Treatment Intensity Stratification in COVID-19 by Fully Automated Analysis of Pulmonary and Cardiovascular Metrics on Initial Chest CT using Deep Learning

Thomas Weikert  (✉ thomas.weikert@usb.ch)
University Hospital Basel  https://orcid.org/0000-0001-9274-053X

Saikiran Rapaka
Siemens Corporate Technology  https://orcid.org/0000-0001-6527-0974

Sasa Grbic
Siemens Corporate Technology

Thomas Re
Siemens Corporate Technology  https://orcid.org/0000-0003-2285-1562

Shikha Chaganti
Siemens Corporate Technology  https://orcid.org/0000-0002-7029-2665

David Winkel
University Hospital Basel  https://orcid.org/0000-0001-7051-8022

Constantinos Anastasopoulos
University Hospital Basel  https://orcid.org/0000-0001-9818-2460

Jens Bremerich
University Hospital Basel  https://orcid.org/0000-0002-1002-8483

Raphael Twerenbold
University Hospital Basel  https://orcid.org/0000-0003-3814-6542

Gregor Sommer
University Hospital Basel  https://orcid.org/0000-0002-8952-0808

Dorin Comaniciu
Siemens Corporate Technology  https://orcid.org/0000-0002-5238-8647

Alexander W. Sauter
University Hospital Basel  https://orcid.org/0000-0002-6707-2258

Research Article

Keywords: COVID-19, Tomography, X-Ray Computed, Therapeutics, Classification, Deep Learning

Posted Date: June 22nd, 2020
Abstract

Objectives

To predict ultimate treatment intensity of COVID-19 patients using pulmonary and cardiovascular metrics fully automatically extracted from initial chest CTs.

Methods

All patients tested positive for SARS-CoV-2 by RT-PCR at our emergency department between March 25 and April 14, 2020 were identified (n=391). For those patients, all initial chest CTs were analyzed (n=85). Multiple pulmonary and cardiovascular metrics were extracted using deep convolutional neural networks. Three clinical treatment intensity groups were defined according to the most intensive treatment of a patient, determined six weeks later: Group 1 (outpatient), group 2 (general ward), and group 3 (intensive care unit; ICU). Univariate analyses were performed to analyze differences between groups. Subsequently, multiple metrics were combined in two binary logistic regression analyses and resulting prediction probabilities used to classify whether a patient needed hospitalization or ICU care. For analysis of discriminatory power, ROC curves were plotted and areas-under-the-curves (AUCs) calculated.

Results

The mean interval between presentation at the emergency department and the chest CT was 1.4 days. Among others, mean percentage of lung volume affected by opacities (PO) and mean total pericardial volume (TPV) increased statistically significantly with higher treatment intensity [from group 1 to 3, standard deviation in brackets: PO: 0.8%(1.5)–11.6%(13.1)–31.6%(20.1); TPV: 733.4ml(231.7)–866.2ml(211.2)–925.7ml(125.5); both: p<0.001]. AUCs were 0.85 (ICU vs. no ICU) and 0.94 (hospitalization vs. no hospitalization).

Conclusions

Metrics fully automatically extracted from initial chest CTs increase with treatment intensity of COVID-19 patients. This information can be exploited to prospectively manage allocation of healthcare resources.

Key Points

- Multiple pulmonary and cardiovascular imaging metrics differ statistically significant between COVID-19 patients with varying ultimate treatment intensity.
- Those metrics include the percentage of lung area affected by opacities and the total pericardial volume and are present on initial CT scans.
- Automated, combined analysis of the metrics using deep convolutional neural networks can contribute to prospectively manage healthcare resources.
Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a global pandemic with over 5.5 million infections and 350,000 deaths worldwide as of 05/28/2020 [1]. The associated infectious disease, named coronavirus disease 2019 (COVID-19), progresses mildly in most cases with nonspecific symptoms like fever and cough [2]. However, severe and critical courses of disease occur in approximately 20% of all patients [3], mostly due to atypical pneumonia and acute respiratory distress syndrome (ARDS) [4]. These patients require critical care including mechanical ventilation and intensive care unit (ICU) treatment, which are scarce resources in a pandemic. At the same time, there are regional differences in utilization of these resources with temporary shortages. Due to the fact that the peak of symptoms usually occurs eight to ten days after onset [5], early prediction of treatment intensity at the time of presentation at the emergency department (ED) might contribute to an improved matching of healthcare resources and needs: Medical experts and equipment could be allocated to patients with a high probability of a severe course of disease.

Chest CT is a primary source of early predictive features in COVID-19, because of its high sensitivity for the detection of the disease and the fact that it’s a primary modality for imaging of pneumonia [6], There is growing, consistent evidence that CT features are associated with disease severity in COVID-19 based on (semi)manual assessment of pulmonary parameters on chest CT [7–11]. This study builds on these approaches and expands them in three aspects: First, by introducing a fully automated evaluation, which is especially relevant in times of a pandemic spread with heavy workloads on healthcare providers. The feasibility of this approach has been demonstrated by Nishiyama et al. for ARDS in a non-COVID-19 patient collective [12]. Second, by explicitly focusing on treatment intensity defined by a patient’s ultimate clinical pathway, because of its relevance for health care management. Third, by including five cardiovascular metrics that are covered in all chest CTs, but rarely interpreted. Noteworthy, preexisting cardiovascular disease is a major risk factor for adverse outcomes in COVID-19 [13]. Thus, it is obvious to exploit this information for treatment intensity prediction.

Our hypothesis is that features derived from initial chest CT representing pulmonary and cardiovascular disease are associated with ultimate treatment intensity of a patient in COVID-19. The aim of this study is to automatically extract multiple features from initial chest CT and identify those that differ statistically significantly between treatment intensity groups. Furthermore, we want to establish a predictive model contributing to improved healthcare resources management during the pandemic.

Materials And Methods

This prospective study was approved by the local ethics committee (ID *BLINDED_FOR_REVIEW*). It is part of a research project that had been registered on ClinicalTrials.gov on 04/29/2020 (Identifier: (*BLINDED_FOR_REVIEW*).

Study population
All results of RT-PCRs for SARS-CoV-2 performed at the ED of our institution between 03/25/2020 - 04/13/2020 were downloaded from our laboratory data system (n = 4083 RT-PCRs in 3706 patients). All positive RT-PCRs were identified (n = 404 in 391 patients). In cases with multiple RT-PCRs, a patient was rated as positive, if at least one of the specimens was positive. For those 391 patients, we searched our RIS/PACS system for chest CTs performed in the study period, which resulted in 111 chest CTs. All follow-up CTs were excluded from analysis (n = 26). This resulted in 85 baseline CT scans that were included into the analysis. There were no exclusions due to bad image quality or processing errors. Figure 1 shows the search strategy.

**Laboratory data and definition of treatment intensity groups**

RT-PCRs for SARS-CoV-2 were performed at our institution using specimens from nasopharyngeal and oropharyngeal swabs. Information on the ultimate clinical pathway of a patient was retrieved from our hospital information system six weeks after CT data collection ended (05/25/2020). Based on this information, the following three groups with gradually higher treatment intensity were defined: group 1 (outpatient treatment), group 2 (inpatient treatment, general ward), and group 3 (admission to ICU). Each patient was assigned to the highest category individually reached (e.g. a patient that had initially been treated on a general ward, but eventually admitted to ICU, was assigned to group 3).

**CT acquisition parameters**

The 85 chest CTs were acquired in supine position on two 128-slice scanners: SOMATOM Definition AS+ (n = 84) and SOMATOM Force (n = 1) (both Siemens Healthineers). Mean tube voltage was 105.2 kVp (standard deviation [SD]: 10.2), tube current 80.9 mAs (SD: 19.9), and pitch factor 1.05 in all cases. Most of the scans were performed without contrast (n = 72), while 13 CTs were performed with a mean of 70.2 mL (SD: 19.2) of contrast agent (Iopromide, Bayer) at an injection rate of 4 mL/sec. Images were reconstructed with 1 mm slice thickness using soft (26f) reconstruction kernels.

**Technical details of the algorithms**

Algorithms developed by a Princeton-based research center (Details *BLINDED_FOR_REVIEW*) were locally deployed on an imaging post-processing platform (*BLINDED_FOR_REVIEW*). Due to their good performance regarding image segmentation, multiple deep convolutional neural networks (DCNNs) as specified below were used.

**Pulmonary metrics**

The 1-mm series in soft kernel reconstruction served as input to an algorithm prototype based on a Deep Image-to-Image Network (DI2IN) for lung segmentation and a subsequent DenseUNet for immediate
segmentation of opacities (ground glass opacities [GGO] and consolidation). They had been trained on data of n=9549 (DI2IN) and n=901 (DenseUNet) patients, independent from our local testing dataset. The DenseUNet defined all voxels with GGO or consolidation as positive/foreground, all other areas of the lung as negative/background. Subsequently, a HU-threshold of -200 was applied to the prediction mask to further differentiate GGO from consolidations. Further technical details and high diagnostic performance of the algorithms had been reported in a previous study [14]. All pulmonary metrics extracted in this study are specified in Table 1.

**Table 1.** Features analyzed by the algorithms with units and definitions.

<table>
<thead>
<tr>
<th>Metric name</th>
<th>Unit</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A) Pulmonary metrics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LungVol</td>
<td>mL</td>
<td>Total volume of the lung</td>
</tr>
<tr>
<td>PO</td>
<td>%</td>
<td>Percent of volume of the lung affected by opacities, equivalent to GGO and consolidation</td>
</tr>
<tr>
<td>PHO</td>
<td>%</td>
<td>Percent of volume of the lung affected by high HU opacities (equal to or above -200 HU), equivalent to consolidations</td>
</tr>
<tr>
<td>HH</td>
<td>%</td>
<td>Percent of volume of the lung with low attenuation, defined as &lt; -950 HU</td>
</tr>
<tr>
<td>LL#</td>
<td>N</td>
<td>Number of lung lesions detected (GGO or consolidations)</td>
</tr>
<tr>
<td>LSS</td>
<td>N</td>
<td>Sum of severity scores for each of the five lung lobes*</td>
</tr>
<tr>
<td>LHOS</td>
<td>N</td>
<td>Sum of severity scores for each of the five lung lobes, consolidation only*</td>
</tr>
<tr>
<td><strong>B) Cardiovascular metrics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPV</td>
<td>mL</td>
<td>Total pericardial volume. This includes the heart as well as pericardial structures like fat and/or effusion</td>
</tr>
<tr>
<td>QCC</td>
<td>mm³</td>
<td>Total volume of coronary calcifications. Only for CT without contrast</td>
</tr>
<tr>
<td>D_AAsc</td>
<td>mm</td>
<td>Diameter of ascending aorta at the level of the right pulmonary artery</td>
</tr>
<tr>
<td>D_Arch</td>
<td>mm</td>
<td>Diameter of aorta, midway through aortic arch region</td>
</tr>
<tr>
<td>D_ADsc</td>
<td>mm</td>
<td>Diameter of descending aorta, midway between the branching of the subclavian artery and celiac artery</td>
</tr>
</tbody>
</table>

* Severity scores calculated as follows. 0: lobe not affected; 1: 1-25% affected; 2: 25-50%; 3: 50-75%; 4: 75-100%

**Cardiovascular metric**
The non-electrocardiogram (ECG)-gated 1-mm series served as only input to an algorithm prototype, which uses a DCNN based on U-Net architecture for segmentation of the thoracic aorta and the total pericardial volume (TPV). The TPV segmentation, which includes the heart as well as pericardial structures such as fat and (if present) pericardial effusion is subsequently used to identify all candidate coronary calcification voxels by applying a threshold of 130 HU. The calcium detection model based on a ResNet then processes each candidate voxel, predicting which ones are true coronary calcifications. The aortic segmentation model is further used to remove any aortic calcifications, which might have been falsely classified as coronary calcifications. The aortic segmentation is also used to compute the aortic centerline. Diameters are computed at key anatomical landmarks. The aorta segmentation model was trained on 1150 annotated CT volumes; the coronary calcium detection model was initially trained on 1261 ECG-gated calcium scoring scans without contrast and then fine-tuned on 579 non-ECG gated, non-contrast enhanced chest CT scans; the TPV segmentation model was trained on 560 Chest CT volumes. Detailed information is provided elsewhere [15, 16]. Of note, the quantification of coronary calcifications (QCC) could only be calculated for series without contrast (n = 72/85). All cardiovascular metrics analyzed in this study are specified in Table 1.

**Statistical Analysis**

Categorical variables are specified as counts and percentages. For continuous variables, means with corresponding standard deviations (SDs) as measures of variance are provided. For comparison of differences between groups, one-way analyses of variance (ANOVAs) for normally distributed continuous variables, Kruskal-Wallis H tests for non-normally distributed, continuous variables, and Chi-squared tests for categorical variables were used. Normal distribution was checked with the Shapiro-Wilk Test and inspection of histograms and QQ-plots. P-values of $\leq 0.05$ were defined as being statistically significant. Statistical analyses were performed with SPSS Statistics for Windows, version 22.0 (IBM Corp). The analysis was performed in two steps:

First, a series of univariate analyses with appropriate post-hoc tests to identify metrics with predictive potential regarding treatment intensity. All pulmonary, cardiovascular and patient characteristic metrics were analyzed for statistically significant differences between the three treatment intensity groups.

Second, to assess the predictive potential of combinations of pulmonary and cardiovascular metrics for predicting whether a patient ultimately needs hospitalization or not and ICU-care or not, two binary logistic regression analyses were performed. Selection criteria for the inclusion of pulmonary and cardiovascular metrics as independent variables were P-values $\leq 0.05$ in step 1 of the analysis and adherence to the assumption of non-multicollinearity [17].

A. Dependent variable: no ICU admission [groups 1 & 2] vs. ICU admission [group 3].
B. Dependent variable: no hospitalization [group 1] vs. hospitalization [groups 2 & 3].
For both approaches, ROC curves with two curves indicating the upper and lower limits of 95% confidence were plotted with the \textit{jroct} tool from Johns Hopkins University School of Medicine \cite{18} and AUCs were calculated, with the prediction probabilities that resulted from the binary logistic regression analyses as parameters. Optimal thresholds were calculated with the method described by Youden \cite{19}.

**Results**

**Patient characteristics**

Both RT-PCR and clinical pathway data were complete. The study included data of 85 patients with a mean age of 58.7 years (SD: 17.7; range: 18 to 92 years; 32/85 [=37.6\%] female). Patient characteristics in the three treatment intensity groups are summarized in Table 2. Age differed statistically significantly between the groups ($\chi^2 = 7.13; p = 0.03$). Post-hoc analysis revealed statistically significant differences for group comparisons 1-2 ($p=0.03$) and 1-3 ($p=0.05$). Sex did not differ statistically significantly between severity groups ($\chi^2 = 0.15; p = 0.93$).

**Table 2.** Patient characteristics in the clinical treatment groups.

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>n</th>
<th>Mean age (SD)</th>
<th>Sex (M/F; % male)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (outpatient)</td>
<td>8</td>
<td>41.4 (18.9)</td>
<td>5/3 (62.5%)</td>
</tr>
<tr>
<td>Group 2 (general ward)</td>
<td>62</td>
<td>60.1 (17.1)</td>
<td>38/24 (61.3%)</td>
</tr>
<tr>
<td>Group 3 (ICU)</td>
<td>15</td>
<td>62.2 (15.3)</td>
<td>10/5 (66.7%)</td>
</tr>
</tbody>
</table>

**Automated analysis**

All datasets were processed without failure. The mean percentage of lung volume affected by opacities, high opacities as well as the two severity scores (LSS and LHOS) steadily increased from group 1 to group 3. The same is true for TPV, QCC and the aortic diameters. Table 3 shows the results of the analyses of the pulmonary and cardiovascular metrics, including results of the univariate analyses for group differences. Figure 2 illustrates the data of three metrics as boxplots.

**Table 3.** Results of the univariate analyses of pulmonary and cardiovascular metrics. Mean values and standard deviations (in brackets) for the three treatment intensity groups as well as statistical testing for group differences. Metrics with P-values $\leq 0.05$ in bold print. Regarding the abbreviations of the metrics, we refer to Table 1.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Statistical test</th>
<th>Measure</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A) Pulmonary metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LungVol</td>
<td>mL</td>
<td>4981.4 (1399.9)</td>
<td>4499.8 (1240.6)</td>
<td>3955.1 (1100.4)</td>
<td>ANOVA</td>
<td>F = 2.00</td>
<td>0.14</td>
</tr>
<tr>
<td>PO</td>
<td>%</td>
<td>0.8 (1.5)</td>
<td>11.6 (13.1)</td>
<td>31.6 (20.1)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 28.8$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>PHO</td>
<td>%</td>
<td>0.2 (0.3)</td>
<td>2.0 (2.9)</td>
<td>9.4 (10.2)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 28.79$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>%LOWHU</td>
<td>%</td>
<td>8.6 (8.6)</td>
<td>6.2 (4.8)</td>
<td>4.0 (3.3)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 7.23$</td>
<td>0.08</td>
</tr>
<tr>
<td>LL#</td>
<td>N</td>
<td>11.1 (10.7)</td>
<td>28.7 (16.6)</td>
<td>25.3 (13.4)</td>
<td>ANOVA</td>
<td>F = 4.52</td>
<td>0.01</td>
</tr>
<tr>
<td>LSS</td>
<td>N</td>
<td>1.4 (1.4)</td>
<td>4.9 (2.9)</td>
<td>9.1 (3.8)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 34.11$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>LHOS</td>
<td>N</td>
<td>0.6 (1.1)</td>
<td>2.3 (1.9)</td>
<td>4.8 (2.4)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 26.04$</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td><strong>B) Cardiovascular metrics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPV</td>
<td>mL</td>
<td>733.4 (231.7)</td>
<td>866.2 (211.2)</td>
<td>925.7 (125.5)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 10.71$</td>
<td>0.02</td>
</tr>
<tr>
<td>QCC</td>
<td>mm$^2$</td>
<td>321.4 (700.3)</td>
<td>335.8 (633.6)</td>
<td>461.9 (915.6)</td>
<td>ANOVA</td>
<td>F = 0.37</td>
<td>0.96</td>
</tr>
<tr>
<td>D_AAsc</td>
<td>mm</td>
<td>30.6 (4.1)</td>
<td>35.3 (4.1)</td>
<td>36.3 (3.0)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 9.45$</td>
<td>0.01</td>
</tr>
<tr>
<td>D_Arch</td>
<td>mm</td>
<td>27.1 (4.5)</td>
<td>29.5 (2.7)</td>
<td>30.6 (2.4)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 7.49$</td>
<td>0.09</td>
</tr>
<tr>
<td>D_ADsc</td>
<td>mm</td>
<td>23.1 (5.2)</td>
<td>26.2 (3.3)</td>
<td>26.8 (2.4)</td>
<td>Kruskal-Wallis</td>
<td>$\chi^2 = 7.95$</td>
<td>0.11</td>
</tr>
</tbody>
</table>

**Step 1: Identification of metrics with predictive potential**
Pulmonary metrics

Regarding pulmonary metrics, the treatment intensity groups differed statistically significantly in the percent of volume affected by opacities (PO) and high HU opacities (PHO) as well as in the total number of infiltrates (LL#) and the severity scores (LSS, LHOS). Details are listed in Table 3. Post-hoc testing revealed that differences in PO were statistically significant at P-values < 0.01 between all groups. The same was true for PHO. Regarding LL#, the only statistically significant difference was found for comparison 1-2 (p = 0.01). Due to confluent areas of infiltration, LL# decreased in group 3. Figure 3 displays one image example per treatment intensity group.

Cardiovascular metrics

TPV and D_AAsc differed statistically significantly between the treatment intensity groups. Post-hoc analysis revealed that the differences in TPV were statistically significant only for the comparison of groups 1 and 3 (P-value = 0.02). For D_AAsc, group comparisons 1-2 (P-value=0.02) and 1-3 (P-value=0.01) were statistically significant. Figure 4 illustrates the output of the cardiovascular algorithms.

Step 2: Comprehensive analyses regarding need of hospitalization and ICU care

According to the criteria described in the methods section, two binary logistic regressions were performed with PO, TPV, D_AAsc and #LL as independent variables. Other metrics like PHO and LSS were not included to meet the assumption of non-multicollinearity. The fit of the overall model was statistically significant (A: $\chi^2 = 20.26; p < 0.001$; B: $\chi^2 = 23.47; p < 0.001$). In detail, an increase in the unit of volume affected by opacities was associated with an increase in the odds of eventually needing ICU care (Wald $\chi^2 = 12.6; p < 0.001$). The AUCs under the ROC curves were 0.85 (ICU vs. no ICU) and 0.94 (hospitalization vs. no hospitalization). Figure 5 displays the ROCs for analyses A and B with 95% C.I.s. Of note, including sex and age into the models did have only marginal influence on the discriminatory power (A: 0.85 B: 0.96). At the optimal cutoff as determined by the Youden index, the sensitivity/specificity was 79%/75% (A) and 84%/91% (B).

CTs with vs. CTs without contrast

Diameters of the aorta were larger in those CT scans performed with contrast compared to those without. This difference was either statistically significant or showed a trend (P-values: <0.01 [D_ADsc], 0.01 [D_Arch], 0.07 [D_AAsc]). Comparing non-contrast CTs to scans with contrast, there were no statistically significant differences regarding all other metrics (P-values ranging from 0.52 [LL#] to 0.99 [LHOS]).

Discussion
This study applied a method for automated quantification of pulmonary and cardiovascular metrics to initial chest CTs of patients with RT-PCR confirmed COVID-19 to predict treatment intensity. It was demonstrated that multiple metrics continuously and statistically significantly increase with higher treatment intensity. As expected, metrics quantifying lung opacities, representing pulmonary infiltrates and therefore acute pathology, were among the strongest discriminators. Moreover, the cardiovascular metric TPV showed a statistically significant increase towards higher treatment intensity, while QCC did not. AUC under the ROCs were 0.85 for discriminating those patients that eventually needed ICU care vs. not and 0.95 for discriminating patients needing hospitalization vs. not.

Our results regarding the relevance of pulmonary metrics in COVID-19 and their association with treatment intensity are in line with reports from China. Li et al. reported an increasing extent of inflammatory pulmonary lesions from light to common to serve/critical clinical manifestation [10]. Guan et al. found no radiographic or CT abnormalities in 157 of 877 patients (17.9%) with non-severe disease and in only 5 of 173 patients (2.9%) with severe disease [20]. Noteworthy, in both studies, CTs had been analyzed manually. Similarly, Lyu et al. [7] and Zhang et al. [11] showed that the number of lung segments and lobes affected by crazy-paving pattern and consolidation increased with case severity, a fact in line with the increase in LSS and LHOS with increased treatment intensity. The opposite approach was taken by Colombi et al., who quantified areas of well-aerated, normal lung to predict adverse outcome in COVID-19 pneumonia [9]. They found that the percentage of well-aerated areas was lower in patients with ICU admission or death (57%) compared to other patients (78%), which is also compatible to our results. Of note, all aforementioned approaches required at least some manual input of a human reader and did include neither quantitative measurements of cardiovascular CT metrics nor immediate segmentation of lung opacities.

The percentage of areas with low attenuation below -950 HU (%LowHU) generally used to quantify lung emphysema [21] did not differ statistically significantly between treatment intensity groups, in contradiction to a meta-analysis that came to the conclusion that pre-existing COPD is associated with a nearly 4-fold higher risk of developing severe COVID-19 [22]. However, given the fact that the load of pulmonary opacities increased with severity of COVID-19 in our study collective, which counteracts the measurement of %LowHU, this metric is not correctly evaluable when facing pneumonia.

As previously shown, preexisting cardiovascular disease is a risk factor for adverse outcome in COVID-19 [23] and at the same time COVID-19 affects the cardiovascular system [24]. Therefore, the suggested approach includes cardiovascular metrics such as TPV as an estimate for heart size. Indeed, higher TPV was associated with higher risk for higher treatment intensity. Increased TPV might represent preexisting cardiac disease, an increased amount of pericardial fat or be mediated through the known impact of age, sex and body size [25]. The difference in mean diameters of the thoracic aorta is most probably explained by differences in age [26], because both mean diameters and age were similar in treatment intensity groups 2 and 3, but lower in group 1. Of note, in all groups all mean diameters of the aorta were within normal ranges.
In a pandemic situation with limited human resources, fully automated approaches are preferred. In this respect, Huang et al. have applied CT-derived opacification measures using deep learning to stratify four clinical subtypes according to their baseline clinical, laboratory and CT findings [27]. They provided further evidence of CT as an important tool for risk stratification in COVID-19 patients and reported percentages of lung areas with opacities ranging from 0% (mild disease) to 49.6% (critical disease), which is confirmed by the analysis at hand. However, radiological findings used to predict outcome were at the same time part of the outcome criteria of this study [28]. Furthermore, we explicitly defined the requirement of ICU-care as a parameter defining treatment intensity group 3, as we wanted to find baseline metrics that are early predictors of high workload for healthcare providers. Additionally, the present study included more complex pulmonary metrics and added cardiovascular metrics to the analysis.

This study has limitations. First, it is a single-site, one-vendor-only study design, which limits the generalizability of the results. However, given the high standardization of chest CT, we are confident that the results will prevail. Furthermore, the algorithm had been trained on a large multi-vendor dataset independent from the data we used. Second, no structured visual assessment of the segmentations was performed. However, the diagnostic accuracy of the algorithm prototypes had been demonstrated in previous studies. Third, the differentiation of consolidations from GGO was based on HU-thresholding. We are aware that this might slightly affect the accuracy of the measurement of consolidations (PHO). Fourth, this study included only patients with COVID-19 that received a CT scan, the diagnostic standard in patients with pulmonary complications from COVID-19 at our center. Therefore, the presented approach might be less relevant in medical centers with chest radiographs as diagnostic standard.

**Conclusion**

To conclude, this study demonstrates that treatment intensity stratification in COVID-19 by fully automatically extraction of pulmonary and cardiovascular metrics from initial chest CTs is feasible. Given its widespread availability, fast acquisition times and high spatial resolution, CT is an image modality of choice in times of pandemic spread of a disease affecting the respiratory system such as COVID-19. This approach can help to improve the allocation of medical resources both on a patient and public health level, complementing predictive models based on laboratory or clinical metrics.

**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARDS</td>
<td>Acute respiratory distress syndrome</td>
</tr>
<tr>
<td>AUC</td>
<td>Area under the curve</td>
</tr>
<tr>
<td>COVID-19</td>
<td>Coronavirus disease 2019</td>
</tr>
<tr>
<td>CON</td>
<td>Consolidation</td>
</tr>
</tbody>
</table>
CT  Computed tomography
D_AAsc  Diameter of ascending aorta
D_ADsc  Diameter of descending aorta
D_Arch  Diameter of the aorta midway through the aortic arch region
DCNN  Deep Convolutional Neural Network
DI2IN  Deep image-to-image network
ECG  Electrocardiogram
GGO  Ground-glass opacity
HH  Percent of volume of the lung with low attenuation < 950 HU
HU  Hounsfield unit
ICU  Intensive Care Unit
LL#  Number of lung lesions detected by the algorithm
LHOS  Sum of severity scores for each of the five lung lobes, consolidations only
LSS  Sum of severity scores for each of the five lung lobes
LungVol  Total volume of lung
PO  Percent of volume of the lung affected by high HU opacities
QCC  Quantification of coronary calcifications
SARS-CoV-2  Severe acute respiratory syndrome coronavirus 2
SD  Standard deviation
RT-PCR  Reverse-transcription polymerase chain reaction
ROC  Receiver operating characteristic
TPV  Total pericardial volume
95% C.I.  95% confidence interval

Declarations
The study was approved by the local ethics committee, the need for informed consent was waived (IRB-approval number 2020-00566).

The authors declare the following competing interests: S. Rapaka, S. Grbic, T. Re, S. Chaganti and D. Comaniciu are employees of Siemens Healthineers, Corporate Technology, Princeton, U.S.A. The company provided the algorithm prototypes used in this research article and had no influence on methodology and the statistical evaluations. All other authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

References


14. (*BLINDED FOR REVIEW*)

15. (*BLINDED FOR REVIEW*)

16. (*BLINDED FOR REVIEW*)


28. Commission GO of NH (2020) Diagnosis and Treatment Protocols for Patients with Novel Coronavirus Pneumonia v5
Figure 1

ROC curves (blue continuous line) with 95% C.I.s (grey continuous lines) for assessing the discriminatory power of multiple pulmonary and cardiovascular metrics automatically extracted from initial chest CT for predicting whether a patient finally required ICU care or not (A) and hospitalization or not (B).
Figure 2

Exemplary chest CT images in transverse plane visualizing the output of the cardiovascular algorithm: input image (A) with colored overlay for segmentation of the pericardium (B), detected coronary calcium in the circumflex artery additionally highlighted with an arrow (C) and for aorta segmentation as foundation for diameter calculations (D).
Figure 3

Exemplary chest CT images in transverse plane for each of the treatment intensity groups 1 (A), 2 (B), and 3 (C) with PO in percent in the right lower corner. (D) shows the output of the pulmonary opacity segmentation algorithm for case (C), with fine delineation of areas with infiltrate-equivalent opacities in red color.
Figure 4

Boxplots comparing pulmonary and cardiovascular metrics in different treatment intensity groups. (A) PO, (B) LSS, and (C) TPV. (*) indicate statistically significant differences in post-hoc testing.

Figure 5

Search strategy.