

Article

Outcome Prediction for SARS-CoV-2 Patients Using Machine Learning Modeling of Clinical, Radiological, and Radiomic Features Derived from Chest CT Images

Lorenzo Spagnoli ^{1,2}, Maria Francesca Morrone ^{1,2} , Enrico Giampieri ^{3,4} , Giulia Paolani ^{1,2}, Miriam Santoro ^{1,2} , Nico Curti ⁴ , Francesca Coppola ⁵, Federica Ciccarese ⁵, Giulio Vara ⁵ , Nicolò Brandi ⁵ , Rita Golfieri ⁵ , Michele Bartoletti ⁶, Pierluigi Viale ⁶ and Lidia Strigari ^{1,*} 

- ¹ Medical Physics Unit, IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna, Italy; lorenzo.spagnoli@studio.unibo.it (L.S.); maria.morrone@studio.unibo.it (M.F.M.); giulia.paolani@studio.unibo.it (G.P.); miriam.santoro@studio.unibo.it (M.S.)
 - ² Medical Physics Specialization School, Alma Mater Studiorum, University of Bologna, 40126 Bologna, Italy
 - ³ National Institute of Nuclear Physics (INFN), 40127 Bologna, Italy; enrico.giampieri@unibo.it
 - ⁴ Experimental, Diagnostic and Specialty Medicine-DIMES, University of Bologna, 40126 Bologna, Italy; nico.curti2@unibo.it
 - ⁵ Department of Radiology, IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna, Italy; francesca_coppola@hotmail.it (F.C.); federica.ciccarese@aosp.bo.it (F.C.); giulio.vara@studio.unibo.it (G.V.); nicolo.brandi@studio.unibo.it (N.B.); rita.golfieri@aosp.bo.it (R.G.)
 - ⁶ Infectious Diseases Unit, Department of Medical and Surgical Sciences, IRCCS Azienda Ospedaliero-Universitaria di Bologna, 40138 Bologna, Italy; m.bartoletti@unibo.it (M.B.); pierluigi.viale@aosp.bo.it (P.V.)
- * Correspondence: lidia.strigari@aosp.bo.it



Citation: Spagnoli, L.; Morrone, M.F.; Giampieri, E.; Paolani, G.; Santoro, M.; Curti, N.; Coppola, F.; Ciccarese, F.; Vara, G.; Brandi, N.; et al. Outcome Prediction for SARS-CoV-2 Patients Using Machine Learning Modeling of Clinical, Radiological, and Radiomic Features Derived from Chest CT Images. *Appl. Sci.* **2022**, *12*, 4493. <https://doi.org/10.3390/app12094493>

Academic Editors: Alessandro Micarelli, Giuseppe Sansonetti and Giuseppe D'Aniello

Received: 11 March 2022

Accepted: 24 April 2022

Published: 28 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Featured Application: The present study demonstrates that semi-automatic segmentation enables the identification of regions of interest affected by SARS-CoV-2 infection for the extraction of prognostic features from chest CT scans without suffering from the inter-operator variability typical of segmentation, hence offering a valuable and informative second opinion. Machine Learning methods allow identification of the prognostic features potentially reusable for the early detection and management of other similar diseases.

Abstract: (1) Background: Chest Computed Tomography (CT) has been proposed as a non-invasive method for confirming the diagnosis of SARS-CoV-2 patients using radiomic features (RFs) and baseline clinical data. The performance of Machine Learning (ML) methods using RFs derived from semi-automatically segmented lungs in chest CT images was investigated regarding the ability to predict the mortality of SARS-CoV-2 patients. (2) Methods: A total of 179 RFs extracted from 436 chest CT images of SARS-CoV-2 patients, and 8 clinical and 6 radiological variables, were used to train and evaluate three ML methods (Least Absolute Shrinkage and Selection Operator [LASSO] regularized regression, Random Forest Classifier [RFC], and the Fully connected Neural Network [FcNN]) for their ability to predict mortality using the Area Under the Curve (AUC) of Receiver Operator characteristic (ROC) Curves. These three groups of variables were used separately and together as input for constructing and comparing the final performance of ML models. (3) Results: All the ML models using only RFs achieved an informative level regarding predictive ability, outperforming radiological assessment, without however reaching the performance obtained with ML based on clinical variables. The LASSO regularized regression and the FcNN performed equally, both being superior to the RFC. (4) Conclusions: Radiomic features based on semi-automatically segmented CT images and ML approaches can aid in identifying patients with a high risk of mortality, allowing a fast, objective, and generalizable method for improving prognostic assessment by providing a second expert opinion that outperforms human evaluation.

Keywords: radiomics; CT images; Machine Learning; SARS-CoV-2; mortality

1. Introduction

Since the beginning of 2020, the SARS-CoV-2 virus (Severe Acute Respiratory Syndrome-Corona Viruses-2) has triggered the outbreak of a world-wide pandemic, leading to restrictive measures of isolation and closure. To face the health emergency, hospitals increased the number of beds in intensive care units (ICUs) and introduced novel indicators for prioritizing patient admission and predicting patient outcome.

A reverse-transcriptase polymerase chain reaction (RT-PCR) assay from nasopharyngeal swabs or bronchoalveolar lavage is the reference test for diagnosing SARS-CoV-2 infection [1]. Chest Computed Tomography (CT) has recently been considered to be a potential non-invasive method for independently confirming the diagnosis of suspected COVID-19 patients with a sensitivity of 97%, specificity of 25%, and accuracy of 68% [2]. Consequently, many COVID-19 patients underwent CT scans to evaluate the extent of the damage and improve prognosis estimation thus increasing the possibility of an overdiagnosis.

In addition, CT-based radiological findings (e.g., Ground Glass Opacity [GGO], Crazy Paving, Lung Consolidation) can detect SARS-CoV-2 virus based on 2D/3D imaging techniques in one or both lungs and can be used as a surrogate of disease severity. These findings were reached by means of a consensus in the European Society of Radiology (ESR) [3].

Furthermore, images can convey a large amount of information which the human eye cannot objectively quantify, providing other potential predictive or prognostic factors related to the COVID-19 disease. For this reason, the field of radiomics uses rigorous mathematical definitions and well-defined approaches [4] to quantitatively describe the image-based properties contained within radiological images, such as texture and shape/volumetric information.

Semi-automatic segmentation has recently been suggested as a tool for quickly sectioning the lungs or the COVID-19 lesions, enabling the extraction of the radiomic features in order to improve the prediction of several clinical endpoints, including ICU admission, need for ventilators [5–10], and severe vs. critical conditions [9]. However, only a limited number of papers have investigated patient mortality as an outcome, often having only a relatively limited patient cohort or a short patient follow-up [5–10].

The former limitation is likely related to the manual nature of the segmentation methods used in the papers published, which represents a very time-consuming task. During the pandemic, various semi-automatic segmentation COVID-19-dedicated tools became available; therefore, the performance of Machine Learning models built on the radiomic features extracted was investigated, using one of these tools for predicting mortality in a high-risk COVID-19-positive group.

2. Materials and Methods

2.1. Study Design

The study, regarding the prognostic value of radiomic features, was conducted and included all patients suitable for analysis, according to the guidelines of the Declaration of Helsinki. The study was approved by the Institutional Review Board (or Ethics Committee) of the IRCCS University Hospital of Bologna (protocol code no. EM949-2020_507/2020/Oss/AOUBo, approved on 16 September 2020).

All patients identified according to the inclusion/exclusion criteria before the Ethics committee approval were included retrospectively, while the remaining population (after 16 September 2020) was included prospectively; informed consent forms were obtained. All the clinical data were retrieved from an ad hoc clinical database for SARS-CoV-2 patient management, while the radiological data and CT chest images were retrieved from structured reports and Digital Image Communication in Medicine (DICOM) files available from the Radiology Information System (RIS) and Picture Archiving and Communication System (PACS), respectively.

2.2. Patient Cohort

The patient cohort was made up of a subset of patients, confirmed positive for COVID-19 using RT-PCR, admitted to the IRCCS University Hospital of Bologna–Polyclinic Sant’Orsola-Malpighi (IRCCS AOSP), redirected from neighboring hospitals from February 2020 to March 2021 since the authors’ Institute is a regional emergency hub capable of managing patients at high risk of SARS-CoV-2 infection [11]. Consequently, the present cohort of hospitalized patients was considered at high risk irrespective of the referring hospital.

Chest CT scan findings (radiological and radiomics) and the clinical data available at patient admission were used to develop a predictive model of patient mortality.

The inclusion criteria were the following: having a chest CT scan with slice thicknesses of between 1 and 1.25 mm without contrast medium acquired after patient admission and recorded on the RIS-PACS of the IRCCS AOSP associated with radiological findings, and a complete set of clinical baseline information including RT-PCR positivity to COVID-19. When multiple CT scans were available, only that closest to the date of admission was analyzed.

The duration of hospitalization is reported in Supplementary Materials Table S1 according to patient outcome as well as period of first diagnosis (first or second wave). Moreover, the days elapsed between the CT scan and the hospitalization date were not statistically significantly different (p -value = 0.29) using a standard t -test comparing patients by outcome. The average days of survival were 21 and 14 in the patients hospitalized in the first and second wave, respectively; this difference showed a trend (p -value = 0.074), indicating that patients with severe disease were better selected during the second wave, albeit with expected improvement in the treatment strategies available over time.

The inclusion criteria were fulfilled by 436 patients, i.e., 286 males (65.6%) and 150 females (34.4%). The main patient characteristics and baseline comorbidities are reported in Table 1. The median age was 68.5 (21–99) years; a hypertension status was recorded in 241 patients. Two-hundred and fifty-one had a fever (Temperature $\geq 38^\circ$) at hospital admission. The choice of using this cutoff for fever was based on the variability of body temperature occurring on the day of admission. Information regarding fever and hypertensive state were included in the routine admission procedure and, hence, were available for all patients; however, no additional details were recorded at admission.

Table 1. Clinical characteristics of the patients as well as the radiological findings obtained upon radiologist inspection of the CT scans.

| Variables | Median (Min–Max) |
|------------------------------|-------------------------------------|
| Age (years) | 68.5 (21–99) |
| Respiratory rate (Breaths/m) | 20 (10–98) |
| Days of hospitalization | 13 (0.25–99) |
| | Yes N (%)—No N (%) |
| Hypertension | 241 (55.3%)—195 (44.7%) |
| History of smoking | 347 (79.5%)—89 (20.5%) |
| Obesity | 363 (83.3%)—73 (16.7%) |
| Sex | Male 286 (65.6%)—Female 150 (34.4%) |
| Fever | 251 (57.6%)—185 (42.4%) |
| Lung Consolidation | 225 (51.6%)—211 (48.4%) |
| Ground Glass Opacity (GGO) | 382 (87.6%)—54 (12.4%) |
| Crazy Paving | 336 (77.1%)—100 (22.9%) |
| Bilateral involvement | 403 (92.4%)—33 (7.6%) |

It is also worth noting that the present cohort presented a large prevalence of obese individuals (83%). This could have been a bias as since an estimation of the visceral fat surface and muscular surface obtained with the segmentation software by segmenting a slice at the height of vertebra T12 of the thoracic region was available, the authors expected some dependency on body composition to arise from the respective radiomic feature, which allowed much more nuance in patient characterization.

The CT scans were obtained using an Ingenuity CT (Philips Medical Systems, Cleveland, OH, USA) in 56% of patients, a Lightspeed VCT (General Electric Healthcare, Chicago, IL, USA) in 41% of patients, and an ICT SP (Philips Medical Systems, Cleveland, OH, USA) in 3% of patients.

The scanners can be considered equivalent as the CT chest acquisition protocols were set to produce comparable image quality as verified during the Quality Assurance (QA) controls. In addition, the acquisition protocols remained unchanged during the entire data collection period.

For the most part, the kilovolt peaks (kVps) were set to 120 kV (91.5% scans), with a few exceptions which were set to 100 kV (5.0% scans) or 140 kV (3.5% scans), according to patient characteristics.

2.3. Image Segmentation

Sophia DDM for Radiomics [12] is a CE/FDA-marked software for SARS-CoV-2 patients which offers a CT-based automated workflow for whole-lung segmentation and disease quantification. It was used for both lung and disease volume of interest (VOI) segmentation as well as for radiomic feature extraction [12].

The segmentation was based on region growing techniques, and used gradient detection and volume stability to regulate the convergence of the process. The majority of the radiomic features were defined and extracted following the workflow as per Image Biomarker Standardization Initiative (IBSI) [4] regulations.

Sophia Radiomics also uses two thresholds which correspond qualitatively to the portion of GGO (from -740 HU to -400 HU) and the range of pixel values representing the vascular tree (from -400 HU to about 1000 HU).

These voxels are counted and kept as a measure of damage volume (in mL). In particular, these ranges are generally appropriate for differentiating GGO from the vascular tree; they may affect the quality of the radiomic features extracted and can be manually modified by the user upon visual inspection, if required.

The software produces one-hundred and seventy-seven features relative to both lungs as a single VOI. In addition, quantification of the visceral fat and abdominal mass surface, as a surrogate of Sarcopenia, was computed using manual segmentation of the abdominal cavity on a single slice at the height of vertebra T12. These surfaces identified via the thresholding method were computed by counting the pixels identified and were expressed in cm^2 . In all cases in which the segmentation obtained semi-automatically was incomplete, the patients were eliminated from the study both in cases of partial imaging scans as well as in cases of widespread infection affecting software segmentation capability. All patients were checked manually after the segmentation process for a final approval of inclusion.

2.4. Patient and Image Characteristics

The dataset was composed of 436 patients, each with a set of assigned features. For convenience, the features were categorized into three subsets: Clinical, Radiomic, and Radiological.

The clinical features available at hospital admission were divided into (a) continuous: age at the time of the CT exam and respiratory rate in breaths/min, and (b) binary: Sex of the patient, obesity status, fever at the hospital admission, hypertension condition, and smoking history.

One-hundred and seventy-nine radiomic features were supported by the segmentation software, the majority of which were described in [4], with the addition of visceral fat and abdominal mass.

The six radiological features included the acquisition parameters (kVp, current, and slice thickness) extracted from the DICOM header and Boolean features (such as the bilaterality of the lung damage, the presence of GGO, lung consolidations, and crazy paving) assessed by expert radiologists and extracted from the structured medical report.

Different models were built using each of the feature groups to compare performance in a single and/or combined fashion and evaluate the potential benefits in terms of prognostic value. The structure of the training and testing, reported below in this section, was the same in all subsets regardless of the input features; the models were named using the same name as the family given in the input features. The outcome investigated was the mortality observed in 78/436 patients.

2.5. Predictive Models

All the analyses were conducted using Python-3 [13], utilizing the scikit-learn libraries [14], imblearn [15], pandas [16], numpy [17], scipy [18], and ELI5 [19], while the plotting was carried out using matplotlib [20] and seaborn [21].

The data were analyzed using Machine Learning (ML) methods, including regression regularized via Least Absolute Shrinkage and Selection Operator (LASSO) [22], the Random Forest classifier [23], or the Fully connected Neural Network (FcNN) [24].

Details regarding the implementation of all the algorithms can be found in the scikit-learn [14] documentation of LASSO cross-validation (LassoCV), Random Forest classifier, and Multi-layer Perceptron Classifier (FcNN Classifier) functions. Lasso CV has been utilized with all the default parameters since they are automatically optimized by means of a built-in cross-validation procedure. The random forest was built using 200 decision trees with balanced class weights, and the FcNN classifier was utilized with $\alpha = 10$, a single hidden layer with five nodes, max number of iterations = 1000, activation function ReLu, and "lbfgs" solver. The RF hyper-parameters were chosen using a parametric scan to explore the main possible combination of values including number of estimators, max depth, max number of features, and oob score. In addition, the impact of dataset dimensionality reduction was also investigated for the RF approach.

Different pre-processing procedures were followed for the different algorithms. Since the present dataset was heavily unbalanced (18% mortality), the Random Forest, which was the most sensitive to imbalances in the dataset, was preceded by a Synthetic Minority Oversampling Technique [25] which created new instances of the minority class using the convex combination of a set of samples in the minority class. The Standard Scaler was used to carry out z-score scaling on all the features before Random Forest implementation.

In the case of LASSO and FcNN, normalization and scaling of the features was achieved using the Box-Cox transformation and the Standard Scaler, respectively. The number of features was reduced by using a threshold of 0.6 in the Spearman correlation. In addition, the single feature which was best correlated with the patient outcome using the Spearman correlation test was re-included in the set of remaining features.

For all the algorithms, evaluation of the models was carried out using a 10-fold cross-validation approach, with stratification with respect to the outcome, to obtain a more realistic evaluation of the model performance, using the "cross-val-predict" scikit-learn function. The data analysis pipeline is represented schematically in Figure 1.

The hyperparameter search for the Lasso was carried out automatically, using an additional stratified 10-fold cross-validation in the training phase. To avoid data leakage, the entire cross-validation procedure was managed using the scikit-learn library.

In all cases, performance was evaluated using the Area Under the Curve (AUC) of the respective Receiver Operator characteristic (ROC) curves as well as sensitivity and specificity.

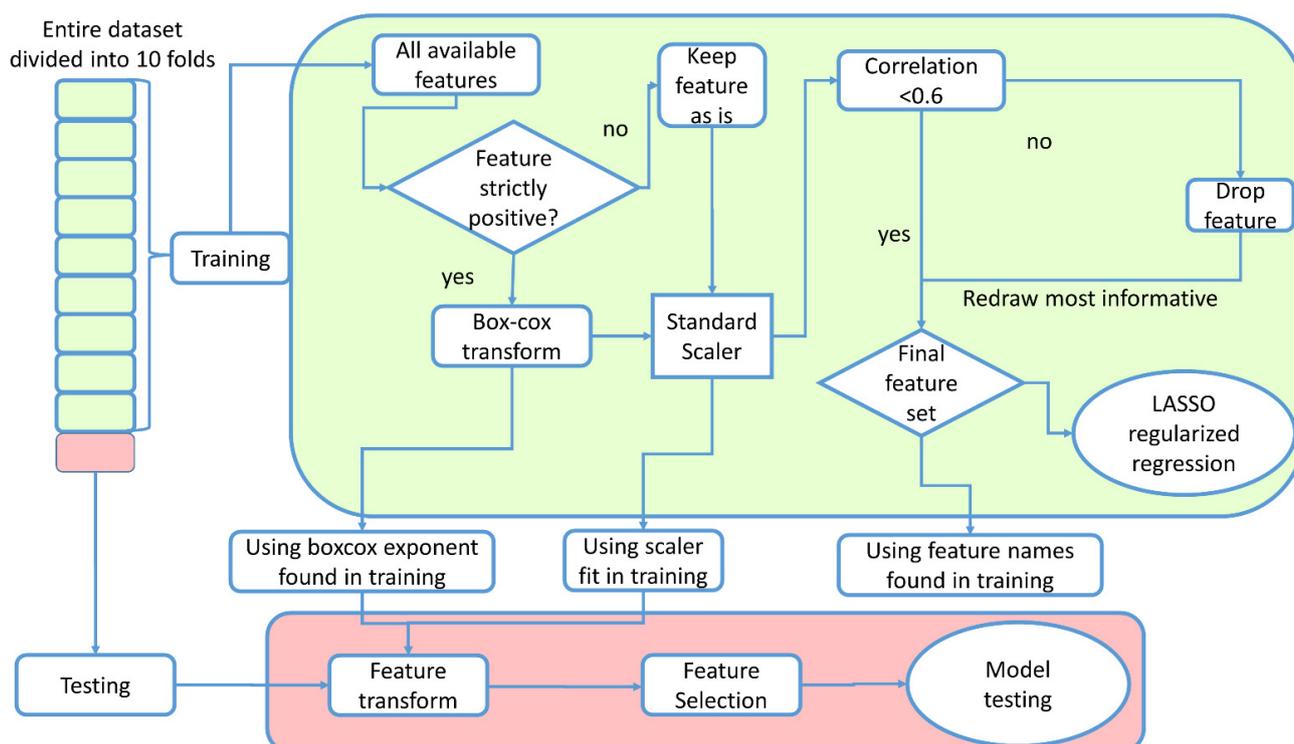


Figure 1. Representation of the 10-fold cross-validation approach used for the training and testing of a single feature group classifier. The input family was selected before entering the green box.

3. Results

The plots in Figures 2, 3 and S1 are ROC curves relative to the LASSO, FcNN, and RFC methods, respectively. In all cases, the fainter lines represent the 10 curves relative to the 10 testing phases. The bold blue line is the average performance, the turquoise bands represent the standard deviation around the mean, and the red line is the performance of a random guesser blindly predicting mortality. The performances of all the developed models are reported in Table 2. To investigate the capabilities of the LASSO model based on all the available features to describe our cohort irrespective of admission rate, the present cohort was divided into two groups according to hospitalization date (before or after 20 July 2020). The AUCs were determined, resulting in 0.73 and 0.76, which were found not to be statistically significantly different in demonstrating the capability of the model to describe the present dataset, irrespective of the wave of belonging. Similar results have been reported in [26] using a semi-quantitative score based on a database including only radiological information.

DeLong's tests were used to compare the ROCs. Without considering the radiological models, only the Lasso clinical and the Lasso radiomic models were statistically different, with a p -value of 0.044.

The relevant features in the Lasso models are reported in Table 3; a graphical representation of the importance of the features in each model is reported in Figure 4.

For the Lasso regularized regression, the importance is expressed by the coefficient of the feature in the linear combination. For the RFC, the importance is the Gini importance built into the implementation of the sklearn function, and for the FcNN, the importance is obtained using a Permutation Importance approach implemented in the ELI5 library (25). It should be noted that the performances, as well as the values of the importance produced by the models, are directly affected by the kind of regularization, or lack thereof, employed in the training. This can also be seen in the performance evaluation of the train dataset, which is obtained as the average over the different folds used for the cross-validation. Regularized models (i.e., LASSO and FcNN) tend to perform better while non-regularized models (i.e., RF) have slightly worse performances. It is also worth mentioning that the lack of balance

in the training labels particularly affects the performance of the RFs, despite the attempts made to reduce these effects.

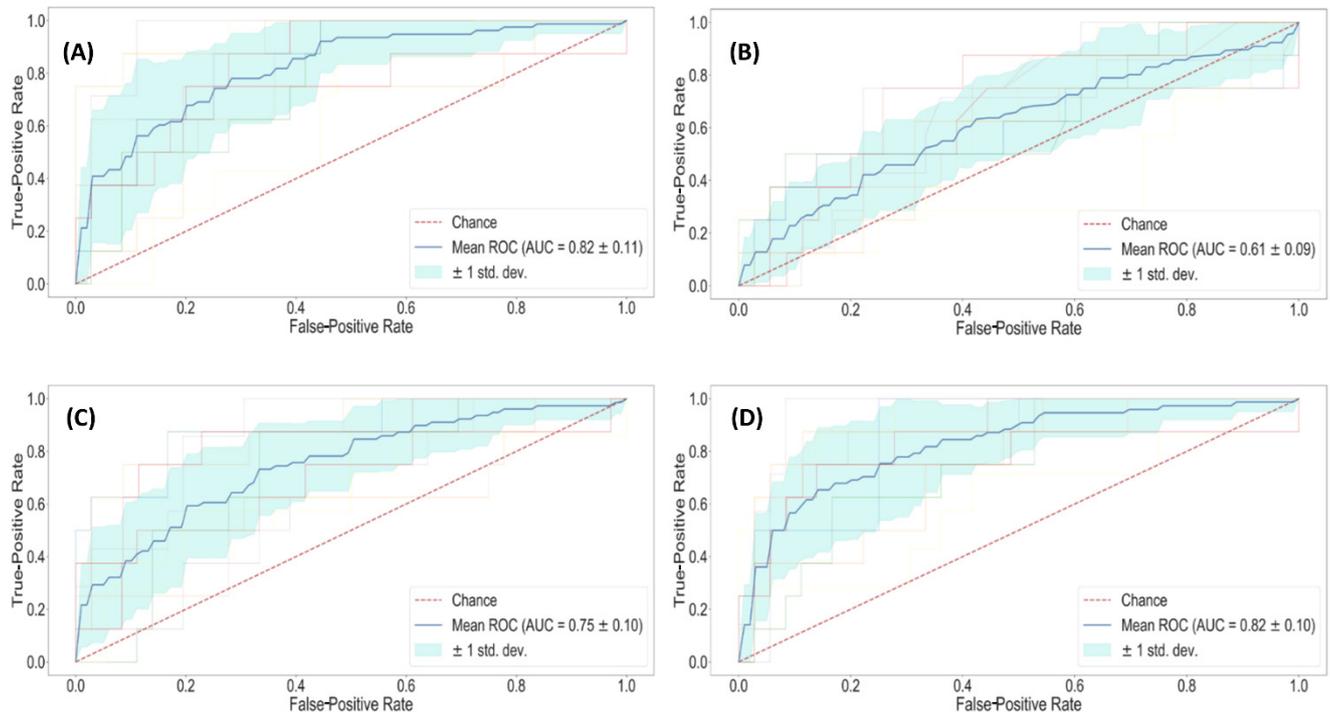


Figure 2. Performance of the Lasso regularized classifier on different input features: (A) Clinical, (B) Radiological, (C) Radiomic, and (D) All available features.

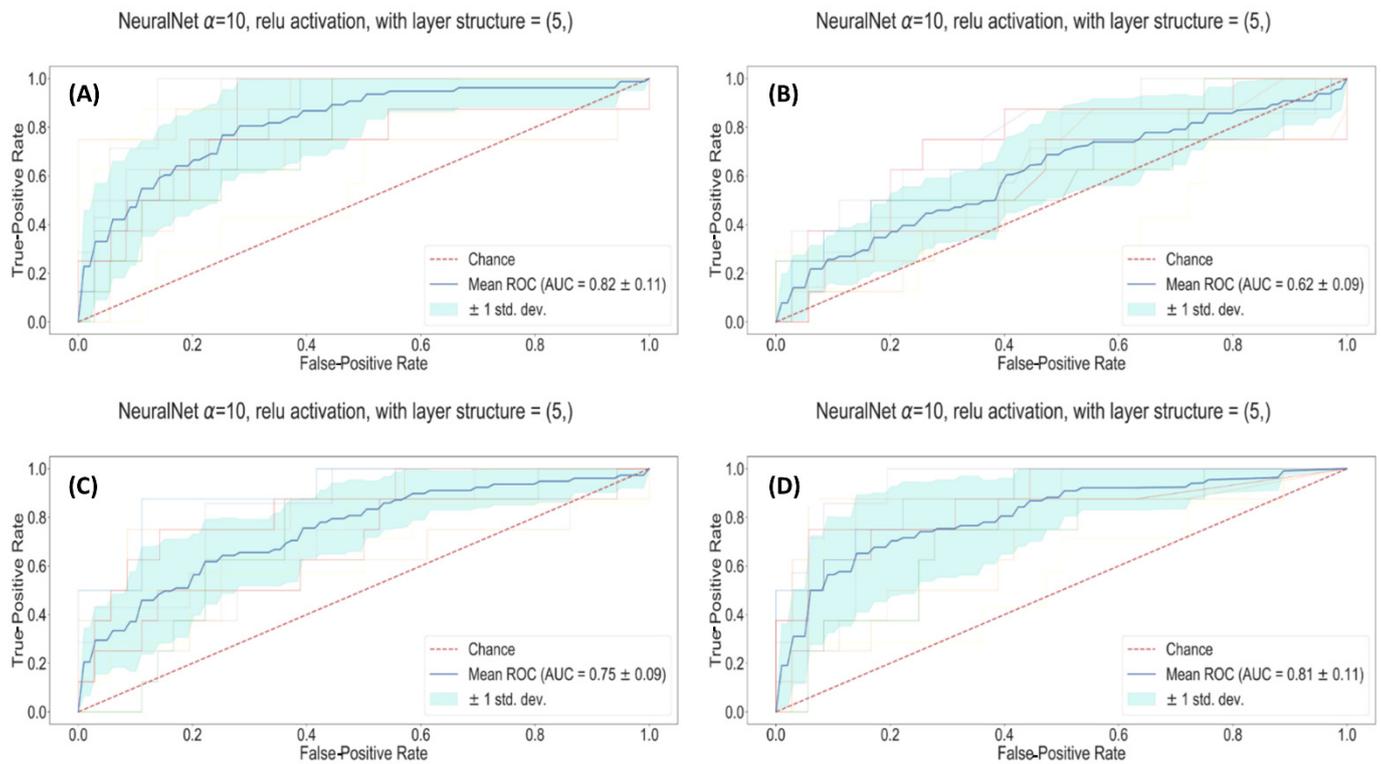


Figure 3. Performance of the FcNN classifier on different input features: (A) Clinical, (B) Radiological, (C) Radiomic, and (D) All available features.

Table 2. Table with all the AUCs obtained using the different models during both training and testing. Sensitivity (Sens) and Specificity (Spec) are relative to the testing phase.

| | Model Name | Training AUC | Testing AUC | Sens | Spec |
|---------------------------------------|--------------|--------------|-------------|------|------|
| Random Forest classifier | Clinical | 0.98 ± 0.01 | 0.63 ± 0.09 | 44% | 83% |
| | Radiomic | 1.00 ± 0.01 | 0.64 ± 0.08 | 41% | 86% |
| | Radiological | 0.93 ± 0.01 | 0.49 ± 0.07 | 19% | 79% |
| | All | 1.00 ± 0.01 | 0.67 ± 0.11 | 44% | 88% |
| Fully connected Neural Network | Clinical | 0.82 ± 0.11 | 0.82 ± 0.01 | 76% | 75% |
| | Radiomic | 0.83 ± 0.09 | 0.75 ± 0.01 | 77% | 64% |
| | Radiological | 0.69 ± 0.09 | 0.62 ± 0.02 | 63% | 56% |
| | All | 0.91 ± 0.11 | 0.81 ± 0.01 | 69% | 83% |
| Lasso regularized classifier | Clinical | 0.84 ± 0.01 | 0.82 ± 0.11 | 69% | 83% |
| | Radiomic | 0.81 ± 0.01 | 0.75 ± 0.10 | 64% | 78% |
| | Radiological | 0.67 ± 0.01 | 0.61 ± 0.09 | 70% | 51% |
| | All | 0.88 ± 0.01 | 0.82 ± 0.10 | 85% | 68% |

Table 3. Set of variables chosen via Lasso regression with respective weights of the linear combination in the tested version of the model. Arranged in descending order of absolute value; in all cases, the Intercept = 0.178899.

| Model Name | Relevant Features (Coefficient) |
|---------------------|--|
| Clinical | Age (years) (0.116771), Respiratory Rate (0.082292), Sex (−0.037591), Fever (−0.022923) |
| Radiomic | 10th intensity percentile (−0.125094), Intensity-based interquartile range (0.103349), Complexity (−0.102924), Cluster prominence (−0.064690), Area density-aligned bounding box (−0.039374), Entropy (0.033002), Number of compartments (GMM) (−0.032441), Asphericity (0.028517), Local intensity peak (0.028478), Global intensity peak (−0.024832), Intensity range (0.012509), Fat surface (0.007267) |
| Radiological | Ground-glass opacity (−0.043875), Lung consolidation (0.038143), X-ray Tube Current (−0.017264), kVp (0.004995) |
| All | Age (years) (0.092963), Intensity-based interquartile range (0.057260), Respiratory Rate (0.049603), Ground-glass opacity (−0.031423), Sex (−0.028895), Complexity (−0.028606), Lung consolidation (0.017272), Fever (−0.016933), X-ray Tube Current (−0.016908), Area density—aligned bounding box (−0.009676), Cluster prominence (−0.006663), Fat surface (0.004984), Number of compartments (GMM) (−0.001448), Local intensity peak (0.000195) |

To clarify the impact of regularization on performance of the RF approach, a dataset of reduced dimensionality obtained using the LASSO approach was implemented. However, the resulting performance in terms of AUC of this second attempt (data not shown) remained very similar to the authors' previous attempt. Thus, the sub-optimal result was likely due to the application of this classifier on a strongly imbalanced dataset [27]. In addition, the RF hyper-parameters were chosen using a parametric scan to explore the main possible combination of values, including number of estimators, max depth, max number of features, and oob score. None of these parameter combinations produced any relevant improvement in the RF models when applied to the dataset being tested.

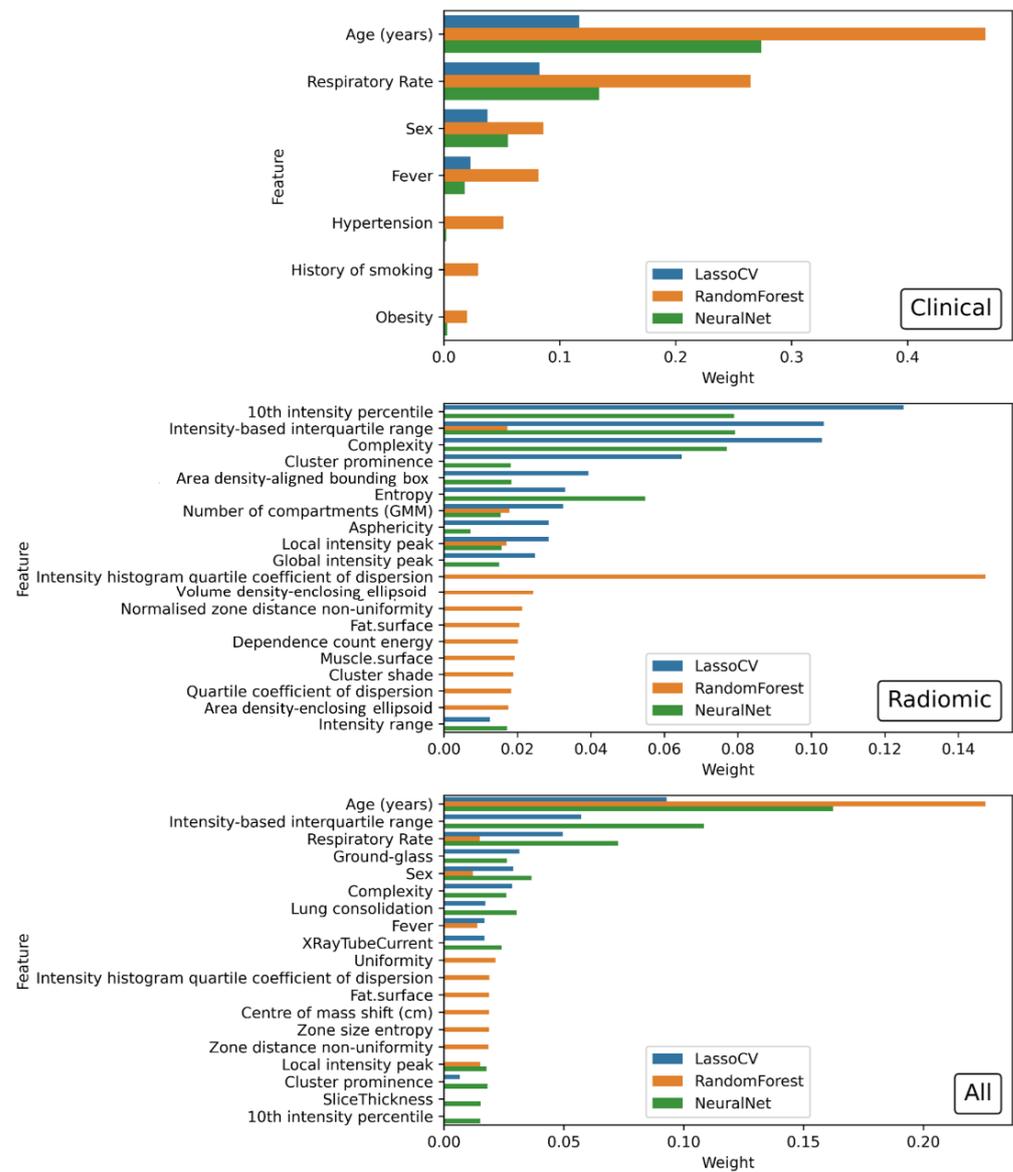


Figure 4. Barplot with the most important features of each model and their respective importance within each model. The barplots for the radiological models are not reported since, in all cases, these did not have any predictive ability at all.

Figure S2 shows an example of how age and SARS-CoV-2 disease affect CT image appearance and grey level inhomogeneities, consequently impacting the values of the radiomic features. One such example is entropy, which did not remain in the final model, being related to patient age. In particular, the entropy values obtained from the images were 8.29, 9.96, 8.22, and 9.97 for the patients illustrated in panels A, B, C, and D, respectively. A and C were both under 70 years of age while B and D were both older. A and B were successfully discharged from the hospital while C and D died from SARS-CoV-2 disease.

These findings suggested investigating the impact of ageing on several relevant radiomic features, as shown in Supplementary Materials Figure S3 (entropy/complexity).

Figure 5 shows the misclassification distribution with respect to patient age (which was, in all cases, the most relevant feature included in the Lasso model). Moreover, from Figure 5, it can be noted that the radiomic and clinical models seem to have different weaknesses while having a slight overlap in patient misclassification.

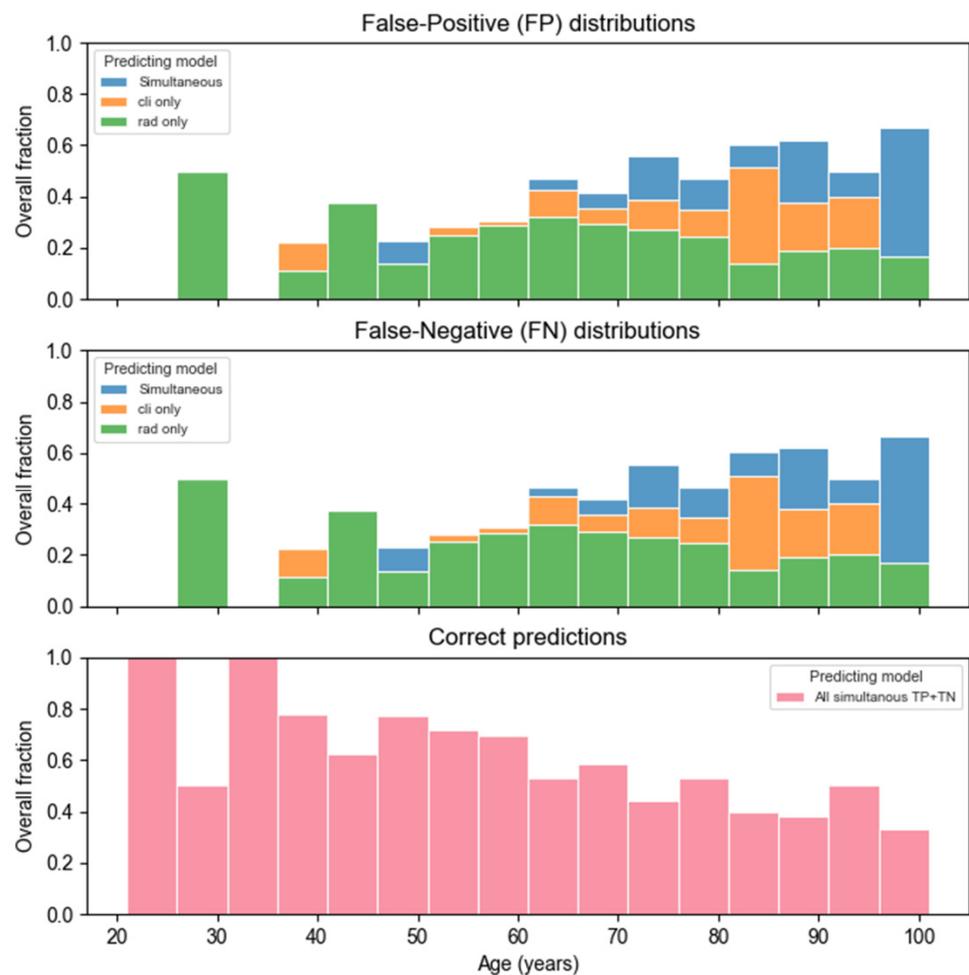


Figure 5. Histogram of different misclassifications individually made by the clinical (Orange), the radiomic (Green), or both simultaneously (Blue) LASSO models. The pink-colored histogram represents the distribution of the simultaneously correct predictions for both the clinical and the radiomic models. False positives are cases in which the model predicts patient mortality and the patient survives; false negatives are cases in which the model predicts that the patient survives, but the patient dies. In all cases, all the bins are normalized with the size of the population in the corresponding age range.

This peculiar behavior suggested that, in the clinical model, the risk for older patients was overestimated (more False Positives) and was somewhat underestimated (more False Negatives) in the younger population, while the opposite was true for the radiomic model. Similar behavior was found for the FcNN classifier, as reported in Supplementary Materials Figure S4.

4. Discussion

The radiological findings extracted from the clinical database and assessed by expert radiologists have, in no instance, proven to be informative regarding the outcome investigated. Correspondingly, in all cases, the radiological model was statistically different from all others as well as the worst performing.

At no point in the analysis did the history as a smoker seem to be relevant within the models in this study, despite what was shown by [8]. This could be due to the present dataset having a high percentage of patients with a smoking history. Although this variable could be indirectly associated with hypertension, respiratory rate, or other clinical variables (i.e., age and sex) in the present dataset, history as a smoker was found not to be correlated above the correlation threshold set at 0.6 before the preprocessing phases.

The set of clinical variables in the present study contained fewer features which attempted to predict prognosis than the majority of those used in the available literature [8,28]. Of note, the clinical model used in the present study had a performance (AUC = 0.82) comparable to that obtained by [28] and slightly worse than that of Shiri et al. [8].

In concordance with what was found in [29], the present model outperformed the radiological assessment obtained by the expert radiologists who took part in the study cited.

As one would have expected from the World Health Organization (WHO) guidelines [30], when included in the present dataset, age was the most relevant variable in the model, followed by respiratory rate and sex.

In this study, three different ML models were investigated in terms of ability to predict the relevant clinical outcome, i.e., death. The combination of the segmentation method with predictive models was chosen with the intent of identifying the most important predictive features while keeping the interpretation of the results as simple as possible and facilitating their application in clinical practice. The authors recognized that Convolutional Neural Networks could be applied to image analysis and segmentation [31]. However, these approaches require large computational power as well as large training datasets and can be of difficult interpretability, often resulting as black boxes [32].

Looking at Figure 4, it can be noted that the features contributing the most to the compared models are the same irrespective of the algorithm adopted which included age (years), respiratory rate, ground glass (GGO), and intensity-based interquartile range.

It should also be noted that the relative importance (weight) of the features in each model was similar in the two models (i.e., Lasso and FcNN) which better described the present cohort. Their performance, as well as the magnitude of the importance estimated, can be attributed to the regularized nature of the methods utilized.

Furthermore, these findings supported the presence of an association between patient outcome, clinical parameters (e.g., age and respiratory rate), and radiological (e.g., GGO) and hidden image properties not noticeable by the human eye but requiring ad hoc computation (e.g., intensity-based interquartile range). All of the above features enabled taking into consideration the deterioration of lung tissues related to SARS-CoV-2 disease as well as the ageing process.

The most relevant radiomic variables in the model used in the present study were related to the Gray-Level distribution and disorder/inhomogeneity in the image (i.e., entropy, complexity, 10th intensity percentile). Some of these features were found in models developed by [28] and were also informative in a univariate analysis carried out by [33].

As expected, looking at the same univariate analysis as in the study of [33], the performance of a more complex model is consistently better than that of a single radiomic variable.

The dimensionality (2D vs. 3D) of the images probably affected performance; in fact, the present models consistently outperformed those obtained using radiographic chest images as in the studies of [6,28,33].

The authors hypothesized that ageing of pulmonary tissue may affect several of the relevant radiomic features left after the LASSO feature reduction, as shown in Supplementary Materials Figure S3. Unfortunately, the current dataset used in the study did not allow discriminating the impact of lung tissue ageing, even when using the Neural Network approach.

Figure S2 highlights how disorder and inhomogeneity in the grey levels are related to damage in the lungs as well as to the age of the patient. To the best of the authors' knowledge, this has not previously been highlighted.

As a final consideration, it is important to note that the semi-automatic segmentation tool significantly reduced human costs in terms of manpower and time with respect to a manual approach. Moreover, the segmentation of a single patient may require from 10 to 60 min when performed manually against the 2–6 min necessary with an automatic tool, depending on computer and software specifics. It is noteworthy that manual segmentation, which is feasible only with small patient cohorts, may achieve a slightly better predictive performance [6,8]. On the other hand, the time utilized by trained radiologists to manually

segment all the chest CT images may be unavailable in a busy department, especially during pandemic events.

Some of the limitations of the present study include the imbalanced nature of the majority of the clinical variables available as well as the reduced number of clinical features available.

However, this may also represent one of the strong points of the study, since it showed that, even with a basic amount of information, it was still possible to obtain acceptable results.

Another similar point is that of the length of time from the date of the CT scan to the outcome. It is a clear limitation since only the first CT was considered, hence concealing all the disease progression after the first scan. However, it showed that it was possible to have a quick and reliable evaluation of patients at admission, allowing better allocation of hospital resources.

Some future prospects in this regard may include an additional analysis of the dataset in a delta-radiomics setting in which disease progression is also included in the patient evaluation by looking at the changes in radiomic features in successive CT scans.

Another interesting prospect would be to additionally investigate the relationship between patient characteristics, such as age, and radiomic variables extracted from various organs.

5. Conclusions

The present study pointed out that semi-automatic segmentation tools allowed the extraction of the radiomic features, which allowed the construction of predictive ML models, having a performance not reaching those obtained using clinical variables but more accurate than the models based on radiological findings. The models developed could provide valuable support to clinicians and radiologists in discerning CT-based RFs representative of the extension and severity of areas affected by SARS-CoV-2.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app12094493/s1>, Figure S1: Performance of the RFC on different input features: (A) Clinical, (B) Radiological, (C) Radiomic, or (D) All available features; Figure S2: CT axial images from different patients. In first column (A,C), the patients are both under 70 years of age while, in the second column (B,D), both are older. Similarly, patients in the same row (A,B and C,D) were alive and deceased at the end of the follow-up, respectively; Figure S3: Behavior of relevant radiomic features (A) Entropy and (B) Complexity as a function of Age. The patients are also represented as a Misclassification group, using the marker in the plot, and by outcome, by color, as Deceased (Red) or Alive (Blue); Figure S4. Histogram of different misclassifications individually made by the clinical model (Orange), the radiomic model (Green), or both simultaneously (Blue) according to the FeNN. The pink-colored histogram represents the distribution of the simultaneously correct predictions for both the clinical and the radiomic models. False positives are cases in which the model predicts patient mortality and the patient survives; false negatives are cases in which the model predicts that the patient survives, but the patient dies. In all cases, all the bins are normalized with the size of the population in the corresponding age range; Table S1: Description of follow-up lengths for various groups of patients in the study.

Author Contributions: Conceptualization, L.S. (Lorenzo Spagnoli), F.C. (Francesca Coppola), P.V., R.G. and L.S. (Lidia Strigari); methodology and formal analysis, L.S. (Lorenzo Spagnoli), M.S., G.P., E.G., N.C. and L.S. (Lidia Strigari); investigation, L.S. (Lorenzo Spagnoli), F.C. (Francesca Coppola), P.V., R.G., E.G., N.C. and L.S. (Lidia Strigari); resources, R.G.; data curation, F.C. (Federica Ciccarese), M.B., G.V., N.B., M.F.M. and F.C. (Francesca Coppola); writing—original draft preparation, L.S. (Lorenzo Spagnoli), E.G., M.F.M., G.P., M.S. and L.S. (Lidia Strigari); writing—review and editing, all authors; supervision, F.C. (Francesca Coppola), E.G. and L.S. (Lidia Strigari); funding acquisition, L.S. (Lidia Strigari) and R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the S. Orsola Polyclinic Foundation (Fondazione Policlinico S. Orsola) of Bologna, Italy. The APC was funded by the Alma Mater Studiorum University of Bologna, Specialization in Radiology Bologna, Italy.

Institutional Review Board Statement: This study regarding the prognostic value of radiomic features was conducted and included all the patients suitable for analysis, according to the guidelines of the Declaration of Helsinki. The study was approved by the Institutional Review Board (or Ethics Committee) of IRCCS University Hospital of Bologna (protocol code No. EM949-2020_507/2020/Oss/AOUBo, approved on: 16 September 2020).

Informed Consent Statement: Written informed consent was obtained from all patients before publishing this paper.

Data Availability Statement: Data will be available after reasonable request to the corresponding author.

Acknowledgments: We would like acknowledge Eng. Stefano Vezzani from the S. Orsola Polyclinic Foundation of Bologna, Italy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Corman, V.M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D.K.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M.L.; et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* **2020**, *25*, 2000045. [[CrossRef](#)] [[PubMed](#)]
2. Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* **2020**, *296*, E32–E40. [[CrossRef](#)] [[PubMed](#)]
3. Revel, M.P.; Parkar, A.P.; Prosch, H.; Silva, M.; Sverzellati, N.; Gleeson, F.; Brady, A. 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.* **2020**, *30*, 4903–4909. [[CrossRef](#)] [[PubMed](#)]
4. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, *295*, 328–338. [[CrossRef](#)]
5. Wang, S.; Dong, D.; Li, L.; Li, H.; Bai, Y.; Hu, Y.; Huang, Y.; Yu, X.; Liu, S.; Qiu, X.; et al. A Deep Learning Radiomics Model to Identify Poor Outcome in COVID-19 Patients With Underlying Health Conditions: A Multicenter Study. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2353–2362. [[CrossRef](#)]
6. Ke, Z.; Li, L.; Wang, L.; Liu, H.; Lu, X.; Zeng, F.; Zha, Y. Radiomics analysis enables fatal outcome prediction for hospitalized patients with coronavirus disease 2019 (COVID-19). *Acta Radiol.* **2021**, *63*, 319–327. [[CrossRef](#)]
7. Xiao, F.; Sun, R.; Sun, W.; Xu, D.; Lan, L.; Li, H.; Liu, H.; Xu, H. Radiomics analysis of chest CT to predict the overall survival for the severe patients of COVID-19 pneumonia. *Phys. Med. Biol.* **2021**, *66*, 10. [[CrossRef](#)]
8. Shiri, I.; Sorouri, M.; Geramifar, P.; Nazari, M.; Abdollahi, M.; Salimi, Y.; Khosravi, B.; Askari, D.; Aghaghazvini, L.; Hajianfar, G.; et al. Machine learning-based prognostic modeling using clinical data and quantitative radiomic features from chest CT images in COVID-19 patients. *Comput. Biol. Med.* **2021**, *132*, 104304. [[CrossRef](#)]
9. Wang, D.; Huang, C.; Bao, S.; Fan, T.; Sun, Z.; Wang, Y.; Jiang, H.; Wang, S. Study on the prognosis predictive model of COVID-19 patients based on CT radiomics. *Sci. Rep.* **2021**, *11*, 11591. [[CrossRef](#)]
10. Li, C.; Dong, D.; Li, L.; Gong, W.; Li, X.; Bai, Y.; Wang, M.; Hu, Z.; Zha, Y.; Tian, J. Classification of Severe and Critical COVID-19 Using Deep Learning and Radiomics. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 3585–3594. [[CrossRef](#)]
11. Gamberini, L.; Coniglio, C.; Cilloni, N.; Semeraro, F.; Moro, F.; Tartaglione, M.; Chiarini, V.; Lupi, C.; Bua, V.; Gordini, G. Remodelling of a regional emergency hub in response to the COVID-19 outbreak in Emilia-Romagna. *Emerg. Med. J.* **2021**, *38*, 308. [[CrossRef](#)] [[PubMed](#)]
12. Bettinelli, A.; Marturano, F.; Avanzo, M.; Loi, E.; Menghi, E.; Mezzenga, E.; Pirrone, G.; Sarnelli, A.; Strigari, L.; Strolin, S.; et al. A Novel Benchmarking Approach to Assess the Agreement among Radiomic Tools. *Radiology* **2022**, 211604. [[CrossRef](#)] [[PubMed](#)]
13. Van Rossum, G.A.D.; Fred, L. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009.
14. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
15. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
16. McKinney, W.A.O. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; Volume 445, pp. 51–56.
17. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)] [[PubMed](#)]
18. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
19. Fan, A.; Jernite, Y.; Perez, E.; Grangier, D.; Weston, J.; Auli, M. ELI5: Long Form Question Answering. *arXiv* **2019**, arXiv:1907.09190.
20. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]

21. Waskom, M.; Botvinnik, O.; O’Kane, D.; Hobson, P.; Lukauskas, S.; Gemperline, D.C.; Augspurger, T.; Halchenko, Y.; Cole, J.B.; Warmenhoven, J.; et al. *Mwaskom/Seaborn: v0.8.1 (September 2017)*; Zenodo: Geneva, Switzerland, 2017.
22. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Society. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
23. Tin Kam, H. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 271, pp. 278–282.
24. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Prentice Hall PTR: Lebanon, IN, USA, 1994.
25. Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, W. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res. (JAIR)* **2002**, *16*, 321–357. [[CrossRef](#)]
26. Balacchi, C.; Brandi, N.; Ciccarese, F.; Coppola, F.; Lucidi, V.; Bartalena, L.; Parmeggiani, A.; Paccapelo, A.; Golfieri, R. Comparing the first and the second waves of COVID-19 in Italy: Differences in epidemiological features and CT findings using a semi-quantitative score. *Emerg. Radiol.* **2021**, *28*, 1055–1061. [[CrossRef](#)] [[PubMed](#)]
27. Boulesteix, A.L.; Janitza, S.; Kruppa, J.; König, I.R. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2012**, *2*, 493–507. [[CrossRef](#)]
28. Bae, J.; Kapse, S.; Singh, G.; Gattu, R.; Ali, S.; Shah, N.; Marshall, C.; Pierce, J.; Phatak, T.; Gupta, A.; et al. Predicting Mechanical Ventilation and Mortality in COVID-19 Using Radiomics and Deep Learning on Chest Radiographs: A Multi-Institutional Study. *Diagnostics* **2021**, *11*, 1812. [[CrossRef](#)] [[PubMed](#)]
29. Homayounieh, F.; Ebrahimian, S.; Babaei, R.; Mobin, H.K.; Zhang, E.; Bizzo, B.C.; Mohseni, I.; Digumarthy, S.R.; Kalra, M.K. CT Radiomics, Radiologists, and Clinical Information in Predicting Outcome of Patients with COVID-19 Pneumonia. *Radiol. Cardiothorac. Imaging* **2020**, *2*, e200322. [[CrossRef](#)]
30. World Health Organization, Coronavirus Disease (COVID-19) Advice for the Public, Mythbusters. Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters> (accessed on 10 March 2022).
31. Manco, L.; Maffei, N.; Strolin, S.; Vichi, S.; Bottazzi, L.; Strigari, L. Basic of machine learning and deep learning in imaging for medical physicists. *Phys. Med.* **2021**, *83*, 194–205. [[CrossRef](#)]
32. Santoro, M.; Strolin, S.; Paolani, G.; Della Gala, G.; Bartoloni, A.; Giacometti, C.; Ammendolia, I.; Morganti, A.G.; Strigari, L. Recent Applications of Artificial Intelligence in Radiotherapy: Where We Are and Beyond. *Appl. Sci.* **2022**, *12*, 3223. [[CrossRef](#)]
33. Varghese, B.A.; Shin, H.; Desai, B.; Gholamrezanezhad, A.; Lei, X.; Perkins, M.; Oberai, A.; Nanda, N.; Cen, S.; Duddalwar, V. Predicting clinical outcomes in COVID-19 using radiomics on chest radiographs. *Br. J. Radiol.* **2021**, *94*, 20210221. [[CrossRef](#)]