Development of Severity and Mortality Prediction Models for COVID-19 Patients at Emergency Department Including the Chest X-Ray

Pilar Calvillo Batllés (✉ pilarcalvillo@gmail.com)

Radiology Department, University and Polytechnic La Fe Hospital, Valencia, Spain. <https://orcid.org/0000-0002-2288-0980>

Leonor Cerdá-Alberich

Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute, Valencia, Spain.

Carles Fonfría-Esparcia

Radiology Department, University and Polytechnic La Fe Hospital, Valencia, Spain.

Ainhoa Carreres-Ortega

Radiology Department, University and Polytechnic La Fe Hospital, Valencia, Spain.

Carlos Francisco Muñoz-Núñez

Radiology Department, University and Polytechnic La Fe Hospital, Valencia, Spain

Laura Trilles-Olaso

Radiology Department, University and Polytechnic La Fe Hospital, Valencia, Spain.

Carlos Baeza-Delgado

Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute, Valencia, Spain.

Jose Sánchez-García

Quantitative Imaging Biomarkers in Medicine (QUIBIM SL), Valencia, Spain.

Ignacio Gómez-Rico Junquero

Universitat Politècnica de Valencia (UPV), Spain.

Ana Gil-Brusola

Microbiology Department, University and Polytechnic La Fe Hospital of Valencia, Spain.

Rodrigo Blanco-Salado

Radiology Department, University Arnau de Vilanova Hospital, Valencia, Spain.

Alberto García-Ortega

Respiratory Department, University and Polytechnic La Fe Hospital. Medical Research Institute Hospital La Fe (IISLAFE), Valencia, Spain.

Marta Montero-Alonso

Internal Medicine, University and Polytechnic La Fe Hospital, Valencia, Spain.

Pilar Estellés-Lerga

Radiology Department, University and Polytechnic La Fe Hospital, Valencia, Spain.

Javier Millán-Soria

Emergency Department, University and Polytechnic La Fe Hospital of Valencia, Spain

Cristina Aguado-Codina

Clinical Analysis Department, University and Polytechnic La Fe Hospital of Valencia

Luis Martí-Bonmatí

Radiology Department, University and Polytechnic La Fe Hospital. Biomedical Imaging Research Group (GIBI230), La Fe Health Research Institute, Valencia, Spain

Keywords: COVID-19, chest X-Ray, prognosis, predictive models, artificial intelligence

Posted Date: October 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-88282/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Objectives: To develop prognosis prediction models for COVID-19 patients attending an emergency department (ED) based on initial chest X-ray (CXR), demographics, clinical and laboratory parameters.

Methods: All symptomatic confirmed COVID-19 patients admitted to our hospital ED between February 24th and April 24th 2020 were recruited. CXR features, clinical and laboratory variables and CXR abnormality indices extracted by a convolutional neural network (CNN) diagnostic tool were considered potential predictors on this first visit. The most serious individual outcome defined the three severity level: 0) home discharge or hospitalization ≤ 3 days, 1) hospital stay >3 days and 2) intensive care requirement or death. Severity and in-hospital mortality multivariable prediction models were developed and internally validated. The Youden index was used for model selection.

Results: A total of 440 patients were enrolled (median 64 years; 55.9% male); 13.6% patients were discharged, 64% hospitalized, 6.6% required intensive care and 15.7% died. The severity prediction model included oxygen saturation/inspired oxygen fraction (SatO₂/FiO₂), age, C-reactive protein (CRP), lymphocyte count, extent score of lung involvement on CXR (ExtScoreCXR), lactate dehydrogenase (LDH), D-dimer level and platelets count, with AUC-ROC=0.94 and AUC-PRC=0.88. The mortality prediction model included age, SatO₂/FiO₂, CRP, LDH, CXR extent score, lymphocyte count and D-dimer level, with AUC-ROC=0.97 and AUC-PRC=0.78. The addition of CXR CNN-based indices slightly improved the predictive metrics for mortality (AUC-ROC=0.97 and AUC-PRC=0.83).

Conclusion: The developed and internally validated severity and mortality prediction models could be useful as triage tools for COVID-19 patients and they should be further validated at different ED.

Introduction

The COVID-19 pandemic is posing a large challenge for health systems, forcing a balance to be found between resource management and safe decision-making with a lower than needed scientific evidence. Uncertainties make necessary the development of specific disease models in order to identify patients by prognosis and severity, requiring hospital or even intensive care. Thoracic imaging has served as a diagnostic tool in emergency department (ED) as it may reveal suggestive COVID-19 patterns of lung involvement [1-5]. However, studies on the utility of the chest X-ray (CXR) for predicting health outcomes are limited [6, 7] and the prognostic studies have mainly been based on chest CT [8-10].

Considering the higher use of CXR, its larger availability and safer use to control the spread of the virus when compared with CT, we aimed to develop two multivariable prediction models for severity and mortality estimations in COVID-19 taking into consideration the radiological, demographic, clinical and laboratory variables registered on the emergency evaluation.

Materials And Methods

The institutional review board approved this study. It followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis: the TRIPOD statement [11]. The risk of bias and applicability was assessed with the Prediction model Risk of Bias Assessment Tool (PROBAST) [12].

Patients

All consecutive symptomatic adult patients visiting the ED of our university hospital between 24 February and 24 April 2020 were included if CXR was performed and Severe Acute Respiratory Syndrome - Coronavirus 2 (SARS-CoV-2) RNA was detected in nasopharyngeal swab or sputum/bronchoalveolar lavage. Patients with simultaneous final diagnosis other than COVID-19 were excluded. Emergency physicians triaged these patients. Oligosymptomatic patients with normal CXR and absence of chronic diseases were discharged at home. Patients admitted at the hospital were treated with standard of care drugs in force

at the time and they were discharged if afebrile for at least 3 days with respiratory symptoms and laboratory parameters improvement (Flowchart of the study in Figure 1).

Chest X-ray review

Initial CXR readings on admission were distributed among five thoracic radiologists blinded to the rest of parameters and outcome. The following items were described (Figure 2):

- Absence (level 0) or presence and density of opacities: only low-density (level 1) or consolidation (+/- low density) (level 2). Lung opacities were considered “low-density” if the attenuation did not conceal the underlying vessels and “consolidation” if the opacification of the parenchyma obscured the underlying vessels.
- Distribution of opacities: peripheral prevalence / central prevalence / both without clear prevalence; unilateral / bilateral; superior / medium / lower fields. For determining the extent of involvement, each lung was divided in upper, medium and inferior field, with a maximum of six fields. The affected lobes were not recorded since a high percentage of the radiographs were portable (27.7%).
- Extension degree and score of lung involvement: the extent was graded as mild (if size opacity was less than 1 field); moderate (1-2 fields involved); extensive (3-4 fields involved) and very extensive (5-6 fields involved). A numerical value was assigned to each field depending on the percentage with increased attenuation: 0 (0%), 1 ($\leq 50\%$), and 2 ($> 50\%$). A total score of the lung involvement extent (ExtScoreCXR) was reached by adding up the six field scores, obtaining a value from 0 to 12.

House-made repository software was used to record all the variables in a structured shared database, with description and imaging reminders aiming to reduce variability between readers and mandatory fill-in fields to optimize data collection.

Demographics, clinical and laboratory variables

Demographics, institutionalization, comorbidities, clinical manifestations, peripheral oxygen saturation (SatO₂), laboratory data -C reactive protein (CRP), lactate dehydrogenase (LDH), lymphocyte count, platelet count, and D-dimer- and coinfections were recorded. We calculated SatO₂/FiO₂ to avoid data loss from patients with SatO₂ obtained under oxygen therapy. FiO₂ is the fraction of inspired oxygen and changes depending on the oxygen flow rate delivered to each patient; for room air it is 0.21.

Computational Imaging (Artificial Intelligence Data)

Indices for “consolidation”, “lung opacity” and “abnormal CXR” extracted from CXR by a Convolutional neural networks (CNN)-based diagnostic tool, QUIBIM Precision CXR v2.0.0 (QUIBIM S.L) with CE mark class IIa, were incorporated into the final model to assess whether they improved its predictive accuracy.

Outcome variables

Three severity levels were defined: home discharge or hospitalization ≤ 3 days (level 0), hospital stay > 3 days (level 1), need for intensive care unit (ICU) stay or death due to COVID-19 (level 2). Both days of hospitalization and days to death were registered. The median of follow-up was 91 days (range 64-124 days).

Univariable analysis.

Correlations between the lung involvement extension on CXR (degrees and score) and the days with symptoms, the SatO₂/FiO₂ and the variable outcomes were investigated. Spearman, Kendall, Rank or Point biserial tests were used depending on the type of the studied variables.

Prognostic predictive models

Different prognostic predictive models were developed using three types of classifiers or ensemble methods (Gradient Boosting, Random Forest and Support Vector Machine) and applying a stratified cross-validation with the 80% of the population. An internal validation was performed with an unseen dataset corresponding to the remaining 20% to assess model generalizability and robustness.

In order to avoid redundant information, features with high correlation (>80%) and low predictive power were excluded from the models. Following this criterion, the extent degrees and the distribution of the opacities in the medium lung field were discarded (Figure 3).

Three models were developed with different predictive variables, the first one containing epidemiological and radiological features. The clinical and laboratory parameters as well as the CNN-derived based data were successively incorporated into the first model to build the second and third models, respectively. A partial under-sampling methodology followed by a synthetic minority over-sampling technique (SMOTE) was used to address the data imbalance problem, very common in Machine Learning environments [13]. Features were standardized accordingly.

Sensitivity, specificity, PPV, NPV, AUC-ROC and precision-recall curves (AUC-PRC) were obtained for each model. The Youden index was used for the optimal threshold selection of the classification model, by maximizing the highest sensibility and NPV for critically ill (or dead) patients and the highest specificity and PPV for the mild severity (or alive) ones. DeLong's test for two correlated ROC curves [14] was used to compare the performance of the models.

Results

Patients

From 445 registered patients, 5 were excluded (1 acute appendicitis, 1 cholangitis, 1 diverticulitis, 1 ictus stoke and 1 cardiac failure). A total final population of 440 was enrolled in the study.

Demographics, comorbidities

The median age was 64 years (range 17-100) and 55.9% were male. 79% suffered ≥ 1 comorbidities; the most frequent were hypertension, dyslipidaemia and diabetes (Table 1).

Clinical and laboratory variables, SatO2/FiO2

Upon their arrival to ED the average of days with symptoms was 6.8 (range 0-30) and the most common symptoms were by this order fever and cough. The average oxygen saturation was 93.7% (range 55-100%). There was loss of ≥ 1 laboratory parameters in 67 patients because they had not been performed (Table 2).

Outcome variables

13.6% patients were discharged at home or hospitalized ≤ 3 days; 64% patients were hospitalized (4-54 days, average 17 days); 6.6% required intensive care (2-65 days, average 18 days in ICU) and 15.7% died (0-51 days after admission, average 10 days).

Chest X-ray review

The median time between CXR and Real-Time reverse transcription Polymerase Chain Reaction (RT-PCR) was 1 day (range 0-30). 65.9% of patients with pending RT-PCR result showed suggestive COVID-19 lung involvement on CXR, anticipating the definitive diagnosis. The ExtScoreCXR was 3.3 ± 3.07 (average $\pm$ SD) (Table 3).

Consolidations suggesting organized pneumonia were already presents in the baseline CXR of 8 patients and were also included for the extent calculation. The lineal pattern was not considered for the model since it was only described in 2

patients with more advanced disease. 11 coinfections were detected in the 3 days before or after the CXR (7 streptococcus pneumonia, 4 respiratory viruses): 4 CXR suggested COVID-19 infection, 4 were normal and 2 showed non-specific opacities; no lobar consolidation was observed. From 76 patients discharged at home 24% were admitted in a second visit to ED. The first CXR was normal in 7 of these patients but the lung involvement was underreported in 11 because the opacities were very slight or difficult to interpret due to the overlapping of the breasts or scapulae; in the second visit all presented progression of the lung involvement.

Univariable analysis. Lung involvement extension in CXR

The lung involvement extent -degrees and score- showed poor correlation with days of symptoms duration ($r=0.198$ and $r=0.176$, respectively, $p\text{-value}<0.001$), strong negative correlation with $\text{SatO}_2/\text{FiO}_2$ ($r=-0.53$ and $r=-0.57$, respectively, $p\text{-value}<0.001$), strong-moderate correlation with the evolution severity level ($r=0.536$ and $r=0.491$, respectively, $p\text{-value}<0.001$), poor correlation with days of hospitalization ($r=0.240$ and $r=0.246$, respectively, $p\text{-value}<0.001$), no significant correlation with ICU stay days, and poor-moderate correlation with mortality ($r=0.277$ and $r=0.310$, respectively, $p\text{-value}<0.001$).

Prognostic prediction models

The $\text{SatO}_2/\text{FiO}_2$, the CNN-based index for lung consolidation, the LDH, the ExtScoreCXR, the age, the lymphocyte count, the CRP, the CNN-based index for lung opacities, the D-dimer level, and the platelets count were, in this order, the most important predictors of severity level outcome for the most severe group of patients (Figure 4). The ROC and the PRC curves of the internal validation performed with an unseen dataset for the severity level prognostic predictive models built with three different combinations of features are shown in Figure 5. The curves per severity level are obtained after applying a one-vs-all classification methodology. An improvement of the AUC-ROC and AUC-PRC is observed for both the most and least critically ill patients, as more features are included in the model. In particular, the larger effect is obtained when adding the clinical and laboratory parameters (micro-average AUC-ROC=0.94, micro-average AUC-PRC=0.88). The addition of the CNN-based indices increases the AUC-PRC value of the patients belonging to the extreme severity levels but has the opposite effect on the mid-severity level ones, resulting in a worsening of the predictive metrics (micro-average AUC-ROC=0.91, micro-average AUC-PRC=0.82). A weighted micro-average statistical approach is used to obtain the values per severity level after threshold optimization of the classification model with the Youden index (Table 4).

Regarding the clinical outcome of mortality, the best model is achieved with a Gradient Boosting classifier with the inclusion of the selected epidemiological, radiological, clinical, and laboratory parameters and the CNN-based indices (AUC-ROC=0.97, AUC-PRC=0.83). The age, the $\text{SatO}_2/\text{FiO}_2$, the lymphocyte count, the LDH, the CNN-based index for lung consolidation on CXR, the platelets count and the D-dimer level were, in this order, the most weighted predictors of in-hospital mortality (Figure 4). A decrease in these metrics is observed when the CNN-based indices are removed from the model (AUC-ROC=0.97, AUC-PRC=0.78) but an improvement on the PPV and specificity of the model is achieved by means of threshold optimization with the Youden index (Table 5). In this case, the age, the $\text{SatO}_2/\text{FiO}_2$, the CRP, the LDH, the ExtScoreCXR, the lymphocyte count and the D-dimer level were, in this order, the most weighted predictors. There were no statistically significant differences in terms of the AUC-ROC between the models with and without CNN-based indices ($p\text{-value}=0.315$), suggesting that the addition of the AI parameters does not offer a significant additional improvement on the model performance. Figure 6 shows the ROC and the PRC curves of the internal validation performed with an unseen dataset for a selection of three classification mortality models built with three different combinations of features.

Discussion

In this study, the presence and extent of lung involvement on the initial CXR of COVID-19 patients has a prognostic value in both univariable and multivariable analysis. In the first developed multivariable models based on age, gender and radiological features, the ExtScoreCXR was the strongest predictor of severity and the second predictor of in-hospital mortality after age. However, the addition of parameters usually registered at admission significantly improved the predictive accuracy of the models. These results demonstrate the greatest usefulness of CXR as a prognostic tool in COVID-19 when considered together

with SatO₂/FiO₂, age and laboratory parameters. Despite a careful analysis of the imaging, slight opacities, overlapping of structures, indeterminate opacities or normal CXR are not uncommon (31.8% in our series). The use of prognostic models attempts the safe decision-making, avoiding discharge at home of patients requiring hospital care and unnecessary hospitalizations or CT overuse.

On the other hand, we also confirmed the strong negative correlation between lung involvement extent and SatO₂/FiO₂. It supports the widely accepted indication of CT pulmonary angiography in case of oxygen desaturation or dyspnea and normal or mild lung involvement on CXR [15], looking for extended slight lung opacities not visible on CXR or pulmonary thrombosis/embolism [16]. In concordance with this, both developed prognostic models also include the SatO₂/FiO₂ as a strong predictor.

The distribution and the density of opacities were not strong enough to remain predictors in the definitive models.

In the literature, a CXR severity score in ED was predictive of risk for hospital admission and intubation of COVID-19 patients aged 21-50 [6]. Including the CXR abnormality as predictor a risk score showed an AUC-ROC of 0.88, but without extent assessment [17]. On admission chest CT, a well aerated lung parenchyma less than 73%, and after adjustment for patient demographics and clinical parameters, was associated with ICU admission or death [10]. The extent of lung involvement was also associated to worse outcomes in severe acute respiratory syndrome [18, 19].

The most reported predictors of severe prognosis in patients with COVID-19 included age, sex, features derived from CT, CRP, LDH, and lymphocyte count [9, 20] and the most published predictors of mortality are older age [21-23] and D-dimer level [20, 21]. These predictors coincide with most of those we have observed in the multivariable analysis and included in the two predictive models (SatO₂/FiO₂, age, CRP, lymphocyte count, ExtScoreCXR, LDH, D-dimer level and platelet count).

Days with symptoms, clinical presentation, institutionalization, comorbidities and the rest of CXR features did not show enough predictive power to be included in the models. The number of days with symptoms on arrival of patients to ED was neither related to the lung involvement extension. In other series neither a significant difference was identified between the severe and non-severe patients, regarding the median days from symptom onset to hospital admission [24]. Tobacco, comorbidities as hypertension, diabetes, cardiovascular disease, respiratory diseases, cancer history and the presence of fever, dyspnoea, haemoptysis and unconsciousness, were also associated to a worse prognosis in some publications [17, 23, 25], but not in our study. The ExtScoreCXR and laboratory parameters have been observed to have a large impact in the model in contrast to the symptoms and comorbidities. This raises the need of performing these tests to all COVID-19 patients with viral symptomatology, regardless of the type of symptoms and chronic diseases.

Regarding the addition of CNN-based diagnostic tool, a non-significant improvement of the predictive metrics of mortality prediction model was observed, probably because only the "consolidation" and "lung opacity" indices, and not the extension, were included as predictors. The extent score CNN-based promises to stage the severity disease on CXR of COVID-19 patients and its weight in a predictive model has to be investigated [26, 27].

The National Early Warning Score 2 (NEWS2), based on vital signs, is the most used score in ED. Its predictive accuracy in COVID-19 patients is higher than other clinical risk scores [28, 29], but the models developed in this study exceed this accuracy (AUC-ROC=0.94 and AUC-ROC=0.97 for severity and mortality respectively), as expected because the addition of other relevant variables.

Recently a multivariable model including CXR at admission was developed to predict critical illness in hospitalized COVID-19 patients [7]. The predictors that remained in the model were male gender, obstructive lung disease, symptom duration > 7 days, neutrophil count, CPR, LDH, distribution of lung disease and CXR score, with an AUC-ROC=0.77. The CPR, LDH and the lung involvement extension on CXR are also included in our final model but there are no further coincidences in the rest of predictors. This is probably explained by the different model development methodologies including a different feature selection strategy, as they used a univariable statistical test and we based our selection on the correlation between parameters and with respect to the clinical outcome. Other discrepancies include the data pre-processing steps, as we included a

combination of some over- and under-sampling techniques as well as data standardization; and the consideration of different model architectures, as they employed a multivariable logistic regression, which relies on transformations for non-linear features. In order to overcome this issue, we tested three different model architectures: Support Vector Machine, Random Forest and Gradient Boosting, which can handle non-linear features as well as their interactions, and perform well in a large feature space.

As potential sources of bias, the severity level is a decision-based clinical outcome, with a certain degree of inter-and intra-observer variability, unlike mortality. However, decisions about hospital or ICU stay and treatment followed an agreed action guide depending on their clinical status, decreasing this variability. Also, the proportion of patients with critical evolution (22%) was within the range published in longer series (15-36%) [30, 31]. The number of comorbidities were not analysed because we observed a strong association to the age, included as predictor. The internal validation was performed with 88 cases. However, it has been reported that a minimum sample size of 100 is recommended in order to achieve a robust validation [32]. In addition, an external validation with cases from other hospitals is desirable to assess the generalizability and the potential use of the developed models in daily clinical practice.

In conclusion, the developed multivariable prognosis prediction models showed a high predictive accuracy that could allow triage of symptomatic COVID-19 patients at ED to improve the decision-making. The application to estimate the severity level and the in-hospital mortality is available on <http://upv.datahub.ege.eu:30054/hulafecovid19models> and it should be validated at different ED.

Abbreviations

COVID-19: 2019 Novel Coronavirus

ED: Emergency Department

CXR: Chest X-ray

CNN: Convolutional neural networks

SatO₂/FiO₂: oxygen saturation/inspired oxygen fraction

PRC: precision-recall curve.

CRP: C-Reactive Protein.

ExtScoreCXR: Extent score of lung involvement on CXR.

LDH: Lactate dehydrogenase.

SARS-CoV-2: Severe acute respiratory syndrome - Coronavirus 2

ICU: Intensive care unit

RT-PCR: Real-Time reverse transcription Polymerase Chain Reaction

Declarations

Competing interests: The authors declare no competing interests.

Conflicts of interest: The authors have not received any funding. Jose Sánchez-García is an employee of Quantitative Imaging Biomarkers in Medicine (QUIBIM SL) whose software has been used in one of the predictive models.

References

1. Chung M, Bernheim A, Mei X et al. (2020) CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology* 295(1):202–7.
2. Bernheim A, Mei X, Huang M et al. (2020) Chest CT Findings in Coronavirus Disease-19 (COVID-19): Relationship to Duration of Infection. *Radiology* 295(3):200463.
3. Pan F, Ye T, Sun P et al. (2020) Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). *Radiology* 295(3):715–21.
4. Bai HX, Hsieh B, Xiong Z et al. (2020) Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT. *Radiology* 296(2):200823.
5. Salehi S, Abedi A, Balakrishnan S, Gholamrezanezhad A. (2020) Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. *Am J Roentgenol* 14;1–7.
6. Toussie D, Voutsinas N, Finkelstein M et al. (2020) Clinical and Chest Radiography Features Determine Patient Outcomes In Young and Middle Age Adults with COVID-19. *Radiology* 14:201754.
7. Schalekamp S, Huisman M, van Dijk RA et al. (2020) Model-based Prediction of Critical Illness in Hospitalized Patients with COVID-19. *Radiology* 13:202723.
8. Yuan M, Yin W, Tao Z, Tan W, Hu Y. (2020) Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. *PloS One* 15(3):e0230548.
9. Wynants L, Van Calster B, Bonten MMJ et al. (2020) Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 7;369:m1328.
10. Colombi D, Bodini FC, Petrini M et al. (2020) Well-aerated Lung on Admitting Chest CT to Predict Adverse Outcome in COVID-19 Pneumonia. *Radiology* 17:201433.
11. Collins GS, Reitsma JB, Altman DG, Moons KGM. (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann Intern Med* 6;162(1):55–63.
12. Wolff RF, Moons KGM, Riley RD et al. (2019) PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 1;170(1):51–8.
13. Leevy JL, Khoshgoftaar TM, Bauder RA, Seliya N. (2018) A survey on addressing high-class imbalance in big data. *J Big Data* 5(1):42.
14. DeLong ER, DeLong DM, Clarke-Pearson DL. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3):837–45.
15. Revel M-P, Parkar AP, Prosch H et al. (2020) COVID-19 patients and the radiology department – advice from the European Society of Radiology (ESR) and the European Society of Thoracic Imaging (ESTI). *Eur Radiol* 20;1–7.
16. Rosovsky RP, Grodzin C, Channick R et al. (2020) Diagnosis and Treatment of Pulmonary Embolism During the COVID-19 Pandemic: A Position Paper from the National PERT Consortium. *Chest* 27 doi: 10.1016/j.chest.2020.08.2064.
17. Liang W, Liang H, Ou L et al. (2020) Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. *JAMA Intern Med* 180(8):1–9.
18. Chau T, Lee P, Choi K et al. (2004) Value of initial chest radiographs for predicting clinical outcomes in patients with severe acute respiratory syndrome. *Am J Med* 15;117(4):249–54.
19. Wan Y-L, Tsay P-K, Cheung Y-C et al. (2007) A Correlation between the Severity of Lung Lesions on Radiographs and Clinical Findings in Patients with Severe Acute Respiratory Syndrome. *Korean J Radiol* 8(6):466–74.
20. Velavan TP, Meyer CG. (2020) Mild versus severe COVID-19: Laboratory markers. *Int J Infect Dis* 95: 304-307.
21. Zhou F, Yu T, Du R et al. (2020) Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet Lond Engl* 395(10229):1054–62.
22. Wu C, Chen X, Cai Y et al. (2020) Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern Med* 180(7):1–11.

23. Zheng Z, Peng F, Xu B et al. (2020) Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *J Infect* 81(2): e16–25.
24. Zhang J, Dong X, Cao Yet al. (2020) Clinical characteristics of 140 patients infected with SARS-CoV-2 in Wuhan, China. *Allergy* 75(7):1730–41.
25. Ruan Q, Yang K, Wang W, Jiang L, Song J. (2020) Clinical predictors of mortality due to COVID-19 based on an analysis of data of 150 patients from Wuhan, China. *Intensive Care Med* 3;1–3.
26. Zhu J, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, Duong TQ. (2020) Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. *PLoS One* 15(7): e0236621.
27. Cohen JP, Dao L, Roth K et al. (2020) Predicting COVID-19 Pneumonia Severity on Chest X-ray With Deep Learning. *Cureus* 12(7): e9448.
28. Jang JG, Hur J, Hong KS, Lee W, Ahn JH. (2020) Prognostic Accuracy of the SIRS, qSOFA, and NEWS for Early Detection of Clinical Deterioration in SARS-CoV-2 Infected Patients. *J Korean Med Sci* 29;35(25): e234.
29. Myrstad M, Ihle-Hansen H, Tveita AA et al. (2020) National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19 – a prospective cohort study. *Scand J Trauma Resusc Emerg Med* 28:66.
30. Wu Z, McGoogan JM. (2020) Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 7;323(13):1239–42.
31. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. (2020) Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA* 26; 323(20):2098.
32. Collins GS, Ogundimu EO, Altman DG. (2016) Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 30;35(2):214–26.

Tables

Table 1. **Characteristics of COVID-19 patients.**

DEMOGRAPHIC INFORMATION		
Age*	Median (interquartile range), (range), years	64 (51-79), (17-100)
Sex*	No. (% of 440)	Male 246 (55,9)
		Female 194 (44)
COMORBIDITIES		No. (% of 440)
Hypertension		191 (43,4)
Dislypemia		125 (28,4)
Diabetes		93 (21,1)
Institutionalization		75 (17)
Cardiovascular disease		56 (12,7)
Immunosuppression		46 (10,5)
Corticotherapy and other immunosuppressors		16 (3,6)
Advanced chronic kidney disease/Dialysis		13 (2,9)
Hematological neoplasm / disease		10 (2,3)
Solid organ trasplant		8 (1,8)
HIV		1 (0,2)
Chronic kidney disease (total)		43 (9,8)
Obesity		42 (9,5)
Cerebrovascular disease		36 (8,2)
Neoplasm		32 (7,3)
Prostate adenocarcinoma		7 (1,6)
Asthma		25 (5,7)
Dementia		19 (4,3)
Current smoker		16 (3,6)
Obstructive sleep apnea		16 (3,6)
Ex-smoker		15 (3,4)
Hypothyroidism		13 (2,9)
Atrial fibrillation		13 (2,9)
Chronic obstructive pulmonary disease		9 (2)
NUMBER OF COMORBIDITIES		No. (% of 440)
None		92 (20,9)
1		87 (19,8)
2		68 (15,4)
3		54 (12,3)
≥ 4 (4-8)		139 (31,5)

Table 2. Clinical presentation of COVID-19 patients.

INTERVAL SINCE SYMPTOMS ONSET	Average +/- sd (range)
Days	6,8 +/- 4,5 (0-30)
SYMPTOMS Total No. 441	No. (%)
Fever	361 (82)
Cough	287 (65,2)
Dyspnea	178 (40,4)
General dyscomphort/Asthenia	171 (38,9)
<u>Gastrointestinal symptoms</u>	103 (23,4)
Myalgias	72 (16,4)
Headache	47 (10,7)
Hyposmia/Dysgeusia	32 (7,3)
OXYGEN SATURATION	Average +/- sd (range)
Sat O2 (air room)	93,7 +/- 5,6 (55-100)
Sat O2 /Fi O2*	443,5 +/-34,4
LABORATORY DATA	Average/ Min / 25% / Median / 75% / Max
Lactate dehydrogenase (U/l) *	306 / 16 / 218,2 / 265,5 / 349,7 / 2146
C reactive protein (mg/l) *	81 / 0,3 / 15,5 / 44,6 / 117,2 / 655,5
Lymphocytes count (x103µ/l) *	1,2 / 0,06 / 0,7 / 1 / 1,4 /12,3
Platelets count (x103µ/l) *	203,7 / 32 / 150 / 193 / 242 / 716
D dimer (ng/l)	1390 / 83 / 350 / 608 / 1015 / 38282

Table 3. Lung involvement on CXR, distribution and extension.

PRESENCE AND DENSITY OF LUNG OPACITIES	No. (%) Total 440
No lung opacities (+/- other findings) - level 0-	86 (19,5)
Low-density opacity(ies) - level 1 -	254 (57,7)
Consolidation/s +/- low-density opacity(ies) -level 2-	100 (22,7)
DISTRIBUTION/LOCATION OF LUNG OPACITIES	No. (%) Total 354
Bilateral	212 (59,9)
Unilateral	142 (39,8)
Peripheral (only or predominantly)	182 (51,1)
Peripheral and central (without predominance)	116 (32,5)
Central (only or predominantly)	56 (15,7)
Lower fields	290 (81,4)
Medium fields	259 (72,7)
Upper fields	178 (50)
EXTENT DEGREES OF LUNG OPACITIES	No. (%) Total 440
No lung opacities (+/- other findings) - level 0-	86 (19,5)
Mild	66 (15)
Moderate	111 (25,2)
Extensive	101 (22,9)
Very extensive	76 (17,3)
SCORE OF LUNG INVOLVEMENT EXTENSION * (ExtScoreCXR)	Average +/- sd (range)
	3,3 +/- 3,07 (0-12)

Table 4. Metrics of the severity predictive models.

Combination of parameters	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC-ROC (%)	AUC-PRC (%)
Epidemiological and Radiological	92.5	68.2	59.8	93.6	85.7	71.3
Epidemiological, Radiological, Clinical and Laboratory	83.1	87.9	77.5	95.1	93.8	87.6
Epidemiological, Radiological, Clinical, Laboratory and CNN-based	86.4	84.1	72.5	96.7	91.4	82.3

Table 5. **Metrics of the mortality predictive models.**

Combination of parameters	Model Architecture	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUC-ROC (%)	AUC-PRC (%)
Epidemiological and Radiological	RF	71.8	90.3	59.5	92.2	86.7	69.3
Epidemiological, Radiological, Clinical and Laboratory	GB	90.0	93.7	69.2	98.3	96.5	77.9
Epidemiological, Radiological, Clinical, Laboratory and CNN-based	GB	90.0	92.1	64.3	98.3	97.1	83.4

Figures

Figure 1

Flowchart of the study.

Figure 2

Chest X-ray features. Up: examples of distribution, density of opacities and extent score. Down (the same CXR): Division of lung fields. Upper fields (suprahilar area) limited by the line that passes under the aortic arch; medium fields (hilar area) and inferior fields (infrahilar area) separated by the line that divides the rest of the lungs into two halves (frequently this line crosses the bifurcation of the right inferior lobar artery). A1: Unilateral central and peripheral low-density opacities (arrow), without predominance. A2: Medium and lower right fields involved $\leq 50\%$; ExtScoreCXR=2. B1: Central right low-density opacity (arrow) and peripheral left consolidation (larger arrows). Peripheral predominance. B2: Medium right and left fields with $\leq 50\%$ of involvement and upper left field with $> 50\%$ of involvement; ExtScoreCXR=4. C1: Bilateral low-density opacities (arrows) and consolidations (larger arrows) without predominance. C2: Upper fields and lower left field with $\leq 50\%$, medium fields and right lower right field with $> 50\%$ of involvement; ExtScoreCXR=9.

Figure 3

Spearman's rank-order correlation matrix for the epidemiological and radiological features.

Figure 4

Importance of model predictors obtained for the severity level (up): SatO2/FiO2 (33%), CNN-based index for lung consolidation on CXR (13%), LDH (12%), ExtScoreCXR (9%), the age (9%), the lymphocyte count (9%), the CRP (7%), CNN-based index for lung opacities on CXR (3%), the D-dimer level (3%) and the platelets count (2%); and mortality (down): age (54%), SatO2/FiO2 (14%), the lymphocyte count (10%), LDH (8%), CNN-based index for lung consolidation on CXR (7%), platelets count (4%) and D-dimer level (3%). Those with an importance of less than 0.01 are excluded from the model.

Figure 5

The ROC (up) and the PRC (down) curves of the internal validation performed with an unseen dataset for the severity predictive models built with three different combinations of parameters. (A: epidemiological and radiological parameters, B: epidemiological, radiological, clinical and laboratory parameters, C: epidemiological, radiological, clinical, laboratory and CNN-based parameters). The curves per severity level are obtained with a one-vs-all classification methodology. The corresponding values of AUC are shown. Precision = true positive / (true positive + false positive). Recall (sensitivity) = true positive / (true positive + false negative).

Figure 6

The ROC (up) and the PRC (down) curves of the internal validation performed with an unseen dataset for the in-hospital mortality predictive models (SVM: Support Vector Machine, RF: Random Forest, GB: Gradient Boosting) built with three

different combinations of parameters (A: epidemiological and radiological parameters, B: epidemiological, radiological, clinical and laboratory parameters, C: epidemiological, radiological, clinical, laboratory and CNN-based parameters). The corresponding values of AUC are shown.