Article

Classification Performance for COVID Patient Prognosis from Automatic AI Segmentation—A Single-Center Study

Riccardo Biondi [1,†], Nico Curti [1,†], Francesca Coppola [2,3,*], Enrico Giampieri [1,*], Giulio Vara [2], Michele Bartoletti [4], Arrigo Cattabriga [2], Maria Adriana Cocozza [2], Federica Ciccarese [2], Caterina De Benedittis [2], Laura Cercenelli [1], Barbara Bortolani [1], Emanuela Marcelli [1], Luisa Pierotti [5], Lidia Strigari [5], Pierluigi Viale [4], Rita Golfieri [2] and Gastone Castellani [6]

[1] eDIMESLab, Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, 40138 Bologna, Italy; riccardo.biondi7@unibo.it (R.B.); nico.curti2@unibo.it (N.C.); laura.cercenelli@unibo.it (L.C.); barbara.bortolani@unibo.it (B.B.); emanuela.marcelli@unibo.it (E.M.)

[2] Department of Radiology, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Via Albertoni 15, 40138 Bologna, Italy; giulio.vara@studio.unibo.it (G.V.); arrigo.cattabriga@studio.unibo.it (A.C.); mariaadriana.cocozza@studio.unibo.it (M.A.C.); federica.ciccarese@aosp.bo.it (F.C.); caterina.de@unibo.it (C.D.B.); rita.golfieri@unibo.it (R.G.)

[3] SIRM Foundation, Italian Society of Medical and Interventional Radiology, Via della Signora 2, 20122 Milan, Italy

[4] Infectious Diseases Unit, IRCCS Sant' Orsola-Malpighi Teaching Hospital, 40138 Bologna, Italy; m.bartoletti@unibo.it (M.B.); pierluigi.viale@unibo.it (P.V.)

[5] Department of Medical Physics, IRCCS Azienda Ospedaliero-Universitaria di Bologna, Via Massarenti 9, S. Orsola-Malpighi Hospital, 40138 Bologna, Italy; luisa.pierotti@aosp.bo.it (L.P.); lidia.strigari@aosp.bo.it (L.S.)

[6] Department of Experimental, Diagnostic and Specialty Medicine, University of Bologna, 40138 Bologna, Italy; gastone.castellani@unibo.it

* Correspondence: francesca.coppola@aosp.bo.it (F.C.); enrico.giampieri@unibo.it (E.G.)

† These authors contributed equally to this work.

Abstract: Background: COVID assessment can be performed using the recently developed individual risk score (prediction of severe respiratory failure in hospitalized patients with SARS-COV2 infection, PREDI-CO score) based on High Resolution Computed Tomography. In this study, we evaluated the possibility of automatizing this estimation using semi-supervised AI-based Radiomics, leveraging the possibility of performing non-supervised segmentation of ground-glass areas. Methods: We collected 92 from patients treated in the IRCCS Sant'Orsola-Malpighi Policlinic and public databases; each lung was segmented using a pre-trained AI method; ground-glass opacity was identified using a novel, non-supervised approach; radiomic measurements were collected and used to predict clinically relevant scores, with particular focus on mortality and the PREDI-CO score. We compared the prediction obtained through different machine learning approaches. Results: All the methods obtained a well-balanced accuracy (70%) on the PREDI-CO score but did not obtain satisfying results on other clinical characteristics due to unbalance between the classes. Conclusions: Semi-supervised segmentation, implemented using a combination of non-supervised segmentation and feature extraction, seems to be a viable approach for patient stratification and could be leveraged to train more complex models. This would be useful in a high-demand situation similar to the current pandemic to support gold-standard segmentation for AI training.

Keywords: radiomics; artificial intelligence; machine and deep learning; medical imaging

1. Introduction

Since the beginning of last year, the world has been facing a health emergency, the pandemic caused by the novel Coronavirus, Sars-CoV2. Even up to the present date, many aspects of the physiopathology of the COVID-19 infection are yet to be fully understood. The diagnostic gold standards for Sars-CoV2 are the reverse transcription-polymerase chain

reaction (rt-PCR) and the gene sequencing of sputum, throat swab and lower respiratory tract secretion [1,2]. These tests have several limitations: a limited testing capacity related to insufficient kits or laboratory supplies; a long reporting time, varying from 6 to 48 h; a great variability in sensitivity, ranging from 37% to 71%. To circumvent these limitations, imaging has emerged as an important tool to guide diagnosis, especially in cases of clinical-laboratory discordances.

Imaging protocols directive from public health authorities are heterogeneous: chest radiography is widely used, although it is not accurate in mild or early COVID-19 infection; there is an improving interest in bedside lung ultrasound, but limited experiences are reported in the literature [3]. Among imaging modalities, Computed Tomography (CT) is the most sensitive (60–98%) acquisition technique, but it has low specificity in the early stage of the disease [4]. For this reason, World Health Organization (WHO) and most radiologic societies do not recommend performing screening CT (WHO characterizes COVID-19 as a pandemic—11 March 2020). High-Resolution Computed Tomography (HRCT) proved nonetheless to be a valuable aid to the clinical and epidemiological management of affected patients, especially during shortages of the necessary reagents, long reporting time, and the high operator-dependency [5–7].

There have been controversial publications about the role of chest CT imaging analysis in the diagnosis and management of COVID-19 [8]. The identification of healthy lung chest CTs from pneumonia cases has been deeply investigated in literature [9,10], but the COVID-19 pandemic has posed the non-trivial problem of classifying different pulmonary diseases. Patchy shadows or ground-glass opacities (GGOs) and consolidations (CSs) are not exclusive of COVID-19 but might be also caused by pulmonary edema, bacterial infection, other viral infection, or alveolar hemorrhage [11].

An initial prospective made by Huang et al. [12] on chest CT scans of patients affected by COVID-19 has shown that the examined subjects have a bilateral GGO and CS. The same medical results were also confirmed by other authors [13,14], who posed the basis for the next quantification studies allowing a better characterization of the COVID-19 features. The classification between early-stage patients and progressive phases has been thoroughly investigated [15–18], and all of them lead to the same conclusion: the main COVID-19 features can be difficult to detect in early stages of the disease, and their correct identification is strongly dependent on the radiologist's expertise.

The role of HRCT in COVID-19 infection management is also controversial due to the radiation exposure problem. Initially, the American College of Radiology (ACR) and the Italian Society of Medical and Interventional radiology (SIRM) guidelines did not recommend HRCT in the diagnostic workup, the former referring to 2018 guidelines for acute respiratory illnesses [19–21]. With the increase of available evidence resulting from clinical trials, a prognostic role of HRCT is now emerging, increasing its value beyond the diagnosis [3,22]. Moreover, a study from Ria et al. evaluated the risk-benefit ratio of radiation exposure in COVID-19 patients, stating that HRCT is justified in patients older than 30 years [23].

The application of automated methods to support the clinicians in the analysis of a large amount of data aims to remove the subjectivity of the measurement and improve the time required for the diagnosis formulation [21,24,25]. Machine learning and deep learning models applied in the medical research field are becoming more popular as the basis for clinical decision support systems. Medical image segmentation plays a pivotal role in the automation of these applications since a correct segmentation of anatomical structures is a crucial step to improve the accuracy of the algorithms and to minimize possible confounders.

In 2015, Mansoor et al. [26] proposed a classification of classical image processing segmentation algorithms, where they divided them into four classes: (1) thresholding-based methods [27]; (2) region-based methods [28,29]; (3) shape-based methods [30]; (4) neighboring anatomy [31]. Deep learning models are outperforming all these techniques [32,33].

Deep learning applications have achieved the most significant results also in the COVID-19 literature [24,27,34–36], with extremely high classification performances and low execution time. The drawback of these methods is the demand for labeled data: the training of a deep learning model requires a vast amount of manual (or semi-automatic) labelled data. This is problematic because data annotation is a very time-consuming operation, dependent on the operator experience.

To overcome this issue, many authors, such as Wang et al. [34], proposed a different approach, combining deep learning features with standard machine learning one, proving the efficiency of this synergy. This approach could also increase the ability to explain the model, potentially leading to an improvement in the understanding of the main features of the COVID-19 disease. Wang et al. highlighted the irregularity and heterogeneous intensities of the lung lesion textures as COVID-19 significant features. Concurrently, in non-COVID-19 patients, Wang et al. found stronger uniformity of Hounsfield values in the chest CT within the lesions. Both these features can be automatically quantified by a machine learning algorithm, allowing stratification of the patients' severity and removing possible false positives. Other identified useful characteristics concern the geometry and shape of the lesions [8,37].

For a more detailed stratification of patients, the Radiological Society of North America (RSNA) released a consensus statement, endorsed by the Society of Thoracic Radiology and the American College of Radiology (ACR), that classifies the CT appearance of COVID-19 pneumonia into four categories: "typical," "atypical," "undetermined," and "negative" [38]. The "typical" pattern is characterized by the presence of round-shaped Ground-Glass Opacities (GGO), usually bilateral with a sub-pleural location on the dorsal basal segments. The GGO can be associated with "Crazy Paving" areas or other signs of organizing pneumonia. The "undetermined" pattern is characterized by the absence of the "typical" pattern findings, with diffuse GGO areas with a perihilar or unilateral distribution, with or without consolidated areas. The "atypical" pattern is characterized by either the absence of the "typical" or "undetermined" signs and the presence of lobar consolidations, "tree in bud," smooth thickening of the septa and pleural effusion; in this presentation, no GGO are detectable. The "negative" pattern is characterized by the absence of pathological findings. The "typical" and "negative" patterns have proven to be very accurate in identifying the disease in patients with suspected COVID-19 infection [3].

Many authors have already proved an occasional discordance between HRCT and rt-PCR. There have been observations of patients with a high clinical suspect of COVID-19 supported by epidemiological criteria and imaging, but with negative rt-PCR [39,40]. On the other hand, there was evidence of patients with a positive rt-PCR and suggestive clinical findings, that did not present pathological findings on HRCT [41]. The clinico-radiological dissociation in asymptomatic individuals requires reconsidering the role of radiological findings in the clinical management of these patients [42].

To improve the reliability of radiological examination, several authors presented Texture Analysis of the CT scans as a valuable tool to aid the diagnosis [43–45] and to identify clinically severe patients [46]. Texture Analysis can identify putative features that are not part of the RSNA criteria, such as enlargement of pulmonary vessels [47–51], and that could be overlooked during the human visual inspection, such as fine characteristics of the GGO areas [52]. Currently, only a few methods exist to automatically process HRCT scans and quantify the extent of the pulmonary involvement [32,53]. From the clinical point of view, the disease can be assessed with the newly developed PREDI-CO (prediction of severe respiratory failure in hospitalized patients with SARS-COV2 infection), which considers clinical parameters predictive of severe respiratory failure in hospitalized patients, and is defined as the sum of the following conditions:

- age $\geq$ 70 years
- obesity BMI $\geq$ 30 kg/m^2
- fever at hospitalization $\geq$ 38 $^\circ$C
- respiratory rate $\geq$ 22 breaths/minute

- lymphocyte count ≤ 900 cells/mm^3
- creatinine ≥ 1 mg/dl
- C-reactive protein ≥ 10 mg/dl
- lactate dehydrogenase ≥ 350 IU/l.

This score outperforms in the stratification of patients the well-established qSOFA, SOFA, CURB-65, and MEWS scores; this is due to the fact that the PREDI-CO score was designed and validated ad-hoc for this purpose [54].

In the current work, we aim to automatize the evaluation of the PREDI-CO score and several radiomic features using a novel, non-supervised image processing pipeline.

2. Materials and Methods

2.1. Patients Selection

This study involves 92 CT scans of patients affected by COVID-19. 10 of these scans come from the public dataset "COVID-19 CT Lung and Infection Segmentation Dataset" published on Zenodo [55]. Left lung, right lung, and infections are labeled by two radiologists and verified by an expert radiologist (with more than 10 years of experience). These scans were used to validate the segmentation performances of the implemented pipeline.

Department of Diagnostic and Preventive Medicine of the IRCCS Policlinic Sant'Orsola-Malpighi provided the remaining 82 scans. The selected patients are composed for the 66.3% by male, the age distribution (min/median/max) is 35/60/89. For each patient, experts ascertained the presence of ground-glass opacities (100%), consolidation (10%), crazy paving (53%), multifocal GGO (multiple locations affected by GGO, 32%), peripheral GGO (presence of GGO areas exclusively far away from the trachea, 23%), and roundish GGO (GGO characterized by round regular shapes, 12%). Moreover, each patient was assigned the PREDI-CO score value estimated by two radiologists.

In IRCCS Policlinic Sant'Orsola-Malpighi, HRCT exams were performed with the following parameters: two different Multi-Slices CT (64 slices, GE VCT or PHILIPS Ingenuity), with keV range 100–120, tube current modulation with a low Quality Index to optimize patient dose, slice thickness range 1–2 mm; images were reconstructed with high-resolution kernel. The CT parameters used in the Zenodo dataset are not available.

2.2. Pipeline Overview

The workflow developed in this work as show in Figure 1 can be split into 3 steps: (1) the segmentation of the lungs; (2) segmentation of the GGO areas; (3) estimation of the radiomic features.
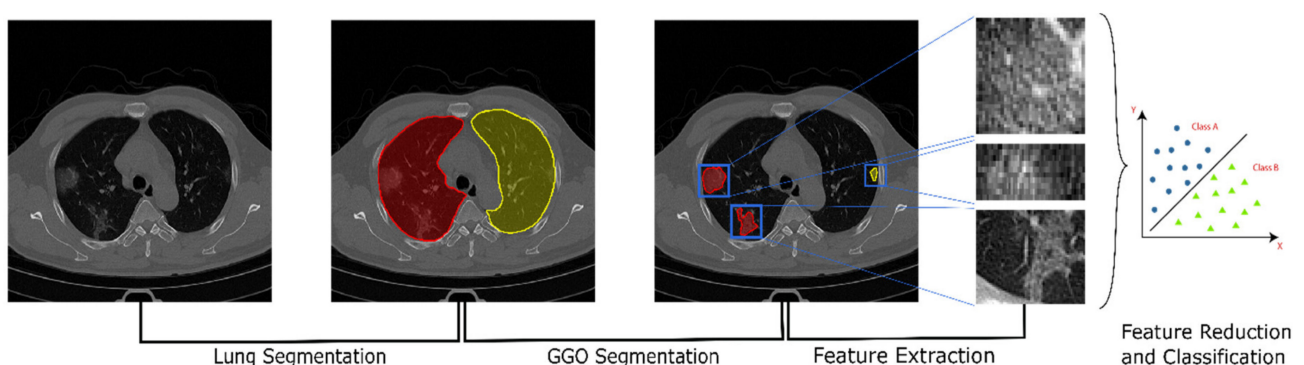

Figure 1. A schematic representation of the proposed pipeline. From the left: raw CT scan; segmentation of the two lungs using a pre-trained U-Net model; segmentation of the GGO areas using k-means clustering; extraction of radiomic and Haralick features; classification and prediction of the clinical characteristics and outcomes.

2.3. Lung Segmentation

Lung segmentation is a pivotal pre-processing step in many image analyses, such as identification and classification of pathologies. Rule-based approach, like thresholding, region growing, etc., usually fails for CT scans of patients with severe Interstitial Lung Disease (ILD) [33]. For this reason, we used a pre-trained publicly available https://github.com/JoHof/lungmaskv0.24 (accessed on 24 March 2021) U-Net model [33] for lung segmentation.

2.4. GGO Segmentation

In the second step, a novel automated pipeline for the segmentation of GGO areas was developed, combining 2 different techniques: vessel artifacts exclusion and k-means clustering. Both these methods are unsupervised learning techniques and have been chosen due to the limited number of available samples. The decision to avoid supervised methods such as Convolutional Neural Networks (CNNs) is based on the likelihood of including strong biases, even including possible data augmentation strategies.

The intensity of pulmonary vessels is very similar to solid components of GGO; therefore, it affects the segmentation, introducing potential false positives [56]. To remove these vessels, we used the vesselness measure, i.e., the presence of multiscale tubular structures. Multiscale Vessel Enhancement Filtering (MVEF) [57] is defined as the likelihood of an image region to contain vessels, and it is estimated using the Frangi filter [58]. The areas with high values of MVEF were identified as vessels and therefore removed from the lung region obtained in the previous step.

After the removal of the vessels, we identified the GGO as areas with a common color texture. To identify these regions, we used k-means clustering, grouping voxels by color and texture similarity, and identifying the tissue corresponding to each cluster [59].

Since GGO involves extended areas, it is informative to include neighborhood voxel information. The color contrast between GGO and healthy areas may change between patients; it is, therefore, useful to consider different gamma corrections of the image. We applied a series of image processing filters to obtain a high-dimensional feature space, including all these features for each individual voxel. For each voxel, we estimated a vector of features obtained by the application of the following filters:

- Gamma corrected image ($\gamma = 1.5$);
- Adaptive Histogram Equalized image, in a radius of 3 voxels;
- Median blurred image with a kernel of radius 3 voxels;
- Standard deviation filtered image with a kernel of radius 1 voxels.

We used the Adaptive Histogram Equalization algorithm for the image standardization: for each slice, the histogram was equalized considering a volume of 3 voxels. The gamma correction is a non-linear operation used to decode the luminance and enhance the low contrast regions. The median blurring allows considering the information about the neighborhood voxels, reducing the effect of outlier voxels. The last feature is the application of a local standard deviation filter; this filter replaces each voxel value with the standard deviation of its neighborhood. This feature is useful as GGO is characterized by highly heterogeneous gray level regions, allowing to filter out bronchial structures and motion artifacts not removed during the lung segmentation.

The k-means clustering is "isotropic" in all directions of space and therefore tends to produce round (rather than elongated) clusters. In this situation, leaving variances unequal is equivalent to put more weight on variables with smaller variance. To avoid this, the features were standardized according to the mean and standard deviation estimated on the training dataset.

We selected 10 scans and applied the k-means clustering for the estimation of the centroids using the Euclidean metric. The k-means clustering is sensitive to the class balance in the training phase (it might give more prominence to more common present structures). Therefore, for the training phase, we considered only a subset of scans with a reasonable amount of GGO areas, excluding in all the cases the (overrepresented) image

background. The k-means cardinality was estimated based on lung anatomy considerations. The resulting clusters were:

- Healthy lung;
- Edges;
- Remaining vessels;
- Noise;
- GGO.

We implemented the whole pipeline using Python, and the source code is publicly available on GitHub (https://github.com/RiccardoBiondi/segmentation, accessed on 24 March 2021). We used SimpleITK [60,61] for the implementation and management of image filters and the OpenCV [62] library for the implementation of the k-means clustering.

We estimated the performances of our segmentation algorithm according to the following scores:

- Sensitivity

$$TPR = \frac{TP}{TP + FN}$$

- Specificity

$$TNR = \frac{TN}{TN + FP}$$

- Precision

$$PPV = \frac{TP}{TP + FP}$$

- F_1 score

$$F_1 = \frac{2TP}{2TP + FP + FN}$$

where *TP, TN, FP*, and *FN* are the True Positives, True Negative, False Positive, and False Negative scores, respectively.

2.5. Feature Extraction

We applied the proposed pipeline on the 82 patients provided by the Department of Diagnostic and Preventive Medicine of the IRCCS Policlinic Sant'Orsola-Malpighi of Bologna. The radiomic features extraction was performed on the identified GGO areas. The extracted features include morphological and texture-based scores:

- Texture;
- Gray Level Distribution;
- GGO Shape;
- Bilaterality (distribution of GGO between left and right lung);
- Peripherality;
- GGO volume

For the scores classification, we included the patient's age as informative features.

We measured the texture properties by computing the Haralick features (Energy, Inertia, Entropy, Inverse Difference Moment, Cluster Shade and Cluster Prominence) from the Gray Level Co-occurrence Matrix (GLCM), computed on the whole identified area [63]. For each identified GGO area, we computed its elongation and roundness [64], obtaining the corresponding distribution for each patient. We computed the distribution of the distances between the lesion and lung centroids, normalized to the semiaxis of the equivalent ellipsoid as a measure of the GGO peripherality. Each distribution was characterized by the minimum, maximum, median, interquartile range (25–75), skewness,

and kurtosis. We estimated the bilaterality distribution using the Matthews coefficient (MCC) defined as follows:

$$MCC = \frac{LGV \cdot LLV - RGV \cdot RLV}{(LGV + RGV)(RLV + LGV)(LLV + RGV)(LLV + RLV)}$$

being RGV = Volume of GGO in the right lung, LGV = Volume of GGO in the left lung, LLV = Left Lung Volume, and RGV = Right Lung Volume. The volume of GGO was normalized according to the total lung volume to overcome possible issues related to anatomical differences between patients.

2.6. Classification

We used the whole set of extracted radiomic features to predict the following GGO characteristics estimated by the expert radiologists:

- Multifocal GGO;
- Presence of Crazy Paving;
- Presence of Consolidation;
- Roundish GGO;
- Peripheral GGO;

Additionally, the two clinical outcomes:

- PREDI-CO score;
- Patient survival.

Not all the above scores were available for all the patients. The GGO characteristics were reported for only 78/82 patients, while the clinical outcomes for only 72 of them. We restricted the score classification on these two subsets of data.

The considered scores are all binary values (True/False), representing the presence/absence of the corresponding characteristic. The only exception is given by the PREDI-CO score, which, by definition, allows an incremental set of values: PREDI-CO score values range from 0 (minimal risk) to 9 (maximal risk). The 47% of patients report a PREDI-CO score of 0 or 1, leading to an overrepresentation of these 2 labels. We grouped multiple labels into the same class to overcome this issue: we applied the cutoff of 1 to dichotomize the PREDI-CO score values into two classes.

We applied a feature selection procedure to filter out redundant and non-informative values for each classification. This step is required to improve the classification performances (remotion of possible confounders) and to make the obtained results more interpretable from a clinical point of view. The selection was performed using a Fisher Exact and a χ^2 tests selecting the 3 features with the highest significance (lower p-values). Both the tests require categorical data, so we dichotomized the features according to their medians.

We used the set of filtered features as input to 4 different classification algorithms to predict the various GGO characteristics and clinical outcomes:

- Logistic Classifier (L1 penalty);
- (Logistic) Ridge Classifier (L2 penalty);
- K-Nearest Neighbors;
- Random Forest Classifier.

The performances of these 4 methods give us an insight into the structure of the data: KNN methods are strongly local, Random Forest relies on the separability of the samples but does not include co-linearities between variables, and linear models (Logistic classifier and Ridge classifier) strongly rely on linear dependencies between the observed values and the predictions.

The classification performances were estimated according to the following metrics:

- Precision
- Sensitivity

- F$_1$ score
- Balanced accuracy

$$BA = \frac{Sensitity + Specificity}{2}$$

using a leave-one-out cross-validation strategy.

The numerosity of the labels of each characteristic is strongly unbalanced (e.g., only 10% of patients show the presence of consolidations) for every characteristic, except for Crazy Paving and PREDI-CO score (after the dichotomization). To compensate for this, we weighted the classification performances of each algorithm according to the inverse of the class frequency.

We used the scikit-learn [65] Python package for the implementation of all the analyses.

3. Results

3.1. GGO Segmentation

We applied the proposed segmentation pipeline on each patient under analysis. For samples, we collected also a manual segmentation performed by an expert radiologist, which was used as the gold standard.

The collected results were evaluated using the previously introduced metric scores and their distribution analyzed as show in Table 1, details in Table S1.

Table 1. A comparison between the score results for the gold standard segmentation in the included databases. For each score, the average value (and corresponding standard deviation at 1 σ) is reported. For each column the maximum value was indicated with bold font.

Cases	Specificity	Sensitivity	Precision	F$_1$ Score
CORONACASES OVERALL	0.9992 ± 0.0005	0.62 ± 0.13	**0.79 ± 0.12**	**0.67 ± 0.07**
GOLD STD OVERALL	**0.9993 ± 0.0003**	**0.74 ± 0.14**	0.67 ± 0.28	0.65 ± 0.18
OVERALL	0.9992 ± 0.0005	0.66 ± 0.15	0.75 ± 0.20	0.67 ± 0.12

We show in Figure 2 the distribution of the individual scores (Sensitivity, Specificity, Precision, and F$_1$ score). The scores were obtained by the average of the whole 15 scans (overall) on the 10 corona cases (CORONACASES OVERALL) and on the five gold standard (GOLD STD OVERALL). In Figures 3 and 4, we show a visual comparison between the achieved segmentation and the ground truth labels (both for corona cases and the gold standard).
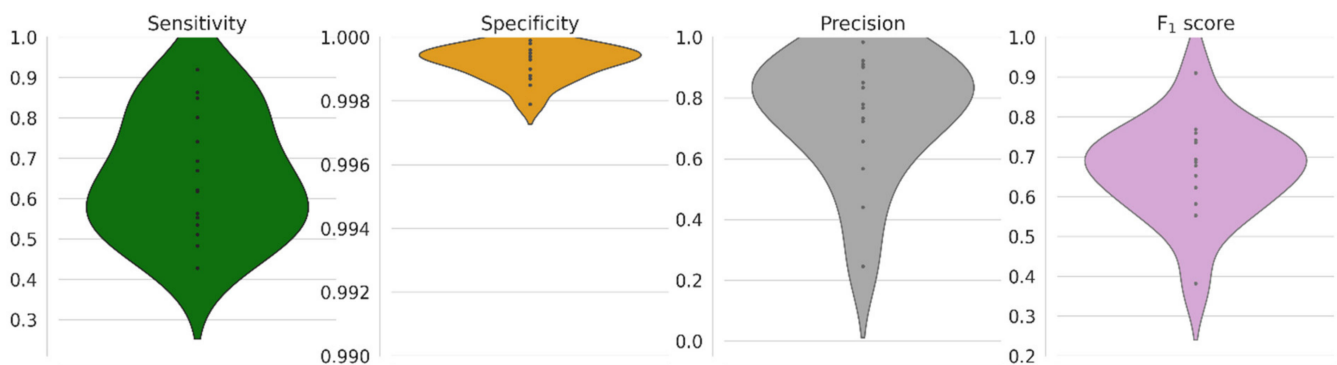
Figure 2. The distribution of segmentation scores obtained by the proposed pipeline. From the left: distribution of Sensitivity, i.e., True Positive rate, Specificity, i.e., True negative rate, Precision, i.e., Positive Predictive Value, and F$_1$ score, i.e., the harmonic mean of precision and sensitivity.
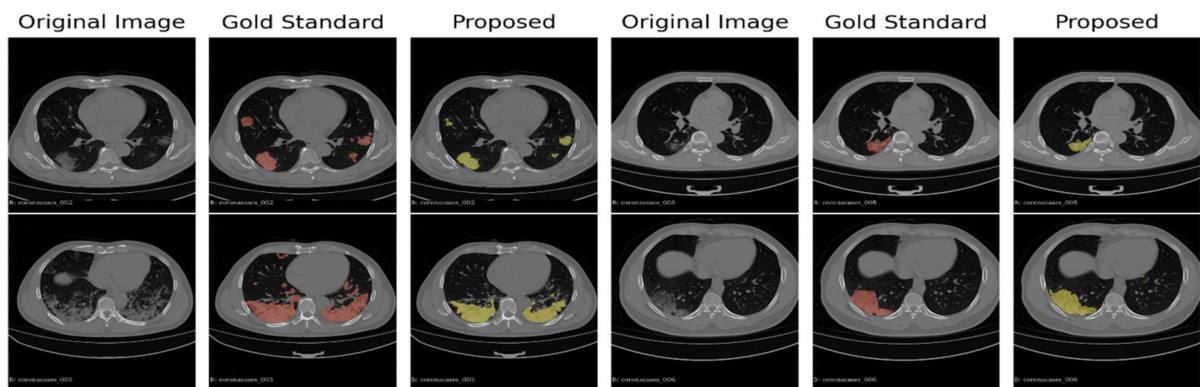
Figure 3. A comparison between the ground truth and the results obtained by the proposed pipeline for the corona cases segmentation. In green are highlighted the GGO areas identified by the experts, and in red, those identified by our segmentation pipeline, respectively.
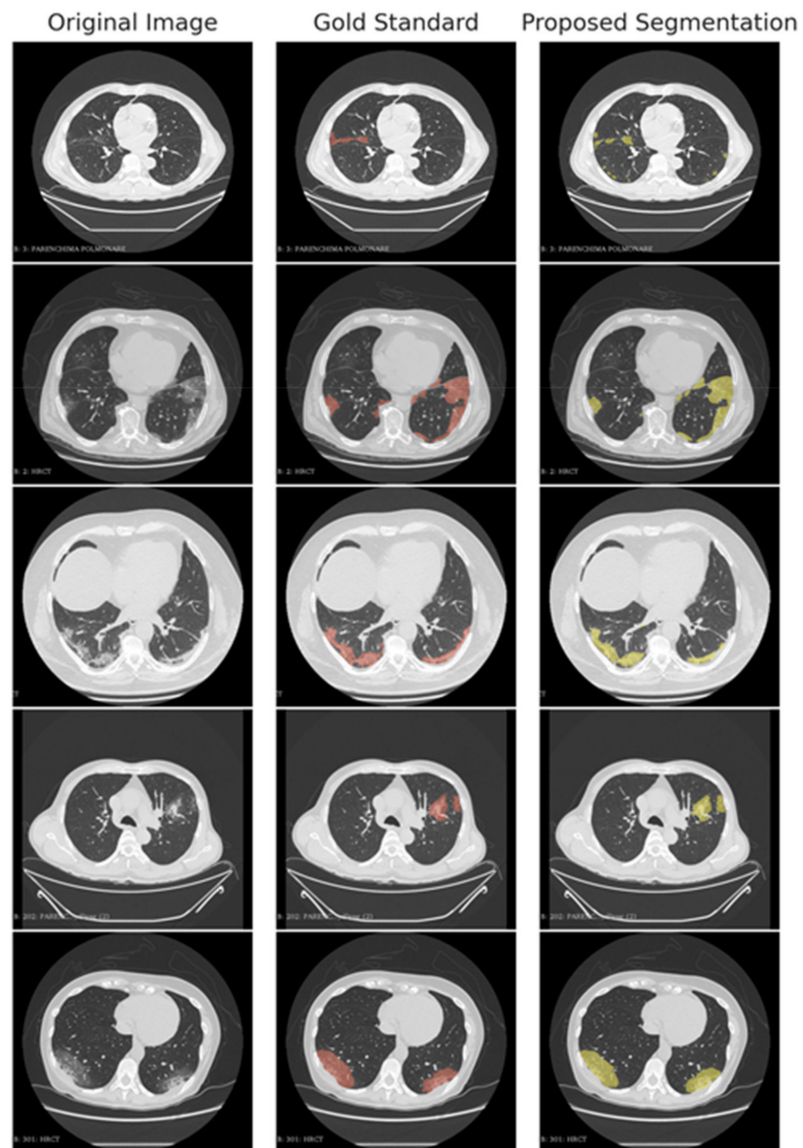
Figure 4. A comparison between the proposed automated segmentation pipeline and a gold standard segmentation manually performed by an expert radiologist.

In Table 2 we report the results of other published pipelines in comparison with our method (details of the employed methods in Table S2). Notice that Jun2020 is a benchmark database for COVID-19 annotation methods and CT scans segmentation (https://gitee.com/junma11/COVID-19-CT-Seg-Benchmark, accessed on 24 March 2020). The evaluation set for both Jun2020 and Muller2020 is COVID-19-CT-Seg, which is the database containing both corona case studies (the one used for annotation) and radiopedia ones (removed because rescaled on 8-bit GL images, which is not compatible with the implemented pipeline). For each technique on each database, only the best results presented in the literature are reported.

Table 2. A comparison between the results of various segmentation techniques applied on the same datasets. For each column the maximum value was indicated with bold font.

Study	Technique	F_1 Score	Sensitivity	Specificity	Precision
Fan2020 [66]	InfNet	0.579	0.870	0.974	0.500
Fan2020 [66]	SemiInfNet	0.597	0.865	0.977	**0.915**
Muller2020 [49]	U-Net	0.761	0.739	**0.999**	-
Jun2020 [52]	3D U-Net	67.3 ± 22.3	-	-	-
Jun2020 [52]	2D U-Net	60.9 ± 24.5	-	-	-
Qingsen2020 [65]	U-Net	**0.726**	**0.751**	-	0.726

3.2. Feature Extraction

We applied the feature extraction step on the GGO areas identified by our segmentation algorithm. In Figure 5 we show the Pearson's correlation matrix between each pair of observed features. The cluster plot highlights the existence of multiple groups of strongly correlated features. The first set of clusters are given by the features related to the roundness, elongation, and distance features. The most prominent cluster is composed of the Haralick features: energy, entropy, inertia, and cluster prominence.

In Figure 6 we show the relation between feature distributions and labels. Using the median of the feature distributions as a threshold, we estimated the percentage of the patients associated with each label who are above this threshold. A result of 0.5 (white-like cells) indicates non-specificity of the variable for that individual label, i.e., a uniform distribution of the feature. A result greater than 0.5 (red-like cells) or smaller than 0.5 (blue-like cells) indicates a high/low percentage of patients for whom the feature values are greater/smaller than the median of the distribution, respectively. This representation allows a visual analysis for the identification and selection of the most informative features related to each label.

Figure 5. The Correlation Matrix between the estimated features. We report for each pair of features the Pearson's correlation coefficient. In red are highlighted the positive correlations, while in blue the negative ones. White-like colors identify the features with low correlation.

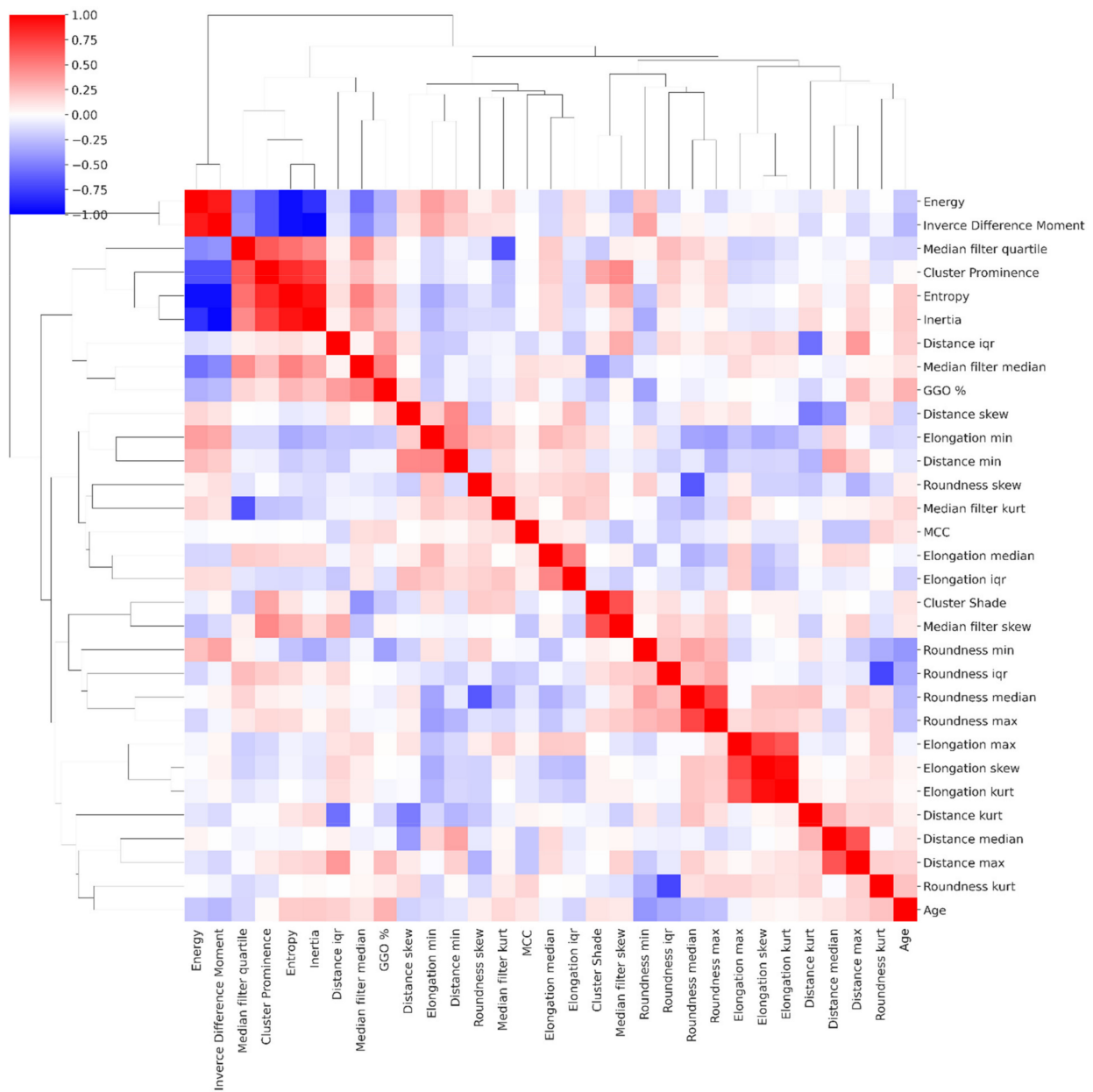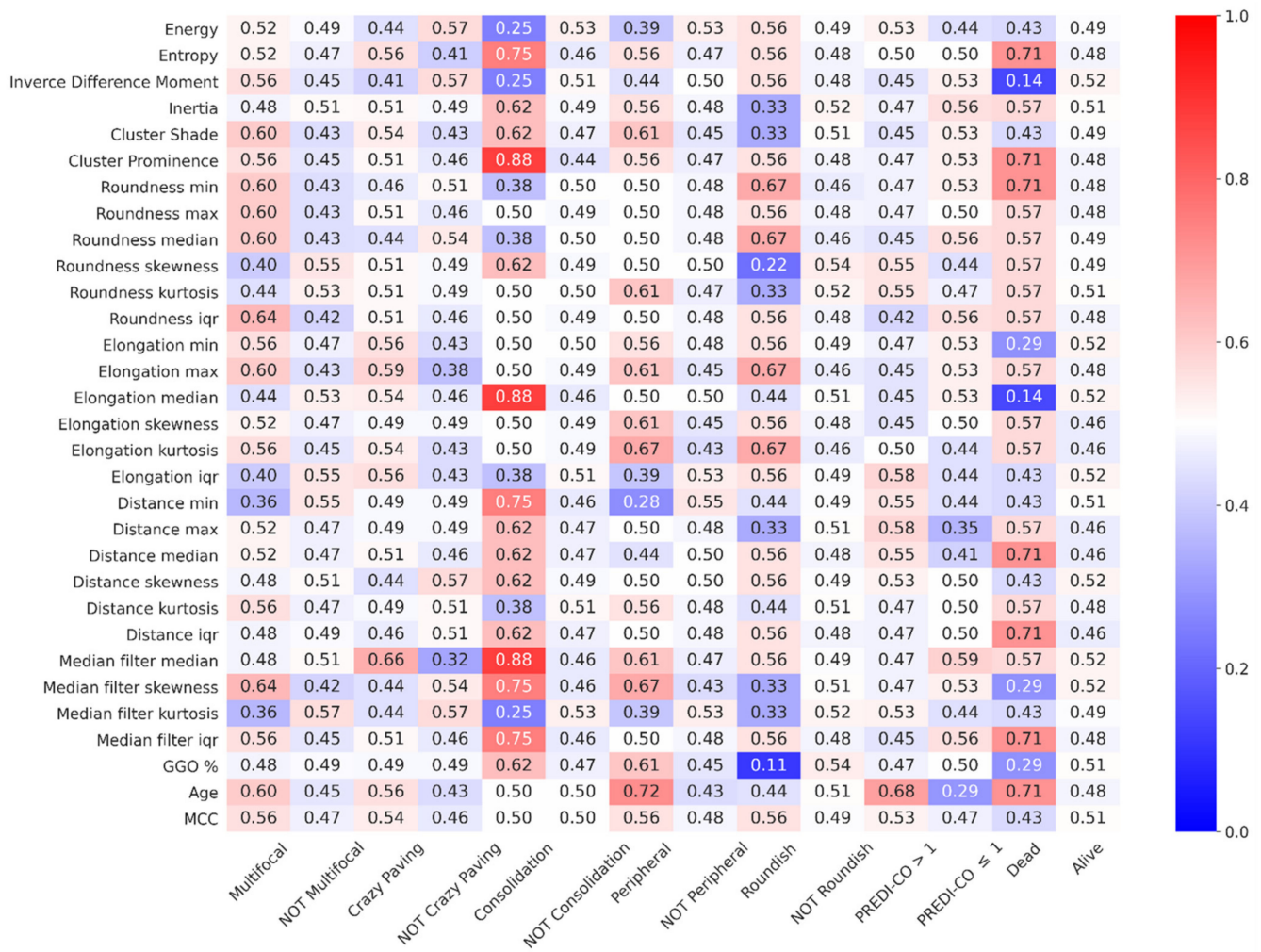|	Multifocal	NOT Multifocal	Crazy Paving	NOT Crazy Paving	Consolidation	NOT Consolidation	Peripheral	NOT Peripheral	Roundish	NOT Roundish	PREDI-CO > 1	PREDI-CO ≤ 1	Dead	Alive
Energy	0.52	0.49	0.44	0.57	0.25	0.53	0.39	0.53	0.56	0.49	0.53	0.44	0.43	0.49
Entropy	0.52	0.47	0.56	0.41	0.75	0.46	0.56	0.47	0.56	0.48	0.50	0.50	0.71	0.48
Inverce Difference Moment	0.56	0.45	0.41	0.57	0.25	0.51	0.44	0.50	0.56	0.48	0.45	0.53	0.14	0.52
Inertia	0.48	0.51	0.51	0.49	0.62	0.49	0.56	0.48	0.33	0.52	0.47	0.56	0.57	0.51
Cluster Shade	0.60	0.43	0.54	0.43	0.62	0.47	0.61	0.45	0.33	0.51	0.45	0.53	0.43	0.49
Cluster Prominence	0.56	0.45	0.51	0.46	0.88	0.44	0.56	0.47	0.56	0.48	0.47	0.53	0.71	0.48
Roundness min	0.60	0.43	0.46	0.51	0.38	0.50	0.50	0.48	0.67	0.46	0.47	0.53	0.71	0.48
Roundness max	0.60	0.43	0.51	0.46	0.50	0.49	0.50	0.48	0.56	0.48	0.47	0.50	0.57	0.48
Roundness median	0.60	0.43	0.44	0.54	0.38	0.50	0.50	0.48	0.67	0.46	0.45	0.56	0.57	0.49
Roundness skewness	0.40	0.55	0.51	0.49	0.62	0.49	0.50	0.50	0.22	0.54	0.55	0.44	0.57	0.49
Roundness kurtosis	0.44	0.53	0.51	0.49	0.50	0.50	0.61	0.47	0.33	0.52	0.55	0.47	0.57	0.51
Roundness iqr	0.64	0.42	0.51	0.46	0.50	0.49	0.50	0.48	0.56	0.48	0.42	0.56	0.57	0.48
Elongation min	0.56	0.47	0.56	0.43	0.50	0.50	0.56	0.48	0.56	0.49	0.47	0.53	0.29	0.52
Elongation max	0.60	0.43	0.59	0.38	0.50	0.49	0.61	0.45	0.67	0.46	0.45	0.53	0.57	0.48
Elongation median	0.44	0.53	0.54	0.46	0.88	0.46	0.50	0.50	0.44	0.51	0.45	0.53	0.14	0.52
Elongation skewness	0.52	0.47	0.49	0.49	0.50	0.49	0.61	0.45	0.56	0.48	0.45	0.50	0.57	0.46
Elongation kurtosis	0.56	0.45	0.54	0.43	0.50	0.49	0.67	0.43	0.67	0.46	0.50	0.44	0.57	0.46
Elongation iqr	0.40	0.55	0.56	0.43	0.38	0.51	0.39	0.53	0.56	0.49	0.58	0.44	0.43	0.52
Distance min	0.36	0.55	0.49	0.49	0.75	0.46	0.28	0.55	0.44	0.49	0.55	0.44	0.43	0.51
Distance max	0.52	0.47	0.49	0.49	0.62	0.47	0.50	0.48	0.33	0.51	0.58	0.35	0.57	0.46
Distance median	0.52	0.47	0.51	0.46	0.62	0.47	0.44	0.50	0.56	0.48	0.55	0.41	0.71	0.46
Distance skewness	0.48	0.51	0.44	0.57	0.62	0.49	0.50	0.50	0.56	0.49	0.53	0.50	0.43	0.52
Distance kurtosis	0.56	0.47	0.49	0.51	0.38	0.51	0.56	0.48	0.44	0.51	0.47	0.50	0.57	0.48
Distance iqr	0.48	0.49	0.46	0.51	0.62	0.47	0.50	0.48	0.56	0.48	0.47	0.50	0.71	0.46
Median filter median	0.48	0.51	0.66	0.32	0.88	0.46	0.61	0.47	0.56	0.49	0.47	0.59	0.57	0.52
Median filter skewness	0.64	0.42	0.44	0.54	0.75	0.46	0.67	0.43	0.33	0.51	0.47	0.53	0.29	0.52
Median filter kurtosis	0.36	0.57	0.44	0.57	0.25	0.53	0.39	0.53	0.33	0.52	0.53	0.44	0.43	0.49
Median filter iqr	0.56	0.45	0.51	0.46	0.75	0.46	0.50	0.48	0.56	0.48	0.45	0.56	0.71	0.48
GGO %	0.48	0.49	0.49	0.49	0.62	0.47	0.61	0.45	0.11	0.54	0.47	0.50	0.29	0.51
Age	0.60	0.45	0.56	0.43	0.50	0.50	0.72	0.43	0.44	0.51	0.68	0.29	0.71	0.48
MCC	0.56	0.47	0.54	0.46	0.50	0.50	0.56	0.48	0.56	0.49	0.53	0.47	0.43	0.51

Figure 6. The relationship between individual features and labels. The values range from 0 (all the feature values are lower than the median of the sample distribution) to 1 (all the values are greater than the median of the sample distribution). A value of 0.5 indicates the non-specificity of the variable for the individual label (uniform distribution of the values).

3.3. Individual Features Analysis

For each classification algorithm, we considered the full set of radiomic features extracted and the three best features identified by the feature selection methods (Fisher and χ^2 tests). For each feature selection criteria, we report the list of the three best features ordered according to their informative power.

3.3.1. Multifocal GGO

A Multifocal GGO label equal to 1 (32% of the patients) identifies the patients with the presence of a multifocal lesion, while a label equal to 0 (68% of the patients) identifies its absence.

The three best features selected by the Fisher criterion are (with the type of feature indicated in parenthesis):

- Skewness of the gray level distribution (Radiomics);
- Interquartile (25–75) of the roundness distribution (Radiomics);
- Kurtosis of the gray level distribution (Radiomics).

The three best features selected by the χ^2 criterion are (with type of feature indicated in parenthesis):

- Kurtosis of the gray level distribution (Radiomics);
- Minimum of the distance distribution (Radiomics);
- Skewness of the gray level distribution (Radiomics).

We show in Table 3 the results obtained in terms of global adjusted accuracy and in Table 4 the precision, sensitivity and F_1 score of the same prediction.

Table 3. Global adjusted accuracy of the models to predict multifocal GGO presence using different feature selection strategies. From the left: adjusted accuracy score obtained using only the 3 best features identified by the Fisher Exact test; the 3 best features identified by the χ2 test; all the radiomic features extracted. For each column the maximum value was indicated with bold font.

	Fisher	χ^2	**All**
Logistic	0.56	0.52	0.42
Ridge	**0.62**	**0.58**	0.45
KNN	0.44	0.44	**0.48**
R. Forest	0.49	0.43	0.47

Table 4. Precision, Sensitivity, and F_1 score for each model and each feature selection strategy to predict multifocal GGO presence. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the maximum value was indicated with bold font.

	Multifocal **1 = Presence** **0 = Absence**	**Fisher PPV**	χ^2 **PPV**	**Fisher TPR**	χ^2 **TPR**	**Fisher F_1**	χ^2 F_1
Logistic	0	0.72	0.69	0.64	0.64	0.68	0.67
	1	0.39	0.34	0.48	0.40	0.43	0.37
Ridge	0	**0.77**	**0.74**	0.64	0.64	0.70	0.69
	1	0.44	0.41	0.60	0.52	0.51	0.46
KNN	0	0.65	0.65	**0.83**	**0.83**	**0.73**	**0.73**
	1	0.10	0.10	0.04	0.04	0.06	0.06
R. Forest	0	0.67	0.64	0.77	0.77	0.72	0.70
	1	0.29	0.14	0.20	0.08	0.24	0.10

3.3.2. Presence of Crazy Paving

A Crazy Paving label equal to 1 (52% of the patients) identifies the patients with the presence of a crazy-paving pattern, while a label equal to 0 (48% of the patients) identifies its absence.

The three best features selected by the Fisher criterion are (with type of feature indicated in parenthesis):

- Median of the gray level distribution (Radiomics);
- Maximum of the Elongation distribution (Radiomics);
- Entropy (Haralick).

The three best features selected by the χ^2 criterion are (with type of feature indicated in parenthesis):

- Median of the gray level distribution (Radiomics);
- Inverse Difference Moment (Haralick);
- Skewness of the gray level distribution (Radiomics).

We show in Table 5 the results obtained in terms of global adjusted accuracy and in Table 6 the precision, sensitivity and F_1 score of the same prediction.

Table 5. Global adjusted accuracy of the models to predict the presence of crazy paving using different feature selection strategies. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F$_1$ score obtained using the 3 best features identified by the Fisher test; F$_1$ score obtained using the 3 best features identified by the χ^2 test. For each column the maximum value was indicated with bold font.

	Fisher	**χ^2**	**All**
Logistic	**0.57**	**0.61**	0.47
Ridge	0.51	0.56	0.45
KNN	0.46	0.50	**0.57**
R. Forest	0.50	0.51	0.51

Table 6. Precision, Sensitivity, and F1 Score for each model and each feature selection strat-egy to predict the presence of crazy paving. From the left: dichotomized score value, Pre-cision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best fea-tures identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F1 score obtained using the 3 best features identified by the Fisher test; F1 score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Crazy Paving **1 = Presence** **0 = Absence**	**Fisher PPV**	**χ^2 PPV**	**Fisher TPR**	**χ^2 TPR**	**Fisher F$_1$**	**χ^2 F$_1$**
Logistic	0	0.56	0.60	0.49	0.57	0.52	0.58
	1	**0.59**	**0.63**	**0.66**	0.66	**0.62**	**0.64**
Ridge	0	0.48	0.55	0.41	0.46	0.44	0.50
	1	0.53	0.57	0.61	0.66	0.57	0.61
KNN	0	0.43	0.47	0.43	0.51	0.43	0.49
	1	0.49	0.43	0.49	0.49	0.49	0.51
R. Forest	0	0.48	0.50	0.35	0.35	0.41	0.41
	1	0.53	0.54	**0.66**	**0.68**	0.59	0.60

3.3.3. Presence of Consolidation

A consolidation label equal to 1 (10% of the patients) identifies the patients with the presence of consolidation, while a label equal to 0 (90% of the patients) identifies its absence.

The three best features selected by the Fisher criterion are (with type of feature indicated in parenthesis):

- Cluster Prominence (Haralick);
- Median of the gray level distribution (Radiomics);
- Median of the elongation distribution (Radiomics).

The three best features selected by the χ^2 criterion are (with type of feature indicated in parenthesis):

- Median of the gray level distribution (Radiomics);
- Median of the elongation distribution (Radiomics);
- Cluster Prominence (Haralick).

We show in Table 7 the results obtained in term of global adjusted accuracy and in Table 8 the precision, sensitivity, and F$_1$ score of the same prediction.

Table 7. Global adjusted accuracy of the models to predict the presence of consolidations using different feature selection strategies. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F$_1$ score obtained using the 3 best features identified by the Fisher test; F$_1$ score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Fisher	χ^2	**All**
Logistic	**0.54**	**0.54**	**0.70**
Ridge	**0.54**	**0.54**	0.69
KNN	0.51	0.51	0.50
R. Forest	0.50	0.50	0.49

Table 8. Precision, Sensitivity, and F$_1$ Score for each model and each feature selection strategy to predict the presence of consolidations. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F$_1$ score obtained using the 3 best features identified by the Fisher test; F$_1$ score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Consolidation **1 = Presence** **0 = Absence**	**Fisher PPV**	χ^2 **PPV**	**Fisher TPR**	χ^2 **TPR**	**Fisher F$_1$**	χ^2 **F$_1$**
Logistic	0	**0.91**	**0.91**	0.59	0.59	0.71	0.71
	1	0.12	0.12	0.50	0.50	0.20	0.20
Ridge	0	**0.91**	**0.91**	0.59	0.59	0.71	0.71
	1	0.12	0.12	0.50	0.50	0.20	0.20
KNN	0	0.90	0.90	**0.90**	0.90	0.90	0.90
	1	0.12	0.12	0.12	0.12	0.12	0.12
R. Forest	0	0.90	0.90	**0.90**	**1.00**	**0.95**	**0.95**
	1	0.00	0.00	0.00	0.00	0.00	0.00

3.3.4. Roundish GGO

A roundish GGO label equal to 1 (12% of the patients) identifies the patients with the presence of a roundish GGO lesion, while a label equal to 0 (88% of the patients) identifies its absence.

The three best features selected by the Fisher criterion are (with type of feature indicated in parenthesis):

- GGO volume percentage (Radiomics);
- Skewness of the roundness distribution (Radiomics);
- Median of the roundness distribution (Radiomics).

The three best features selected by the χ^2 criterion are (with type of feature indicated in parenthesis):

- GGO volume percentage (Radiomics);
- Skewness of the roundness distribution (Radiomics);
- Median of the roundness distribution (Radiomics).

We show in Table 9 the results obtained in term of global adjusted accuracy and in Table 10 the precision, sensitivity and F$_1$ score of the same prediction.

Table 9. Global adjusted accuracy of the models to predict the presence of roundish GGO using different feature selection strategies. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Fisher	χ^2	All
Logistic	**0.60**	**0.60**	**0.59**
Ridge	0.58	**0.60**	0.56
KNN	0.43	0.43	0.50
R. Forest	0.45	0.45	0.50

Table 10. Precision, Sensitivity, and F_1 Score for each model and each feature selection strategy to predict the presence of roundish GGO lesions. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Roundish 1 = Presence 0 = Absence	Fisher PPV	χ^2 PPV	Fisher TPR	χ^2 TPR	Fisher F_1	χ^2 F_1
Logistic	0	**0.92**	**0.92**	0.64	0.64	0.75	0.75
	1	0.17	0.17	0.56	0.56	0.26	0.26
Ridge	0	0.91	**0.92**	0.61	0.64	0.73	0.75
	1	0.16	0.17	0.56	0.56	0.24	0.26
KNN	0	0.87	0.87	**0.86**	**0.87**	0.86	0.87
	1	0.00	0.00	0.00	0.00	0.00	0.00
R. Forest	0	0.87	0.87	0.90	0.90	**0.94**	**0.89**
	1	0.00	0.00	0.00	0.00	0.00	0.00

3.3.5. Peripheral GGO

A peripheral GGO label equal to 1 (23% of the patients) identifies the patients with the presence of a peripheral GGO lesion, while a label equal to 0 (77% of the patients) identifies its absence.

The three best features selected by the Fisher criterion are (with the type of feature indicated in parenthesis):

- Patient age (Clinical);
- Minimum of the distance distribution (Radiomics);
- Skewness of the gray level distribution (Radiomics).

The three best features selected by the χ^2 criterion are (with type of feature indicated in parenthesis):

- Minimum of the distance distribution (Radiomics);
- Patient age (Clinical);
- Skewness of the elongation distribution (Radiomics).

We show in Table 11 the results obtained in terms of global adjusted accuracy and in Table 12 the precision, sensitivity, and F_1 score of the same prediction.

Table 11. Global adjusted accuracy of the models to predict the presence of peripheral GGO using different feature selection strategies. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Fisher	χ^2	**All**
Logistic	0.47	0.47	**0.62**
Ridge	**0.52**	**0.52**	0.61
KNN	**0.52**	**0.52**	0.51
R. Forest	0.42	0.42	0.53

Table 12. Precision, Sensitivity, and F_1 Score for each model and each feature selection strategy to predict the presence of peripheral GGO lesions. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Peripheral **1 = Presence** **0 = Absence**	**Fisher** **PPV**	χ^2 **PPV**	**Fisher** **TPR**	χ^2 **TPR**	**Fisher** F_1	χ^2 F_1
Logistic	0	0.74	0.74	0.43	0.43	0.55	0.55
	1	0.21	0.21	0.50	0.50	0.30	0.30
Ridge	0	**0.79**	**0.79**	0.43	0.43	0.56	0.56
	1	0.24	0.24	0.61	0.61	0.35	0.35
KNN	0	0.78	0.78	**0.87**	**0.87**	**0.82**	**0.82**
	1	0.27	0.27	0.17	0.17	0.21	0.21
R. Forest	0	0.74	0.74	0.83	0.83	0.78	0.78
	1	0.00	0.00	0.00	0.00	0.00	0.00

3.4. Primary Outcomes

3.4.1. PREDI-CO Score

The prediction of the PREDI-CO score was performed considering the dichotomized values obtained by thresholding the labels according to the cutoff of 1: values ≤ 1 (47% of the patients) were labelled as 0 and values > 1 (53% of the patients) were labelled as 1. The considered dataset includes only two patients with a PREDI-CO score ≥ 6 proving the strong unbalancing of the classes and highlighting the prominence of non-severe patients.

The three best features selected by the Fisher criterion are (with type of feature indicated in parenthesis):

- Patient age (Clinical);
- Median of the distance distribution (Radiomics);
- Interquartile (25–75) of the roundness distribution (Radiomics).

The three best features selected by the χ^2 criterion are (with type of feature indicated in parenthesis):

- Patient age (Clinical);
- Interquartile (25–75) of the elongation distribution (Radiomics);
- Maximum of the distance distribution (Radiomics).

We show in Table 13 the results obtained in terms of global adjusted accuracy and in Table 14 the precision, sensitivity and F_1 score of the same prediction.

Table 13. Global adjusted accuracy of the models to predict the dichotomized PREDI-CO score using different feature selection strategies. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	Fisher	**χ^2**	**All**
Logistic	**0.70**	**0.70**	**0.59**
Ridge	**0.70**	**0.70**	0.52
KNN	0.62	0.60	0.50
R. Forest	0.61	0.61	**0.59**

Table 14. Precision, Sensitivity, and F_1 Score for each model and each feature selection strategy to predict the dichotomized PREDI-CO score. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the max-imum value was indicated with bold font.

	PREDI-CO **1 = PREDI-CO > 1** **0 = PREDI-CO $\leq$ 1**	**Fisher PPV**	**χ^2 PPV**	**Fisher TPR**	**χ^2 TPR**	**Fisher F_1**	**χ^2 F_1**
Logistic	0	0.67	0.67	0.71	0.71	0.69	0.69
	1	0.72	**0.72**	0.68	0.68	**0.70**	**0.70**
Ridge	0	0.67	0.67	0.71	0.71	0.69	0.69
	1	0.72	**0.72**	0.68	0.68	**0.70**	**0.70**
KNN	0	0.56	0.54	**0.85**	0.85	0.67	0.66
	1	**0.75**	0.72	0.39	0.34	0.52	0.46
R. Forest	0	0.60	0.60	0.53	0.53	0.56	0.56
	1	0.62	0.62	0.68	0.68	0.65	0.65

3.4.2. Patient Survival

A survival label equal to 1 (10% of the patients) identifies the patients survived to the COVID-19, while a label equal to 0 (90% of the patients) identifies death patients.

The three best features selected by the Fisher criterion are (with type of feature indicated in parenthesis):

- Inverse difference moment (Haralick);
- Median of the elongation distribution (Radiomics);
- Median of the distance distribution (Radiomics).

The three best features selected by the χ^2 criterion are (with type of feature indicated in parenthesis):

- Inverse difference moment (Haralick);
- Median of the elongation distribution (Radiomics);
- Skewness of the elongation distribution (Radiomics).

We show in Table 15 the results obtained in terms of global adjusted accuracy and in Table 16 the precision, sensitivity and F_1 score of the same prediction.

Table 15. Global adjusted accuracy of the models to predict mortality using different feature selection strategies. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the maximum value was indicated with bold font.

	Fisher	χ^2	All
Logistic	0.41	0.48	0.43
Ridge	**0.62**	**0.70**	0.41
KNN	0.41	0.42	**0.50**
R. Forest	0.46	0.48	**0.50**

Table 16. Precision, Sensitivity, and F_1 Score for each model and each feature selection strategy to predict patient survival. From the left: dichotomized score value, Precision obtained using the 3 best features identified by the Fisher test, Precision obtained using the 3 best features identified by the χ^2 test; Sensitivity obtained using the 3 best features identified by the Fisher test, Sensitivity obtained using the 3 best features identified by the χ^2 test; F_1 score obtained using the 3 best features identified by the Fisher test; F_1 score obtained using the 3 best features identified by the χ^2 test. For each column the maximum value was indicated with bold font.

	Survival 1 = Dead 0 = Alive	Fisher PPV	χ^2 PPV	Fisher TPR	χ^2 TPR	Fisher F_1	χ^2 F_1
Logistic	0	0.88	0.90	0.82	0.82	0.85	0.85
	1	0.00	0.08	0.00	0.14	0.00	0.10
Ridge	0	**0.93**	**0.95**	0.82	0.82	0.87	0.88
	1	0.20	0.25	0.43	0.57	0.27	0.35
KNN	0	0.89	0.89	0.83	0.85	0.86	0.87
	1	0.00	0.00	0.00	0.00	0.00	0.00
R. Forest	0	0.90	0.90	**0.94**	**0.95**	**0.92**	**0.93**
	1	0.00	0.00	0.00	0.00	0.00	0.00

4. Discussion

4.1. GGO Segmentation

The examples reported in Figures 3 and 4 show how the non-supervised segmentation method proposed in this paper is able to approximate the gold standard results with satisfactory results.

This result has two strong implications for the Radiomics of the COVID-19 patients. First, given that the amount of information required for the k-means method training is considerably lower than CNN methods, while still retaining good results, this segmentation can be implemented with in-patient training. Secondly, this method can be used with success as a first segmentation method to be used as training for other, more specific methods. We remark that all the proposed techniques are voxel-based algorithms: this kind of method requires the whole patient's scan as input, drastically reducing the dimensionality of the dataset. As a reference, a 3D U-Net-based method [66] required two order of magnitude training samples to achieve comparable results.

It is worth noting that the various segmentation scores are dependent on the class balance, and therefore tend to penalize this kind of segmentation where one class (the GGO class in our case) is substantially under-represented. This can be confirmed by confronting the results of the proposed segmentation with published methods such as those reported in Table 1.

4.2. Individual Features Analysis

We noticed an evident improvement in the prediction converting the clinical outcomes to dichotomized classes. A second improvement in the quality of the prediction was obtained by the application of the feature reduction techniques, which allow a stabilization of the results. Most of the predictive power for each feature can be synthesized in 3 to 4 features per variable. Of all the variables, particular prominence was observed for the Radiomic features. These features were the most important ones in most of the predictions. Of all the predicted characteristics and outcomes, only the peripherality of GGO and PREDI-CO score requires the inclusion of the age value. For the PREDI-CO score, this is not unexpected as it is one of the components of this score. As for the peripherality of GGO lesions, this is an interesting result as they are an important predictor of clinical outcomes and not intuitively associated with patients' age.

If one considers the results for the different predictors (linear penalized, KNN, Random Forest), one can observe that in general, KNN and Random Forests achieve similar performances, while the penalized linear methods consistently perform better. This can be interpreted as the effect of a progressive non-dichotomic behavior in the system. These linear models were also the ones that gained the least from the pre-selection of the features. This can be explained as the L1 and L2 penalization already reducing the effect of the numerosity of the provided features. Methods such as KNN, based on features metrics, are particularly affected by the features numerosity and thus are the ones that have the greatest improvement by feature selection. Linear penalized methods, on the other end, include an implicit feature selection internally and could be even penalized by a reduction in the number of considered features.

Most of these features have a strong class unbalance (down to around 10% of samples in one group, such as in the Consolidation and Roundish GGO), and therefore, the prediction score tends to be strongly unbalanced, with a strong penalization for the prediction of the least represented class. When this unbalance is not present, such as in the case of the prediction of the PREDI-CO score, one can observe good, balanced prediction scores.

The prediction scores for the PREDI-CO are also higher than for similarly balanced classes (such as Crazy Paving). This indicates that the extracted features, albeit not optimal for predicting individuals' components of the score, are indeed able to predict the score as a global value. It is interesting to notice that the variables considered as the most important in the prediction (both for the Fisher and χ^2 method) alongside the age (a well-known risk factor) are (1) the distance between the GGO area and the trachea and (2) the irregularity of the GGO lesion shape. This follows the clinical hypothesis that the spreading of the damaged area toward the peripheral area of the lungs leads to the worst prognosis for the patient.

The results obtained in this work cannot overcome the performances of the already published artificial intelligence techniques. The main limitation of this work is related to the number of available samples: semi-supervised learning algorithms are designed to work with small datasets but require better labeling of them compared to supervised methods.

A second criticality is given by the preliminary choice of the number of clusters for the k-means algorithm: in our work, we identified only five putative clusters for the tissue segmentation. This degree of freedom determines the quality of the areas used for the radiomic feature extraction, and therefore, it could affect the efficiency of the prediction models. In contrast, the manipulation of this single degree of freedom could help to achieve better results on in-patient segmentations.

The clinical characteristics and outcomes considered in this work are scores estimated by the expert radiologists for the description of the state of the patient, but they do not consider the real severity of him. This intrinsic limit does not allow a prediction of the real outcome of the patient, allowing only an undirected evaluation.

5. Conclusions

In the present work, we highlighted the possibility of obtaining a reliable automated segmentation using non-supervised approaches and using this segmentation in a prediction pipeline for patient prognosis.

Artificial Intelligence for diagnostic uses, such as a clinical decision support system, is recognized as an approach rich of potential outcomes but is limited by the requirement of human-driven data curation. With this work, we aimed to prove that semi-supervised approaches to segmentation are promising, as they would combine the best effort of highly trained physicians to develop true gold standard segmentation and the expertise of data analysts to augment that segmentation in full-blown models.

The current COVID-19 pandemic highlights the criticality of relying on high specialized clinicians for time-demanding tasks, as the same experts that can generate gold-standard segmentation for AI training are also the ones responsible for patient diagnosis and care. Improving methods for semi-supervised learning in Radiomics would allow for more effective use of the time and energy of these experts while capitalizing on AI training to support them in patient's diagnosis and treatment.

While the results presented in this work are not yet at the accuracy level necessary for assisted diagnostic use, we surmise that this approach would be helpful in developing a solid triage system, which would help to prioritize the resources available and direct them were most effective.

Author Contributions: M.B., A.C., F.C. (Francesca Coppola), A.C., F.C. (Federica Ciccarese), L.P., G.V., B.B., and L.C.: Investigation and Data Curation; R.B., N.C., E.G., and B.B.: Software, Methodology, Data Analysis; P.V., L.S., R.G., F.C. (Francesca Coppola), E.M., and G.C.: Conceptualization and Supervision; C.D.B., R.B., N.C., G.V., and E.G.: Original draft preparation. All authors contributed to reviewing and editing of the manuscript. All authors have read and agreed to the published version of the manuscript.

References

1. Loeffelholz, M.J.; Tang, Y.W. Laboratory diagnosis of emerging human coronavirus infections—The state of the art. *Emerg. Microbes Infect.* **2020**, *9*, 747–756. [CrossRef]
2. Fu, Z.; Tang, N.; Chen, Y.; Ma, L.; Wei, Y.; Lu, Y.; Ye, K.; Liu, H.; Tang, F.; Huang, G.; et al. CT features of COVID-19 patients with two consecutive negative RT-PCR tests after treatment. *Sci. Rep.* **2020**, *10*, 11548. [CrossRef]
3. Ciccarese, F.; Coppola, F.; Spinelli, D.; Galletta, G.L.; Lucidi, V.; Paccapelo, A.; De Benedittis, C.; Balacchi, C.; Golfieri, R. Diagnostic accuracy of north america expert consensus statement on reporting CT findings in patients suspected of having COVID-19 infection: An italian single-center experience. *Radiol. Cardiothorac. Imag.* **2020**, *2*, e200312. [CrossRef]

4. Byrne, D.; Neill, S.B.O.; Müller, N.L.; Müller, C.I.S.; Walsh, J.P.; Jalal, S.; Parker, W.; Bilawich, A.M.; Nicolaou, S. RSNA Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19: Interobserver Agreement Between Chest Radiologists. *Can. Assoc. Radiol. J.* **2021**, *72*, 159–166. [CrossRef] [PubMed]

5. Esbin, M.N.; Whitney, O.N.; Chong, S.; Maurer, A.; Darzacq, X.; Tjian, R. Overcoming the bottleneck to widespread testing: A rapid review of nucleic acid testing approaches for COVID-19 detection. *RNA* **2020**, *26*, 771–783. [CrossRef]

6. Basso, D.; Aita, A.; Navaglia, F.; Franchin, E.; Fioretto, P.; Moz, S.; Bozzato, D.; Zambon, C.F.; Martin, B.; Dal Prà, C.; et al. SARS-CoV-2 RNA identification in nasopharyngeal swabs: Issues in pre-analytics. *Clin. Chem. Lab. Med.* **2020**, *58*, 1579–1586. [CrossRef] [PubMed]

7. Rubin, G.D.; Ryerson, C.J.; Haramati, L.B.; Sverzellati, N.; Kanne, J.P.; Raoof, S.; Schluger, N.W.; Volpi, A.; Yim, J.J.; Martin, I.B.K.; et al. The Role of Chest Imaging in Patient Management During the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. *Chest* **2020**, *158*, 106–116. [CrossRef] [PubMed]

8. 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]

9. Akbari, Y.; Hassen, H.; Al-ma'adeed, S.; Zughaier, S. COVID-19 Lesion Segmentation Using Lung CT Scan Images: Comparative Study Based on Active Contour Models. *Res. Square* **2020**. [CrossRef]

10. Fusco, R.; Granata, V.; Mazzei, M.A.; Meglio, N.D.; Roscio, D.D.; Moroni, C.; Monti, R.; Cappabianca, C.; Picone, C.; Neri, E.; et al. Quantitative imaging decision support (QIDSTM) tool consistency evaluation and radiomic analysis by means of 594 metrics in lung carcinoma on chest CT scan. *Cancer Control* **2021**, *28*, 1073274820985786. [CrossRef]

11. Collins, J.; Stern, E. Ground-glass opacity at CT: The ABCs. *Am. J. Roentgenol.* **1997**, *169*, 355–367. [CrossRef]

12. Huang, C.; Wang, Y.; Li, X.; Ren, L.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [CrossRef]

13. Wang, C.; Shi, B.; Wei, C.; Ding, H.; Gu, J.; Dong, J. Initial CT features and dynamic evolution of early-stage patients with COVID-19. *Radiol. Infect. Dis.* **2020**, *7*, 195–203. [CrossRef] [PubMed]

14. Zhang, C.; Yang, G.; Cai, C.; Xu, Z.; Wu, H.; Guo, Y.; Xie, Z.; Shi, H.; Cheng, G.; Wang, J. Development of a quantitative segmentation model to assess the effect of comorbidity on patients with COVID-19. *Eur. J. Med. Res.* **2020**, *25*, 49. [CrossRef] [PubMed]

15. Adair, L.B.; Ledermann, E.J. Chest CT findings of early and progressive phase COVID-19 infection from a US patient. *Radiol. Case Rep.* **2020**, *15*, 819–824. [CrossRef] [PubMed]

16. Yang, S.; Shi, Y.; Lu, H.; Xu, J.; Li, F.; Qian, Z.; Jiang, Y.; Hua, X.; Ding, X.; Song, F.; et al. Clinical and CT features of early stage patients with COVID-19: A retrospective analysis of imported cases in Shanghai, China. *Eur. Respir. J.* **2020**, *55*, 2000407. [CrossRef] [PubMed]

17. Xu, M.; Qi, S.; Yue, Y.; Teng, Y.; Xu, L.; Yao, Y.; Qian, W. Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset. *Biomed. Eng. Online* **2019**, *18*, 239. [CrossRef] [PubMed]

18. Neri, E.; Coppola, F.; Larici, A.R.; Sverzellati, N.; Mazzei, M.A.; Sacco, P.; Dalpiaz, G.; Feragalli, B.; Miele, V.; Grassi, R. Structured reporting of chest CT in COVID-19 pneumonia: A consensus proposal. *Insights Imag.* **2020**, *11*, 92. [CrossRef] [PubMed]

19. ACR Website. Position Statement Section. Available online: https://www.acr.org/Advocacy-and-Economics/ACR-Position-Statements/Recommendations-for-Chest-Radiography-and-CT-for-Suspected-COVID19-Infection (accessed on 3 June 2021).

20. Kirsch, J.; Ramirez, J.; Mohammed, T.L.; Amorosa, J.K.; Brown, K.; Dyer, D.S.; Ginsburg, M.E.; Heitkamp, D.E.; Jeudy, J.; Macmahon, H.; et al. ACR Appropriateness Criteria® acute respiratory illness in immunocompetent patients. *J. Thorac. Imaging* **2011**, *15*, W42–W44. [CrossRef]

21. Neri, E.; Miele, V.; Coppola, F.; Grassi, R. Use of CT and artificial intelligence in suspected or COVID-19 positive patients: Statement of the Italian Society of Medical and Interventional Radiology. *Radiol. Med.* **2020**, *125*, 505–508. [CrossRef]

22. Zhang, B.; Ni-Jia-Ti, M.Y.; Yan, R.; An, N.; Chen, L.; Liu, S.; Chen, L.; Chen, Q.; Li, M.; Chen, Z.; et al. CT-based radiomics for predicting the rapid progression of coronavirus disease 2019 (COVID-19) pneumonia lesions. *Br. J. Radiol.* **2021**, *94*, 20201007. [CrossRef]

23. Ria, F.; Fu, W.; Chalian, H.; Abadi, E.; Segars, P.W.; Fricks, R.; Khoshpouri, P.; Samei, E. A comparison of COVID-19 and imaging radiation risk in clinical patient populations. *J. Radiol. Prot.* **2020**, *40*, 1336. [CrossRef]

24. Jin, S.; Wang, B.; Xu, H.; Luo, C.; Wei, L.; Zhao, W.; Hou, X.; Ma, W.; Xu, Z.; Zheng, Z.; et al. AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks. *medRxiv* **2020**, *98*, 106897.

25. Cattabriga, A.; Cocozza, M.A.; Vara, G.; Coppola, F.; Golfieri, R. Lung CT Segmentation to Identify Consolidations and Ground Glass Areas for Quantitative Assesment of SARS-CoV Pneumonia. *J. Vis. Exp.* **2020**, *166*. [CrossRef]

26. Mansoor, A.; Foster, B.; Xu, Z.; Papadakis, G.; Folio, L.; Udupa, J.; Mollura, D. Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends. *Radiogr. Rev. Publ. Radiol. Soc. N. Am. Inc.* **2015**, *35*, 1056–1076. [CrossRef] [PubMed]

27. Oulefki, A.; Agaian, S.; Trongtirakul, T.; Laouar, A.K. Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. *Pattern Recognit.* **2020**, *114*, 107747. [CrossRef] [PubMed]

28. Nakagomi, K.; Shimizu, A.; Kobatake, H.; Yakami, M.; Fujimoto, K.; Togashi, K. Multi-shape graph cuts with neighbor prior constraints and its application to lung segmentation from a chest CT volume. *Med. Image Anal.* **2013**, *17*, 62–77. [CrossRef]

29. Dai, S.; Lu, K.; Dong, J.; Zhang, Y.; Chen, Y. A novel approach of lung segmentation on chest CT images using graph cuts. *Neurocomputing* **2015**. [CrossRef]
30. Li, B.; Christensen, G.; Mclennan, G.; Hoffman, E.; Reinhardt, J. Establishing a normative atlas of the human lung: Inter-subject warping and registration of volumetric CT. *Acad. Radiol.* **2003**, *10*, 255–265. [CrossRef]
31. Dey, N.; Rajinikanth, V.; Fong, S.J.; Kaiser, M.S.; Mahmud, M. Social-group-optimization assisted Kapur's entropy and morphological segmentation for automated detection of COVID-19 infection from computed tomography images. *Cogn. Comput.* **2020**, *12*, 1–13. [CrossRef]
32. Abdel-Basset, M.; Chang, V.; Hawash, H.; Chakrabortty, R.K.; Ryan, M. FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. *Knowl. Based Syst.* **2021**, *212*, 106647. [CrossRef] [PubMed]
33. Hofmanninger, J.; Prayer, F.; Pan, J.; Röhrich, S.; Prosch, H.; Langs, G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *Eur. Radiol. Exp.* **2020**, *4*, 50. [CrossRef]
34. Wang, H.; Wang, L.; Lee, E.H.; Zheng, J.; Zhang, W.; Halabi, S.; Liu, C.; Deng, K.; Song, J.; Yeom, K.W. Decoding COVID-19 pneumonia: Comparison of deep learning and radiomics CT image signatures. *Eur. J. Nucl. Med. Mol. Imaging* **2020**, *48*, 1–9. [CrossRef]
35. Gozes, O.; Frid-Adar, M.; Greenspan, H.; Browning, P.D.; Zhang, H.; Ji, W.; Bernheim, A.; Siegel, E. Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis. *arXiv* **2020**, arXiv:2003.05037.
36. Müller, D.; Rey, I.S.; Kramer, F. Automated chest CT image segmentation of COVID-19 lung infection based on 3D u-net. *arXiv* **2020**, arXiv:2007.04774.
37. Bernheim, A.; Mei, X.; Huang, M.; Yang, Y.; Fayad, Z.A.; Zhang, N.; Diao, K.; Lin, B.; Zhu, X.; Li, K.; et al. Chest CT findings in coronavirus disease-19 (COVID-19): Relationship to duration of infection. *Radiology* **2020**, *295*, 200463. [CrossRef]
38. Simpson, S.; Kay, F.U.; Abbara, S.; Bhalla, S.; Chung, J.H.; Chung, M.; Henry, T.S.; Kanne, J.P.; Kligerman, S.; Ko, J.P.; et al. Radiological Society of North America Expert Consensus Statement on Reporting Chest CT Findings Related to COVID-19. Endorsed by the Society of Thoracic Radiology, the American College of Radiology, and RSNA—Secondary Publication. *J. Thorac. Imaging* **2020**, *35*, 219–227. [CrossRef] [PubMed]
39. Xie, X.; Zhong, Z.; Zhao, W.; Zheng, C.; Wang, F.; Liu, J. Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing. *Radiology* **2020**, *296*, E41–E45. [CrossRef]
40. Huang, P.; Liu, T.; Huang, L.; Liu, H.; Lei, M.; Xu, W.; Hu, X.; Chen, J.; Liu, B. Use of Chest CT in Combination with Negative RT-PCR Assay for the 2019 Novel Coronavirus but High Clinical Suspicion. *Radiology* **2020**, *295*, 22–23. [CrossRef] [PubMed]
41. Fang, Y.; Zhang, H.; Xie, J.; Lin, M.; Ying, L.; Pang, P.; Ji, W. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology* **2020**, *296*, E115–E117. [CrossRef]
42. Inui, S.; Fujikawa, A.; Jitsu, M.; Kunishima, N.; Watanabe, S.; Suzuki, Y.; Umeda, S.; Uwabe, Y. Chest CT findings in cases from the cruise ship diamond princess with coronavirus disease (COVID-19). *Radiol. Cardiothorac. Imag.* **2020**, *2*, e200110. [CrossRef]
43. Belkhatir, Z.; Estépar, R.S.J.; Tannenbaum, A.R. Supervised Image Classification Algorithm Using Representative Spatial Texture Features: Application to COVID-19 Diagnosis Using CT Images. *medRxiv* **2020**. [CrossRef]
44. Liu, C.; Wang, X.; Liu, C.; Sun, Q.; Peng, W. Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning. *Biomed. Eng. Online* **2020**, *19*, 66. [CrossRef]
45. Zeng, Q.Q.; Zheng, K.I.; Chen, J.; Jiang, Z.H.; Tian, T.; Wang, X.B.; Ma, H.L.; Pan, K.H.; Yang, Y.J.; Chen, Y.P.; et al. Radiomics-based model for accurately distinguishing between severe acute respiratory syndrome associated coronavirus 2 (SARS-CoV-2) and influenza A infected pneumonia. *MedComm* **2020**. [CrossRef] [PubMed]
46. Wei, W.; Hu, X.W.; Cheng, Q.; Zhao, Y.M.; Ge, Y.Q. Identification of common and severe COVID-19: The value of CT texture analysis and correlation with clinical characteristics. *Eur. Radiol.* **2020**, *30*, 6788–6796. [CrossRef] [PubMed]
47. Caruso, D.; Zerunian, M.; Polici, M.; Pucciarelli, F.; Polidori, T.; Rucci, C.; Guido, G.; Bracci, B.; De Dominicis, C.; Laghi, A. Chest CT Features of COVID-19 in Rome, Italy. *Radiology* **2020**, *296*, E79–E85. [CrossRef] [PubMed]
48. Ye, Z.; Zhang, Y.; Wang, Y.; Huang, Z.; Song, B. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): A pictorial review. *Eur. Radiol.* **2020**, *30*, 4381–4389. [CrossRef] [PubMed]
49. Albarello, F.; Pianura, E.; Di Stefano, F.; Cristofaro, M.; Petrone, A.; Marchioni, L.; Palazzolo, C.; Schininà, V.; Nicastri, E.; Petrosillo, N.; et al. 2019-novel Coronavirus severe adult respiratory distress syndrome in two cases in Italy: An uncommon radiological presentation. *Int. J. Infect. Dis.* **2020**, *93*, 192–197. [CrossRef]
50. Varga, Z.; Flammer, A.J.; Steiger, P.; Haberecker, M.; Andermatt, R.; Zinkernagel, A.S.; Mehra, M.R.; Schuepbach, R.A.; Ruschitzka, F.; Moch, H. Endothelial cell infection and endotheliitis in COVID-19. *Lancet* **2020**, *395*, 1417–1418. [CrossRef]
51. Zhang, H.; Hung, C.L.; Min, G.; Guo, J.P.; Liu, M.; Hu, X. GPU-Accelerated GLRLM Algorithm for Feature Extraction of MRI. *Sci. Rep.* **2019**, *9*, 10883. [CrossRef]
52. Hu, X.; Ye, W.; Li, Z.; Chen, C.; Cheng, S.; Lv, X.; Weng, W.; Li, J.; Weng, Q.; Pang, P.; et al. Non-invasive evaluation for benign and malignant subcentimeter pulmonary ground-glass nodules (≤1 cm) based on CT texture analysis. *Br. J. Radiol.* **2020**, *93*, 20190762. [CrossRef] [PubMed]
53. Fang, X.; Kruger, U.; Homayounieh, F.; Chao, H.; Zhang, J.; Digumarthy, S.R.; Arru, C.D.; Kalra, M.K.; Yan, P. Association of AI quantified COVID-19 chest CT and patient outcome. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 435–44524. [CrossRef] [PubMed]

54. Bartoletti, M.; Giannella, M.; Scudeller, L.; Tedeschi, S.; Rinaldi, M.; Bussini, L.; Fornaro, G.; Pascale, R.; Pancaldi, L.; Pasquini, Z.; et al. Development and validation of a prediction model for severe respiratory failure in hospitalized patients with SARS-CoV-2 infection: A multicentre cohort study (PREDI-CO study). *Clin. Microbiol. Infect.* **2020**, *26*, 1545–1553. [CrossRef] [PubMed]

55. Jun, M.; Cheng, G.; Yixin, W.; Xingle, A.; Jiantao, G.; Ziqi, Y.; Minqing, Z.; Xin, L.; Xueyuan, D.; Shucheng, C.; et al. *COVID-19 CT Lung and Infection Segmentation Dataset*; CERN: Geneva, Switzerland, 2020. [CrossRef]

56. Yokota, K.; Maeda, S.; Kim, H.; Tan, J.K.; Ishikawa, S.; Tachibana, R.; Hirano, Y.; Kido, S. Automatic detection of GGO regions on CT images in LIDC dataset based on statistical features. In Proceedings of the 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS), Kitakyushu, Japan, 3–6 December 2014; pp. 1374–1377.

57. Frangi, R.; Niessen, W.J.; Vincken, K.; Viergever, M. Multiscale vessel enhancement filtering. *Med. Image Comput. Comput. Assist. Interv.* **2000**. [CrossRef]

58. Sato, Y.; Nakajima, S.; Shiraga, N.; Atsumi, H.; Yoshida, S.; Koller, T.; Gerig, G.; Kikinis, R. Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. *Med. Image Anal.* **1998**, *2*, 143–168. [CrossRef]

59. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms SODA '07, Society for Industrial, Applied Mathematics, New Orleans, LA, USA, 7–9 January 2007; pp. 1027–1035.

60. Yaniv, Z.; Lowekamp, B.C.; Johnson, H.J.; Beare, R. Simple ITK Image-Analysis Notebooks: A Collaborative Environment for Education and Reproducible Research. *J. Digit. Imaging* **2018**, *31*, 290–303. [CrossRef] [PubMed]

61. Lowekamp, B.C.; Chen, D.T.; Ibáñez, L.; Blezek, D. The Design of SimpleITK. *Front. Neuroinf.* **2013**, *7*, 45. [CrossRef] [PubMed]

62. Bradski, G. The OpenCV Library. *Dr. Dobb J. Softw. Tools* **2000**, *120*, 122–125.

63. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *3*, 610–621. [CrossRef]

64. Lehmann, G. La Bel Object Representation and Manipulation with ITK. 2007. Available online: http://hdl.handle.net/1926/584 (accessed on 4 June 2021).

65. Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; et al. API design for machine learning software: Experiences from the scikit-learn project. *arXiv* **2013**, arXiv:1309.0238.

66. Yan, Q.; Wang, B.; Gong, D.; Luo, C.; Zhao, W.; Shen, J.; Shi, Q.; Jin, S.; Zhang, L.; You, Z. COVID-19 Chest CT Image Segmentation—A Deep Convolutional Neural Network Solution. *arXiv* **2020**, arXiv:2004.10987.