Combined Model of Quantitative Evaluation of Chest Computed Tomography and Laboratory Values for Assessing the Prognosis of Coronavirus Disease 2019

Prognoseabschätzung von COVID-19 mithilfe eines kombinierten Modells aus quantitativer Auswertung von Computertomografien des Thorax und Laborwerten

Authors
Gregor Scharf1, Stefanie Meiler1, Florian Zeman2, Jan Schaible1, Florian Poschenrieder1, Charlotte Knobloch1, Henning Kleine3, Sophie Elisabeth Scharf4, Julien Dinkel5,6, Christian Stroszczynski1, Niels Zorger7, Okka Wilke aHamer1,8

Affiliations
1 Institut für Röntgendiagnostik, Universitätsklinikum Regensburg, Germany
2 Zentrum für Klinische Studien, Universitätsklinikum Regensburg, Germany
3 Klinik für Pneumologie, Krankenhaus Barmherzige Brüder Regensburg, Germany
4 Fakultät für Sozialwissenschaften, Universität Mannheim, Germany
5 Klinik und Poliklinik für Radiologie, Klinikum der Universität München, Germany
6 Abteilung für Radiologie, Asklepios Fachkliniken München-Gauting, Germany
7 Institut für Radiologie, Krankenhaus Barmherzige Brüder Regensburg, Germany
8 Abteilung für Radiologie, Fachklinik Donaustauf, Germany

Key words
COVID-19, computed X-Ray tomography, pneumonia, artificial intelligence, lung volume measurements, laboratory tests

received 08.05.2021
accepted 20.12.2021
published online 2022

Bibliography
Fortschr Röntgenstr
DOI 10.1055/a-1731-7905
ISSN 1438-9029
© 2022. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14, 70469 Stuttgart, Germany

Correspondence
Dr. Gregor Scharf
Institut für Röntgendiagnostik, Universitätsklinikum Regensburg, Franz-Josef-Strauß-Allee 11, 93053 Regensburg, Germany
Tel.: +49/941/94 41 74 99
gregor.scharf@gmx.de

ZUSAMMENFASSUNG
Ziel Evaluation der prognostischen Aussagekraft der quantitativen Auswertung von Thorax-CTs, Laborwerten und deren Kombination bei COVID-19.


Ergebnisse 89 Patienten wurden zwischen 07.03.20 und 27.04.20 in die Hauptkohorte eingeschlossen (Durchschnittsalter 60,3 Jahre), 38 Patienten hatten ein negatives Outcome. In der univariaten Regressionsanalyse waren alle CT-Parameter, CRP, relative Lymphozytenzahl (RLC), Troponin und LDH mit einem negativen Outcome assoziiert. In der multivariaten Regressionsanalyse waren PO, CRP und RLC unabhängige Prädiktoren eines negativen Outcomes. Die Kombination dieser Werte zeigte eine starke Vorhersagekraft mit einem C-Index von 0,87. Eine Punkteskala basierend auf diesen Werten ermöglichte die Einteilung der Patienten in 4 Gruppen mit einem Risiko von 7 %, 30 %, 67 % und 100 % für ein negatives Outcome. Die Validierungskohorte bestand aus 28 Patienten zwischen 05.05.20 und 13.11.20. Ein negatives Outcome hatten 6 % der Patienten mit 0 Punkten, 50 % mit 1 Punkt und 100 % bei 2 oder 3 Punkten.

Schlussfolgerung Die Kombination aus PO, CRP und RLC zeigte eine sehr hohe Vorhersagekraft für ein negatives Outcome. Eine auf diesen Werten basierende 4-Punkte-Skala...
Introduction

The novel coronavirus SARS-CoV-2 has spread from Wuhan city, China to most other countries worldwide. More than one year after the beginning of the pandemic, many European countries are facing a “fourth wave”. Although no longer recommended as a first-line test, CT of the chest still plays an important role in the diagnostic workup and follow-up of Coronavirus Disease 2019 (COVID-19) [1]. The typical CT findings in COVID-19 pneumonia are multiple ground-glass opacities and consolidation predominantly located in the periphery of the mid and lower lung zones [2–4]. It has been shown that the burden of pulmonary opacifications correlates with prognosis [5–7]. While a rough assessment of pulmonary involvement can be done by subjective eyeballing, human assessment is not as reliable in exact grading [8]. Computer-assisted quantitative analysis of CT images might facilitate and accelerate the precise assessment of pulmonary changes related to COVID-19 pneumonia. Also, quantifiable results might enable the calculation of accurate scores, for example, regarding individual prognosis. Furthermore, reader variability could be diminished.

To date, many studies have investigated the application of artificial intelligence (AI) in COVID-19, but only a few have examined AI-based analyses of CT images to predict prognosis [8–14]. These studies used automatic or semiautomatic procedures, different radiological parameters, and different segmentation techniques. As expected, it was confirmed that the extent of lung opacities was a prognostic marker for an unfavorable clinical outcome. Two studies only distinguished between different densities of opacifications [9, 14]. However, this differentiation might be important because several studies demonstrated that ground glass opacities are the predominant finding in patients with a favorable outcome, whereas consolidation predominates in cases with an adverse outcome [15, 16]. Other groups investigated and determined the prognostic relevance of a number of different laboratory values [17–20]. Zheng et al. [21] evaluated a combined model of clinical and subjective CT features for the prediction of an adverse outcome in patients with COVID-19, while Schalekamp et al. [22] developed a predictive risk model for critical illness in...
patients with COVID-19 based on laboratory findings and visual, semi-quantitative analysis of chest radiographs. The potential benefit of a combined model of quantitative analysis of CT images and laboratory values has not yet been evaluated. However, combining a marker of actual lung involvement with a marker of systemic disease seems reasonable for determining severity of disease.

The aim of our study was to evaluate if the AI-based analysis of chest CT images stratified according to density and extent of opacifications correlates with the clinical outcome of patients suffering from COVID-19 pneumonia. Furthermore, we evaluated if a combined model including CT findings and laboratory data enhances prognostic power.

Materials and methods

This study was approved by the institutional ethics committee. Written informed consent was waived due to the retrospective nature of this study.

Patient recruitment

Cohort for model development

120 consecutive patients from one academic tertiary care hospital and one general tertiary care hospital in Bavaria, Germany with COVID-19 infection confirmed by Polymerase Chain Reaction (PCR) were included in a first step for model development. Chest CT scans were acquired between March 07 and April 27, 2020. The patients had been admitted for CT scan either on first admission or after transfer from another hospital. The indication for the CT scan was based on clinical assessment. In case of more than one CT per patient, only the first CT scan was considered for analysis. Exclusion criteria were a negative chest CT for COVID-19 pneumonia (2 patients), admission to intensive care unit (ICU) > 3 days before CT scan (9 patients), incorrect lung segmentation (16 patients), incomplete scan data (3 patients). Thus, 89 patients were finally included into the analysis.

Validation cohort

The validation cohort consisted of 32 patients from the same academic tertiary care hospital with PCR-confirmed COVID-19 infection with the same exclusion criteria as the first cohort. Chest CT scans were acquired between May 05, 2020 and November 13, 2020. Again, only the first CT scan from each patient was considered for analysis. One patient was excluded due to incorrect automatic lung segmentation, 3 patients were excluded due to admission to ICU > 3 days before CT scan. Thus, 28 patients were ultimately included in the validation cohort.

The flowchart of patient inclusion is shown in Fig. 1.

Epidemiology and laboratory data

Patient characteristics (age, sex), date of symptom onset, and laboratory data were extracted from electronic patient records. With respect to laboratory data, C-reactive protein (CRP), white blood cell count (WBC), relative lymphocyte count (RLC), relative eosinophil count (REC), troponin, NT-proBNP, fibrinogen, interleukin-6 (IL-6), D-dimer, lactate dehydrogenase (LDH), creatine kinase (CK), CK-MB, and lactate were included in the analysis. Laboratory values were recorded as close as possible to the date of the CT scan, but not more than 3 days before or after.

Definition of outcome

For analysis of prognosis, two compound outcomes were defined: positive outcome was defined as either discharge or regular ward care; negative outcome was defined as need for mechanical ventilation, treatment in the ICU, extracorporeal membrane oxygenation (ECMO), or death.
Image acquisition

CT scans were performed using multi-detector spiral CT scanners (SOMATOM Definition Flash/SOMATOM Sensation 16, Siemens Healthcare GmbH) in supine position during end-inspiration. Intravenous contrast material was administered at the discretion of the radiologist considering the individual study indication (e.g., if pulmonary embolism or superinfection was suspected). Automatic tube voltage selection was applied with a reference tube voltage of 120 kV. The tube current was regulated by an automatic tube current modulation technique with the reference mAs being 40–110. The tube current was 100 kV and 140 kV when the dual-energy technique was applied. Collimation width was 0.625 mm.

Statistical methods

Continuous variables are presented as means ± standard deviation for normal distributed or as median (q1–q3) for non-normal distributed variables. Categorical variables are presented as absolute numbers and percentages. Univariable logistic regression analyses were performed to identify predictors for an unfavorable outcome. A multivariable logistic regression model was built according to the following criteria: 1) only predictors with a p-value < 0.1 in the univariable model and 2) with no more than 10 % missing values were added. 3) If two predictors were highly correlated with r > 0.8 (multicollinearity), only one was included in the model. 4) A backward selection was performed and only predictors with p < 0.05 remained in the final model. Odds’ ratios (OR) and corresponding 95 % confidence intervals are presented as effect estimates for all logistic regression models. ORs are presented as per 1 unit change for each predictor, if not otherwise specified. Finally, a prediction score was established by dichotomizing each predictor of the multivariable model. The optimal cut-off for each predictor was chosen according to the Youden-Index in the univariable case.

Model validation

The performance (calibration and discrimination) of the final score was evaluated by a calibration plot and the concordance statistic (c-index). An internal validation of the score to obtain stable optimism-corrected estimates was performed using bootstrap validation with n = 1000 replications. The model was further validated on the validation cohort by calculating the score for each patient and assessing the predictive accuracy for a negative outcome. A p-value < 0.05 was considered statistically significant for all analyses. All analyses were performed using R, version 4.1.1 (The R Foundation for Statistical Computing).

Results

Patient data, outcome

Cohort for model development

A total of 89 CT examinations were evaluated. The average age of the patients was 63.3 years (SD 14.4). 45.6 % of the 89 patients were female. 68 CT scans were performed without contrast agent, 21 were contrast-enhanced for clinical reasons. The mean time period from the onset of symptoms to the CT scan was 9.6 days (SD 5.7). 51 patients (57.3 %) had a positive outcome, 38 patients (42.7 %) a negative outcome. The average timespan from CT scan to negative outcome was 1.6 days (SD 2.5; 1–8 days).

Validation cohort

28 CT scans were evaluated. The average age of the patients was 53.8 years (SD 13.4). 32.1 % patients were female. 7 CT scans (25 %) were contrast-enhanced. The mean time from the onset of symptoms to the CT scan was 7 days (SD 3.3). 18 patients (64.3 %) had a positive outcome, 10 patients (35.7 %) a negative outcome.
The average timespan between CT scan and negative outcome was 0 days (SD 1.6; -3–3 days).

Patient data is summarized in ▶ Table 1.

Univariable logistic regression analyses

Quantitative CT analysis

The results of quantitative CT analysis are summarized in ▶ Table 2. All of the 6 quantitative CT values were significantly associated with a negative outcome.

Laboratory values

A significant association with a negative outcome was found for CRP, RLC, troponin, and LDH. The laboratory values and their predictive value regarding patient outcome are summarized in ▶ Table 3.

Multivariable logistic regression analysis

The final multivariable logistic regression model after variable selection as described in the methods section for the prediction of a negative outcome included PO (OR 1.05 per 1%; 95 %-CI 1.02–1.09; P = 0.002), CRP value (OR 1.01; 95 %-CI 1.00–1.02; P = 0.027), and RLC (OR 0.95; 95 %-CI 0.91–0.99; P = 0.011) (▶ Table 4). The C-index for the model was 0.87 (95 %-CI: 0.80–0.95), indicating a good discrimination between positive and negative outcome.

Based on these findings, we developed a score to facilitate the application of the multivariable model in the clinical routine by dichotomizing each predictor. The cut-off values were chosen at 39 % for PO, 80 mg/l for CRP and 15 % for RLC. A 4-point score (0–3) was established: 1 point each if the cut-off for PO or CRP was exceeded or if RLC fell below the cut-off. Based on the total score, the risk for a negative course of disease in our cohort was as follows: 7 % for a score of 0, 30 % for a score of 1, 67 % for a score of 2, and 100 % for a score of 3 (▶ Table 5).

Internal validation

To validate the score, various methods were applied:

Predictive performance

In our dataset, the C-index (AUC) for the prediction of a negative outcome was high (0.87). Somers’ Dxy rank correlation between the predicted probabilities and the observed responses was 0.744.

Bootstrap validation

An internal validation of the score was performed to obtain stable optimism-corrected estimates by using bootstrap validation with n = 1000 repetitions. The optimism corrected c-index was 0.87, and the corrected Dxy was 0.744.

Calibration plot

A calibration plot comparing the predicted probabilities to the actual probabilities is shown in ▶ Fig. 2. Both the apparent and bias-corrected probabilities are shown in comparison to the ideal line. The score shows good overall performance.

External validation

Applying the score to the validation cohort, the predictive performance was similar to the model development cohort.

<table>
<thead>
<tr>
<th>Cohort for model development</th>
<th>All patients</th>
<th>Positive outcome</th>
<th>Negative outcome</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>89 (100)</td>
<td>51 (57.3)</td>
<td>38 (42.7)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>35 (39.3)</td>
<td>21 (41.1)</td>
<td>14 (36.8)</td>
<td>0.68</td>
</tr>
<tr>
<td>Age</td>
<td>60.3 ± 14.4</td>
<td>60.8 ± 15.2</td>
<td>59.7 ± 13.5</td>
<td>0.72</td>
</tr>
<tr>
<td>Time between symptom onset and CT (days)</td>
<td>9.6 ± 5.7</td>
<td>10.0 ± 6.0</td>
<td>8.8 ± 5.4</td>
<td>0.43</td>
</tr>
<tr>
<td>Time between CT and negative outcome (days)</td>
<td>1.6 ± 2.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validation cohort</th>
<th>All patients</th>
<th>Positive outcome</th>
<th>Negative outcome</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>28 (100)</td>
<td>18 (64.3)</td>
<td>10 (35.7)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>9 (32.1)</td>
<td>6 (33.3)</td>
<td>3 (30)</td>
<td>0.91</td>
</tr>
<tr>
<td>Age</td>
<td>53.8 ± 13.4</td>
<td>53.1 ± 15.1</td>
<td>53.2 ± 10.4</td>
<td>0.72</td>
</tr>
<tr>
<td>Time between symptom onset and CT (days)</td>
<td>7 ± 3.3</td>
<td>6 ± 3.2</td>
<td>8.7 ± 3.1</td>
<td>0.06</td>
</tr>
<tr>
<td>Time between CT and negative outcome (days)</td>
<td>0 ± 1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data are presented as n (%) or mean ± standard deviation; positive outcome: discharge or regular ward care; negative outcome: need for mechanical ventilation, treatment in intensive care unit, ECMO, or death.
A negative outcome was seen for 6% of the patients with a score of 0, 50% for a score of 1, and 100% for a score of 2 or 3 (Table 5).

**Discussion**

In this study, it was found that AI-based automatic lung segmentation and quantification of opacities in COVID-19 pneumonia is feasible in most cases (Fig. 3, 4). All calculated values (LV, VO, PO, VHO, PHO, MHUL) correlated with patient outcome. The percentage of opacified lung volume showed the best suitability for the prediction of a bad outcome with an AUC of 0.83. The percentage of opacity and percentage of high opacity demonstrated such a high degree of correlation that an evaluation stratified according to the density of opacities did not add any additional value.
information, i.e., separate investigation of ground glass opacity and consolidation was not helpful in our cohort. With respect to laboratory values, a higher CRP, troponin, and LDH as well as a lower RLC were significantly associated with a negative clinical outcome. Multivariate regression analysis revealed a higher PO, higher CRP value, and lower RLC to be the best predictors for an adverse outcome. To enhance the predictive power, a model combining the aforementioned values was established. The C-index for the combined model was exceptionally high (0.87). To facilitate integration of the combined model into the clinical routine, a 4-point score was developed applying the cut-off values of 39% for PO, 80 mg/l for CRP, and 15% for RLC. There is a great need for a prognostic score tailored to patients suffering from COVID-19 pneumonia because it was shown that an established score like the CURB-65 did not perform well [13]. The strength of the proposed score is the combination of a powerful imaging feature with readily available laboratory data. All three components can be quickly and easily determined in any hospital regardless of the existing infrastructure. The score was able to separate patients into clearly defined risk groups ranging from 7% for a score of 0

<table>
<thead>
<tr>
<th>Predictor</th>
<th>OR (95%-CI)</th>
<th>p</th>
<th>c-index (95%-CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO (%)</td>
<td>1.05 (1.02–1.09)</td>
<td>0.002</td>
<td>0.87 (0.80–0.95)</td>
</tr>
<tr>
<td>CRP (mg/l)</td>
<td>1.01 (1.00–1.02)</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>RLC (%)</td>
<td>0.95 (0.91–0.99)</td>
<td>0.011</td>
<td></td>
</tr>
</tbody>
</table>

PO: percentage of opacity; OR: odds ratio; CI: confidence interval; c-index, concordance index.
to 100 % for a score of 3. The accuracy of the score was confirmed in a validation cohort with excellent predictive performance.

The correlation of the burden of opacifications on chest CT with clinical outcome was demonstrated in earlier studies [8–10, 13]. However, cut-off values were reported to be much lower compared to our results (Colombi et al.: 27 %, Grodecki et al.: 1.8 % for consolidation and 13.5 % for ground glass opacities, Lanza et al.: 23 %, Gieraerts et al.: 20 %). The most probable explanation for this discrepancy is the different composition of the cohorts. CT scans in our cohort were performed later in the course of disease (mean 9.6 days after symptom onset in our cohort as opposed to 5–7 days in the cited studies). Furthermore, the percentage of adverse outcomes was high in our cohort (42.7 %) compared to the other studies (15.6–46 %). The possible link is the fact that one of the recruiting hospitals in our study was a tertiary referral hospital equipped with an ECMO unit. This factor results in a bias towards seriously diseased patients in our cohort. In line with this circumstance, age was not a prognostic factor in our cohort as it has been in most other investigations [25, 26]. Taken together, the results of our investigation tend to include seriously rather than to mildly diseased patients. This is not necessarily a disadvantage because a viable prognostic score is especially needed for these at-risk patients.

There are some limitations to the study. The CT images were acquired in two different hospitals and were heterogeneous regarding examination protocols. 21 of 89 CT scans were contrast-enhanced for clinical reasons. The lung segmentation algorithm performed equally well in both subgroups. Nevertheless, heterogeneous examination protocols might have affected the results of automatic image analysis and could have introduced a bias. It has been shown earlier that results of lung densitometry are affected by several technical parameters [28]. However, reconstruction algorithm, slice thickness, CT dose, CT scanner, and iterative reconstruction have a minimal effect on mean lung density measurements [29]. The most important factor is the application of contrast media. Heussel et al. showed that the median lung density increased by 18 HU after contrast application in a cohort of patients with pulmonary emphysema [27]. The median lung density in our cohort differed substantially between the subgroups (165 HU). Therefore, the difference created by contrast material seems to be negligible in our cohort of patients with COVID-19 pneumonia.

Automatic lung segmentation was incorrect in 16 cases, especially in the case of CT scans with large areas of subpleural consolidation or pneumothorax (Fig. 5). The version of the software prototype used in this study did not allow for manual correction of lung segmentation so that these CTs had to be excluded from evaluation. Meanwhile the software was further improved to allow manual corrections by the user. Thus, exclusion of cases may no longer be necessary. Software-based lung analysis was not tested for reproducibility. However, the group of Gieraerts et al. used the

<table>
<thead>
<tr>
<th>Score</th>
<th>Cohort for model development (N = 83)</th>
<th>Validation cohort (N = 28)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive outcome</td>
<td>Negative outcome</td>
</tr>
<tr>
<td>0</td>
<td>25 (93 %)</td>
<td>2 (7 %)</td>
</tr>
<tr>
<td>1</td>
<td>14 (70 %)</td>
<td>6 (30 %)</td>
</tr>
<tr>
<td>2</td>
<td>7 (33 %)</td>
<td>14 (67 %)</td>
</tr>
<tr>
<td>3</td>
<td>0 (0 %)</td>
<td>15 (100 %)</td>
</tr>
</tbody>
</table>

Positive outcome: discharge or regular ward care; negative outcome: need for mechanical ventilation, treatment in intensive care unit, ECMO, or death.
same software and was able to show a high intrareader reproducibility [8].

Further research is necessary to validate our findings in a larger validation cohort. Furthermore, it remains to be assessed if our prognostic model could be enhanced by integrating more parameters, like comorbidities or distribution of lung abnormalities in the score.

In summary, in our cohort of patients with advanced COVID-19 pneumonia, the automatic software-based determination of PO on chest CT images was a highly accurate prognostic factor. The combination of PO with two readily available laboratory parameters (CRP and RLC) further enhanced the prognostic power and made it possible to establish an easy and quick to employ 4-point score for individual risk stratification.

**CLINICAL RELEVANCE**

- A high PO in chest CT in COVID-19 pneumonia is a highly prognostic factor for a negative outcome.
- The combination of PO with two readily available laboratory parameters (CRP and RLC) further enhances the prognostic power.
- A 4-point scoring system based on these values allows quick and easy to employ individual risk stratification.

**ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>artificial intelligence</td>
</tr>
<tr>
<td>CK</td>
<td>creatine kinase</td>
</tr>
<tr>
<td>CK-MB</td>
<td>creatine kinase MB</td>
</tr>
<tr>
<td>COVID-19</td>
<td>coronavirus disease 2019</td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
</tr>
<tr>
<td>CTPA</td>
<td>CT pneumonia analysis</td>
</tr>
<tr>
<td>ECMO</td>
<td>extracorporeal membrane oxygenation</td>
</tr>
<tr>
<td>ICU</td>
<td>intensive care unit</td>
</tr>
<tr>
<td>IL-6</td>
<td>interleukin 6</td>
</tr>
<tr>
<td>LDH</td>
<td>lactate dehydrogenase</td>
</tr>
<tr>
<td>LV</td>
<td>lung volume</td>
</tr>
<tr>
<td>MHUL</td>
<td>mean Hounsfield units of the whole lung</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PHO</td>
<td>percentage of high opacity</td>
</tr>
<tr>
<td>PO</td>
<td>percentage of opacity</td>
</tr>
<tr>
<td>REC</td>
<td>relative eosinophil count</td>
</tr>
<tr>
<td>RLC</td>
<td>relative lymphocyte count</td>
</tr>
<tr>
<td>VHO</td>
<td>volume of high opacity</td>
</tr>
<tr>
<td>VO</td>
<td>volume of opacity</td>
</tr>
<tr>
<td>WBC</td>
<td>leukocyte count</td>
</tr>
</tbody>
</table>

**Conflict of Interest**

The authors declare that they have no conflict of interest.

References


