Journal of Digital Imaging (2021) 34:320–329
https://doi.org/10.1007/s10278-021-00430-9

Multicenter Assessment of CT Pneumonia Analysis Prototype for Predicting Disease Severity and Patient Outcome

Fatemeh Homayounieh[1] · Marcio Aloisio Bezerra Cavalcanti Rockenbach[2] · Shadi Ebrahimian[1] · Ruhani Doda Khera[1] · Bernardo C. Bizzo[1,2] · Varun Buch[2] · Rosa Babaei[3] · Hadi Karimi Mobin[3] · Iman Mohseni[3] · Matthias Mitschke[4] · Mathis Zimmermann[4] · Felix Durlak[4] · Franziska Rauch[4] · Subba R Digumarthy[1] · Mannudeep K. Kalra[1]

Received: 29 July 2020 / Revised: 8 January 2021 / Accepted: 2 February 2021 / Published online: 25 February 2021
© Society for Imaging Informatics in Medicine 2021

Abstract

To perform a multicenter assessment of the CT Pneumonia Analysis prototype for predicting disease severity and patient outcome in COVID-19 pneumonia both without and with integration of clinical information. Our IRB-approved observational study included consecutive 241 adult patients (> 18 years; 105 females; 136 males) with RT-PCR-positive COVID-19 pneumonia who underwent non-contrast chest CT at one of the two tertiary care hospitals (site A: Massachusetts General Hospital, USA; site B: Firoozgar Hospital Iran). We recorded patient age, gender, comorbid conditions, laboratory values, intensive care unit (ICU) admission, mechanical ventilation, and final outcome (recovery or death). Two thoracic radiologists reviewed all chest CTs to record type, extent of pulmonary opacities based on the percentage of lobe involved, and severity of respiratory motion artifacts. Thin-section CT images were processed with the prototype (Siemens Healthineers) to obtain quantitative features including lung volumes, volume and percentage of all-type and high-attenuation opacities (≥ −200 HU), and mean HU and standard deviation of opacities within a given lung region. These values are estimated for the total combined lung volume, and separately for each lung and each lung lobe. Multivariable analyses of variance (MANOVA) and multiple logistic regression were performed for data analyses. About 26% of chest CTs (62/241) had moderate to severe motion artifacts. There were no significant differences in the AUCs of quantitative features for predicting disease severity with and without motion artifacts (AUC 0.94–0.97) as well as for predicting patient outcome (AUC 0.7–0.77) ($p > 0.5$). Combination of the volume of all-attenuation opacities and the percentage of high-attenuation opacities (AUC 0.76–0.82, 95% confidence interval (CI) 0.73–0.82) had higher AUC for predicting ICU admission than the subjective severity scores (AUC 0.69–0.77, 95% CI 0.69–0.81). Despite a high frequency of motion artifacts, quantitative features of pulmonary opacities from chest CT can help differentiate patients with favorable and adverse outcomes.

Keywords COVID-19 pneumonia · CT · Patient outcome · Motion artifacts · Deep learning

Abbreviations

CT	Computed tomography
COVID-19	Coronavirus disease 2019
IRB	Institutional ethical board
RT-PCR	Reverse transcription polymerase chain reaction
ICU	Intensive care unit
HU	Hounsfield unit
MANOVA	Multivariable analyses of variance
AUC	Area under the curve
CI	Confidence interval
DL	Deep learning
HIPPA	Health Insurance Portability and Accountability Act
DICOM	Digital imaging and communication in medicine
kV	Kilovolt
LDH	Lactate dehydrogenase

✉ Fatemeh Homayounieh
fhomayounieh@mgh.harvard.edu

Extended author information available on the last page of the article

Introduction

In a global health crisis precipitated by a high prevalence infectious disease, it is critical to understand the associated morbidity and mortality as well as to anticipate and prepare resources needed to mitigate the crises [1–3]. Assessment of

disease severity regardless of its etiology requires knowledge of at-risk patient demographics, their underlying comorbidities, symptoms, vital signs, and laboratory and imaging findings. Such clinical information coupled with the epidemiologic statistics and simulations help understand available and needed healthcare resources to mitigate and minimize the impact of healthcare crises. Such healthcare resources include clinical personnel as well as available hospital and ICU beds, personal protective equipment like masks, shields, and gowns to life support devices such as mechanical ventilators and dialysis units. Even the most advanced nations on our planet can become overwhelmed in a pandemic without such knowledge and careful planning [1, 3].

The ongoing pandemic from the novel coronavirus disease of 2019 (COVID-19) is a textbook example of a healthcare crisis that requires such planning and information [4]. The reverse transcriptase-polymerase chain reaction (RT-PCR) assay is the diagnostic mainstay; imaging use is variable, based on availability of RT-PCR assay, and extends from diagnosis to assessment of disease severity and complications [4–11]. Within weeks of the outbreak, there were published data on features and severity assessment of the disease on chest radiography and CT, the most frequent imaging procedures in the hospital-admitted patients [4–12].

In anticipation of the huge caseload and need to diagnose and quantify disease burden, the deep learning (DL) community had an early start on the pandemic [13–25]. Several studies reported DL algorithms on diagnosis, differentiation from other pneumonia, severity assessment, and mortality prediction based on imaging features [13–25]. However, most studies lack assessment with vastly different scanner technologies and geographic regions as well as the effect of frequent respiratory motion artifacts on the relative performance of DL-generated features versus subjective severity assessment and clinical/laboratory data [13–25]. In this context, we assembled a database of clinical and imaging findings from two sites with dissimilar races, geography, and healthcare to assess a DL-based CT Pneumonia Analysis prototype (Siemens Healthineers, Erlangen, Germany) which was trained on a separate multicenter COVID-19 data. We performed a multicenter assessment of the CT Pneumonia Analysis prototype for predicting disease severity and patient outcome in COVID-19 pneumonia both without and with integration of clinical information.

Methods

Approvals and Disclosures

Our retrospective study was performed following institutional ethical board (IRB) approvals with waiver of written informed consent at both participating sites. De-identified clinical and imaging data were used in compliance with guidelines outlined in the Health Insurance Portability and Accountability Act (HIPAA). We did not receive any research grant or support pertaining to the prototype from a for-profit vendor (Siemens Healthineers) described in the manuscript. Our institution has received unrelated research grants from GE Healthcare, Lunit Inc., Riverain Tech and Siemens Healthineers. Four coauthors (MM, MZ, FD, and FR), employees of Siemens Healthineers, were included to ensure veracity of technical description of the prototype; they did not participate in subject recruitment, data collection, or data analysis parts of the study.

Patients

Our study included 241 adult patients with RT-PCR-positive COVID-19 pneumonia from two tertiary care hospitals (site A: Massachusetts General Hospital, USA; and site B: Massachusetts General Hospital, Iran). Site A contributed 124 de-identified patients (mean age (± standard deviation) 76 ± 10 years; 64 females and 60 males). Site B contributed data from 117 de-identified patients (mean age 61 ± 17 years; 41 females, 75 males). These represented consecutive patients who underwent non-contrast chest CT for clinically indicated reasons such as false-negative or pending RT-PCR assay for COVID-19 pneumonia, assessment of moderate or severe pneumonia, and suspected complications. Patients with post-contrast chest CT were excluded since the prototype was trained for evaluation of non-contrast chest CT and did not recommend use of post-contrast CT.

For each patient, study coinvestigators recorded the following information from their medical records: patient age, gender, past medical history (presence of hypertension, diabetes, cancer, immunosuppressive disease, asthma/chronic obstructive pulmonary disease, and ischemic heart disease), white blood cell counts, platelet counts, and lactate dehydrogenase (LDH). In addition, we recorded whether patients required intensive care unit (ICU at both study sites) and mechanical ventilation (only available for site A) during the course of their hospital admission.

Non-contrast Chest CT

Site A: Using standard-of-care department protocol, all non-contrast chest CT examinations were performed on one of the following scanners: 64–92-detector-row, dual-source CT (Siemens Definition or Force, Siemens Healthineers, Forchheim, Germany), 64-detector-row, single-source CT (Siemens Definition Edge), and GE Discovery 750 HD (GE Healthcare, Waukesha, Wisconsin, USA). The scan factors included 100–120 kV,

automatic exposure control (CARE Dose 4D, Siemens: quality reference mAs of 100; Auto mA, GE: 25–35 noise index), 0.9–0.984:1 pitch, and 0.5-s gantry rotation time. Images were reconstructed with iterative reconstruction techniques (Admire, Siemens: iterative reconstruction strength of 2 for section thickness of 1 mm; ASIR, GE: 40% strength of iterative reconstruction technique for section thickness of 1.25 mm).

Site B: All non-contrast chest CT examinations were performed in accordance with the standard of care protocol with a 16-slice, multidetector-row CT scanner (Siemens SOMATOM Emotion 16, Siemens Healthineers, Forchheim, Germany). The scan factors included 110–130 kV, 30–50 mAs (with fixed tube current), 1.5:1 pitch, 16 × 1.2 mm detector configuration, and 1-s gantry rotation time. Filtered back projection reconstruction images with 2-mm section thickness and B20f (standard soft tissue) kernel were used for image analyses.

Qualitative Evaluation

Two thoracic subspecialty radiologists (SD with 16-year experience, MK with 14-year experience) reviewed all 241 chest examinations in consensus (RadiAnt Dicom Viewer, Medixant, Poznan, Poland) in both lung (window level −600 HU, window width 1500 HU) and soft tissue (window level 50 HU, window width 350 HU) windows. Both radiologists were allowed to change or adjust the display windows according to their preference and anatomy of interest.

In each of the five lung lobes, the radiologists separately recorded presence, type (1 = ground-glass; 2 = mixed defined as ground-glass with consolidation and/or interlobular septal thickening; 3 = consolidation), and extent of pulmonary opacities based on the percentage of lobe involved (0—no opacities; 1—less than 5% lobe volume involved; 2—5–25% lobe involved; 3—26–50% lobe involved; 4—51–75% lobe involved; 5—greater than 75% lobar involvement). This scoring system for COVID-19 pneumonia was described in prior publications [7, 8]. Overall subjective severity score was obtained by adding the lobar involvement scores, and then classified into two groups for statistical analyses (severe: > 15; non-severe: ≤ 15).

Radiologists also recorded presence and severity of respiratory motion artifacts within the lungs for all CTs on a 4-point scale (0: no motion artifacts, 1: minimal motion affecting less than 10% of the lungs; 2: moderate artifacts affecting 10–50% of the lungs without compromising assessment of pulmonary opacities, 3: severe artifacts affecting > 50% of the lungs and limiting evaluation of pulmonary findings).

Quantitative Features From the Prototype

Deidentified DICOM images of patients were processed (FH with 2 years of post-doctoral research experience) with the deep learning-based prototype, which is an offline, standalone software. The research prototype is not approved by United States Food and Drug Administration for clinical use. The prototype was trained and validated separately for detection (on 1371 chest CT exams with COVID-19, other viral pneumonia, and ground-glass and consolidative opacities of other etiologies) and quantification (on 1000 chest CT with COVID-19 pneumonia, 131 with interstitial lung diseases, 113 bacterial pneumonia, and 559 normal CT scans). The training and validation chest CT exams did not belong to either of the two sites included in our study. The details of the prototype are described in a previous publication [26].

For lung and lobe segmentation, the algorithm first applies multi-scale deep reinforcement learning to detect anatomic landmarks such as carina and sternal tip. Then, the algorithm resamples the isolated lung region of interest to a 2-mm isotropic volume and processed with a deep image-to-image network (DI2IN) to create lung segmentation. Lastly, the segmented lung mask is reconfigured to the original resolution of CT input data. For the COVID-19-related abnormality segmentation, we trained a DenseUNet with anisotropic kernels to covert CT 3D image volume to a semantic segmentation mask. Then, a single label is used to define all lung voxels with ground-glass or consolidative opacities as positive voxels. The remaining regions are defined as negative within the network trained as an end-to-end segmentation system. The algorithm filters the output 3D segmentation by the lung segmentation. All automatically segmented volume masks were reviewed to verify their accuracy. Manual editing was required for only 2/241 chest CT examinations included in our study. In these two exams, generated contours included subcutaneous emphysema and pneumothorax as lung parenchyma in one patient and stomach air as part of the left lower lung in another patient. Upon confirmation of the segmented contours, the prototype estimates several quantitative features related to the presence of pulmonary opacities (binary score based on presence of opacity), opacity scores based on percentage of lobe involved (score 0: 0%, 1: 1–25%, 2: 26–50%, 3: 51–75%, 4 > 75% of lobe involved), lung volume (in ml), volume and percentage of all-attenuation opacities within a given lung region (as absolute volume and relative percentage of opacities), volume and percentage of high-attenuation opacities (as absolute volume and percentage of pulmonary opacities with attenuation ≥ −200 HU), and mean HU and standard deviations for lung parenchyma as well as pulmonary opacities within given lung regions. The given lung regions imply separate values for both lungs combined as well as for each lung and each lung lobe.

Statistical Analyses

Data were recorded and analyzed for descriptive statistics with Microsoft EXCEL (Microsoft Inc., Redmond, Washington, USA). We calculated linear correlation coefficients between radiologists' severity score and quantitative features with Microsoft EXCEL. Multivariable analysis of variance (MANOVA) was performed to determine differences in quantitative features and type of pulmonary opacities as recorded from radiologists' assessment. Multiple logistic regression analyses were performed with R Statistical Computing software (https://www.R-project.org, R Foundation for Statistical Computing, Vienna, Austria, accessed on 6.20.2020) to assess if severity scores determined by radiologists and the quantitative features could predict patient outcome (death versus recovery) and need for ICU admission. Areas under the curve (AUC with 95% confidence interval) were set as the output information for the regression analyses. We used p-value of less than 0.05 as a statistically significant difference.

Results

Motion Artifacts

More than a quarter of chest CT examinations (62/241; 26%) had moderate to severe motion artifacts. Sixteen chest CT (site A—3/124, 2%; site B—13/117, 11%) had severe motion artifacts and 46 chest CT (site A—11/124, 9%; site B—35/117, 30%) had moderate artifacts. The remaining chest CT had either mild (site A—23/124, 19%; site B—47/117, 40%) or no (site A—87/124, 70%; site B—22/117, 19%) motion artifacts. There was a significant statistical difference in the frequency and severity of artifacts between the two sites ($p < 0.0001$). There were no significant differences in the AUCs of quantitative features for predicting disease severity (severity scores ≤ or > 15) in chest CT examinations with and without motion artifacts (site A: AUC 0.97 vs 0.97, site B: AUC 0.95 vs 0.94) as well as for predicting patient outcome (site A: AUC 0.7 vs 0.72, site B: AUC 0.76 vs 0.77) ($p > 0.5$). The prediction of ICU admission was better with exclusion of chest CTs with severe motion artifacts at Site A (AUC 0.77 vs 0.82), but it did not change at site B (AUC 0.75 vs 0.76). The subsequent results are limited to chest CT examinations without severe motion artifacts.

Extent and Type of Pulmonary Opacities

There was a similar distribution of patients with high (> 15; site A—36%, 43/121; site B—31%, 32/104) and low (≤ 15; site A—64%, 78/121, site B—69%, 72/104) subjective severity scores at the two sites. The combination of LDH and platelet count had highest AUC of 0.74 (p-value < 0.005) for distinguishing patients with low and high subjective severity scores for pulmonary opacities on chest CT. Among different quantitative features, percentage of pulmonary opacities had the highest AUC for differentiating high and low subjective severity score (site A: AUC 0.97, 95% CI 0.97–0.99; site B: AUC 0.93, 95% CI 0.93–0.98).

There was a moderate to strong direct linear correlation between the subjective severity scores and the quantitative features for chest CT data from both sites. Table 1 summarizes the correlation coefficients for entire lung volumes, separately for right and left lung and lung lobes.

The opacity scores mean HU of the lungs as well as mean HU of pulmonary opacities were significantly different for ground-glass, mixed, and consolidative opacities on chest CT examinations from both sites ($p < 0.0001$) (Table 2). The prototype-estimated mean HU of pulmonary opacities for ground-glass, mixed, and consolidative opacities were −555 HU, −457 HU, and −399 HU, respectively ($p < 0.0001$). The average and standard deviations of quantitative features for different patient outcomes are summaries in Table 3.

Table 1 Summary of correlation coefficients between the radiologists' subjective severity scores and quantitative features obtained from the prototype. The lobar range refers to the minimum to maximum correlation coefficients for all lobes with the severity scores

	Site A				Site B			
Quantitative Features	Entire lung	Left lung	Right Lung	Lobar range	Entire lung	Left lung	Right Lung	Lobar range
Opacity score	0.90	0.88	0.89	0.79–0.87	0.86	0.81	0.87	0.73–0.81
Volume of opacities	0.86	0.84	0.84	0.74–0.82	0.85	0.78	0.84	0.69–0.79
Percentage of opacities	0.89	0.86	0.88	0.80–0.89	0.85	0.78	0.86	0.68–0.82
Volume of high-attenuation opacities	0.74	0.71	0.71	0.51–0.69	0.62	0.57	0.63	0.47–0.61
Percentage of high-attenuation opacities	0.73	0.72	0.7	0.54–0.74	0.60	0.55	0.61	0.45–0.57

Table 2 Summary of clinical, subjective, and quantitative features from prototype in patients with best predictive area under the curve for differentiating those with and without ICU admission and those with recovery versus death from COVID-19 pneumonia. All AUCs were statistically significant ($p = 0.04$–< 0.0001)

		Site A			Site B		
	Features	Best subset	AUC	95% CI	Best subset	AUC	95% CI
With vs. without ICU admission	Radiologists	Subjective severity score	0.77	0.77–0.81	Subjective severity score	0.69	0.69–0.74
	AI	Volume of opacity + percentage of high opacity	0.82	0.79–0.82	Opacity score + percentage of high opacity	0.76	0.73–0.79
	Clinical	WBC count	0.64	0.64–0.69	Patient age	0.68	0.68
	Clinical + radiologists	Subjective severity score	0.77	0.77–0.81	Subjective severity score + patient age	0.76	0.7–0.77
	Clinical + AI	Volume of opacity + percentage of high opacity	0.82	0.79–0.82	Opacity score + percentage of high opacity + patient age	0.8	0.73–0.83
Recovered vs. deceased	Radiologists	Subjective severity score	0.68	0.67–0.68	Subjective severity score	0.68	0.67–0.68
	AI	Volume of opacity	0.72	0.7–0.72	Percentage of high opacity + volume of opacity	0.77	0.7–0.84
	Clinical	LDH	0.69	0.69	No features selected	-*	-
	Clinical + radiologists	Subjective severity score + LDH	0.72	0.69–0.73	Subjective severity score + patient age	0.76	0.68–0.8
	Clinical + AI	Volume of opacity + LDH	0.74	0.71–0.77	Percentage of high opacity + patient age	0.8	0.69–0.87

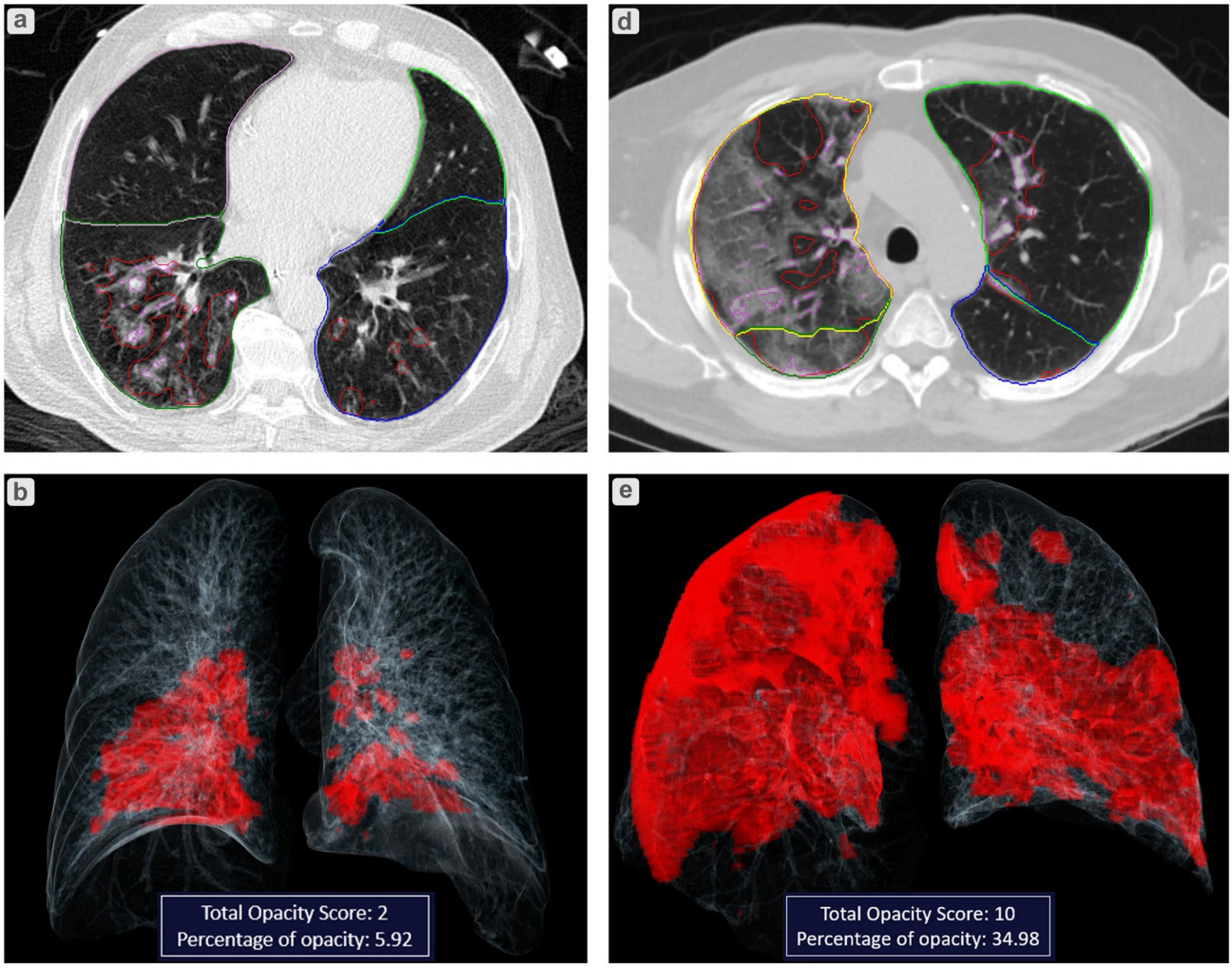
*p-value > 0.06 (AUC 0.71)

ICU Admission

There were differences in the distribution of patients admitted to the non-ICU medical unit (site A—47%, 57/121; site B—65%, 68/104) and ICU (site A—53%, 64/121; site B—35%, 36/104) at the two participating sites (Figs. 1 and 2). At both sites, the combination of the volume of all-attenuation opacities and the percentage of high-attenuation opacities (AUC 0.76–0.82, 95%

Table 3 Site-specific distribution of subjective severity scores and quantitative features for patients with different outcomes

	Site A				Site B			
	ICU admission		Patient outcome		ICU admission		Patient outcome	
Features	Medical floor	ICU	Recovered	Died	Medical floor	ICU	Recovered	Died
Opacity score	5.7 ± 4.7	11 ± 5.3	6.8 ± 4.9	10 ± 5.8	6.2 ± 3.8	9.7 ± 5.2	6.7 ± 4.2	10 ± 5.3
p-value	0.000		0.001		0.000		0.000	
Lung volume	3458 ± 1194	3048 ± 847	3366 ± 1115	3073 ± 919	3950 ± 1059	3663 ± 1100	3980 ± 1052	3361 ± 1053
p-value	0.029		0.127		0.175		0.011	
Volume of opacity	485 ± 528	1141 ± 789	645 ± 662	1083 ± 795	649 ± 558	1198 ± 914	744 ± 655	1228 ± 951
p-value	0.000		0.001		0.000		0.004	
Percentage of opacity	17 ± 19	39 ± 26	22 ± 22	39 ± 27	19 ± 19	36 ± 26	21 ± 20	39 ± 28
p-value	0.000		0.000		0.000		0.001	
Volume of high opacity	120 ± 217	338 ± 320	155 ± 210	345 ± 358	85 ± 108	228 ± 230	108 ± 140	244 ± 244
p-value	0.000		0.000		0.000		0.001	
Percentage of high opacity	4.3 ± 7.9	12 ± 12	5.4 ± 7.2	13 ± 13	2.6 ± 3.7	7.3 ± 8.3	3.2 ± 4.6	8.2 ± 9.1
p-value	0.000		0.000		0.000		0.000	
Mean HU total	−714 ± 108	−599 ± 134	−690 ± 111	−603 ± 149	−734 ± 87	−656 ± 125	−728 ± 92	−628 ± 129
p-value	0.000		0.000		0.000		0.000	
Mean HU of opacity	−497 ± 103	−423 ± 115	−476 ± 109	−433 ± 120	−532 ± 82	−483 ± 104	−526 ± 89	−469 ± 99
p-value	0.000		0.042		0.007		0.007	
Radiologist score	9 ± 5	15 ± 6	11 ± 6	15 ± 7	11 ± 4	15 ± 6	11 ± 4	14 ± 6
p-value	0.000		0.000		0.001		0.030	



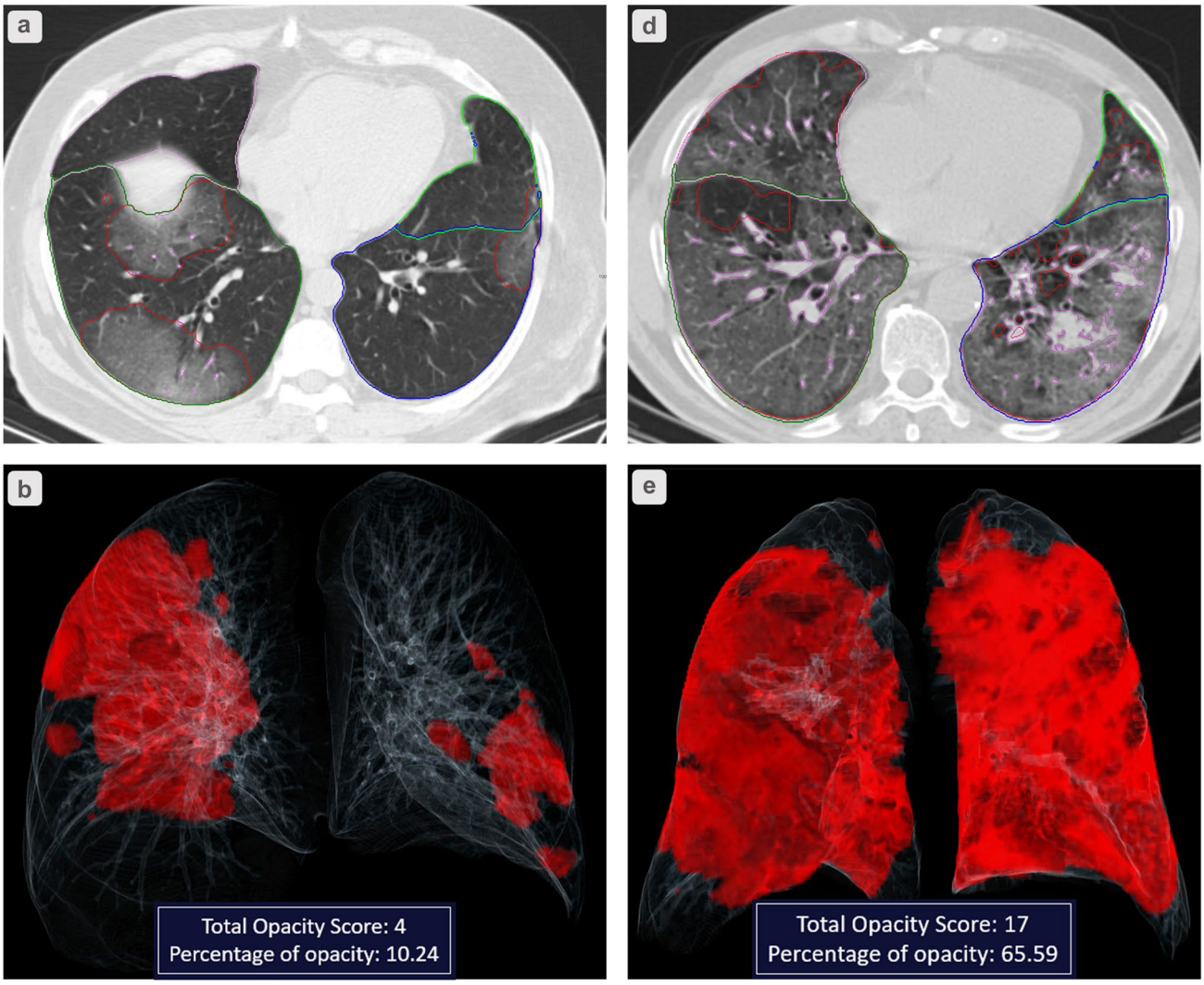
c

	Both Lungs	Left lung	Right lung
Opacity Score	2	1	1
Lung volume (ml)	5448	2528	2920
Volume of opacity (ml)	322	66	256
Percentage of opacity	5.92	2.61	8.78
Volume of high opacity (ml)	21	3	18
Percentage of high opacity	0.38	0.11	0.61
Mean HU total	-789	-794	-785
Mean HU of opacity	-589	-606	-585

f

	Both Lungs	Left lung	Right lung
Opacity Score	10	3	7
Lung volume (ml)	4236	2155	2081
Volume of opacity (ml)	1482	514	968
Percentage of opacity	34.98	23.83	46.53
Volume of high opacity (ml)	230	80	150
Percentage of high opacity	5.41	3.7	7.18
Mean HU total	-705	-746	-662
Mean HU of opacity	-518	-539	-506

Fig. 1 Two patients with COVID-19 pneumonia from site A. **a–c** A 70-year-old male was admitted to the ICU with bilateral lower lobe predominant ground-glass opacities **a** on the transverse chest CT image. The opacities are displayed in red color on the volume rendered 3D image **b** and quantified with different metrics from the prototype **c**. The patient had full recovery. **d–f** A 72-year-old male with hypertension, hyperlipidemia, and cardiac arrhythmia who died after 7 days of hospital admission. There were bilateral, multi-lobar ground-glass opacities on transverse chest CT image **d** which are displayed in red color on the volume rendered image **e** and quantified with the prototype **f**



c	Both Lungs	Left lung	Right lung
Opacity Score	4	1	3
Lung volume (ml)	6318	2902	3416
Volume of opacity (ml)	647	66	581
Percentage of opacity	10.24	2.27	17.02
Volume of high opacity (ml)	7	1	6
Percentage of high opacity	0.11	0.03	0.18
Mean HU total	-824	-841	-810
Mean HU of opacity	-634	-687	-628

f	Both Lungs	Left lung	Right lung
Opacity Score	17	7	10
Lung volume (ml)	5171	2299	2872
Volume of opacity (ml)	3392	1638	1754
Percentage of opacity	65.59	71.26	61.06
Volume of high opacity (ml)	225	172	53
Percentage of high opacity	4.36	7.5	1.85
Mean HU total	-655	-584	-711
Mean HU of opacity	-582	-514	-646

Fig. 2 Two patients with COVID-19 pneumonia from site B. **a**–**c** A 54-year-old male with diabetes mellitus was admitted on the medical floor (no ICU admission) and had full recovery. Transverse chest CT image **a** demonstrated multifocal ground-glass opacities in bilateral lungs as displayed with red color on the volume rendered 3D image **b** and quantified with on the prototype **c**. **d**–**f** A 74-year-old male with diabetes mellitus and ischemic heart disease died from complications related to COVID-19 pneumonia after 12 days of ICU admission. Transverse chest CT image **d** demonstrate diffuse ground-glass opacities with scattered areas with consolidation which are displayed in extensive red color on the volume rendered 3D image **e** and quantified in the table from prototype **f**

CI 0.73–0.82) had higher AUC for predicting ICU admission than the subjective severity scores (AUC 0.69–0.77, 95% CI 0.69–0.81). Among clinical and laboratory variables, white blood cell count (AUC 0.64, 95% CI 0.64–0.69) in site A was the best predictor for ICU admission; patient age (AUC 0. 68, 95% CI 0.68) was the best predictor for ICU admission in site B. Addition of clinical/laboratory data did not result in a significant change in the AUCs of either subjective severity scores or the quantitative features ($p > 0.05$) (Table 2).

For site A, the percentage of all-attenuation opacities in the entire lungs (both lungs combined) was the best feature for predicting the need of mechanical ventilation (AUC 0.83, 95% CI 0.83–0.85). The subjective severity score had performance (AUC 0.82, 95% CI 0.82–0.83) for differentiating those with and without mechanical ventilation. Clinical features (best feature being LDH) had significantly lower AUC (0.71, 95% CI 0.70–0.76) for such differentiation. As noted above, data on mechanical ventilation were not available for site B.

Prediction of Final Outcome

There were significant differences in distribution of patients who recovered (site A: 59%, 71/121; site B: 82%, 85/104) versus those who died from COVID-19 pneumonia (site A: 41%, 50/121; site B: 18%, 19/104) ($p < 0.01$) (Figs. 1 and 2). While the volume of all-attenuation opacities was the best predictor of the final outcome for site A (AUC 0.72, 95% CI 0.70–0.72), the combination of percentage of high-attenuation opacities and volume of all-attenuation opacities of left lower lobe was the best subset (AUC 0.77, 95% CI 0.70–0.84) for site B. The quantitative features for both sites were better than both the subjective severity scores and type of pulmonary opacities from radiologists' assessment (best AUC 0.68, 95% CI 0.67–0.68) for predicting the final outcome. At site A, LDH (AUC 0.69, 95% CI 0.69) could predict patient outcome in site A; no demographic, clinical, or laboratory variable could predict patient outcome for site B.

Discussion

Quantitative features obtained from the DL-based prototype were superior to both the subjective severity scores from radiologists' assessment as well as clinical and laboratory data for prediction of patient outcomes (death, ICU admission, and mechanical ventilation) in patients with COVID-190 pneumonia. Jiang et al. reported 70–80% accuracy for an AI framework based on patient symptoms and laboratory values for assessing disease severity and outcome in COVID-19 [21]. The lower performance of clinical and laboratory data in our study (maximum AUC of 0.69), might have been related to differences in patient demographics, disease severity and/or management strategies.

As reported in prior studies, the percentage and volume of pulmonary opacities were best features for predicting patient outcomes at both sites included in our study [22–24]. Lanza et al. reported that the percentage of compromised lung volume (between −50 and 100 HU) was the most accurate outcome predictor for risk of oxygen support, intubation and in-hospital death in a single-center study of 222 patients [22]. In a study with 176 patients, the volume and ratio of regions with ground-glass opacities (between −700 and −300 HU) obtained from a DL algorithm were the best of the 30 quantitative features obtained with the best AUC of 0.91 for classifying the patient into severe and non-severe COVID-19 pneumonia [23]. Matos et al. reported an AUC of up to 0.92 for their DL models using clinical/laboratory data and CT feature (volume of disease) to predict patient outcome (need for mechanical ventilation or death) [24]. There could be several reasons for the differential performance of our DL algorithm (maximum AUC of 0.82) including the higher frequency of motion artifacts in our study and differences in patient population, disease severity, comorbidities, and treatment strategies. The differences in performance may also be related to significant variations in training and test datasets used in our study as opposed to the other publications where testing and training datasets originated from the same or similar sources [21–24].

The strong predictive value of quantitative features such as opacity scores, volume and percentage of pulmonary opacities for the radiologists, assessed severity and type of pulmonary opacities emphasizes the robust performance of the prototype. This finding is consistent with another study of 126 patients with COVID-19, which reported significant differences ($p < 0.01$) between CT lung opacification percentage between patients with different severities of covid-19 pneumonia [25].

The primary implication of our study is the ability to classify and quantify the type and severity of pulmonary opacities on chest CT with a single-click processing on the prototype used in our study. The predictive information from the prototype on patient outcome (death versus recovery), ICU admission, and mechanical ventilation can help in patient management and resource planning in case of a high-prevalence pandemic. Although radiologists describe the type and distribution of pulmonary opacities, their semantic interpretation does not include qualitative or quantitative assessment of diffuse or multifocal processes such as COVID-19 pneumonia. The subjective severity scores in our study have been reported previously [7–9] but are inefficient, prone to subjective variations, and not used in routine clinical interpretation due to time and difficulty in grading findings based on the percentage of involved lobes. In such context, the addition of quantitative features derived from the prototype following regulatory approvals can provide useful information with

no or minimum additional work or time. Both the user and the developing community must recognize, train, and test their DL models on data from different imaging sites as well as with different CT technical factors including exams with and without motion artifacts to understand how such differences can affect their models.

There are a few limitations in our study. First, we did not perform a power analyses to determine the sample size to test the prototype. However, our data from both sites included all consecutive subjects with RT-PCR-positive COVID-19 pneumonia who underwent non-contrast, thin-section chest CT and had a known outcome in terms of death or recovery, and hospital admission. Second, there were variations in data variables available from the two participating sites. We did not have information on the use of mechanical ventilation in patients from site B. Third, at the height of the pandemic in site A, a few medical floor beds were converted into ICU functionality to accommodate ICU patient overflow. Although this could have affected the classification of patients with and without ICU admission, we classified patients on medical floor beds converted to ICU functionality as those with ICU admission.

Fourth, the differences in mortality associated with patients from the two participating sites was likely related to infrequent use of chest CT at site A relative to site B where all hospital-admitted patients with suspected or known COVID-19 pneumonia regardless of the disease severity and presence of complications underwent chest CT. However, these variations in practice and mortality did not affect the performance of the prototype at either site. Fifth, neither the prototype nor the radiologists evaluated the chest CT for vascular complications of COVID-19 (such as pulmonary thromboembolism) or presence of coronary calcification, pleural effusions, or mediastinal or hilar lymphadenopathy which could have provided additional information on classification of different patient outcomes.

Sixth, some differences in performance of the prototype at the two sites can be attributed to the differences in the scanner technologies. Site A used more advanced and newer CT scanners as compared with site B. Indeed, respiratory motion artifacts were more frequent in site B than in site A. However, there was no difference in the performance of the radiologists or the prototype both with and without chest CT with motion artifacts. The differences in CT scanner technologies and vendors also helped us assess generalizability of the prototype.

In conclusion, the deep learning-based CT Pneumonia Analysis prototype enables a single-click lung segmentation and determination of patient outcome and need for ICU admission in patients with COVID-19 pneumonia. These findings were generalizable at the two high-prevalence sites from Iran and Northeast United States. A strong correlation between the quantitative information from the prototype and radiologists' qualitative assessment of disease severity suggest that the prototype can provide additional quantitative information to the current radiology reports which do not contain information on distribution and extent of pulmonary opacities.

References

1. Barrett K, Khan YA, Mac S, Ximenes R, Naimark DMJ, Sander B. Estimation of COVID-19-induced depletion of hospital resources in Ontario, Canada. CMAJ. 2020;192(24):E640-E646.
2. Moghadas SM, Shoukat A, Fitzpatrick MC, et al. Projecting hospital utilization during the COVID-19 outbreaks in the United States. *Proc Natl Acad Sci U S A*. 2020;117(16):9122-9126. https://doi.org/10.1073/pnas.2004064117
3. Wu G, Yang P, Xie Y, et al. Development of a Clinical Decision Support System for Severity Risk Prediction and Triage of COVID-19 Patients at Hospital Admission: an International Multicenter Study [published online ahead of print, 2020 Jul 2]. *Eur Respir J*. 2020;2001104. https://doi.org/10.1183/13993003.01104-2020
4. Kalra MK, Homayounieh F, Arru C, et al. Chest CT practice and protocols for COVID-19 from radiation dose management perspective. Eur Radiol. https://doi.org/10.1007/s00330-020-07034-x
5. Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases. Radiology. 2020 Feb 26:200642.
6. Fang Y, Zhang H, Xie J, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR [published online ahead of print, 2020 Feb 19]. Radiology. 2020;200432. https://doi.org/10.1148/radiol.2020200432
7. Bernheim A, Mei X, Huang M, et al. Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection. Radiology. 2020 Feb 20:200463.
8. Li K, Wu J, Wu F, et al. The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. Invest Radiol. 2020 Feb 29.
9. Yang R, Li X, Liu H, et al. Chest CT severity score: An imaging tool for assessing severe COVID-19. Radiol Cardiothorac Imaging 2020 2:2.
10. Zhao W, Zhong Z, Xie X, Yu Q, Liu J. Relation Between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study. AJR Am J Roentgenol. 2020;214(5):1072-1077. https://doi.org/10.2214/AJR.20.22976
11. Zhang R, Ouyang H, Fu L, et al. CT features of SARS-CoV-2 pneumonia according to clinical presentation: a retrospective analysis of 120 consecutive patients from Wuhan city [published online ahead of print, 2020 Apr 11]. Eur Radiol. 2020;1–10. https://doi.org/10.1007/s00330-020-06854-1
12. Ng MY, Lee E, Yang J, et al. Imaging profile of the COVID-19 infection: radiologic findings and literature review. Radiol Cardiothorac Imaging. 2020;2(1)
13. Li L, Qin L, Xu Z, et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT [published online ahead of print, 2020 Mar 19]. Radiology. 2020;200905. https://doi.org/10.1148/radiol.2020200905
14. Wu X, Hui H, Niu M, et al. Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study. Eur J Radiol. 2020;128:109041. https://doi.org/10.1016/j.ejrad.2020.109041

15. Javaheri T, Homayounfar M, Amoozgar Z, et al. (2020). CovidCTNet: An Open-Source Deep Learning Approach to Identify Covid-19 Using CT Image. ArXiv, abs/2005.03059.
16. Qin L, Yang Y, Cao Q, et al. A predictive model and scoring system combining clinical and CT characteristics for the diagnosis of COVID-19 [published online ahead of print, 2020 Jul 1]. Eur Radiol. 2020;https://doi.org/10.1007/s00330-020-07022-1.
17. Ardakani AA, Kanafi AR, Acharya UR, Khadem N, Mohammadi A. Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Comput Biol Med. 2020;121:103795. https://doi.org/10.1016/j.compbiomed.2020.103795
18. Ko H, Chung H, Kim KW, et al. COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image [published online ahead of print, 2020 Jun 21]. J Med Internet Res. 2020;https://doi.org/10.2196/19569.
19. Yang S, Jiang L, Cao Z, et al. Deep learning for detecting corona virus disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study. Ann Transl Med. 2020;8(7):450. https://doi.org/10.21037/atm.2020.03.132
20. Singh D, Kumar V, Vaishali, Kaur M. Classification of COVID-19 patients from chest CT images using multi-objective differential evolution-based convolutional neural networks. Eur J Clin Microbiol Infect Dis. 2020;39(7):1379–1389. https://doi.org/10.1007/s10096-020-03901-z
21. Jiang X, Coffee M, Bari A, X. et al. Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. (2020). CMC-Computers, Materials & Continua, 63(1), 537–551.
22. Lanza E, Muglia R, Bolengo I, et al. Quantitative chest CT analysis in COVID-19 to predict the need for oxygenation support and intubation [published online ahead of print, 2020 Jun 26]. Eur Radiol. 2020;1–9. https://doi.org/10.1007/s00330-020-07013-2
23. Z Tang, W Zhao, X Xie, et al. Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images. (2020) ArXiv: 2003.11988
24. 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
25. 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
26. S Chaganti, A Balachandran, G Chabin, et al. Quantification of tomographic patterns associated with COVID-19 from chest CT. ArXiv preprint, arXiv: 2004.01279, 2020.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Fatemeh Homayounieh[1] · Marcio Aloisio Bezerra Cavalcanti Rockenbach[2] · Shadi Ebrahimian[1] · Ruhani Doda Khera[1] · Bernardo C. Bizzo[1,2] · Varun Buch[2] · Rosa Babaei[3] · Hadi Karimi Mobin[3] · Iman Mohseni[3] · Matthias Mitschke[4] · Mathis Zimmermann[4] · Felix Durlak[4] · Franziska Rauch[4] · Subba R Digumarthy[1] · Mannudeep K. Kalra[1]

Marcio Aloisio Bezerra Cavalcanti Rockenbach
marcio.rockenbach@mgh.harvard.edu

Shadi Ebrahimian
sebrahimian@mgh.harvard.edu

Ruhani Doda Khera
rdodakhera@mgh.harvard.edu

Bernardo C. Bizzo
BBIZZO@mgh.harvard.edu

Varun Buch
Varun.Buch@MGH.HARVARD.EDU

Rosa Babaei
rosa.babaei@gmail.com

Hadi Karimi Mobin
hadi.karimimobin@gmail.com

Iman Mohseni
imannn24@yahoo.com

Matthias Mitschke
matthias.mitschke@siemens-healthineers.com

Mathis Zimmermann
mathis.zimmermann@siemens-healthineers.com

Felix Durlak
felix.durlak.ext@siemens-healthineers.com

Franziska Rauch
franziska.rauch.ext@siemens-healthineers.com

Subba R Digumarthy
sdigumarthy@mgh.harvard.edu

Mannudeep K. Kalra
MKALRA@mgh.harvard.edu

1 Department of Radiology, Massachusetts General Hospital and the Harvard Medical School, Boston, MA, USA

2 MGH & BWH Center for Clinical Data Science, Boston, MA, USA

3 Department of Radiology, Firoozgar Hospital and Iran University of Medical Sciences, Tehran, Iran

4 Diagnostic Imaging, Siemens Healthcare GmbH, Erlangen, Germany