

# 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 Scharf<sup>1</sup> , Stefanie Meiler<sup>1</sup>, Florian Zeman<sup>2</sup>, Jan Schaible<sup>1</sup>, Florian Poschenrieder<sup>1</sup>, Charlotte Knobloch<sup>1</sup>, Henning Kleine<sup>3</sup>, Sophie Elisabeth Scharf<sup>4</sup>, Julien Dinkel<sup>5, 6</sup>, Christian Stroszczyński<sup>1</sup>, Niels Zorger<sup>7</sup>, Okka Wilkea Hamer<sup>1, 8</sup>

### 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 10.03.2022

### Bibliography

Fortschr Röntgenstr 2022; 194: 737–746

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.

**Material und Methoden** Retrospektive Auswertung von Patienten mit einer mittels PCR bestätigten COVID-19-Pneumonie und Thorax-CT zwischen 07.03.20 und 13.11.20. Mithilfe eines Software-Prototypen wurden das Volumen und der prozentuale Anteil (PO) von Dichteanhebungen der Lunge und die durchschnittlichen HU der gesamten Lunge quantifiziert. 13 Laborwerte wurden erhoben. Als negatives Outcome wurden definiert: Tod, Einweisung auf Intensivstation, mechanische Beatmung oder extrakorporale Membranoxygenierung. Behandlung auf Normalstation oder Entlassung wurden als positives Outcome definiert. Mittels logistischer Regression wurde die prognostische Wertigkeit der CT-Parameter und der Laborwerte bestimmt. Auf Grundlage der unabhängigen Prädiktoren wurde ein Scoring-System zur Vorhersage der Prognose entwickelt. Der Score wurde in einer separaten Validierungskohorte überprüft.

**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

erleichtert die Risikostratifizierung im klinischen Alltag und zeigte eine gute Übereinstimmung in der Validierungskohorte.

#### Kernaussagen:

- Eine hohe PO im Thorax-CT ist mit einem negativen Outcome bei COVID-19 assoziiert.
- PO, CRP und RLC sind unabhängige Prädiktoren eines negativen Outcomes und ihre Kombination hat eine hohe Vorhersagekraft.
- Eine 4-Punkte-Skala, die auf diesen Werten basiert, erlaubt eine schnelle Risikostratifizierung im klinischen Alltag.

#### ABSTRACT

**Purpose** To assess the prognostic power of quantitative analysis of chest CT, laboratory values, and their combination in COVID-19 pneumonia.

**Materials and Methods** Retrospective analysis of patients with PCR-confirmed COVID-19 pneumonia and chest CT performed between March 07 and November 13, 2020. Volume and percentage (PO) of lung opacifications and mean HU of the whole lung were quantified using prototype software. 13 laboratory values were collected. Negative outcome was defined as death, ICU admittance, mechanical ventilation, or extracorporeal membrane oxygenation. Positive outcome was defined as care in the regular ward or discharge. Logistic regression was performed to evaluate the prognostic value of CT parameters and laboratory values. Independent predictors were combined to establish a scoring system for prediction of prognosis. This score was validated on a separate validation cohort.

**Results** 89 patients were included for model development between March 07 and April 27, 2020 (mean age: 60.3 years). 38 patients experienced a negative outcome. In univariate

regression analysis, all quantitative CT parameters as well as C-reactive protein (CRP), relative lymphocyte count (RLC), troponin, and LDH were associated with a negative outcome. In a multivariate regression analysis, PO, CRP, and RLC were independent predictors of a negative outcome. Combination of these three values showed a strong predictive value with a C-index of 0.87. A scoring system was established which categorized patients into 4 groups with a risk of 7 %, 30 %, 67 %, or 100 % for a negative outcome. The validation cohort consisted of 28 patients between May 5 and November 13, 2020. A negative outcome occurred in 6 % of patients with a score of 0, 50 % with a score of 1, and 100 % with a score of 2 or 3.

**Conclusion** The combination of PO, CRP, and RLC showed a high predictive value for a negative outcome. A 4-point scoring system based on these findings allows easy risk stratification in the clinical routine and performed exceptionally in the validation cohort.

#### Key Points:

- A high PO is associated with an unfavorable outcome in COVID-19
- PO, CRP, and RLC are independent predictors of an unfavorable outcome, and their combination has strong predictive power
- A 4-point scoring system based on these values allows quick risk stratification in a clinical setting

#### Citation Format

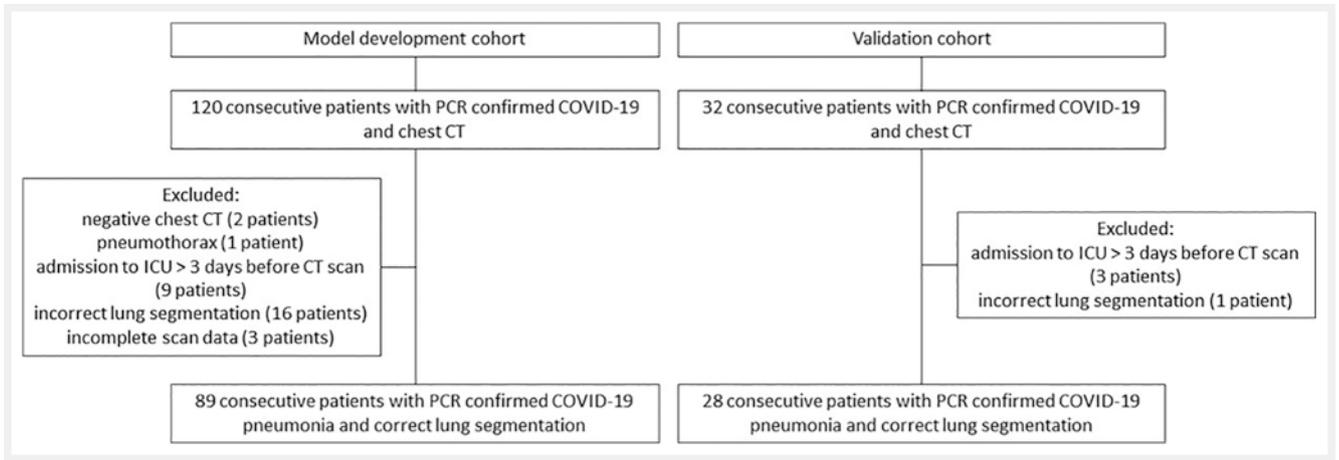
- Scharf G, Meiler S, Zeman F et al. Combined Model of Quantitative Evaluation of Chest Computed Tomography and Laboratory Values for Assessing the Prognosis of Coronavirus Disease 2019. *Fortschr Röntgenstr* 2022; 194: 737–746

## 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 indi-

vidual 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



► **Fig. 1** Patient selection process.

► **Abb. 1** Patientenselektion.

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), pneumothorax (1 patient), incomplete CT scan data (3 patients) and incorrect automatic lung

segmentation (16 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–0.75 mm. Axial planes were reconstructed with a slice thickness of 0.75–1.5 mm (81 CTs) and 2–5 mm (8 CTs) in the lung kernel and with a slice thickness of 0.75–1 mm (82 CTs) and 2–5 mm (7 scans) in the soft-tissue kernel using either filtered back-projection or iterative reconstructions. Dual-energy acquisitions were reconstructed as a weighted average with 60% from the 140 kVp and 40% from the 100 kVp data. Additional sagittal and coronal MPRs were reconstructed with a slice thickness of 1–3 mm using a lung and a soft-tissue kernel. The pictures were sent to a picture archiving and communication system (PACS, Syngo Imaging, Siemens Healthcare GmbH).

## Image evaluation

Axial reconstructions of the CT scans in the soft-tissue kernel were processed in syngo.via (Siemens Healthcare GmbH) using the CT Pneumonia Analysis prototype (CTPA). This software performs automatic lung segmentation and differentiates between normal lung tissue and opacified lung tissue by a deep learning-based algorithm. This software has been validated before including CT scans with a slice thickness of > 3.0 mm [23]. Opacified lung tissue is subdivided into two categories, namely density below and above a cutoff-value of –200 HU. This threshold is preset by the software and represents the approximate transition of ground glass to consolidation. The automatic segmentation of the lung parenchyma and opacities was reviewed for correctness by a senior radiologist who was blinded to clinical data and laboratory data. The software calculated the following values: lung volume (LV, ml), volume of opacity (VO, ml), percentage of opacity (= VO/LV, PO, %), volume of high (> –200HU) opacity (VHO, ml), percentage of high opacity (= VHO/LV, PHO, %), and mean HU of the whole lung (MHUL, HU).

## Statistical considerations

### Sample size

For both cohorts, the maximal available sample size within the predefined periods was included. With a sample size of 89 patients with 38 patients experiencing a negative outcome, a prediction model with 3 predictors had a sufficient number of events per variable (EPV)[24] to get a stable and not overfitted regression model. The validation cohort with 28 patients (10 negative outcomes) was large enough to give first evidence about the general performance of the developed score.

## Statistical methods

Continuous variables are presented as means  $\pm$  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. Odd's 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 60.3 years (SD 14.4). 35 (39.3%) 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). 9 (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.

► **Table 1** Patient demographics.

► **Tab. 1** Demografische Daten.

	All patients	Positive outcome	Negative outcome	p
<b>Cohort for model development</b>				
Number of patients	89 (100)	51 (57.3)	38 (42.7)	
Female	35 (39.3)	21 (41.1)	14 (36.8)	0.68
Age	60.3 ± 14.4	60.8 ± 15.2	59.7 ± 13.5	0.72
Time between symptom onset and CT (days)	9.6 ± 5.7	10.0 ± 6.0	8.8 ± 5.4	0.43
Time between CT and negative outcome (days)			1.6 ± 2.5	
<b>Validation cohort</b>				
Number of patients	28 (100)	18 (64.3)	10 (35.7)	
Female	9 (32.1)	6 (33.3)	3 (30)	0.91
Age	53.8 ± 13.4	53.1 ± 15.1	55.2 ± 10.4	0.72
Time between symptom onset and CT (days)	7 ± 3.3	6 ± 3.2	8.7 ± 3.1	0.06
Time between CT and negative outcome (days)			0 ± 1.6	

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.

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'  $D_{xy}$  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  $D_{xy}$  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. A

► **Table 2** Predictive value of quantitative CT analysis on patient outcome.► **Tab. 2** Prädiktive Aussagekraft der quantitativen CT-Analyse bezüglich des Outcomes.

Variable	All patients	Positive outcome	Negative outcome	OR (95 %-CI)	p
LV (ml)	4083 (3049–4984)	4486 (3594–5187)	3461 (2739–4408)	0.94 per 100 ml (0.90–0.98)	0.001
VO (ml)	604 (293–1496)	427 (173–923)	1423 (671–1936)	1.16 per 100 ml (1.08–1.25)	<0.001
PO (%)	14.4 (6.8–47)	10.0 (4.0–16.5)	49.4 (18.5–61.4)	1.07 per 1 % (1.04–1.10)	<0.001
VHO (ml)	79 (32–237)	53 (13–110)	242 (75–537)	1.92 per 100 ml (1.34–2.74)	<0.001
PHO (%)	2.2 (0.7–7.6)	1.2 (0.3–3.0)	7.8 (2.1–16.3)	1.23 per 1 % (1.02–1.39)	<0.001
MHUL	–718 (–784–609)	–764 (–794–684)	–599 (–718–513)	1.13 per 10 units (1.07–1.20)	<0.001

LV: lung volume; VO: volume of opacity; PO: percentage of opacity; VHO: volume of high opacity; PHO: percentage of high opacity; MHUL: mean HU of the whole lung; MHUO: mean HU of opacity. Data are presented as median (q1–q3); positive outcome: discharge or regular ward care; negative outcome: need for mechanical ventilation, treatment in intensive care unit, ECMO, or death; OR: odds ratio; CI: confidence interval.

► **Table 3** Predictive value of laboratory values on patient outcome.► **Tab. 3** Prädiktive Vorhersagekraft der Laborwerte auf das Outcome.

Laboratory value	Number of patients with value available (%)	All patients	Positive outcome	Negative outcome	OR (95 %-CI)	p
CRP (mg/l)	86 (97)	61.7 (31.7–145)	40.3 (18–61.3)	130 (85–195)	1.01 (95 %-CI:1.01,1.02)	<0.001
WBC (cells/nl)	85 (96)	6.6 (4.7–8.2)	5.9 (4.4–7.2)	7.2 (5.4–10.2)	1.02 (95 %-CI:0.99,1.07)	0.4
RLC (%)	83 (93)	13.6 (9.9–20.6)	18.1 (12.1–25)	11.7 (9–14.4)	0.95 (95 %-CI:0.9,0.99)	0.05
REC (%)	80 (90)	0.1 (0–0.8)	0.1 (0–0.8)	0 (0–0.6)	0.98 (95 %-CI:0.67,1.39)	0.91
Troponin (ng/l)	47 (53)	12 (7.3–34.5)	8.4 (6–14)	30 (8–69.3)	1.06 (95 %-CI:1.02,1.12)	0.03
NT-proBNP (pmol/l)	35 (39)	387 (138.5–1112)	224 (115–469)	455 (186–2576)	1 (95 %-CI:1,1)	0.19
Fibrinogen (mg/dl)	27 (30)	579.3 (512.5–645.5)	601.2 (588–614.3)	579.3 (504–650)	1 (95 %-CI:0.99,1.01)	0.85
IL-6 (pg/ml)	36 (40)	90.4 (34.4–186.8)	35.1 (30.7–62.9)	129.8 (74.7–233.4)	1.01 (95 %-CI:1,1.03)	0.06
D-dimer (mg/l)	68 (76)	284 (12.9–970.8)	263 (154.5–578.3)	393.5 (10.5–1083)	1 (95 %-CI:1,1)	0.11
LDH (U/l)	82 (92)	336.5 (262–473.5)	289 (243.5–356)	485 (360.5–611.5)	1.01 (95 %-CI:1.01,1.02)	<0.001
CK (U/l)	53 (60)	145 (73–309)	89 (57.5–243.3)	186 (132–436)	1 (95 %-CI:1,1.01)	0.06
CK-MB (ng/ml)	23 (26)	9.4 (1.5–18.2)	2.3 (1.4–7.9)	11.4 (3.4–20.1)	1.07 (95 %-CI:0.97,1.21)	0.2
Lactate (mmol/l)	73 (82)	0.16 (0.09–0.89)	0.12 (0.09–0.83)	0.18 (0.09–1)	1.01 (95 %-CI:0.94,1.09)	0.8

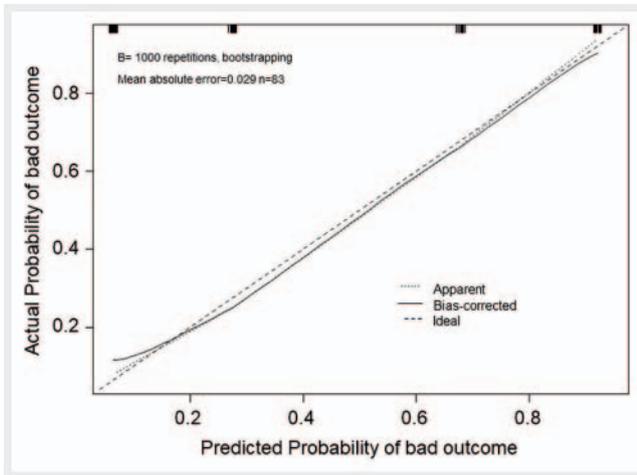
Data are presented as median (q1–q3); positive outcome: discharge or regular ward care; negative outcome: need for mechanical ventilation, treatment in intensive care unit, ECMO, or death; OR: odds ratio; CI: confidence interval.

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



► **Fig. 2** Calibration plot comparing the predicted probability of a negative outcome to the actual probability.

► **Abb. 2** Kalibrierungsplot mit dem Vergleich der vorhergesagten und tatsächlichen Wahrscheinlichkeit eines negativen Outcomes.

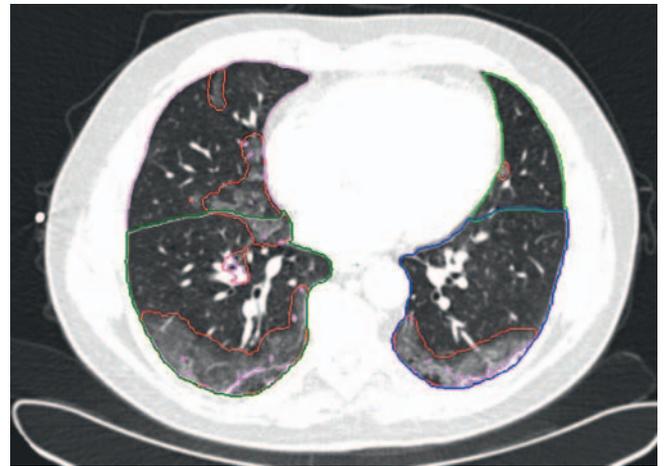
► **Table 4** Multivariable logistic regression model for the prediction of a negative patient outcome.

► **Tab. 4** Multivariates Modell zur Vorhersage eines negativen Outcomes.

Predictor	OR (95%-CI)	p	c-index (95%-CI)
PO (%)	1.05 (1.02–1.09)	0.002	0.87 (0.80–0.95)
CRP (mg/l)	1.01 (1.00–1.02)	0.027	
RLC (%)	0.95 (0.91–0.99)	0.011	

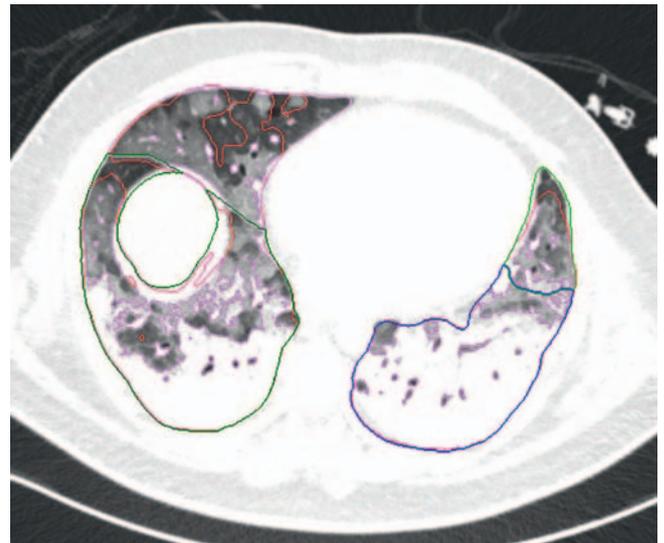
PO: percentage of opacity; OR: odds ratio; CI: confidence interval; c-index, concordance index.

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



► **Fig. 3** CT image of a 65-year-old patient with mild COVID-19 pneumonia and correct automatic segmentation. The blue, light green, violet, and dark green lines show the segmentation of the different lung lobes. Opacified lung areas are marked by a red line, and high opacity is marked by a violet line.

► **Abb. 3** CT eines 65-jährigen Patienten mit einer milden COVID-19-Pneumonie und korrekter Segmentierung. Die blaue, hellgrüne, violette und dunkelgrüne Linie markieren die Segmentierung der Lungenlappen. Dichte angehobene Areale sind von einer roten Linie markiert, Konsolidierungen von einer violetten Linie.



► **Fig. 4** CT image of a 39-year-old patient with severe COVID-19 pneumonia and correct automatic segmentation. Color coding same as in Fig. 3.

► **Abb. 4** CT eines 39-jährigen Patienten mit schwerer COVID-19-Pneumonie und korrekter Segmentierung. Farbige Markierungen wie in Abb. 3.

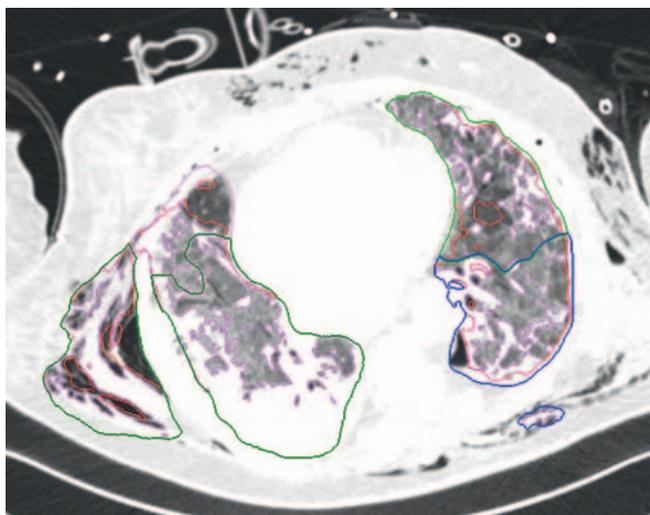
be quickly and easily determined in any hospital regardless of the existing infrastructure. The score was able to separate patients into clearly defined risk groups ranging from 7% for a score of 0

► **Table 5** Predictive value of established score on outcome.

► **Tab. 5** Prognostische Wertigkeit des entwickelten Scores.

Score	Cohort for model development (N = 83)		Validation cohort (N = 28)	
	Positive outcome	Negative outcome	Positive outcome	Negative outcome
0	25 (93%)	2 (7%)	17 (94%)	1 (6%)
1	14 (70%)	6 (30%)	1 (50%)	1 (50%)
2	7 (33%)	14 (67%)	0 (0%)	4 (100%)
3	0 (0%)	15 (100%)	0 (0%)	4 (100%)

Positive outcome: discharge or regular ward care; negative outcome: need for mechanical ventilation, treatment in intensive care unit, ECMO, or death.



► **Fig. 5** CT image of a 69-year-old patient with severe COVID-19 pneumonia and pneumothorax. Erroneous lung segmentation most likely caused by pneumothorax. Color coding same as in Fig. 3. This patient was excluded from the analysis.

► **Abb. 5** CT eines 69-jährigen Patienten mit schwerer COVID-19-Pneumonie und Pneumothorax. Fehlerhafte Lungensegmentierung, am ehesten bedingt durch den Pneumothorax. Farbige Markierungen wie in Abb. 3. Dieser Patient wurde von der Auswertung ausgeschlossen.

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

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

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

### Conflict of Interest

The authors declare that they have no conflict of interest.

### References

- [1] American College of Radiology. ACR Recommendations for the use of Chest Radiography and Computed Tomography (CT) for Suspected COVID-19 Infection (28.12.2020). Im Internet (Stand: 28.12.2020): <https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection>
- [2] Li X, Fang X, Bian Y et al. Comparison of chest CT findings between COVID-19 pneumonia and other types of viral pneumonia: a two-center retrospective study. *Eur Radiol* 2020; 30: 5470–5478. doi:10.1007/s00330-020-06925-3
- [3] Roberto G, Roberta F, Paola BM et al. Coronavirus disease 2019 (COVID-19) in Italy: features on chest computed tomography using a structured report system. *Scientific reports* 2020; 10: 17236. doi:10.1038/s41598-020-73788-5
- [4] Zhou Z, Guo D, Li C et al. Coronavirus disease 2019: initial chest CT findings. *Eur Radiol* 2020; 30: 4398–4406. doi:10.1007/s00330-020-06816-7
- [5] Feng Z, Yu Q, Yao S et al. Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics. *Nat Commun* 2020; 11: 4968. doi:10.1038/s41467-020-18786-x
- [6] Francone M, Iafra F, Masci GM et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. *Eur Radiol* 2020. doi:10.1007/s00330-020-07033-y
- [7] Jalaber C, Lapotre T, Morcet-Delattre T et al. Chest CT in COVID-19 pneumonia: A review of current knowledge. *Diagnostic and interventional imaging* 2020; 101: 431–437. doi:10.1016/j.diii.2020.06.001
- [8] Gieraerts C, Dangis A, Janssen L et al. Prognostic Value and Reproducibility of AI-assisted Analysis of Lung Involvement in COVID-19 on Low-Dose Submillisievert Chest CT: Sample Size Implications for Clinical Trials. *Radiology: Cardiothoracic Imaging* 2020; 2: e200441. doi:10.1148/ryct.2020200441
- [9] Grodecki K, Lin A, Cadet S et al. Quantitative Burden of COVID-19 Pneumonia on Chest CT Predicts Adverse Outcomes: A Post-Hoc Analysis of a Prospective International Registry. *Radiology: Cardiothoracic Imaging* 2020; 2: e200389. doi:10.1148/ryct.2020200389
- [10] Colombi D, Bodini FC, Petrini M et al. Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia. *Radiology* 2020; 296: E86–E96. doi:10.1148/radiol.2020201433
- [11] Liu F, Zhang Q, Huang C et al. CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics* 2020; 10: 5613–5622. doi:10.7150/thno.45985
- [12] Shen C, Yu N, Cai S et al. Quantitative computed tomography analysis for stratifying the severity of Coronavirus Disease 2019. *Journal of pharmaceutical analysis* 2020; 10: 123–129. doi:10.1016/j.jpha.2020.03.004
- [13] Lanza E, Muglia R, Bolengo I et al. Quantitative chest CT analysis in COVID-19 to predict the need for oxygenation support and intubation. *Eur Radiol* 2020. doi:10.1007/s00330-020-07013-2
- [14] Wang Y, Chen Y, Wei Y et al. Quantitative analysis of chest CT imaging findings with the risk of ARDS in COVID-19 patients: a preliminary study. *Annals of translational medicine* 2020; 8: 594. doi:10.21037/atm-20-3554
- [15] Yu Q, Wang Y, Huang S et al. Multicenter cohort study demonstrates more consolidation in upper lungs on initial CT increases the risk of adverse clinical outcome in COVID-19 patients. *Theranostics* 2020; 10: 5641–5648. doi:10.7150/thno.46465
- [16] Meiler S, Schaible J, Poschenrieder F et al. Can CT performed in the early disease phase predict outcome of patients with COVID 19 pneumonia? Analysis of a cohort of 64 patients from Germany. *European journal of radiology* 2020; 131: 109256. doi:10.1016/j.ejrad.2020.109256

- [17] Tan L, Wang Q, Zhang D et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal Transduct Target Ther* 2020; 5: 33. doi:10.1038/s41392-020-0148-4
- [18] Yu H-H, Qin C, Chen M et al. D-dimer level is associated with the severity of COVID-19. *Thromb Res* 2020; 195: 219–225. doi:10.1016/j.thromres.2020.07.047
- [19] Ghahramani S, Tabrizi R, Lankarani KB et al. Laboratory features of severe vs. non-severe COVID-19 patients in Asian populations: a systematic review and meta-analysis. *European journal of medical research* 2020; 25: 30. doi:10.1186/s40001-020-00432-3
- [20] Bonetti G, Manelli F, Patroni A et al. Laboratory predictors of death from coronavirus disease 2019 (COVID-19) in the area of Valcamonica, Italy. *Clinical chemistry and laboratory medicine* 2020; 58: 1100–1105. doi:10.1515/cclm-2020-0459
- [21] Zheng Y, Xiao A, Yu X et al. Development and Validation of a Prognostic Nomogram Based on Clinical and CT Features for Adverse Outcome Prediction in Patients with COVID-19. *Korean journal of radiology* 2020; 21: 1007–1017. doi:10.3348/kjr.2020.0485
- [22] Schalekamp S, Huisman M, van Dijk RA et al. Model-based Prediction of Critical Illness in Hospitalized Patients with COVID-19. *Radiology* 2021; 298: E46–E54. doi:10.1148/radiol.2020202723
- [23] Chaganti S, Grenier P, Balachandran A et al. Automated Quantification of CT Patterns Associated with COVID-19 from Chest CT. *Radiology: Artificial Intelligence* 2020; 2: e200048. doi:10.1148/ryai.2020200048
- [24] Harrell FE jr. *Regression modeling strategies*. Springer Series in Statistics. 2. Aufl. New York: Springer; 2015
- [25] Zhou F, Yu T, Du R et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 2020; 395: 1054–1062. doi:10.1016/S0140-6736(20)30566-3
- [26] Bonanad C, García-Blas S, Tarazona-Santabalbina F et al. The Effect of Age on Mortality in Patients With COVID-19: A Meta-Analysis With 611583 Subjects. *J Am Med Dir Assoc* 2020; 21: 915–918. doi:10.1016/j.jamda.2020.05.045
- [27] Heussel CP, Kappes J, Hantusch R et al. Contrast enhanced CT-scans are not comparable to non-enhanced scans in emphysema quantification. *European journal of radiology* 2010; 74: 473–478. doi:10.1016/j.ejrad.2009.03.023
- [28] Mascalchi M, Camiciottoli G, Diciotti S. Lung densitometry: why, how and when. *J Thorac Dis* 2017; 9: 3319–3345. doi:10.21037/jtd.2017.08.17
- [29] Coxson HO. Sources of variation in quantitative computed tomography of the lung. *Journal of Thoracic Imaging* 2013; 28: 272–279. doi:10.1097/RTI.0b013e31829efbe9