Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology* 2018;286:800–9. doi: [10.1148/radiol.2017171920](https://doi.org/10.1148/radiol.2017171920).
- [20] Wang G, Liu X, Li C, Xu Z, Ruan J, Zhu H, et al. A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Trans Med Imaging* 2020;39:2653–63. doi: [10.1109/TMI.2020.3000314](https://doi.org/10.1109/TMI.2020.3000314).
- [21] Zhao S, Li Z, Chen Y, Zhao W, Xie X, Liu J, et al. SCOAT-Net: a novel network for segmenting COVID-19 lung opacification from CT images. *Pattern Recognit* 2021;119:108109. doi: [10.1016/j.patcog.2021.108109](https://doi.org/10.1016/j.patcog.2021.108109).
- [22] Hosmer DW, Lemeshow S. *Applied logistic regression*. 2nd ed. New York: Wiley; 2000.
- [23] Cellina M, Orsi M, Bombaci F, Sala M, Marino P, Oliva G. Favorable changes of CT findings in a patient with COVID-19 pneumonia after treatment with tocilizumab. *Diagn Interv Imaging* 2020;101:323–4. doi: [10.1016/j.diii.2020.03.010](https://doi.org/10.1016/j.diii.2020.03.010).
- [24] Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B, et al. Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology* 2020;200905. doi: [10.1148/radiol.2020200905](https://doi.org/10.1148/radiol.2020200905).
- [25] Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi T-N, Dangeard S, et al. AI-Driven CT-based quantification, staging and short-term outcome prediction of COVID-19 pneumonia. *Infect Dis (except HIV/AIDS)* 2020. doi: [10.1101/2020.04.17.20069187](https://doi.org/10.1101/2020.04.17.20069187).
- [26] Hamard A, Frandon J, Larbi A, Goupil J, De Forges H, Beregi J-P, et al. Impact of ultra-low dose CT acquisition on semi-automated RECIST tool in the evaluation of malignant focal liver lesions. *Diagn Interv Imaging* 2020;101:473–9. doi: [10.1016/j.diii.2020.05.003](https://doi.org/10.1016/j.diii.2020.05.003).
- [27] Roth H, Xu Z, Diez CT, Jacob RS, Zember J, Molto J, et al. Rapid artificial intelligence solutions in a pandemic - the COVID-19-20 lung CT lesion segmentation challenge. *Res Sq* 2021 rs.3.rs-571332. doi: [10.21203/rs.3.rs-571332/v1](https://doi.org/10.21203/rs.3.rs-571332/v1).
- [28] Yang D, Xu Z, Li W, Myronenko A, Roth HR, Harmon S, et al. Federated semi-supervised learning for COVID region segmentation in chest CT using multi-

- national data from China, Italy, Japan. *Med Image Anal* 2021;70:101992. doi: [10.1016/j.media.2021.101992](https://doi.org/10.1016/j.media.2021.101992).
- [29] Fan D-P, Zhou T, Ji G-P, Zhou Y, Chen G, Fu H, et al. Inf-Net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging* 2020;39:2626–37. doi: [10.1109/TMI.2020.2996645](https://doi.org/10.1109/TMI.2020.2996645).
- [30] Wang Y, Zhang Y, Liu Y, Tian J, Zhong C, Shi Z, et al. Does non-COVID-19 lung lesion help? investigating transferability in COVID-19 CT image segmentation. *Comput Methods Programs Biomed* 2021;202:106004. doi: [10.1016/j.cmpb.2021.106004](https://doi.org/10.1016/j.cmpb.2021.106004).
- [31] Belfiore MP, Urraro F, Grassi R, Giacobbe G, Patelli G, Cappabianca S, et al. Artificial intelligence to codify lung CT in Covid-19 patients. *Radiol Med (Torino)* 2020;125:500–4. doi: [10.1007/s11547-020-01195-x](https://doi.org/10.1007/s11547-020-01195-x).
- [32] Liu F, Zhang Q, Huang C, Shi C, Wang L, Shi N, et al. CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics* 2020;10:5613–22. doi: [10.7150/thno.45985](https://doi.org/10.7150/thno.45985).
- [33] Jung J, Hong H, Goo JM. Ground-glass nodule segmentation in chest CT images using asymmetric multi-phase deformable model and pulmonary vessel removal. *Comput Biol Med* 2018;92:128–38. doi: [10.1016/j.combiomed.2017.11.013](https://doi.org/10.1016/j.combiomed.2017.11.013).