EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Development of a Clinical Decision Support System for Severity Risk Prediction and Triage of COVID-19 Patients at Hospital Admission: an International Multicenter Study

Guangyao Wu, Pei Yang, Yuanliang Xie, Henry C. Woodruff, Xiangang Rao, Julien Guiot, Anne-Noelle Frix, Renaud Louis, Michel Moutschen, Jiawei Li, Jing Li, Chenggong Yan, Dan Du, Shengchao Zhao, Yi Ding, Bin Liu, Wenwu Sun, Fabrizio Albarello, Alessandra D'Abramo, Vincenzo Schininà, Emanuele Nicastri, Mariaelena Occhipinti, Giovanni Barisione, Emanuela Barisione, Iva Halilaj, Pierre Lovinfosse, Xiang Wang, Jianlin Wu, Philippe Lambin

Please cite this article as: Wu G, Yang P, Xie Y, *et al*. Development of a Clinical Decision Support System for Severity Risk Prediction and Triage of COVID-19 Patients at Hospital Admission: an International Multicenter Study. *Eur Respir J* 2020; in press (https://doi.org/10.1183/13993003.01104-2020).

This manuscript has recently been accepted for publication in the *European Respiratory Journal*. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJ online.

Title page

Title: Development of a Clinical Decision Support System for Severity Risk Prediction and Triage of COVID-19 Patients at Hospital Admission: an International Multicenter Study

Authors: Guangyao Wu[1,#], MD; Pei Yang[2,#,] MD; Yuanliang Xie[2], MD , Henry C. Woodruff, PhD[1,3]; Xiangang Rao[4], MD; Julien Guiot[5], MD, PhD; Anne-Noelle Frix[5], MD; Renaud Louis[5], MD, PhD; Michel Moutschen[6], MD, PhD; Jiawei Li[7], MD; Jing Li[8], MD; Chenggong Yan[1,9], MD; Dan Du[2], MD; Shengchao Zhao[2], MD; Yi Ding[2], MD; Bin Liu[2], MD; Wenwu Sun[10], MD; Fabrizio Albarello[11], MD; Alessandra D'Abramo[11], MD; Vincenzo Schininà[11], MD; Emanuele Nicastri[11], MD; Mariaelena Occhipinti[12], MD; Giovanni Barisione[13], MD; Emanuela Barisione[14], MD; Iva Halilaj[1], MSc; Pierre Lovinfosse[15], MD, PhD; Xiang Wang[2], MD; Jianlin Wu[16], MD, PhD; Philippe Lambin[1,3], MD, PhD

[1] The D-Lab, Department of Precision Medicine, GROW - School for Oncology, Maastricht University Medical Center+, Maastricht, The Netherlands

[2] Department of Radiology, The Central Hospital of Wuhan, Huazhong University of Science and Technology, Wuhan, China

[3] Department of Radiology and Nuclear Medicine, GROW- School for Oncology and Developmental Biology, Maastricht University Medical Center+, Maastricht, The Netherlands

[4] Department of Ultrasound, The Central Hospital of Huangshi, Huangshi, China

[5] Department of Respiratory Medicine, CHU of Liège, Liège, Belgium

[6] Department of Infectiology, CHU of Liège, Liège, Belgium

[7] Department of Radiology, China Resources Wuhan Iron and Steel Hospital, Wuhan, China

[8] Department of Radiology, The Central Hospital of Shaoyang, Shaoyang, China

[9] Department of Medical Imaging Center, Nanfang Hospital, Southern Medical University, Guangzhou, China

[10] Department of Intensive Care Unit, The Central Hospital of Wuhan, Huazhong University of Science and Technology, Wuhan, China

[11] National Institute for Infectious Diseases – IRCCS, Lazzaro Spallanzani, Via Portuense, Rome, Italy

[12] Department of Biomedical, Clinical and Experimental Sciences "Mario Serio", University of Florence, Florence, Italy

[13] Unit of Respiratory Pathophysiology, Respiratory Diseases and Allergy Clinic, Department of Internal

Medicine and Medical Specialties, University of Genoa, IRCCS Ospedale Policlinico San Martino, Genoa, Italy

[14] Unit of Interventional Pulmonology, IRCCS Ospedale Policlinico San Martino, Genoa, Italy

[15] Nuclear Medicine and Oncological Imaging, Department of Medical Physics, CHU of Liège, Liège, Belgium

[16] Department of Radiology, Affiliated Zhongshan Hospital of Dalian University, Dalian, China

[#] **Guangyao Wu and Pei Yang are joint first authors**

Correspondence to:

Guangyao Wu, The D-Lab, Department of Precision Medicine, GROW - School for Oncology, Maastricht University Medical Center+, 6229 ER, Maastricht, The Netherlands **g.wu@maastrichtuniversity.nl**

Xiang Wang, Department of Radiology, The Central Hospital of Wuhan, Huazhong University of Science and Technology, Wuhan, 430014, China **drwangxiang385@163.com**

Summary: An internationally validated model, nomogram, and online- calculator for severity risk assessment and triage of COVID-19 patients at hospital admission.

Background: The outbreak of the coronavirus disease 2019 (COVID-19) has globally strained medical resources and caused significant mortality.

Objective: To develop and validate machine-learning model based on clinical features for severity risk assessment and triage for COVID-19 patients at hospital admission.

Method: 725 patients were used to train and validate the model including a retrospective cohort of 299 hospitalized COVID-19 patients at Wuhan, China, from December 23, 2019, to February 13, 2020, and five cohorts with 426 patients from eight centers in China, Italy, and Belgium, from February 20, 2020, to March 21, 2020. The main outcome was the onset of severe or critical illness during hospitalization. Model performances were quantified using the area under the receiver operating characteristic curve (AUC) and metrics derived from the confusion-matrix.

Results: The median age was 50.0 years and 137 (45.8%) were men in the retrospective cohort. The median age was 62.0 years and 236 (55.4%) were men in five cohorts. The model was prospectively validated on five cohorts yielding AUCs ranging from 0.84 to 0.89, with accuracies ranging from 74.4% to 87.5%, sensitivities ranging from 75.0% to 96.9%, and specificities ranging from 57.5% to 88.0%, all of which performed better than the pneumonia severity index. The cut-off values of the low, medium, and high-risk probabilities were 0.21 and 0.80. The online-calculators can be found at www.covid19risk.ai.

Conclusion: The machine-learning model, nomogram, and online-calculator might be useful to access the onset of severe and critical illness among COVID-19 patients and triage at hospital admission.

Introduction

In December 2019, a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; earlier named as 2019-nCoV), emerged in Wuhan, China [1]. The disease caused by SARS-CoV-2 was named coronavirus disease 2019 (COVID-19). As of May 15, 2020, more than 4 490 000 COVID-19 patients have been reported globally, and over 300 000 patients have died [2]. The outbreak of COVID-19 has developed into a pandemic [3].

Among COVID-19 patients, around 80% present with mild illness whose symptoms usually disappear within two weeks [4]. However, around 20% of the patients may proceed and necessitate hospitalization and increased medical support. The mortality rate for the severe patients is around 13.4% [4]. Therefore, risk assessment of patients preferably in a quantitative, non-subjective way, is extremely important for patient management and medical resource allocation. General quarantine and symptomatic treatment at home or mobile hospital can be used for most non-severe patients, while a higher level of care and fast track to the intensive care unit (ICU) is needed for severe patients. Previous studies have summarized the clinical and radiological characteristics of severe COVID-19 patients, while the prognostic value of different variables is still unclear [5, 6].

Several scoring systems that are in common clinical use (e.g. Sequential Organ Failure Assessment score, Confusion-Urea-Respiratory rate-Blood pressure-Age 65, Acute Physiology and Chronic Health Evaluation, etc.) could be applied to the triage problem, albeit each with their own problems and limitations, such as the need for laboratory variables that are hard to obtain at hospital admission [7]. The pneumonia severity index (PSI) stands out as it is used to assess the probability of severity and mortality among adult patients with community-acquired pneumonia and to help hospitalization management [8].

A better solution could possibly be found using machine-learning, a branch of artificial intelligence that learns from past data in order to build a prognostic model [9]. In recent years, machine learning has been developed as a useful tool to analyze large amounts of data from medical records or images [10]. Previous modeling studies focused on forecasting the potential international spread of COVID-19 [11].

Therefore, our objective is to develop and validate a prognostic machine-learning model based on clinical, laboratory, and radiological variables of COVID-19 patients at hospital admission for severity risk assessment during hospitalization, and compare the performance with that of PSI as a representative clinical assessment method. Our ambition is to develop a multifactorial decision support system with different datasets to facilitate risk prediction and triage (home or mobile hospital quarantine, hospitalization, or ICU) of the patient at hospital admission.

Methods

Patients

The institutional review board approved this study (2020-71), which followed the Standards for Reporting of Diagnostic Accuracy Studies statement [12], and the requirement for written informed consent was waived. 299 adult confirmed COVID-19 patients from the central hospital of Wuhan were included consecutively and retrospectively between December 23, 2019 and February 13, 2020. The inclusion criteria were: (1) patients with a confirmed COVID-19 disease, (2) patients present at hospital for admission. The exclusion criteria were: (1) patients already with a severe illness at hospital admission; (2) time interval > 2 days between admission and examinations; and (3) no data available or delayed results as described below. The patients included from this center were divided into two datasets according to the entrance time of hospitalization, 80% for training (239 patients from December 23, 2019, to January 28, 2020) and 20% for internal validation (60 patients from January 29 to February 13, 2020). The five test datasets were collected between February 20, 2020 and March 31, 2020 from other eight centers (**Supplementary**) in China, Italy, and Belgium under the same inclusion and exclusion criteria (**Figure 1**).

Patients were labelled as having a "severe disease" if at least one of the following criteria were met during hospitalization [6, 13]: (a) respiratory failure requiring mechanical ventilation; (b) shock; (c) ICU admission; (d)organ failure; or (e) death. Patients were labelled as having a "non-severe disease" if none of the above-mentioned criteria were met during the whole hospitalization process until deemed recovered and discharged from the hospital.

Data collection

Clinical, laboratory, radiological characteristics and outcome data were obtained in the case record form shared by the International Severe Acute Respiratory and Emerging Infection Consortium from the electronic medical records [14]. A confirmed case with COVID-19 was defined as a positive result of high-throughput sequencing or real-time reverse-transcriptase polymerase-chain-reaction assay for nasal and pharyngeal swab specimens. After consultation with respiratory specialists and review of the recent COVID-19 literature, a set of clinical, laboratory, and radiological characteristics was identified and the data collected from the electronic medical system. The clinical characteristics included basic information (5 variables), comorbidities (11 variables), and symptoms (13 variables). All clinical characteristics were obtained when the patients were admitted to hospital for the first time. 42 laboratory results were recorded, including complete blood count, white blood cell differential count, D-dimer, C-reactive protein (CRP), cardiac enzymes, procalcitonin, liver function test, kidney

function test, B-type natriuretic peptide and electrolyte test. The arterial blood gas was not taken into account due to missing data for most early-stage patients. The metric conversion of laboratory results was performed using an online conversion table [15]. A detailed list of variables can be found in **Tables 1 and 2**.

The semantic CT characteristics (including ground-glass opacity, consolidation, vascular enlargement, air bronchogram, and lesion range score) were independently evaluated on all datasets by two radiologists (PY [a radiologist with 5 years' experience in chest CT images] and YX [a radiologist with 20 years' experience in chest CT images]), who were blinded to clinical and laboratory results. Any disagreement was resolved by a consensus read. Lesion range was identified as areas of ground-glass opacity or consolidation and was graded with a 6-point scale according to the lesion volume proportion in each single lobe: 0 = no lung parenchyma involved, 1 = up to 5% of lung parenchyma involved, 2 = 5-25%, 3 = 26-50%, 4 = 51-75%, and 5 = 76-100% of lung parenchyma involved. The lesion volume proportion was automatically calculated by Shukun Technology Pneumonia Assisted Diagnosis System (Version 1.17.0), and the final score is a total score from five lobes (**Figure 2**). Detailed CT acquisition and reconstruction parameters are presented in the **Supplementary**.

Feature selection and modeling

All feature selection and model training were performed in the training dataset alone to prevent information leakage. An overview of the functions used is given in **Supplementary Table S1**. In order to reduce feature dimensionality, features showing high pairwise Spearman correlation ($r > 0.8$) and the highest mean correlation with all remaining features were removed, followed by application of the Boruta algorithm to select important features [16]. The Boruta algorithm combines feature rank based on the random forest classification algorithm and selection frequency based on multiple iterations of the feature selection procedure. Recursive feature elimination based on bagged tree models with a cross-validation technique (10 folds, 10 times) was performed to select the best performing combination of features. In order to balance the positive and negative sample size, an adaptive synthetic sampling approach for imbalanced learning (ADASYN) was used during feature selection and modeling [17]. The feature selection process was used for clinical, laboratory, and CT semantic models alone, and in combination.

Logistic regression models based on selected features were trained and the validation dataset was used to internally validate the prognostic performance of the models. Four models were trained: Model 1 contained only baseline clinical features without symptoms; Model 2 used all selected clinical features; Model 3 used selected semantic CT features, age, and sex; Model 4 employed all selected clinical, laboratory, and CT features.

The prognostic performances of the best model were compared with other models on the training dataset, due

to a bigger sample size. The performance of the best model and PSI scoring were gauged on the datasets via the receiver operator characteristic (ROC) and confusion matrix. In order to gauge the level of overfitting, the outcomes were randomized on the best model and the entire process repeated, from feature selection to model building and evaluation.

The patients from the training datasets were divided into low, medium and high risk according to the first quartile (25th percentile) and the third quartile (75th percentile) of probabilities from the best performing model. Nomograms and on-line calculators were used to provide the interpretability of the best trained models. The test datasets were used to gauge the prognostic performance and the validity for the best model.

Statistical analysis

Baseline data were summarized as median, and categorical variables as frequency (%). Differences between the severe group and the non-severe group were tested using the Mann-Whitney test for continuous data and Fisher's exact test for categorical data. Feature correlations were measured using the Spearman correlation coefficient. We determined the area under the ROC curve (AUC) with its 95% confidence interval (CI) and tested AUC difference between Models 1-3 and Model 4 by the DeLong method [18], measures of prognostic performance included the AUC, and metrics derived from the confusion matrix - accuracy, sensitivity, specificity, positive prediction value (PPV), and negative prediction value (NPV). A calibration-plot based on the Hosmer-Lemeshow test was used to estimate the goodness-of-fit and consistency of the model on the test datasets. All p values were two-sided, and $p < 0.05$ was regarded as significant. All statistical analyses, modeling, and plotting were performed in R (version 3.5.3), and the detailed package characteristics are listed in **Supplementary Table S1**.

Results

Demographic and Clinical characteristics

Of 299 hospitalized COVID-19 patients in retrospective cohort, the median age was 50.0 years (interquartile range, 35.5-63.0; range, 20-94 years) and 137 (45.8%) were men. All the clinical characteristics and CT findings are summarized in **Table 1**, and more details of laboratory findings can be seen in **Table 2**. Of 426 hospitalized COVID-19 patients in 5 cohorts as test datasets, the median age was 62.0 years (interquartile range, 50.0-72.0; range, 19-94 years) and 236 (55.4%) were men.

Feature selection

Among the clinical features, age, hospital employment, body temperature, and the time of onset to admission were selected. Lymphocyte (proportion), neutrophil, (proportion), CRP, lactate dehydrogenas (LDH), creatine

kinase (CK), urea, and calcium were selected from the laboratory feature set. Only the lesion range score was selected from CT semantic features. When putting these three category features together to select features, age, Lymphocyte (proportion), CRP, LDH, CK, urea and calcium were finally included in the combination model.

Models performance in the training and validation dataset

Model performance was as follows. The Model 1 based on age and hospital employment showed an AUC of 0.74 (95% CI, 0.69-0.79) on the training dataset and an AUC of 0.83 (95% CI, 0.72-0.94) on the validation dataset. The Model 2 with the clinical features of age hospital employment, body temperature, and the time of onset yield an AUC of 0.78 (95% CI, 0.73-0.83) on the training dataset and an AUC of 0.74 (95% CI, 0.59-0.89) on the validation dataset. The Model 3 based on age and lesion range score on CT, had an AUC of 0.75 (95% CI, 0.70-0.80) on the training dataset and an AUC of 0.83 (95% CI, 0.72-0.94) on the validation dataset.

When pooling these three categories of features, the combination model (Model 4) selected 7 features (age, lymphocyte [proportion], CRP, LDH, CK, urea, and calcium), which achieved the highest AUC of 0.86 (95% CI, 0.82-0.90) on the training dataset and an AUC of 0.90 (95% CI, 0.82-0.98) on the validation dataset. The AUC value of Model 4 was significantly higher than Model 1 ($p = 0.001$), Model 2 ($p = 0.033$), and Model 3 ($p = 0.003$) on the training dataset. The cut-off values from reclassification of low, medium, and high-risk probabilities were 0.21 and 0.80.

External validation

Model 4 was validated on the five test datasets, which showed AUCs ranging from 0.84 to 0.93 with accuracies ranging from 74.4% to 87.5%, sensitivities ranging from 75.0% to 96.9%, specificities ranging from 57.5% to 88.0%, PPVs ranging from 71.4% to 84.1%, and NPVs ranging from 73.9% to 93.9% (**Table 3**). The ROC, confusion-matrix, and calibration plots are shown in **Figure 3**. The results of randomizing the outcomes and re-running the analysis yielded AUC of 0.50 (95% CI, 0.44-0.55) for the Model 4.

Clinical use

Based on the selected features from the best models, a nomogram was established to quantitatively assess the severity risk of illness (**Figure 4**). The developed online-calculators can be found at www.covid19risk.ai. Compared to PSI scoring, Model 4 showed higher AUCs, accuracies, sensitivities, and NPVs on the five test datasets (**Table 3**). There were significant difference for the proportion of severe patients among low, medium, and high-risk groups in the five test datasets (**Figure 5**).

Discussion

This international multicenter study analyzed individually and in combination, clinical, laboratory and radiological characteristics for COVID-19 patients at hospital admission, to retrospectively develop and prospectively validate a prognostic model and tool to assess the severity of the illness, and its progression, and to compare these with PSI scoring. We found that COVID-19 patients that developed a severe illness were often of an advanced age, accompanied by multiple comorbidities, presenting with chest tightness, and had abnormal laboratory results and broader lesion range on lung CT on admission. Using simpler linear regression models yielded better prognostic performance than PSI scoring in the test datasets. We believe these models could be useful for risk assessment and triage.

Previous studies have reported that age and underlying comorbidities (such as hypertension, diabetes, and cardiovascular diseases) may be risk factors for the COVID-19 patients requiring intensive care unit (ICU) [19 20]. In this study, we found that the elderly COVID-19 patients who were male, non-hospital staff, suffering from hypertension, diabetes, cardiopathy disease, chronic obstructive pulmonary disease, cerebrovascular disease, renal disease, hepatitis B virus infection, lower body temperature, and chest tightness were more vulnerable to develop a severe illness in the early stages of the disease. Among these features, age, hospital staff, body temperature, and the time of onset to admission had certain prognostic abilities. Age was the most important feature, which may interact with other features, which was why only age was selected into our combination model (Model 4) from these features. Zhou and colleagues have confirmed that SARS-CoV-2 uses the same cell entry receptor (angiotensin-converting enzyme II [ACE2]) with SARS-CoV [21]. However, whether COVID-19 patients with hypertension and diabetes have higher severe illness risk, which is due to treatment with ACE2- increasing drugs is still unknown [22]. Hospital staff had a lower risk of progression, possibly due lower age, higher levels of education, and more medical knowledge once infected although the unbalanced nature of this type of data has to be taken into account.

Furthermore, early studies have shown that COVID-19 patients with severe illness had more laboratory abnormalities such as CRP, D-dimer, lymphocyte, neutrophil, and LDH, than those patients with non-severe illness, which were associated with the prognosis [19, 20, 23]. In our study, we also found that the severe group had numerous laboratory abnormalities in complete blood cell count, white cell differential count, D-dimer, CRP, liver function, renal function, procalcitonin, B-type natriuretic peptides, and electrolytes. Among these abnormalities, lymphocyte proportion, neutrophil proportion, CRP, LDH, CK, urea, and calcium were significant prognostic factors, which suggest that COVID-19 may cause damage to multiple organ systems when developing

into a severe illness. However, current pathological findings of COVID-19 suggest that there is no evidence that SARS-CoV-2 can directly impair the other organs such as liver, kidney and heart [24].

Current reports have shown that thin-slice chest CT is a powerful tool in clinical diagnosis due to the high sensitivity and the ability to monitor the development of the disease [25, 26]. In addition, a previous study reported that ground-glass opacity and consolidation were the most common CT findings for COVID-19 patients with pneumonia, while being nonspecific [27]. Clinical observations showed that there were significantly more consolidation lesions in ICU patients on admission, while more ground-glass opacity lesions were observed in non-ICU patients [28]. In our study, we found that vascular enlargement, air-bronchogram, and lesion range score differ significantly between non-severe and severe groups. Among these features, only the lesion range score had prognostic power, but not enough to be selected for the combination model. This indicates that while these early stage CT semantic features could have diagnostic value, they have limited ability to prognose the onset of severe illness in COVID-19 patients.

The Chinese National Health Committee added some warning indicators for severe or critical cases in the updated diagnosis and treatment plan for COVID-19 patients (version 7) [29], which includes progressive reduction of peripheral blood lymphocytes, a progressive increase of IL-6, CRP and lactate, and rapid progression of lung CT findings in a short period. In this study, we used age, lymphocyte fraction, CRP, LDH, CK, urea, and calcium scores from clinical, laboratory, and radiological exams recorded at hospital admission to train a model for the prediction of the onset of severe illness. Our model combining these features from multiple sources showed a favorable performance when validated in the five external datasets from China, Italy, and Belgium. In addition, the model is able to stratify COVID-19 patients into low, medium, and high-risk groups for developing severe illness. We propose that this model with its higher prediction performance and simplicity than PSI score could be used for a preliminary screening and triage tool at hospital admission for the potential to develop severe illness. Furthermore, the model could be used for the selection and/or stratification of patients in clinical trials in order to homogenize the patient population. Follow-up laboratory tests are needed to assess the severity risk with a higher accuracy.

As one of the coronaviruses family infecting humans, SARS-CoV-2 has similar etiologic, clinical, radiological and pathological features to those of severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus [23, 30, 31]. Therefore, we believe that developing a reliable early warning model based on presently clinical, radiological, and pathological data is necessary for current outbreaks and possible future outbreaks of coronaviruses.

Our study has several limitations. First, selection bias is unavoidable and the limited and unbalanced sample size. Second, patients from different races and ethnicities may have diverse clinical and laboratory results, and the self-medication of patients before admission may affect the clinical and laboratory results. Third, the threshold to go to the hospital and hospitalization management can vary from country to country, we are also aware that RNA viruses can mutate rapidly and that could have an impact of the performance of the models. We therefore propose that those models should be continuously updated to achieve a better performance for example using privacy-preserving distributed learning approaches [32, 33]. Fourth, the CT features used for this study are semantic features from the first CT scan, and radiomics or deep learning approaches may improve its prognostic performance, and follow-up CT scan may yield more information. Fifth, due to the large number of predictors included in the analysis, and the complexity of feature selection and modelling, overfitting is always possible. We have mitigated this with the use of external validation cohorts, and by rerunning the analysis on randomized outcomes to arrive at a "chance" (AUC=0.5) result.

Conclusions

Elderly COVID-19 patients and non-hospital staff seem more vulnerable to develop a severe illness after hospitalization as per defining criteria, which can cause a wide range of laboratory and CT anomalies. Furthermore, our model based on lactate dehydrogenase, C-reactive protein, calcium, age, lymphocyte proportion, urea, and creatine kinase might be a more useful preliminary screening and triage tool than pneumonia severity index for risk assessment of COVID-19 patients at hospital admission.

Author Contributors: G. Wu, P. Yang, Y. Xie, X. Wang, and P. Lambin conceived and designed the study. G. Wu and P. Yang contributed to the literature search. P. Yang, X. Rao, J. Li, J. Li, D. Du, S. Zhao, Y. Ding, B. Liu, W. Sun, F. Albarello, A. D'Abramo, V. Schininà, E. Nicastri, J. Wu, M. Occhipinti, G. Barisione, E. Barisione, J. Guiot, A. Frix, M. Moutschen, R. Louis, P. Lovinfosse, and C. Yan contributed to data collection. G. Wu, H. Woodruff, and P. Lambin contributed to data analysis. G. Wu, H. Woodruff, and P. Lambin contributed to data interpretation. G. Wu and C. Yan contributed to the tables and figures. G. Wu, I Halilaj, and P. Lambin contributed to build a website. G. Wu, P. Yang, H. Woodruff, and P. Lambin contributed to writing of the report.

Conflict of Interest Disclosures: Dr Philippe Lambin reports, within the submitted work, minority shares in The Medical Cloud Company and outside the submitted work grants/sponsored research agreements from Varian medical, Oncoradiomics, ptTheragnostic/DNAmito, Health Innovation Ventures. He received an advisor/presenter fee and/or reimbursement of travel costs/external grant writing fee and/or in kind manpower contribution from Oncoradiomics, BHV, Varian, Elekta, ptTheragnostic and Convert pharmaceuticals. Dr. P. Lambin has shares in the company Oncoradiomics SA, Convert pharmaceuticals SA and and is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248, PCT/NL2014/050728) licensed to Oncoradiomics and one issue patent on mtDNA (PCT/EP2014/059089) licensed to ptTheragnostic/DNAmito, three non-patented invention (softwares) licensed to ptTheragnostic/DNAmito, Oncoradiomics and Health Innovation Ventures and three non-issues, non licensed patents on Deep Learning-Radiomics and LSRT (N2024482, N2024889, N2024889). Dr Henry C. Woodruff has (minority) shares in the company Oncoradiomics. Dr Mariaelena Occhipinti reports grants from Menarini Foundation and Novartis, outside the submitted work. The other authors declare no competing interests.

Funding/Support: This work was supported from ERC advanced grant (ERC-ADG-2015, n° 694812 - Hypoximmuno), European Program H2020 (ImmunoSABR - n° 733008, PREDICT - ITN - n° 766276, CHAIMELEON - n° 952172, EuCanImage – n° 952103), TRANSCAN Joint Transnational Call 2016 (JTC2016 “CLEARLY”- n° UM 2017-8295), China Scholarships Council (n° 201808210318), and Interreg V-A Euregio Meuse-Rhine (“Euradiomics” - n° EMR4). This work was supported by the Dutch Cancer Society (KWF Kankerbestrijding), Project number 12085/2018–2.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

References

1. WHO. Coronavirus disease (COVID-19) outbreak. https://www.who.int/emergencies/diseases/ novel-coronavirus-2019 (Accessed March 07, 2020)
2. WHO. Coronavirus disease 2019 (COVID-19) Situation Report – 116. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (Accessed May 15, 2020)
3. WHO. Events as they happen. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen (Accessed March 15, 2020)
4. WHO. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf (Accessed March 07, 2020)
5. Young BE, Ong SWX, Kalimuddin S, et al. Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. *JAMA*. 2020;323(15):1488-1494.
6. Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med.* 2020;382(18):1708-1720.
7. Zou XJ, Li SS, Fang MH, et al. Acute Physiology and Chronic Health Evaluation II Score as a Predictor of Hospital Mortality in Patients of Coronavirus Disease 2019. *Critical Care Medicine*. 2020. doi: 10.1097/CCM.0000000000004411
8. Fine MJ, Auble TE, Yealy DM, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med*. 1997; 336(4):243-250.
9. Fralick M, Colak E, Mamdani M. Machine Learning in Medicine. *N Engl J Med*. 2019;380(26):2588-2589.
10. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762.
11. Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*. 2020;395(10225):689-697.
12. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. *Clin Chem.* 2003;49(1):1-6.
13. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and Treatment of Adults with Community-acquired

Pneumonia. An Official Clinical Practice Guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med.* 2019;200(7):e45-e67.

14. ISARIC-WHO. https://isaric.tghn.org/covid-19-clinical-research-resources (Accessed March 07, 2020)
15. LabCorp. https://www.labcorp.com/resource/si-unit-conversion-table (Accessed March 07, 2020)
16. Kursa MB, Rudnicki WR. Feature selection with the boruta package. *J Stat Softw.* 2010;36(11):1-13.
17. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *In: Proceedings of the International Joint Conference on Neural Networks*. 2008; pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969.
18. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837-845.
19. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* 2020;395(10223):497-506.
20. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA.* 2020;323(11):1061-1069.
21. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature.* 2020;579(7798):270-273.
22. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?. *Lancet Respir Med*. 2020;8(4):e21.
23. Cao B, Wang Y, Wen D, et al. A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N Engl J Med.* 2020;382(19):1787-1799.
24. Xu Z, Shi L, Wang Y, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med.* 2020;8(4):420-422.
25. Song F, Shi N, Shan F, et al. Emerging 2019 Novel Coronavirus (2019-nCoV) Pneumonia. *Radiology*. 2020;295(1):210-217.
26. Wang Y, Dong C, Hu Y, et al. Temporal Changes of CT Findings in 90 Patients with COVID-19 Pneumonia: A Longitudinal Study. *Radiology*. 2020. doi: 10.1148/RADIOL.2020200843
27. Chung M, Bernheim A, Mei X, et al. CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology*. 2020;295(1):202-207.
28. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis.* 2020;20(4):425-434.

29. National Health Commission & State Administration of Traditional Chinese Medicine. 2020. http://en.nhc.gov.cn/ (Accessed March 07, 2020)
30. Chan JWM, Ng CK, Chan YH, et al. Short term outcome and risk factors for adverse clinical outcomes in adults with severe acute respiratory syndrome (SARS). *Thorax.* 2003;58(8):686-689.
31. Memish ZA, Perlman S, Van Kerkhove MD, Zumla A. Middle East respiratory syndrome. *Lancet.* 2020;395(10229):1063-1077.
32. Zerka F, Barakat S, Walsh S,et al. Systematic Review of Privacy-Preserving Distributed Machine Learning From Federated Databases in Health Care. *JCO Clin Cancer Inform.* 2020;4:184-200.
33. Deist TM, Dankers FJWM, Ojha P,et al. Distributed learning on 20 000+ lung cancer patients - The Personal Health Train. *Radiother Oncol.* 2020;144:189-200.

Table 1. Clinical characteristics and radiological findings of patients confirmed with COVID-19

Basic information	Non-severe group (n=228)	Severe group (n=71)	p value*
Age	43.0 (33.0-61.0)	62.0 (52.5-71.5)	< 0.001
No. of men	95 (41.7)	42 (59.2)	0.014
No. with smoking history	17 (7.5)	14 (19.7)	0.006
No. of hospital staff	86 (37.7)	4 (5.6)	< 0.001
Time of onset of illness, days	4.0 (2.0- 7.0)	4.0 (2.0-7.0)	0.963
Comorbidities			
No. with hypertension	44 (19.3)	31 (43.7)	<0.001
No. with diabetes	19 (8.3)	18 (25.4)	<0.001
No. with hyperlipidemia	11 (4.8)	5 (7.0)	0.545
No. with cardiopathy disease	2 (0.88)	8 (11.3)	< 0.001
No. with chronic obstructive pulmonary disease	8 (3.5)	13 (18.3)	< 0.001
No. with cerebrovascular disease	6 (2.6)	16 (22.5)	< 0.001
No. with kidney disease	5 (2.2)	11 (15.5)	< 0.001
No. with fatty liver	28 (12.3)	12 (16.9)	0.322
No. of Hepatitis B virus carrier	2 (0.88)	5 (7.0)	0.009
No. with cancer history	12 (5.3)	4 (5.6)	1
No. with surgical history	30 (13.2)	15 (21.1)	0.127
Symptoms			
Fever	172 (75.4)	47 (66.2)	0.128
Body temperature, °C	37.8 (37.3-38.4)	37.5 (36.8-38.0)	0.027
Cough	153 (67.1)	46 (64.8)	0.774
Sputum	59 (25.9)	25 (35.2)	0.133
Weakness	102 (44.7)	31 (43.7)	0.892
Diarrhea	25 (11.0)	8 (11.3)	1
Vomiting	16 (7.0)	9 (12.7)	0.144
Chest tightness	50 (21.9)	30 (42.3)	0.001
Dyspnoea	11 (4.8)	4 (5.6)	0.760
Muscular soreness	61 (26.8)	17 (23.9)	0.757
Chill	38 (16.7)	11 (15.5)	1
Conjunctival congestion	1 (0.44)	1 (1.4)	0.419
Headache or dizziness	34 (14.9)	12 (16.9)	0.708
Radiological findings			
Main findings			0.928
Normal	4 (1.8)	1 (1.4)	
Ground-glass opacity only	132 (57.9)	39 (54.9)	
Consolidation only	22 (9.6)	6 (8.5)	
Mixed	70 (30.7)	25 (35.2)	
Vascular enlargement	66 (28.9)	36 (50.7)	< 0.001
Air-bronchogram	49 (21.5)	30 (42.3)	0.001
Lesion range score	4.5 (2.0-7.0)	6.0 (4.0-10.5)	0.001

Data are median (IQR) and N (%) where N is the total number of patients with available data. p values comparing non-severe and severe groups were obtained Fisher's exact test or Mann-Whitney U test.

Table 2. Laboratory results of patients with COVID-19 at hospital admission

Laboratory results	Non-severe group (n=228)	Severe group (n=71)	p value*
Complete blood cell count			
White blood cell count, × 10^9 /L	4.5 (3.3-5.8)	5.5 (4.0-7.6)	< 0.001
Red blood cell count, × 10^{12}/L	4.4 (4.1-4.7)	4.3 (4.0-4.7)	0.217
Hemoglobin, g/L	130.5 (121.0-142.0)	132.0 (117.0-142.5)	0.968
Platelets, × 10^9 /L	174.0 (140.8-214.5)	149.0 (116.5-183.5)	< 0.001
Hematocrit, %	39.5 (36.6-42.7)	38.9 (36.0-42.6)	0.472
Mean corpuscular volume, fL	90.5 (87.5- 93.4)	90.5 (87.8-94.7)	0.526
Mean corpuscular hemoglobin, pg	30.0 (28.7-30.9)	30.0 (29.1-31.5)	0.266
Mean corpuscular hemoglobin concentration, g/dL	330.0 (323.0-336.0)	329.0 (323.0-337.0)	0.684
Red blood cell distribution width standard deviation, fL	39.1 (36.4-41.1)	40.0 (15.3-41.7)	0.193
Red blood cell distribution width coefficient of variation, %	12.7 (12.1-14.9)	13.0 (12.5-38.3)	0.015
Platelet distribution width, %	12.9 (10.7-16.3)	12.4 (10.8-15.8)	0.371
Platelet large cell ratio,%	24.0 (19.7-30.1)	26.8 (21.1-32.3)	0.028
Mean platelet volume, fL	9.8 (9.2-10.6)	10.2 (9.4-10.9)	0.016
Thrombocytocrit, %	0.17 (0.14-0.21)	0.15 (0.13-0.18)	0.002
White cell differential count			
Neutrophil, %	64.5 (56.6-74.7)	76.3 (67.1-85.7)	< 0.001
Lymphocyte, %	26.1 (17.8-26.8)	15.3 (7.6-22.8)	< 0.001
Monocyte, %	7.7 (5.7-9.8)	6.9 (4.5-8.9)	0.061
Eosinophil, %	0.10 (0.00-0.60)	0.00 (0.00-0.31)	0.007
Basophil, %	0.20 (0.10-0.30)	0.20 (0.10-0.30)	0.558
Neutrophil count, × 10^9 /L	2.9 (1.9-3.9)	4.2 (2.7-6.1)	< 0.001
Lymphocyte count, × 10^9 /L	1.1 (0.78-1.5)	0.77 (0.49-1.14)	< 0.001
Monocyte count, × 10^9 /L	0.33 (0.24-0.44)	0.37 (0.23-0.52)	0.215
Eosinophil count, × 10^9 /L	0.01 (0.00-0.03)	0.00 (0.00-0.01)	0.087
Basophil count, × 10^9 /L	0.01 (0.01-0.01)	0.01 (0.01-0.02)	0.364
D-dimer, mg/L	0.45 (0.24-0.90)	0.73 (0.46-1.83)	< 0.001
C-reactive protein, mg/dL	1.4 (0.46-3.1)	3.9 (2.4-7.2)	< 0.001
Cardiac Enzymes			
Aspartate amino transferase, U/L	22.1 (17.3-31.1)	33.6 (22.2-42.8)	< 0.001
Alpha-hydroxybutyric dehydrogenase, U/L	143.0 (113.0-174.1)	189.0 (157.5-268.5)	< 0.001
Lactate dehydrogenase, U/L	183.5 (142.8-231.2)	252.0 (199.7-331.8)	< 0.001
Creatine kinase, U/L	78.5 (45.0-132.4)	108.4 (59.8-248.0)	0.001
Liver function			
Alanine aminotransferase, U/L	21.6 (13.4-33.4)	25.8 (15.7-38.8)	0.202
Aspartate transaminase, U/L	22.3 (17.1-31.2)	31.1 (21.0-40.1)	< 0.001
Gamma-glutamyl transpeptidase, U/L	22.3(13.7-42.8)	36.8 (22.7-51.5)	< 0.001
Kidney function			
Urea, mmol/L	4.1 (3.2-5.1)	6.3 (4.6-8.2)	< 0.001
Creatinine, μmol/L	64.7 (52.8-75.1)	77.3 (63.8-90.8)	< 0.001
Procalcitonin, ng/mL	0.05 (0.04-0.09)	0.11 (0.06-0.31)	< 0.001
B-type natriuretic peptide, pg/mL	59.4 (22.4-106.7)	156.0 (56.4-486.7)	< 0.001
Electrolyte			
Potassium, mmol/L	4.1 (4.0-4.3)	4.1 (3.8-4.3)	0.353
Sodium, mmol/L	141.1 (140.0-142.2)	139.9 (137.8-141.5)	< 0.001
Chloride, mmol/L	103.9 (102.5-105.6)	102.7 (100.9-105.1)	0.006
Calcium, mmol/L	2.3 (2.2-2.4)	2.2 (2.1-2.3)	< 0.001
Phosphate, mmol/L	1.0 (0.9-1.1)	0.95 (0.80-1.1)	0.005

Data are median (IQR). p values comparing non-severe and severe groups were obtained using the Mann-Whitney U test.

Table 3. The prognostic performance of the combination model (Model 4) on five test datasets

Dataset	Tool	AUC (95% CI)	Accuracy (95% CI)	Sensitivity	Specificity	PPV	NPV
Test 1	Model	0.88 (0.75-1.0)	80.6% (64.0-91.8%)	75.0%	85.0%	80.0%	81.0%
	PSI	0.87 (0.74-1.0)	77.8% (60.9-89.9%)	56.3%	95.0%	90.0%	73.1%
Test 2	Model	0.88 (0.80-0.95)	78.9% (69.0-86.8%)	75.5%	82.9%	84.1%	73.9%
	PSI	0.86 (0.79-0.94)	65.6% (54.8-75.3%)	38.8%	97.6%	95.0%	57.1%
Test 3	Model	0.93 (0.83-1.0)	87.5% (71.0-96.5%)	85.7%	88.0%	66.7%	95.7%
	PSI	0.89 (0.77-1.0)	75.0% (56.6-88.5%)	0.00%	96.0%	0.00%	77.4%
Test 4	Model	0.84 (0.76-0.93)	74.4% (64.2-83.1%)	90.0%	55.0%	71.4%	81.5%
	PSI	0.78 (0.68-0.87)	67.8% (57.1-77.3%)	62.2%	73.3%	70.0%	66.0%
Test 5	Model	0.89 (0.85-0.94)	79.2% (72.5-84.9%)	96.9%	57.5%	73.6%	93.9%
	PSI	0.71 (0.63-0.78)	62.9% (55.4-70.0%)	42.9%	87.5%	80.8%	55.6%
Mean (Test 1-5)	Model	0.88	80.1%	84.6%	73.7%	75.2%	85.2%
	PSI	0.82	69.8%	40.0%	89.9%	67.2%	65.8%

PSI, pneumonia severity index; AUC, area under the receiver operating characteristic curve; CI, confidence interval; PPV, positive prediction value; NPV, negative prediction value.

Figure legends

Figure 1. Flowchart of the patient selection process

Figure 2. Chest CT images of two patients with COVID-19 pneumonia

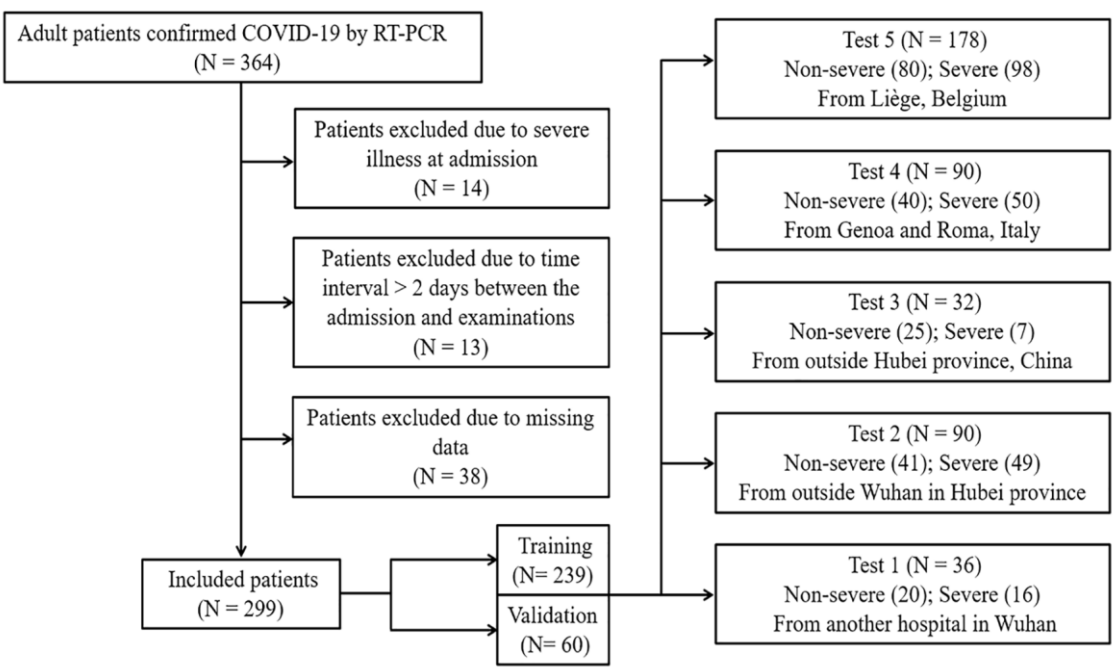
Figure 2-A, 48 year-old man, the focal ground-glass opacities in the bilateral lung lobes (yellow arrow) were automatically segmented (orange areas) and calculated the lesion volume in each lobe (right superior lobe: 0.2%, right middle lobe: 0.3%, right inferior lobe: 0.1%, left superior lobe: 0.9%, and left inferior lobe: 9.4%). The lesion range score was 6 (1+1+1+1+2).

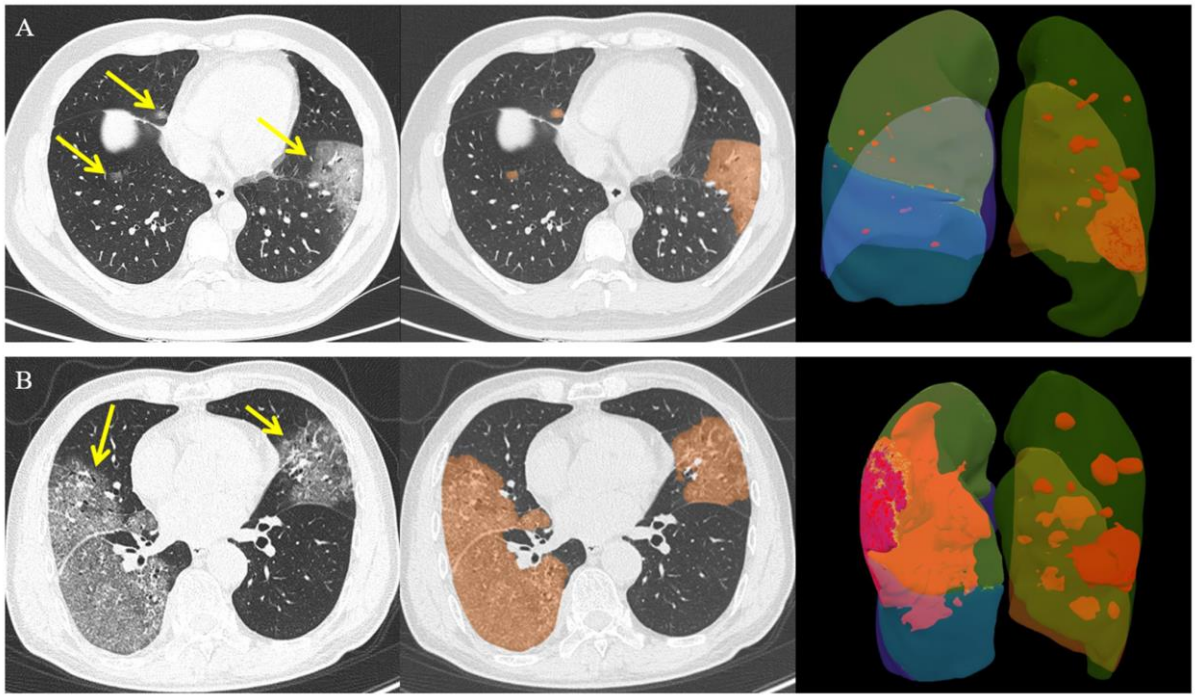
Figure 2-B, 70 year-old man, the peripheral ground-glass opacities in the bilateral lung lobes (yellow arrow)were automatically segmented (orange areas) and calculated the lesion volume in each lobe (right superior lobe: 32.1%, right middle lobe: 16.4%, right inferior lobe: 62.7%, left superior lobe: 12.8%, and left inferior lobe: 7.3%). The lesion range score was 13 (3+2+4+2+2).

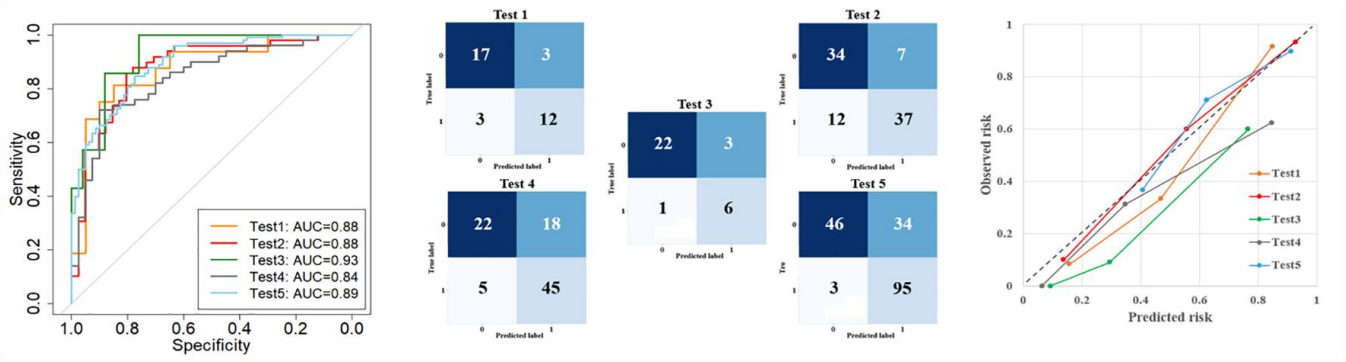
Figure 3. The receiver operator characteristic curve, confusion matrix, and calibration curve for the test datasets

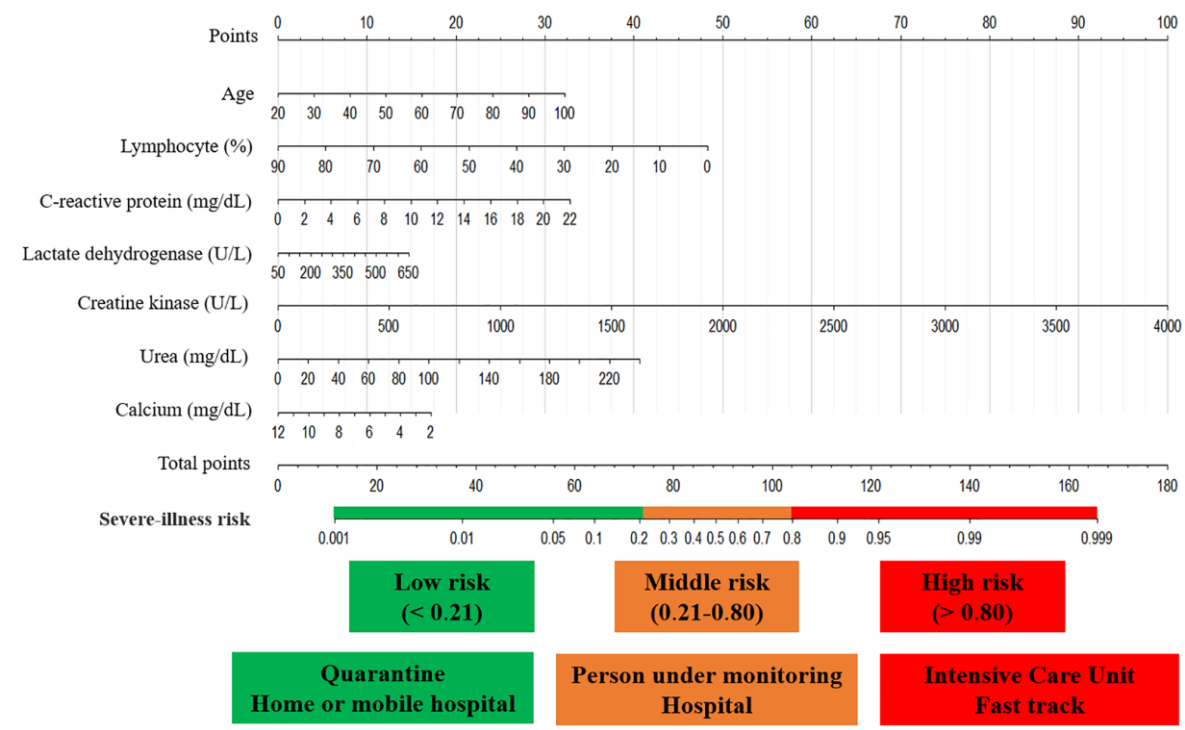
Figure 4. Severe-illness risk nomogram and triage tool for clinicians

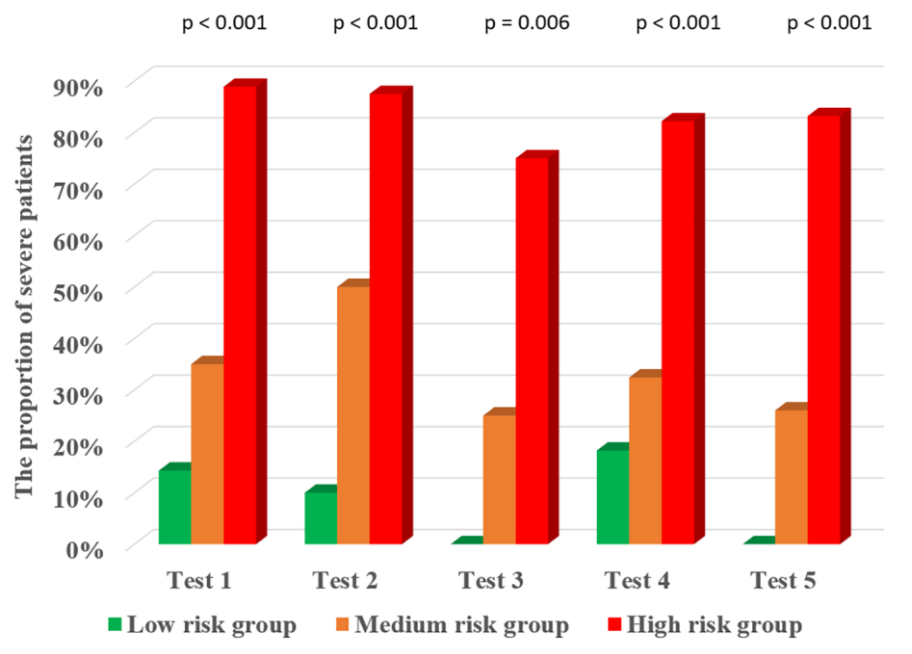
Figure 5. A histogram plot of the proportion of severe patients in low, medium, and high-risk groups of the test datasets


Adult patients confirmed COVID-19 by RT-PCR
(N = 364)
Patients excluded due to severe illness at admission
(N = 14)
Patients excluded due to time interval > 2 days between the admission and examinations
(N = 13)
Patients excluded due to missing data
(N = 38)
Included patients
(N = 299)
Training
(N= 239)
Validation
(N= 60)
Test 5 (N = 178)
Non-severe (80); Severe (98)
From Liège, Belgium
Test 4 (N = 90)
Non-severe (40); Severe (50)
From Genoa and Roma, Italy
Test 3 (N = 32)
Non-severe (25); Severe (7)
From outside Hubei province, China
Test 2 (N = 90)
Non-severe (41); Severe (49)
From outside Wuhan in Hubei province
Test 1 (N = 36)
Non-severe (20); Severe (16)
From another hospital in Wuhan


A
B


Sensitivity
Specificity
Test1: AUC=0.88
Test2: AUC=0.88
Test3: AUC=0.93
Test4: AUC=0.84
Test5: AUC=0.89
Test 1
Test 2
Test 3
Test 4
Test 5
True label
Predicted label
Observed risk
Predicted risk
Test1
Test2
Test3
Test4
Test5


Points
0 10 20 30 40 50 60 70 80 90 100
Age
20 30 40 50 60 70 80 90 100
Lymphocyte (%)
90 80 70 60 50 40 30 20 10 0
C-reactive protein (mg/dL)
0 2 4 6 8 10 12 14 16 18 20 22
Lactate dehydrogenase (U/L)
50 200 350 500 650
Creatine kinase (U/L)
0 500 1000 1500 2000 2500 3000 3500 4000
Urea (mg/dL)
0 20 40 60 80 100 140 180 220
Calcium (mg/dL)
12 10 8 6 4 2
Total points
0 20 40 60 80 100 120 140 160 180
Severe-illness risk
0.001 0.01 0.05 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.95 0.99 0.999
Low risk (< 0.21)
Middle risk (0.21-0.80)
High risk (> 0.80)
Quarantine Home or mobile hospital
Person under monitoring Hospital
Intensive Care Unit Fast track


p < 0.001
p < 0.001
p = 0.006
p < 0.001
p < 0.001
The proportion of severe patients
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Test 1
Test 2
Test 3
Test 4
Test 5
Low risk group
Medium risk group
High risk group

Supplement

Test datasets

Test 1: from China Resources Wuhan Iron and Steel Hospital, Wuhan, China (non-severe patients: 20, severe patients: 16). Test 2: from Huangshi Central Hospital, Huangshi, China (non-severe patients: 41, severe patients: 49). Test 3: from Shaoyang Central Hospital, Shaoyang, China (non-severe patients: 16, severe patients: 3), Southern Hospital of Southern Medical University, Guangzhou, China (non-severe patients: 5, severe patients: 1), and Affiliated Zhongshan Hospital Dalian University, Dalian, China (non-severe patients: 4, severe patients: 3). Test 4: from National Institute for Infectious Diseases – IRCCS, Roma, Italy (non-severe patients: 18, severe patients: 16) and from IRCCS Ospedale Policlinico San Martino, Genoa, Italy (non-severe patients: 22, severe patients: 34). Test 5: from CHU of Liège, Liège, Belgium ((non-severe patients: 80 severe patients: 98).

CT acquisition and reconstruction parameters

Chest CT scans were performed using one of the CT scanners (uCT 780, United Imaging, China and Brilliance iCT 128, Philips Medical Systems, the Netherlands) with patients in the supine position. The scanning range was from the level of the upper thoracic inlet to the inferior level of the costophrenic angle. For CT acquisition, the tube voltage was 120kVp with automatic tube current modulation, a field of view (FOV) of 350 × 350 mm, and a matrix size of 512 × 512. All images were reconstructed into a slice thickness of 1 mm and an interval of 1 mm.

Table S1. A list of R packages

Purposes	Functions	Packages	Versions
Demographic statistics	‘fisher.test’, ‘wilcox.test’	‘stats’	3.6.3
Spearman's rank correlation	‘cor’	‘stats’	3.6.3
Sample balance	‘ADAS’	‘smotefamily’	1.3.1
Feature selection	‘Boruta’, ‘rfe’	‘caret’, ‘Boruta’	6.0-86, 6.0.0
Modeling, nomogram	‘lrm’, ‘nomogram’	‘rms’	5.1-4
ROC plots, AUC values, and test	‘roc’	‘pROC’	1.16.2, 3.3.0
Diagnosis values	‘confusionMatrix’	‘caret’	6.0-86
Calibration	‘plotCalibration’	‘PredictABEL’	1.2-4

ROC, receiver operating characteristic curve; AUC, area under the roc curve.