



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.



Cardiothoracic Imaging

Comparison of deep learning, radiomics and subjective assessment of chest CT findings in SARS-CoV-2 pneumonia

Chiara Arru^a, Shadi Ebrahimian^a, Zeno Falaschi^b, Jacob Valentin Hansen^c, Alessio Pasche^b, Mads Dam Lyhne^c, Mathis Zimmermann^d, Felix Durlak^d, Matthias Mitschke^d, Alessandro Carriero^b, Jens Erik Nielsen-Kudsk^c, Mannudeep K. Kalra^{a,*}, Luca Saba^e

^a Massachusetts General Hospital, 55 Fruit Street, Boston, MA, USA

^b Ospedale Maggiore della Carità, Novara, Italy

^c Department of Cardiology, Department of Clinical Medicine, Aarhus University Hospital, Palle Juul Jensens Boulevard 99, 8200 Aarhus N, Denmark

^d Siemens Healthcare GmbH, Diagnostic Imaging, Erlangen, Germany

^e Azienda Ospedaliera Universitaria di Cagliari, Cagliari, Italy

ARTICLE INFO

Keywords:

SARS-CoV-2

Chest CT

Pneumonia

Deep learning

Radiomics

ABSTRACT

Purpose: Comparison of deep learning algorithm, radiomics and subjective assessment of chest CT for predicting outcome (death or recovery) and intensive care unit (ICU) admission in patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection.

Methods: The multicenter, ethical committee-approved, retrospective study included non-contrast-enhanced chest CT of 221 SARS-CoV-2 positive patients from Italy (n = 196 patients; mean age 64 ± 16 years) and Denmark (n = 25; mean age 69 ± 13 years). A thoracic radiologist graded presence, type and extent of pulmonary opacities and severity of motion artifacts in each lung lobe on all chest CTs. Thin-section CT images were processed with CT Pneumonia Analysis Prototype (Siemens Healthineers) which yielded segmentation masks from a deep learning (DL) algorithm to derive features of lung abnormalities such as opacity scores, mean HU, as well as volume and percentage of all-attenuation and high-attenuation (opacities >−200 HU) opacities. Separately, whole lung radiomics were obtained for all CT exams. Analysis of variance and multiple logistic regression were performed for data analysis.

Results: Moderate to severe respiratory motion artifacts affected nearly one-quarter of chest CTs in patients. Subjective severity assessment, DL-based features and radiomics predicted patient outcome (AUC 0.76 vs AUC 0.88 vs AUC 0.83) and need for ICU admission (AUC 0.77 vs AUC 0.80 vs 0.82). Excluding chest CT with motion artifacts, the performance of DL-based and radiomics features improve for predicting ICU admission.

Conclusion: DL-based and radiomics features of pulmonary opacities from chest CT were superior to subjective assessment for differentiating patients with favorable and adverse outcomes.

1. Introduction

Managing healthcare resources is crucial in a high prevalence communicable disease such as the ongoing severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. It requires attention to several issues from screening and treatment to anticipation and

management of hospital personnel and resources. Patients with severe disease often require hospitalization with intensive care unit (ICU) admission and mechanical ventilation and high demands of resources.^{1–3} Improved assessment of disease severity and prediction of ICU admission and outcome is warranted as it potentially could help anticipating need of hospital resources.

Abbreviations: DL, Deep learning; ANOVA, Analysis of variance; RT-PCR, Reverse transcriptase-polymerase chain reaction; CXR, Chest X-ray; CT, Computed tomography.

* Corresponding author at: 75 Blossom Court, Boston, MA 02114, USA.

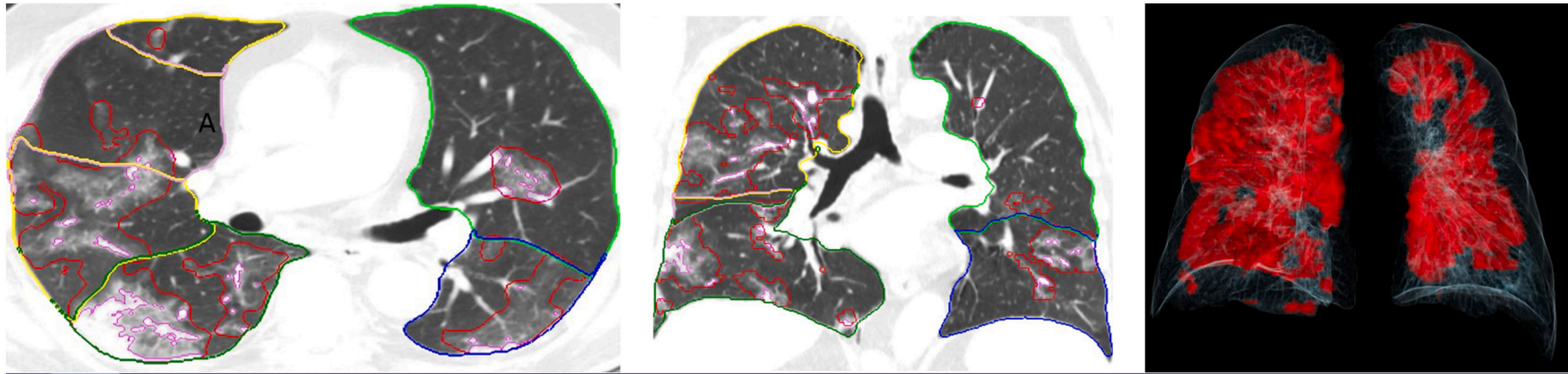
E-mail addresses: sebrahimian@mg.harvard.edu (S. Ebrahimian), jvh@clin.au.dk (J.V. Hansen), mads.dam@clin.au.dk (M.D. Lyhne), mathis.zimmermann@siemens-healthineers.com (M. Zimmermann), felix.durlak.ext@siemens-healthineers.com (F. Durlak), matthias.mitschke@siemens-healthineers.com (M. Mitschke), mkalra@mg.harvard.edu (M.K. Kalra).

<https://doi.org/10.1016/j.clinimag.2021.06.036>

Received 22 February 2021; Received in revised form 23 June 2021; Accepted 28 June 2021

Available online 1 July 2021

0899-7071/© 2021 Published by Elsevier Inc.



L/R	Lobes	Both lungs	Left upper lobe	Left lower lobe	Right upper lobe	Right middle lobe	Right lower lobe
	Affected	Yes	Yes	Yes	Yes	Yes	Yes
	Opacity Score	8	1	2	2	1	2
	Lung volume (ml)	5481.79	1564.36	871.70	1209.14	639.89	1196.69
	Volume of opacity (ml)	1179.89	90.70	259.03	457.04	57.53	315.58
	Percentage of opacity	21.52	5.80	29.72	37.80	8.99	26.37
	Volume of high opacity (ml)	135.95	9.58	47.30	41.37	2.37	35.34
	Percentage of high opacity	2.48	0.61	5.43	3.42	0.37	2.95
	Mean HU total	-740.73	-798.30	-662.85	-709.91	-822.51	-709.59
	Mean HU of opacity	-551.06	-556.83	-485.86	-568.30	-666.25	-556.95
	Standard deviation total	195.97	152.74	232.11	206.16	132.62	198.92
	Standard deviation of opacity	233.77	227.28	249.09	223.00	187.58	231.79

Fig. 1. Transverse and coronal sections of a non-contrast chest CT with contours outlining lungs, lobes and parenchymal opacities in a 73-year-old male. The table summarizes the list of DL variables obtained from the prototype. The volume rendered image demonstrates (top right side) displays the involved lung parenchyma in red color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Summary of DL-based features and subjective severity assessment scores in chest CT with different grades of respiratory motion artifacts (mild, moderate and severe). Both subjective severity assessment and DL-based features suggested extensive pulmonary opacities in patients with moderate to severe artifacts as compared to those with mild or no motion artifacts.

Variables	No motion (n = 167)	Mild (n = 1)	Moderate (n = 36)	Severe (n = 17)	p-Value
Volume of opacity	705 ± 726	854	983 ± 768	1227 ± 707	0.016
Percentage of opacity	18 ± 19	18	32 ± 26	42 ± 23	<0.001
Mean HU of opacity	−534 ± 122	−430	−458 ± 142	−466 ± 101	0.004
Subjective severity assessment	12 ± 5	0	13 ± 6	15 ± 5	0.009

Table 2

Summary of patient demographics, subjective assessment, DL-based and radiomics features based for need for ICU admission in patients with SARS-CoV-2 pneumonia.

	Need for ICU admission	
	With motion artifacts n = 221	Without motion artifacts n = 167
Mean age (years)	65 ± 16.4	62 ± 16.4
Gender M/F	138/83	106/61
Subjective assessment	Extent of opacity (AUC 0.768; p < 0.00001)	Extent of opacity + lymphadenopathy (AUC 0.805; p = 0.02)
DL-based features	Volume of opacity (AUC 0.772; p < 0.00001)	Percentage of opacity (AUC 0.801; p value < 0.0001)
Radiomics	Wavelet-LLL glszm Zone Entropy + wavelet-HLH glcm MCC (AUC 0.784; p = 0.002)	Wavelet-LLL glszm Zone Entropy (AUC 0.822; p < 0.0001)
DL + Radiomics	Volume of opacity + wavelet-HLL glszm Zone Entropy + original glrlm Gray Level Variance + Mean HU of opacity (AUC 0.812; p = 0.01)	Wavelet-LLL glszm Zone Entropy (AUC 0.822; p < 0.0001)

The current interpretation of radiology report in patients with or without SARS-CoV-2 is qualitative with semantic description of type and extent of distribution of findings as focal, multifocal, unilateral, bilateral, or diffuse. Some studies described a subjective severity scoring system based on chest CT images and reported good predictive value for outcome and need for mechanical ventilation or ICU admission.^{1–3} Others redeployed Radiographic Assessment of Lung Edema (RALE) score for assessing severity of lung involvement in patients with SARS-CoV-2 infection on chest radiography.⁴ These subjective scoring systems are time-consuming, inefficient, and likely prone to substantial inter- and intra-radiologist variations since they are not part of routine interpretation in most practices.⁶ To obtain quantitative data on the extent or severity of pulmonary involvement, investigators applied radiomics and deep learning (DL)-based algorithms in patients with SARS-CoV-2 infection.^{6–11} However, to the best of our knowledge, no prior studies have compared the performance of subjective severity scores and quantitative radiomics and DL algorithms for prediction of patient outcome and need for ICU admission. Likewise, it is unknown how individual performance of these methods varies in absence or presence of respiratory motion artifacts, which are common in often short of breath patients with SARS-CoV-2 infection.¹⁷ The purpose of our study was to compare deep learning algorithm, radiomics and subjective assessment of chest CT for predicting outcome and intensive care unit (ICU) admission in patients with SARS-CoV-2 infection.

2. Methods

2.1. Ethical committee approvals and disclosures

The retrospective study received approval from the respective human research ethical committees for analysis and sharing of de-identified data. We did not receive any research funding for the CT Pneumonia Analysis Prototype (Siemens Healthineers). One participating hospital (Massachusetts General Hospital) received unrelated research funding from GE Healthcare, Lunit Inc., Riverain Tech, and Siemens Healthineers. MZ, FD, and MM are employees of Siemens Healthineers who did not participate in the study subject selection or data analysis. All coauthors had unrestricted access to the manuscript.

2.2. Research subjects

The inclusion criteria for our study were age greater than 18 years, thin-section, non-contrast-enhanced chest CT (≤ 2 mm section thickness), positive reverse transcription polymerase chain reaction (RT-PCR assay) for SARS-CoV-2 infection, information on ICU admission, outcome (death versus recovery from SARS-CoV-2 infection), and ability to process image datasets with the DL algorithm and radiomics. Chest CT images from three patients were excluded since these could not be processed with the DL algorithm and the radiomics software.

Of the 221 consecutive patients who met the inclusion criteria at the two tertiary care hospitals there were 196 patients from Ospedale Maggiore della Carita', Novara, Italy (Site 1) and 25 patients from Aarhus University Hospital, Aarhus, Denmark (Site 2). The respective mean (\pm standard deviation) ages of patients from Sites 1 and 2 were 64 \pm 16.6 years and 69 \pm 13.3 years. There were more male than female patients at both sites (Site 1: 122 males; 74 females; Site 2: 16 males, 9 females). Details of patient age, gender, ICU admission for SARS-CoV-2 infection, and outcome (death versus recovery) were recorded.

2.3. Non-contrast-enhanced chest CT

At both sites, all non-contrast chest CT examinations were performed using respective standard of care routine chest CT protocols. All CTs were clinically indicated and performed for assessing extent or severity or complications (non-vascular) of SARS-CoV-2 pneumonia. Only the index chest CT exam was included for each patient.

Patients at both sites were instructed to lie supine and scanned during a breath-hold in full inspiration. At Site 1, all patients were scanned on a 128-slice multidetector-row CT (Philips Ingenuity Core, Philips Healthcare, Netherlands) with 120 kV, 225 mAs (using automatic exposure control technique – Z-DOM, Philips), 1.1:1 pitch factor, 0.5-second gantry rotation time, and 64 * 0.625 mm detector configuration. Thin-section images were reconstructed at 1 mm thickness using a soft tissue reconstruction kernel (Filter B).

All chest CTs at Site 2 were performed on a 16-slice, multidetector-row CT (Siemens SOMATOM Emotion 16, Siemens Healthineers, Forchheim, Germany) using the following scan parameters: 110–130 kV, 30–50 mAs (with fixed tube current), 1.5:1 pitch, 1-second gantry rotation time, and 16 \times 1.2 mm detector configuration. Images were reconstructed with 2-mm section thickness using B20f (standard soft tissue) kernel.

All CT DICOM image data were de-identified, exported offline, and transferred to Site 3 (Massachusetts General Hospital) for subjective and quantitative analyses.

2.4. Subjective assessment

A thoracic subspecialty radiologist (MKK with 14-year experience in thoracic radiology) analyzed all chest CTs from both sites on a DICOM viewing application (RadiAnt Dicom Viewer, Medixant, Poznan, Poland). For subjective severity, the radiologist evaluated pulmonary

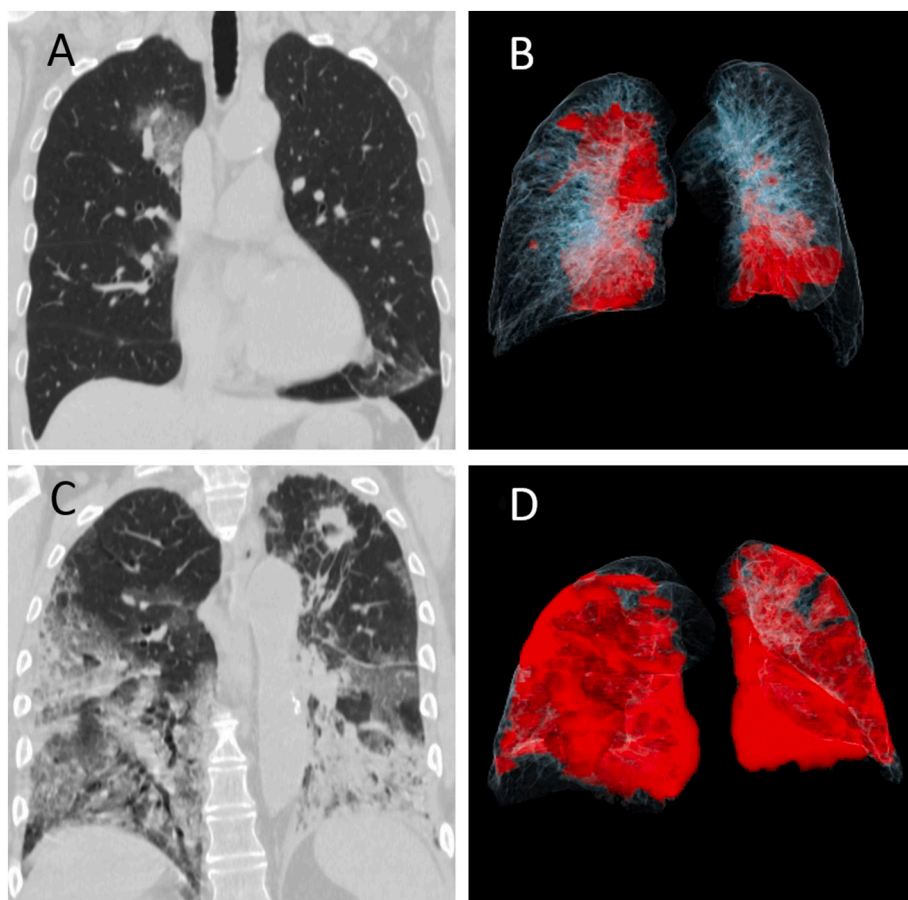


Fig. 2. Chest CT images of two patients with RT-PCR positive COVID-19 pneumonia. (A, B) A 69-year-old male managed without ICU admission had multifocal groundglass opacities in right lung and the left lower lobe on coronal multiplanar image (A), which is rendered in red color on the accompanying movie of volume rendered image dataset (B). (C, D) A 76-year-old-male who was admitted to the ICU and subsequently died from complications related to COVID-19 pneumonia. The patient had extensive consolidative opacities in the left lung and mixed attenuation opacities in the right lung on the coronal image (C) which are annotated in red color in the volume rendered movie (D). Incidentally, the patient had a cavitary nodule in the left apex which was concerning for lung cancer (no histopathology proof). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Summary of patient demographics, subjective assessment, DL-based and radiomics features in patients with different disease outcomes (death versus recovery).

Disease outcome		With motion artifacts n = 221	Without motion artifacts n = 167
Mean age (years)		65 ± 16.4	62 ± 16.4
Gender (M/F)		138/83	106/61
Subjective assessment	Extent of opacity (AUC 0.758; p < 0.0001)		Lymphadenopathy + type of opacity + pleural effusion (AUC 0.864; p = 0.024)
DL-based features	Percentage of opacity + standard deviation of opacity (AUC 0.841; p = 0.0154)		Standard deviation + Volume of high opacity (AUC 0.883; p = 0.001)
Radiomics	Wavelet LLL gldm Small Dependence High Gray Level Emphasis + wavelet-LHL glrlm High Gray Level Run Emphasis (AUC 0.827; p = 0.009)		Wavelet-LLL ngtdm contrast + wavelet-LHH gldm Large Dependence High Gray Level Emphasis+ original shape Spherical Disproportion (AUC 0.815; p = 0.009)
DL + Radiomics	Wavelet-LLL gldm Small Dependence High Gray Level Emphasis + Volume of opacity (AUC 0.836; p < 0.001)		Standard deviation (AUC 0.877; p < 0.0001)

opacities for type (3-point score: 1 = ground-glass; 2 = consolidation; 3 = mixed attenuation opacities which included a combination of groundglass opacities with consolidation, interlobular septal thickening,

and/or nodules) and extent of lung lobe involved (5-point score, 1 = <5%; 2 = 5–25%; 3 = 26–49%; 4 = 50–74%; 5 = >75%).¹² Other recorded findings included presence of pleural effusion and mediastinal or hilar lymphadenopathy. Respiratory motion artifacts were graded as none, mild (affecting less than a lobe of the lung and without impaired evaluation), moderate (involving two or more lung lobes and with minor impairment in evaluation of lung findings), and severe (artifacts associated with substantially impaired evaluation of lung findings).

2.5. DL-based CT Pneumonia Analysis Prototype

The research software on CT Pneumonia Analysis Prototype (Siemens Healthineers) provides quantitative DL-features on the presence and extent of pulmonary opacities related to SARS-CoV-2 pneumonia in chest CT images. The prototype was trained on 1000 chest CT examinations for detection of SARS-CoV-2 pneumonia and on 1371 chest CTs for quantification of pulmonary opacities. These did not belong to the three sites included in our study. As described in prior publications,^{13,14} the prototype automatically estimates presence and extent of pulmonary opacities in both lungs combined, and in each lung and lung lobe separately.¹⁵ The estimated DL-based features include a. presence or absence of pulmonary opacities; b. pulmonary opacity scores (score 1: <25%; score 2: 26–50%; score 3: 51–75%; score 4: >75% of lung lobe involvement); c. volume of pulmonary opacities (ml); d. percentage of lungs affected by opacities; e. mean Hounsfield units (HU) of pulmonary opacities; f. standard deviation of HU of pulmonary opacities; g. volume (ml) of high attenuation opacities with ≥–200 HU; h. percentage of lungs affected by high attenuation opacities (Fig. 1).

A study coinvestigator (CA with one-year post-doctoral research experience) processed 221 chest CT with the prototype. Two chest CT examinations could not be processed with the prototype and were

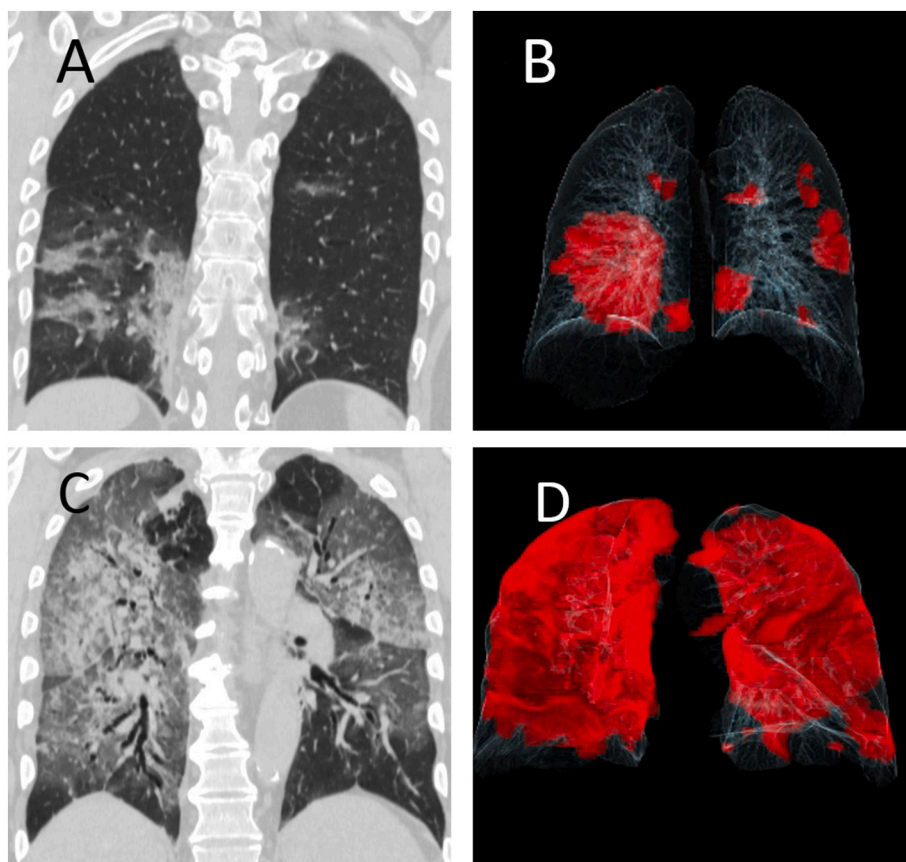


Fig. 3. Chest CT images of two patients with RT-PCR positive COVID-19 pneumonia. (A, B) A 41-year-old male with full recovery. Coronal multiplanar image shows multifocal mixed opacities in left upper and bilateral lower lobes (right greater than left) which are displayed in red in the accompanying movie of volume rendered image dataset (B). (C, D) A 72-year-old-male who died from complications related to COVID-19 pneumonia. The patient had diffuse mixed attenuation opacities in bilateral lungs on coronal image (C) which are annotated in red color in the volume rendered movie (D). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

Summary of DL algorithm derived opacity scores as well as mean HU and standard deviations of opacities for groundglass, consolidative and mixed opacities on a lung-lobe basis (* all values in average/lobe are $p < 0.0001$).

Type of opacity		RUL	RML	RLL	LUL	LLL	Average/lobe*
Opacity score	Groundglass	0.9 ± 0.8	0.9 ± 0.9	1.3 ± 1.1	0.7 ± 0.8	1.1 ± 0.9	1 ± 0.9
	Consolidation	1 ± 0.8	1.1 ± 0.7	2 ± 1	1.8 ± 1.2	2.2 ± 0.8	1.7 ± 1
	Mixed opacities	1.8 ± 1.1	1.7 ± 1	2.1 ± 1.1	1.7 ± 0.9	2 ± 1.1	1.9 ± 1.1
Mean HU of opacity	Groundglass	-586 ± 147	-534 ± 208	-562 ± 161	-599 ± 167	-736 ± 162	-576 ± 170
	Consolidation	-410 ± 74	-461 ± 36	-403 ± 137	-366 ± 111	-355 ± 151	-395 ± 118
	Mixed opacities	-500 ± 133	-567 ± 114	-468 ± 131	-535 ± 124	-472 ± 131	-500 ± 132
Standard deviation of opacity	Groundglass	182 ± 76	168 ± 81	178 ± 75	163 ± 64	169 ± 65	172 ± 73
	Consolidation	210 ± 40	251 ± 39	239 ± 46	244 ± 27	249 ± 62	240 ± 47
	Mixed opacities	217 ± 58	201 ± 48	227 ± 47	207 ± 48	218 ± 50	216 ± 51

excluded. These exams had either incomplete image datasets or corrupted DICOM headers.

2.6. Radiomics

On a separate prototype (Radiomics, Siemens Healthineers), a study coinvestigator (CA) processed thin-section images of 221 chest CTs to obtain radiomics for the entire lungs. The radiomics prototype performs automatic segments and estimates radiomics over bilateral lungs. Radiomics features are thoroughly described in accessed at <https://pyradiomics.readthedocs.io/en/latest/features.html>. The derived radiomics features include first-order, shape, gray level co-occurrence matrix, gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighboring gray tone difference matrix (NGTDM), and gray level dependence matrix features (GLDM).⁵

2.7. Statistical analysis

Statistical analysis for outcome prediction was limited to either the entire study sample (221 patients) or patients from Site 1 (196 patients). No separate statistical testing was performed for Site 2 since very few patients ($n = 4-8$) had adverse outcomes and ICU admission.

Data were analyzed with Microsoft EXCEL (Microsoft Inc., Redmond, Washington, USA) and SPSS Statistical software (SPSS Version 24, IBM Inc., Chicago, Illinois, USA). We used pivot tables within the EXCEL files to obtain descriptive statistics. Chi square test and analysis of variance (ANOVA) were performed with SPSS Statistics software to determine whether the differences in DL-based features among type and extent of pulmonary opacities were statistically significant. In addition, we used the R Statistical Computing (<https://www.R-project.org>, R Foundation for Statistical Computing, Vienna, Austria, accessed on 4.15.2020) built in the Radiomics prototype to perform multiple logistic regression for predicting patient outcome and ICU admission. Area under the curve (AUC with 95% confidence interval) was the output information for the

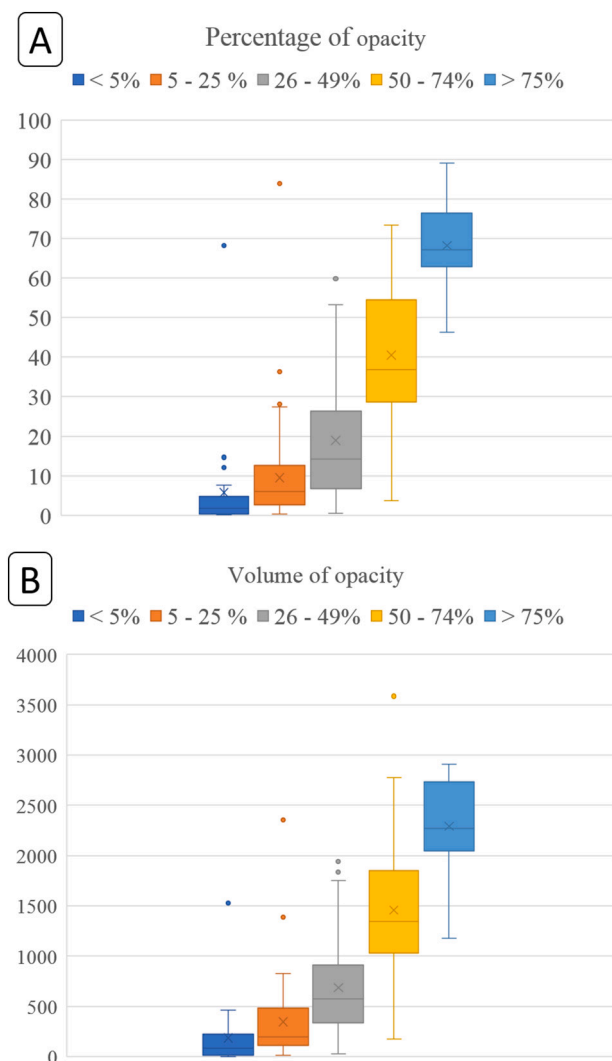


Fig. 4. Box whisker plots for percentage (A) (top graph: y-axis denotes percentage of lung affected by opacities) and volume of opacities (B) (bottom graph: y-axis denotes absolute lung volume affected by opacities in mL). The different color boxes along the x-axis represent subjective severity assessment into different subjective percentage categories of lungs affected by opacities. The horizontal lines within each box represent median values whereas the upper and lower bounds of each box are first and third quartiles. The whiskers denote minimum and maximum values. The cross marks (x) represent the average values.

regression analysis. A p-value less than 0.05 was considered as statistically significant difference.

3. Results

Of the 23/221 patients (10.4%) who passed away from SARS-CoV-2 pneumonia, 20 patients died at Site 1 (20/196; 10.2%) and 3 patients at Site 2 (3/25; 12%). Forty-seven patients required ICU admission (Site 1: 39/196 patients, 19.9%; Site 2: 8/25, 32%) while others had non-ICU hospital admission (174/221).

Of the 221 chest CTs, there were respiratory motion artifacts in almost one-quarter of patients with following distribution: none (167/221; 75.6%), mild (1/221; 0.4%), moderate (36/221; 16.3%), and severe (17/221; 7.7%). Chest CT with moderate to severe motion artifacts had more extensive pulmonary opacities as compared to those with no or mild motion artifacts ($p < 0.016$; Table 1) There were no significant differences in the severity of motion artifacts in patients with and

without ICU admission ($p = 0.148$) and adverse outcomes (0.079).

3.1. ICU admission

Subjective severity assessment had an AUC of 0.77 (95% CI 0.77–0.79) on the entire dataset from both sites for predicting ICU admission; the AUC did not change when chest CTs with respiratory motion artifacts were excluded. Addition of lymphadenopathy to subjective severity assessment increased the AUC to 0.80 on chest CTs without motion artifacts ($p < 0.001$).

Among the DL-based features, volume of pulmonary opacities had the best predictive value (AUC 0.77) for determining ICU admission in the entire dataset (Sites 1 and 2) (Table 2) (Fig. 2). Without the chest CTs with motion artifacts, a DL-based feature, percentage of opacities was the best predictor with an AUC of 0.80. As summarized in Table 1, higher order radiomics also had similar performance with 0.78 AUC for the entire dataset and 0.82 AUC for chest CTs without motion artifacts. When DL-based and radiomics features were combined in the multiple logistic regression analysis, the AUC for predicting ICU admission increased to 0.81 in the entire database ($p = 0.01$) and to 0.82 for chest CTs without severe motion artifacts ($p = 0.006$).

Stratified analysis of Site 1 data for predicting ICU admission demonstrated no difference in performance of subjective severity assessment (AUC 0.80–0.81), DL-based features (AUC 0.76–0.79), and radiomics (AUC 0.77–0.78) for chest CTs with and without motion artifacts ($p > 0.5$).

3.2. Patient outcome

The subjective severity assessment could predict patient outcome (death versus recovery) in the entire dataset of chest CTs from both sites with and without motion artifacts (AUC 0.76) at a threshold of 13 and higher for severity score (sensitivity = 82%, specificity = 60% for ICU admission and sensitivity = 95% and specificity = 57% for patient outcome). The AUC (0.86) of subjective severity assessment improved when combined with pleural effusion and intrathoracic lymphadenopathy ($p = 0.024$), especially from chest CTs without motion artifacts. In the presence of lymphadenopathy and pleural effusions, a severity score of 16 and higher had the highest sensitivity of 89% and specificity of 72% for prediction of ICU admission and 67% and 61% for disease outcome prediction.

Both DL-based (AUC 0.84) and radiomics (AUC 0.83) features outperformed subjective severity assessment for predicting patient outcome when all chest CTs (Sites 1 and 2) with and without motion artifacts were included ($p = 0.01$ –0.009) (Table 3) (Fig. 3). DL-based features (AUC 0.88) were superior to both subjective severity assessment (AUC 0.86) and radiomics (AUC 0.82) when chest CTs with motion artifacts were excluded. Combined analysis of DL-based and radiomics (AUC 0.84–0.87) features did not improve differentiation of patients with and without favorable outcome as compared to their separate performance (AUC 0.82–0.88).

Upon stratified analysis of Site 1 data for outcome prediction, DL-based features (AUC 0.86 for chest CT with motion artifacts; AUC 0.91 with exclusion of chest CT with motion artifacts) were superior to both subjective assessment (corresponding AUCs 0.79 and 0.84) and radiomics (corresponding AUCs 0.78–0.80) ($p < 0.001$).

Thresholds for DL and radiomic best features for ICU admission and disease outcome are summarized in Table 5 in Supplementary.

3.3. Types and extent of pulmonary opacities

There were significant differences between the opacities scores, mean HU of the lungs and mean HU of pulmonary opacities for groundglass, mixed, and consolidative opacities on chest CT examinations ($p < 0.0001$) (Table 4).

There was a strong correlation between the subjective assessment of

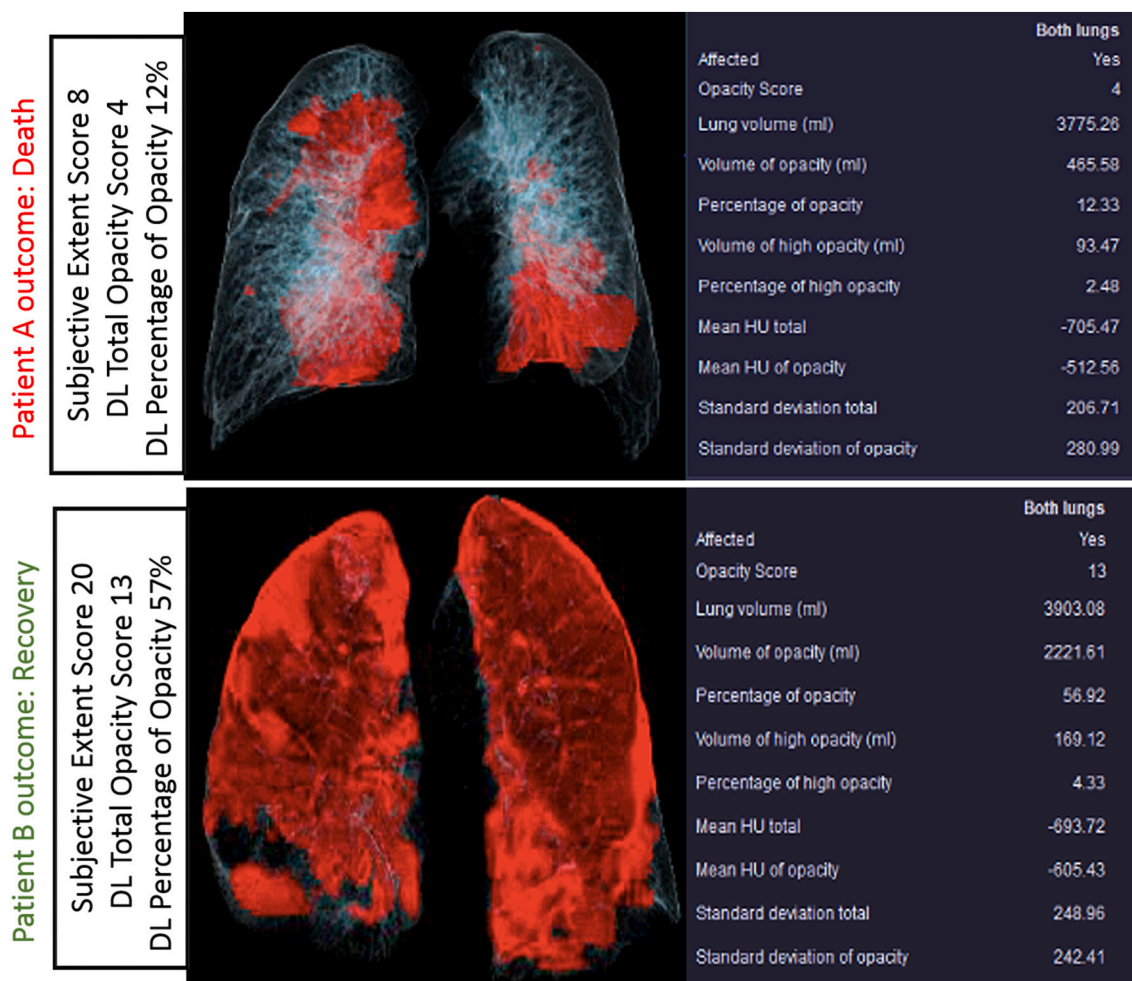


Fig. 5. Volume rendered images and tabular summaries of DL variables in RT-PCR-positive SARS-CoV-2 infections in two patients with different outcomes (top row images from Patient A: 74-year-old man passed away; Patient B: 61-year-old woman survived). Examples demonstrate that some patients with extensive pulmonary involvement survive (patient B) while others (patient A) die with much less pulmonary opacities.

pulmonary opacities and DL-based features on volume ($r^2 = 0.735$) and percentage ($r^2 = 0.728$) of pulmonary opacities. The distribution of volume and percentage of pulmonary opacities for different subjective severity assessment scores is summarized in Fig. 4. Some patients with extensive pulmonary opacities on subjective assessment (as well as low opacity and volume scores on analysis with the prototype) had adverse outcome, while a few patients with low subjective extent score died from complications related to SARS-CoV-2 pneumonia (Figs. 5 and 6).

4. Discussion

Almost one-quarter of chest CT examinations in patients with SARS-CoV-2 pneumonia had moderate or severe respiratory motion artifacts that limited evaluation of lung findings. The presence or absence of respiratory motion artifacts on chest CTs did not change the subjective severity assessment, likely due to non-subtle nature of pulmonary findings in the included patients or from the ability to interpret findings in presence of motion. However, prediction of ICU admission with both the DL-based and radiomics features was better in patients without than in those with motion artifacts on their chest CTs. Prediction of outcome (i.e. death versus recovery) from DL-based features also improved when chest CTs with respiratory motion artifacts were excluded. No prior studies have reported on the incidence of motion artifacts in patients with SARS-CoV-2, however, a pre-SARS-CoV-2 pandemic study described a high incidence of motion artifacts and expiratory phase

scanning in about 1/3 of all chest CTs.¹⁶ Given the change in quantitative pixel values with motion artifacts, it is not surprising that exclusion of motion-impaired chest CTs improved performance of both DL-based and radiomics features. However, none of the described DL or radiomics approaches including the ones used in our study check CT images for presence of motion artifacts.

Prior studies describe separate use of DL algorithms, volume of disease, and radiomics for diagnosis, disease severity, treatment response, outcome (death), oxygen supplement, intubation and ICU admission in patients with SARS-CoV-2 pneumonia.^{5,6,17–19} Although performance of our DL algorithm and radiomics approach is similar to prior reports, besides the influence of motion artifacts, we document both the comparative and additive value of DL-based and radiomics features in prediction of outcome and need for ICU admission. The previously reported subjective grading of disease extent in each lobe, a tedious and time-consuming process, we demonstrate that quantitative lung lobe-level information on volume and percentage of affected lungs is superior for assessing disease severity and predicting patient outcome. Prior studies have also reported adverse outcomes in patients with mixed or consolidative opacities^{20,21}; mean HU of opacities derived from DL algorithm in our study can discriminate between groundglass, mixed attenuation and consolidative opacities.

The primary implication of our study is the relative superiority of both DL and radiomics, either alone or in combination, over subjective severity assessment of SARS-CoV-2 pneumonia. Given the semantic,

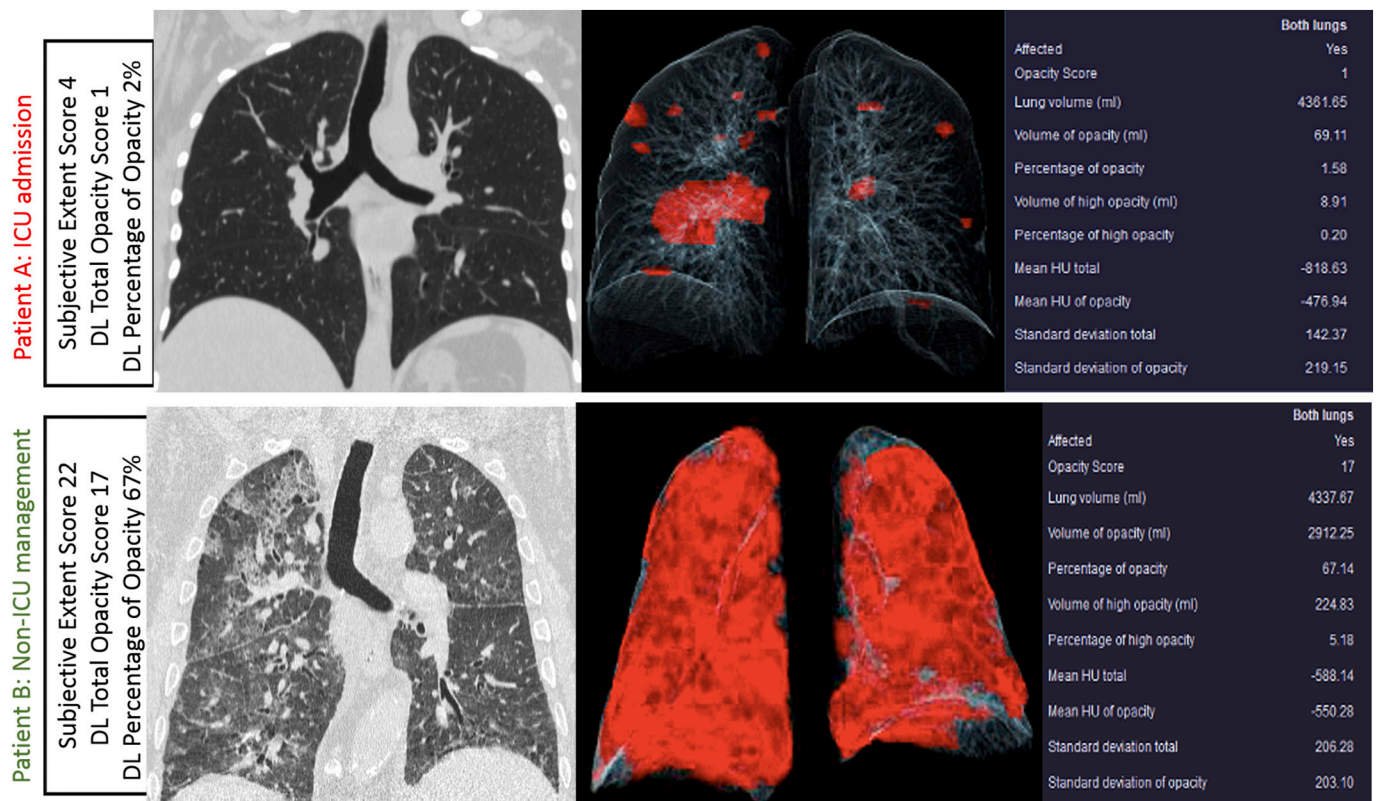


Fig. 6. Coronal multiplanar and volume rendered images along with tabular summaries of DL variables in RT-PCR-positive SARS-CoV-2 infections in two patients with managed with (Top row images from patient A: 46-year-old man was admitted to the ICU and survived) without (bottom row images from patient B: 61-year-old woman was managed without ICU admission and survived) ICU admission. These examples demonstrate that some patients with extensive pulmonary involvement do not require ICU admission.

non-quantitative radiology reporting in most radiology practices, inclusion of DL outputs can add quantitative information on the severity of lung involvement within each lung lobe. Besides the prognostic value of such information, it can provide an objective overview of both extent and attenuation of pulmonary findings in patients with SARS-CoV-2 pneumonia.

Respiratory motion artifacts, present in a substantial proportion of chest CTs in patients with SARS-CoV-2 pneumonia, adversely affect performance of both radiomics and DL algorithm for assessing disease severity. Users must therefore ensure that the algorithms or radiologists perform a quality check on chest CTs before processing them with either technique since motion artifacts can confound results. Also, CT technologists and radiologists must instruct or check patients to assess their compliance with breath-holding instructions. A fast scanning protocol must be used in patients unable to hold their breath for the duration of chest CT. It is also likely, though not specifically assessed in our study, that other artifacts such as beam hardening or metal streaking artifacts, which affect pixel values also impair performance of both DL and radiomics.

Despite high AUCs of subjective assessment, DL-based and radiomics features, a few patients with low scores (less pulmonary opacities) had adverse outcomes. This finding may imply the need to include other clinical or imaging information which impact patient outcome beyond the extent and/or severity of pulmonary opacities. There are reports on prognostic importance of clinical information such as comorbid conditions as well as vital signs and laboratory values.^{5,6,8} Apart from incorporation of the clinical markers, predictive power of DL-based algorithms such as the prototype used in our study may benefit from inclusion of other imaging findings on chest CT like pleural effusions and intrathoracic lymphadenopathy.

The main limitation of our study is its retrospective nature which

precludes determination of the impact of the prototype on prospective patient and resource management. We did not perform a power analysis to determine the needed sample size but included consecutive chest CTs in patients with SARS-CoV-2 pneumonia at both sites. The skewed distribution of SARS-CoV-2 pneumonia at the two participating sites limited statistical analysis on data from Site 2 ($n = 26$ patients). Although this limits the evaluation of wider generalizability of our prototype, the results are encouraging given the fact that the prototype was neither trained nor validated with data from the three sites included in our study. We cannot fully explain why chest CTs from three patients could not be processed with the DL and radiomics prototypes. We did not have access to information related to onset of patient symptoms, comorbidities, vital signs, and laboratory data which could have served as additional predicates or improved the performance of subjective severity assessment, CT Pneumonia Analysis or radiomics prototypes to predict outcome or ICU admission. Chest CT appearance SARS-CoV-2 pneumonia can vary based on the time difference between the onset of symptoms and chest CT examinations, as well as the presence of comorbid conditions (such as cancer or autoimmune diseases); lack of these information can confound our data.

We did not assess effect of other artifacts (such as beam hardening artifacts from arms by the side of the body or metallic prosthesis) on performance of the prototypes. A substantial number of patients with SARS-CoV-2 infection undergo contrast-enhanced chest CT to assess for complications such as pulmonary emboli. Since we do not acquire non-contrast phase images prior contrast-enhanced chest CT, we did not assess how contrast-enhancement influences performance of the prototypes. Also, we did not estimate variations in inter- and intra-radiologists as well as our deep learning models or radiomics with a washout period. However prior studies have documented that there is up to moderate interobserver agreement across radiologists with different

expertise levels.²²

In conclusion, both deep learning-based CT Pneumonia Analysis prototype and radiomics are superior to subjective severity assessment of SARS-CoV-2 pneumonia on chest CT in prediction of patient outcome and the need for ICU admission. In presence of motion artifacts, which are frequent in patients with pneumonia, the prototype outperformed both radiomics and subjective severity assessment. The deep learning-based features can also enable differentiation between different types of pulmonary opacities in patients with SARS-CoV-2 infection.

Declaration of competing interest

We did not receive any research funding for the CT Pneumonia Analysis Prototype (Siemens Healthineers). One participating hospital (Massachusetts General Hospital) received unrelated research funding from GE Healthcare, Lunit Inc., Riverain Tech, and Siemens Healthineers. FD, MZ and MM are employees of Siemens Healthineers who did not participate in the study subject selection or data analysis. Other authors do not have any disclosure.

Acknowledgment

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.clinimag.2021.06.036>.

References

- [1] Francone M, Iafra F, Masci GM, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. *Eur Radiol* 2020. <https://doi.org/10.1007/s00330-020-07033-y>.
- [2] Mahdjoub E, Mohammad W, Lefevre T, et al. Admission chest CT score predicts 5-day outcome in patients with COVID-19. *Intensive Care Med* 2020;46:1648–50.
- [3] Li Y, Yang Z, Ai T, et al. Association of “initial CT” findings with mortality in older patients with coronavirus disease 2019 (COVID-19). *Eur Radiol* 2020. <https://doi.org/10.1007/s00330-020-06969-5>.
- [4] Cozzi D, Albanesi M, Cavigli E, et al. Chest X-ray in new coronavirus disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. *Radiol Med* 2020;125(8):730–7. <https://doi.org/10.1007/s11547-020-01232-9>.
- [5] Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* 2016;278(2):563–77.
- [6] Huang L, Han R, Ai T, et al. Serial quantitative chest CT assessment of COVID-19: deep-learning approach. *Radiol Cardiothorac Imaging* 2020;2(2):e200075. Published 2020 Mar 30, <https://doi.org/10.1148/ryct.2020200075>.
- [7] Tang Z, Zhao W, Xie X, et al. Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. (2020) *ArXiv*: 2003.11988.
- [8] Ebrahimian S, Homayounieh F, Rockenbach MA, Putha P, Raj T, Dayan I, Bizzo BC, Buch V, Wu D, Kim K, Li Q. Artificial intelligence matches subjective severity assessment of pneumonia for prediction of patient outcome and need for mechanical ventilation: a cohort study. *Sci Rep* 2021;11(1):1–10.
- [9] Wu Q, Wang S, Li L, Wu Q, Qian W, Hu Y, Li L, Zhou X, Ma H, Li H, Wang M. Radiomics analysis of computed tomography helps predict poor prognostic outcome in COVID-19. *Theranostics* 2020;10(16):7231.
- [10] Matos J, Paparo F, Mussetto I, et al. Evaluation of novel coronavirus disease (COVID-19) using quantitative lung CT and clinical data: prediction of short-term outcome. *Eur Radiol Exp* 2020;4(1):39. Published 2020 Jun 26, <https://doi.org/10.1186/s41747-020-00167-0>.
- [11] Lanza E, Muglia R, Bolengo I, et al. Quantitative chest CT analysis in COVID-19 to predict the need for oxygenation support and intubation. *Research Square*; 2020. <https://doi.org/10.21203/rs.3.rs-30481/v1>.
- [12] Yang R, Li X, Liu H, et al. Chest CT severity score: an imaging tool for assessing severe COVID-19. *Radiol Cardiothorac Imaging* Mar 2020. <https://doi.org/10.1148/ryct.2020200047>.
- [13] Bogdan Georgescu and Shikha Chaganti and Gorka Bastarrika Aleman and Eduardo Jose Mortani Barbosa Jr. and Jordi Broncano Cabrero and Guillaume Chabin and Thomas Flohr and Philippe Grenier and Sasa Grbic and Nakul Gupta and François Mellot and Savvas Nicolaou and Thomas Re and Pina Sanelli and Alexander W. Sauter and Youngjin Yoo and Valentin Ziebandt and Dorin Comaniciu. Machine learning automatically detects COVID-19 using chest CTs in a large multicenter cohort. *ArXiv preprint*, arXiv 2006.04998. 2020.
- [14] Chaganti, Shikha & Balachandran, Abishek & Chabin, Guillaume & Cohen, Stuart & Flohr, Thomas & Prof, apl & Georgescu, Bogdan & Grenier, Philippe & Prof, & Grbic, Sasa & Liu, Siqi & Mellot, François & Murray, Nicolas & Nicolaou, Savvas & Parker, William & Re, Thomas & Sanelli, Pina & Sauter, Alexander & Xu, Zhoubing & Comaniciu, Dorin. (2020). Automated quantification of CT patterns associated with COVID-19 from chest CT. *ArXiv preprint*, arXiv 2004.01279, 2020.
- [15] Chaganti S, Balachandran A, Chabin G, et al. Quantification of tomographic patterns associated with COVID-19 from chest CT. *ArXiv preprint*, arXiv: 2004.01279. 2020.
- [16] Doda Kherra R, Singh R, Homayounieh F, et al. Deploying clinical process improvement strategies to reduce motion artifacts and expiratory phase scanning in chest CT. *Sci Rep* 2019;9:11858. <https://doi.org/10.1038/s41598-019-48423-7>.
- [17] Jiang X, Coffee M, Bari AX, et al. Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Comput Mater Contin* 2020; 63(1):537–51.
- [18] Lassau N, Ammari S, Chouzenoux E et al. AI-based multi-modal integration of clinical characteristics, lab tests and chest CTs improves COVID-19 outcome prediction of hospitalized patients. *Nat Commun* doi: <https://doi.org/10.1101/2020.05.14.20101972>.
- [19] Wei W, Hu XW, Cheng Q, Zhao YM, Ge YQ. Identification of common and severe COVID-19: the value of CT texture analysis and correlation with clinical characteristics [published online ahead of print, 2020 Jul 1]. *Eur Radiol* 2020. <https://doi.org/10.1007/s00330-020-07012-3>.
- [20] Zhao Wei, Zhong Zheng, Xie Xingzhi, et al. Relation between chest CT findings and clinical conditions of coronavirus disease (COVID-19) pneumonia: a multicenter study. *AJR Am J Roentgenol* 2020;214(5):1072–7.
- [21] Cui N, Zou X, Xu L. Preliminary CT findings of coronavirus disease 2019 (COVID-19). *Clin Imaging* 2020;65:124–32. <https://doi.org/10.1016/j.clinimag.2020.04.042>.
- [22] Bellini D, Panvini N, Rengo M, et al. Diagnostic accuracy and interobserver variability of CO-RADS in patients with suspected coronavirus disease-2019: a multireader validation study. *Eur Radiol* 2021;31:1932–40.