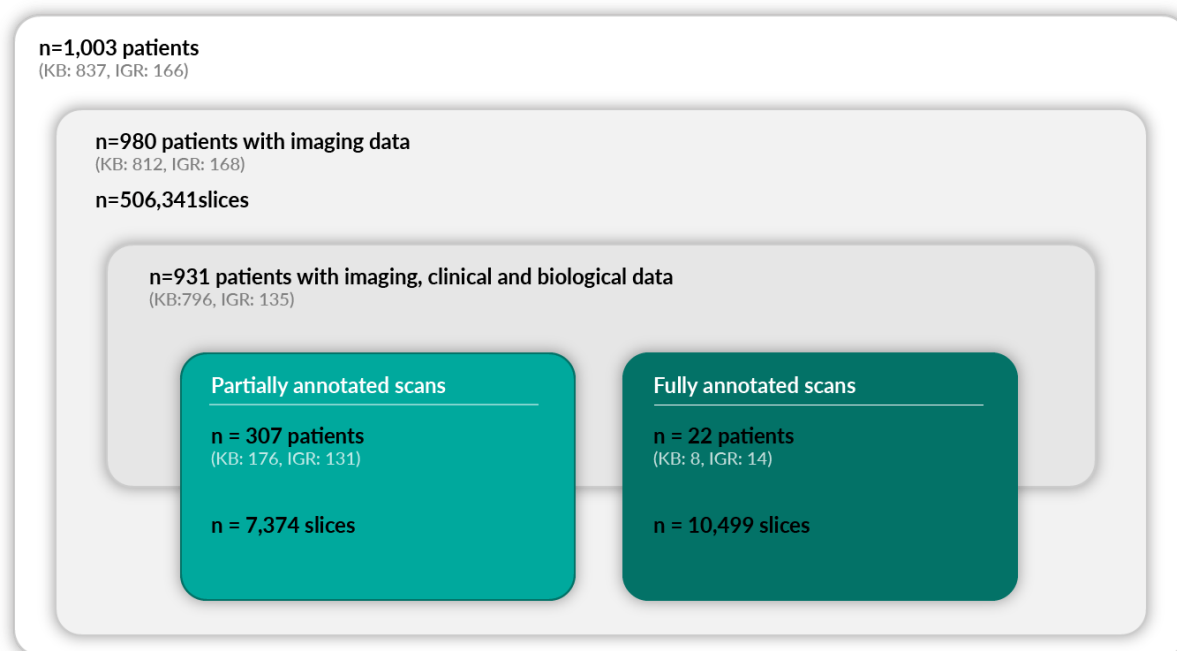
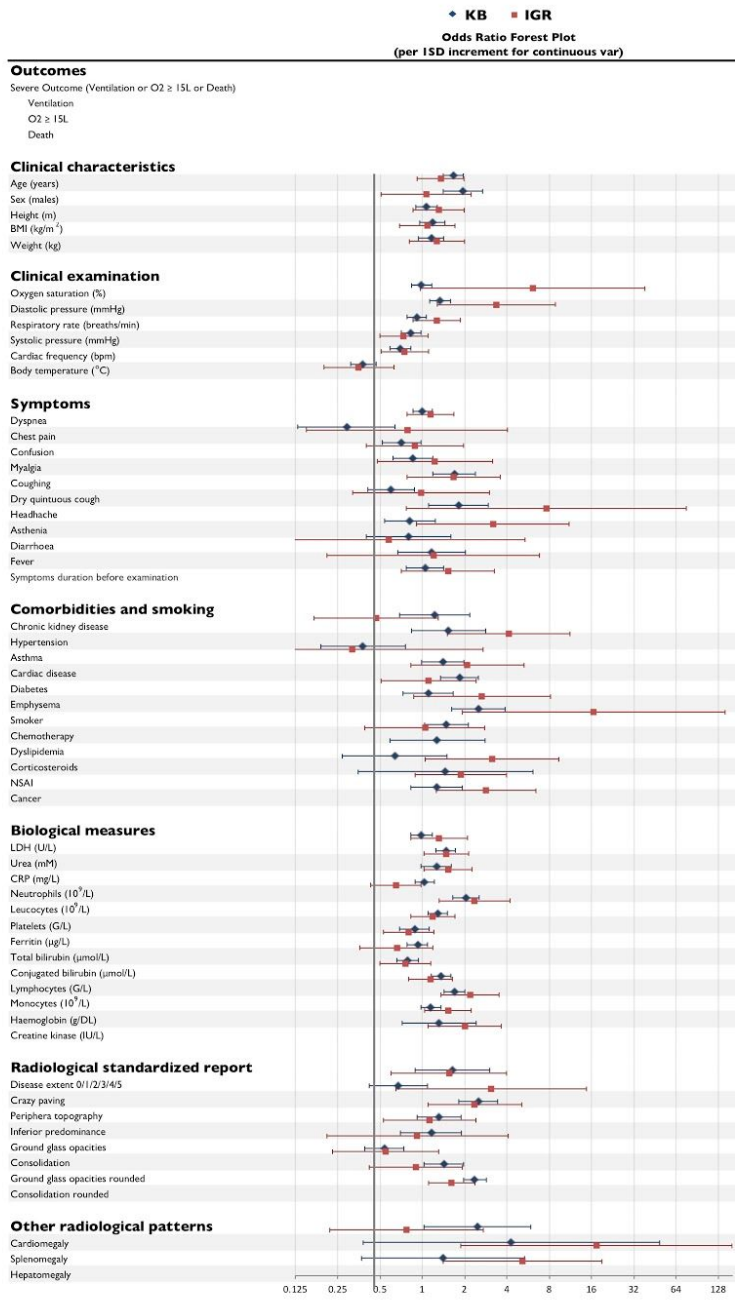


Supplementary material of “AI-based  
*multi-modal* integration of clinical  
characteristics, lab tests and chest  
CTs improves COVID-19 outcome  
prediction of *hospitalized* patients”

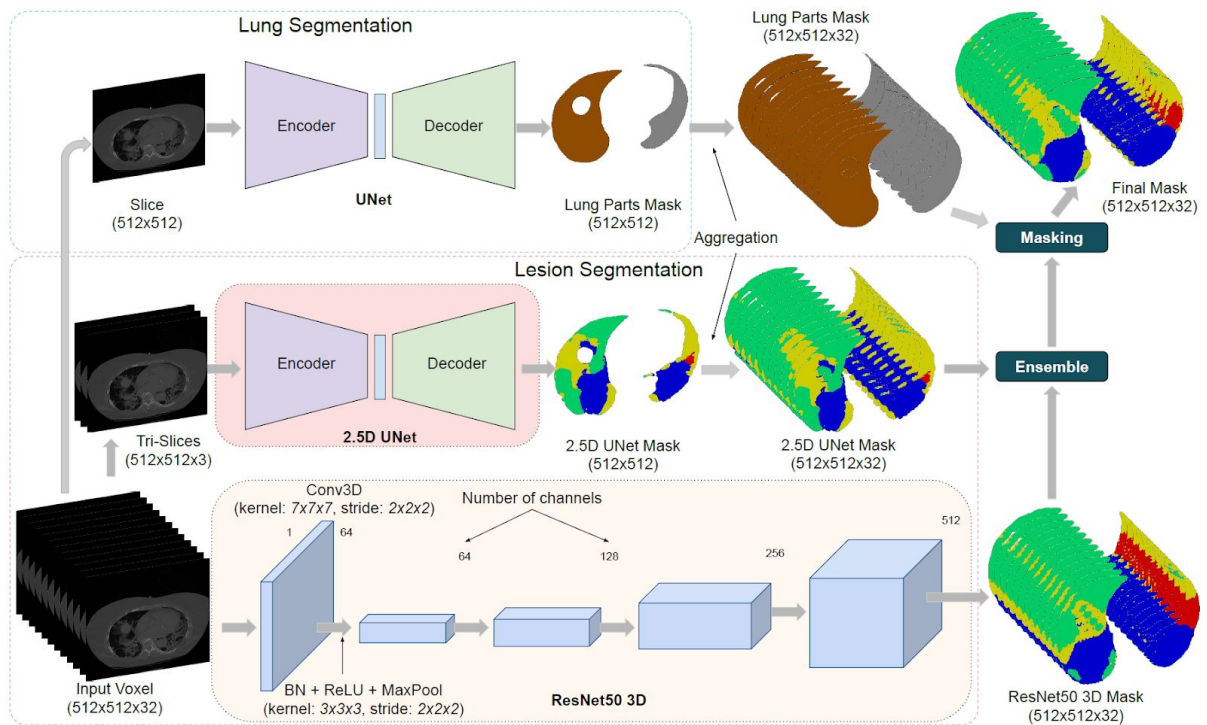
# Supplementary Figures and Tables



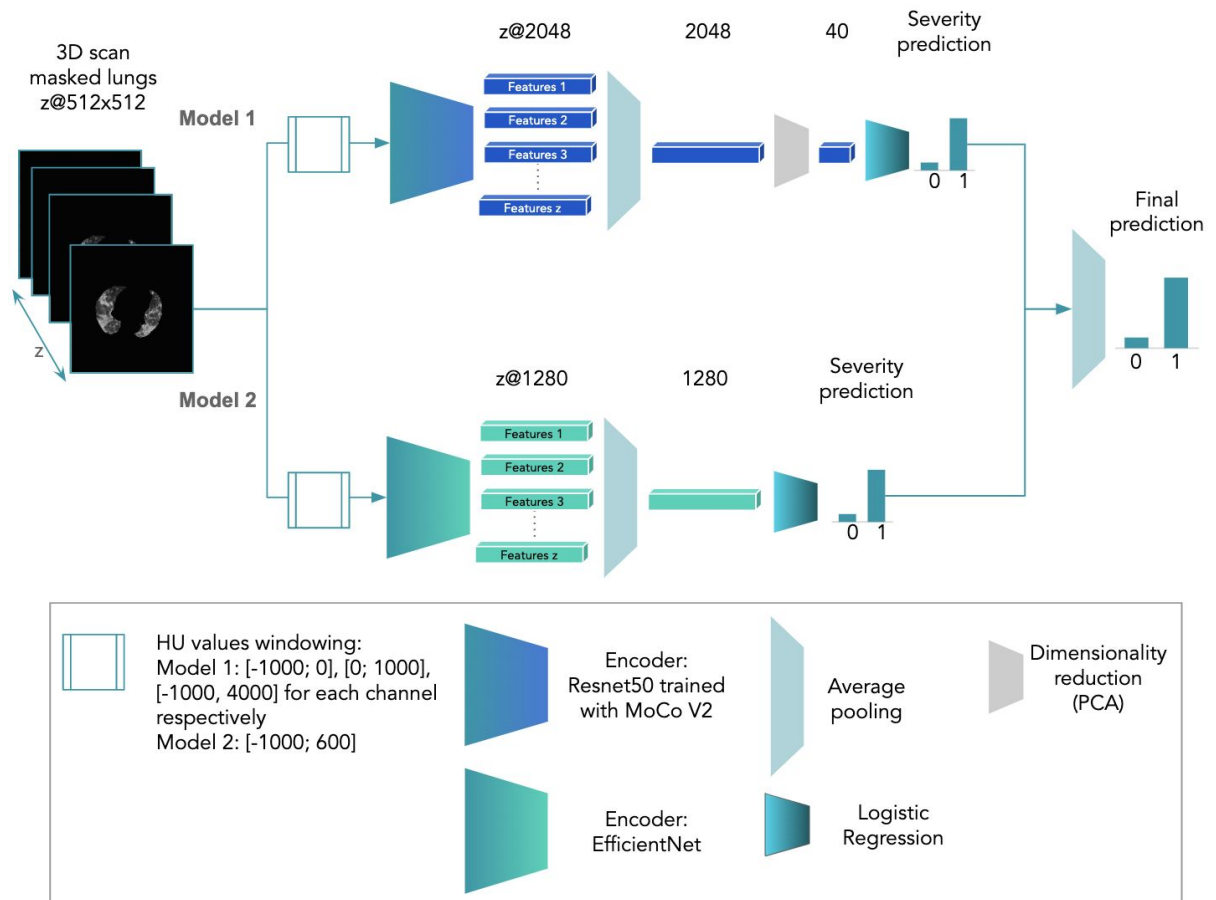
**Supp Fig. 1: Description of the retrospective cohort.** Number of patients and repartition per hospital for different all patients, patients



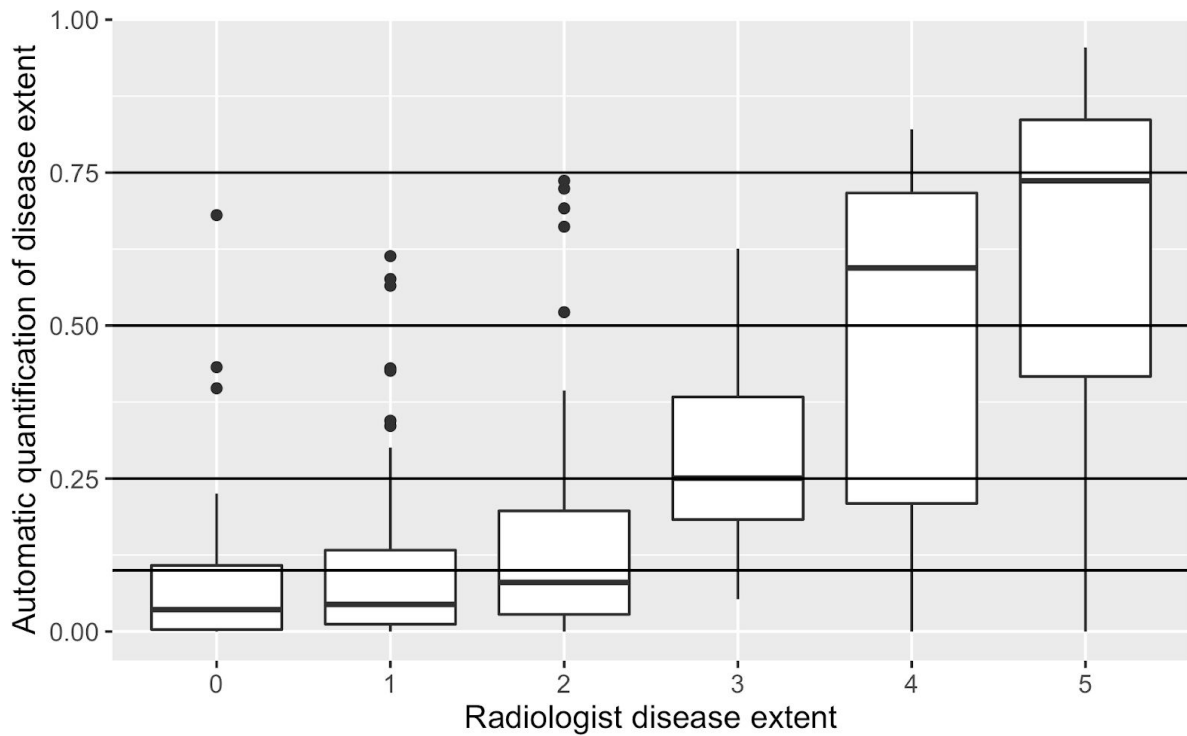
**Supp Fig. 2: Forest plot for the different variables measured at baseline.** For continuous variables, odds ratios are computed for an increase of one standard deviation of the continuous variable. KB odds ratios are in blue, IGR are in red.



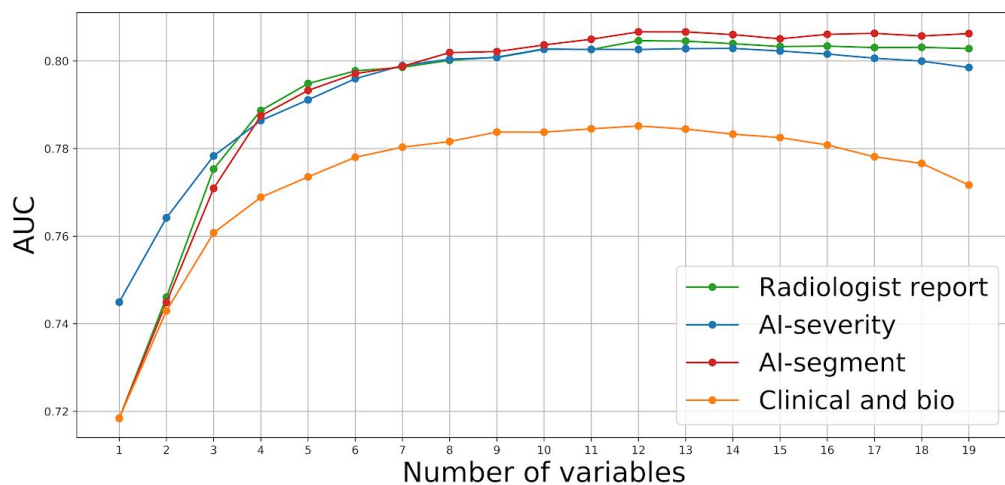
**Supp Fig. 3: AI-segment architecture** - Proposed pipeline to generate lesion volumetry estimates from patient CT scans employing ensemble of segmentation networks. Normalized patient scans are provided to our trained 2.5D U-Net and 3D ResNet50. The masks predicted from both models are then merged by arithmetic mean. In parallel, we segment left-right lungs from the patient scans using a dedicated U-Net. Finally, the left-right lung mask is used to mask-out lesions in left and right lungs from the ensemble output. This pipeline utilizes the complementary features learned by a weak model (2.5D U-Net) and a strong one (3D ResNet50).



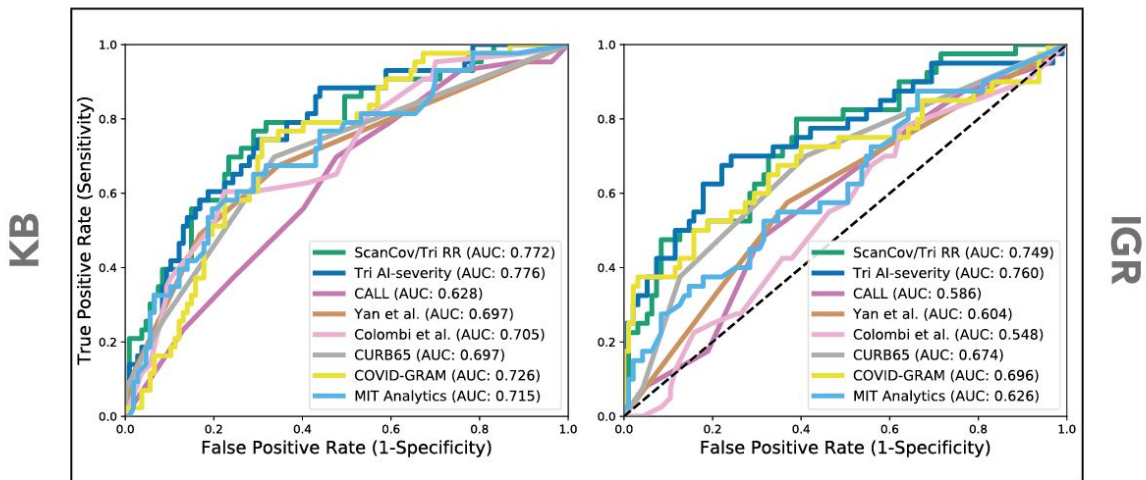
**Supp Fig. 4: AI-severity model to predict severity from 3D chest CT scans.** Two different pipelines were used: one using Resnet50 (trained with MocoV2 on 1 million public CT scan slices) as encoder (model 1) and one using EfficientNet B0 as encoder (model 2).



**Supp Fig. 5: Boxplot to compare automatic quantification of disease extent by *AI-segment* to disease extent as estimated by a radiologist.** The coding of disease extent in the radiologist report is as follows: 0 (0% of lesions), 1 (<10% of lesions), 2 (between 10 and 25% of lesions), 3 (between 25 and 50% of lesions), 4 (between 50 and 75% of lesions), 5 (more than 75% of lesions). The lower and upper hinges correspond to the first and third quartiles. The upper whisker extends from the hinge to the largest value no further than  $1.5 \times \text{IQR}$  from the hinge (where IQR is the inter-quartile range). The lower whisker extends from the hinge to the smallest value at most  $1.5 \times \text{IQR}$  of the hinge. Data beyond the end of the whiskers are called "outlying" points and are plotted individually.



**Supp Fig. 6: AUC curve as a function of the number of clinical and biological information added to the multimodal model.** Variables included in the models consist of CT scan variables only and then a greedy algorithm adds clinical or biological variables iteratively. At each step of the algorithm, the variable that results in the largest increase of AUC score is added.



**Supp Fig. 7: Receiver operating characteristic (ROC) curves of the models that predict severity.** Models were evaluated on two distinct validation sets consisting of 150 patients from KB (left panels) and 137 patients from IGR (right panels). ROC curves were obtained using the severity outcome defined as an oxygen flow rate of 15 L/min or higher, the need for mechanical ventilation, or death.

	<b>GGO</b>	<b>Crazy paving</b>	<b>Consolidation</b>
Accuracy (1% thresh.)	0.7951	0.7684	0.6167
F1 Score (1% thresh.)	0.8848	0.6452	0.7473
Accuracy (2% thresh.)	0.7876	0.7692	0.6667
F1 Score (2% thresh.)	0.8800	0.6182	0.7848

**Supp Table 1: Detection accuracy and F1 scores of *AI-segment* when considering the radiologist report as ground truth.** The binary decision used to compute the score is “presence or not of a lesion type”. Accuracy and F1 score are averaged over the IGR validation set. We compared, for each patient of the IGR validation set, detection obtained using *AI-segment* to the information provided in the standardized radiologist report. When using *AI-segment*, a lesion type is considered as present when its relative volume w.r.t. the full volume of both lung, is above a certain threshold indicated into parenthesis in the 1st column of the table.

<b>Variable</b>	<b>Center</b>	<b>Odds ratio</b> (95%lower - 95% upper)	<b>P-value</b>	<b>P-value Stouffer</b>
GGO AI	KB	0.64 (0.54,0.76)	4.28e-07	1.94e-07
GGO AI	IGR	0.77 (0.54,1.10)	0.15	
Crazy Paving AI	KB	1.47 (1.20,1.79)	0.00015	6.70e-05
Crazy Paving AI	IGR	1.31 (0.92,1.87)	0.13	
Consolidation AI	KB	1.46 (1.23,1.73)	1.59e-05	7.61e-06
Consolidation AI	IGR	1.27 (0.89,1.82)	0.19	
Disease extent AI	KB	2.11 (1.74,2.55)	2.97e-14	7.66e-16
Disease extent AI	IGR	1.90 (1.30,2.79)	0.00091	

**Supp Table 2: Association between severity and amount of lesions inferred by *AI-segment*.** For disease extent, we consider the proportion of the lung volume. For the other three variables (GGO, consolidation, crazy paving), we normalize them by disease extent so that each variable measures the proportion of the corresponding lesion.



Model description	KB	IGR	KB CV
<b>O<sub>2</sub>≥15L/min or Ventilation or Death</b>			
Radiologist report	0.713 (0.621 - 0.804)	0.647 (0.545 - 0.749)	0.702 (0.580 - 0.827)
AI-severity	0.756 (0.666 - 0.846)	0.748 (0.657 - 0.839)	0.745 (0.621 - 0.842)
AI-segment	0.666 (0.565 - 0.767)	0.696 (0.596 - 0.795)	0.712 (0.566 - 0.808)
Clinical and bio	0.736 (0.649 - 0.822)	0.745 (0.648 - 0.842)	0.784 (0.700 - 0.880)
ScanCov (Trimodal Radiologist report)	0.772 (0.688 - 0.855)	0.749 (0.658 - 0.839)	0.798 (0.720 - 0.903)
Trimodal AI-severity	0.776 (0.696 - 0.856)	0.760 (0.665 - 0.856)	0.803 (0.699 - 0.886)
Trimodal AI-segment	0.744 (0.658 - 0.830)	0.764 (0.672 - 0.857)	0.804 (0.708 - 0.894)
CALL	0.628 (0.533 - 0.723)	0.586 (0.483 - 0.689)	
Yan et al. (2020)	0.697 (0.607 - 0.788)	0.604 (0.508 - 0.701)	
Colombi et al. (2020)	0.705 (0.613 - 0.797)	0.548 (0.444 - 0.653)	
CURB65	0.697 (0.604 - 0.789)	0.674 (0.579 - 0.770)	
COVID-GRAM	0.726 (0.642 - 0.811)	0.696 (0.587 - 0.804)	
MIT analytics	0.715 (0.623 - 0.807)	0.626 (0.521 - 0.732)	
<b>Death</b>			
Radiologist report	0.656 (0.548 - 0.764)	0.652 (0.532 - 0.772)	0.659 (0.499 - 0.815)
AI-severity	0.752 (0.660 - 0.845)	0.753 (0.644 - 0.862)	0.710 (0.630 - 0.835)
AI-segment	0.635 (0.517 - 0.753)	0.662 (0.532 - 0.793)	0.660 (0.532 - 0.796)
Clinical and bio	0.773 (0.685 - 0.861)	0.888 (0.824 - 0.952)	0.800 (0.690 - 0.915)
ScanCov (Trimodal Radiologist report)	0.814 (0.732 - 0.896)	0.837 (0.755 - 0.920)	0.792 (0.695 - 0.925)
Trimodal AI-severity	0.815 (0.742 - 0.889)	0.890 (0.829 - 0.950)	0.800 (0.718 - 0.906)
Trimodal AI-segment	0.778 (0.694 - 0.862)	0.897 (0.837 - 0.957)	0.797 (0.708 - 0.926)
CALL	0.711 (0.606 - 0.815)	0.648 (0.544 - 0.753)	
Yan et al. (2020)	0.646 (0.538 - 0.754)	0.706 (0.598 - 0.815)	
Colombi et al. (2020)	0.788 (0.700 - 0.876)	0.585 (0.468 - 0.702)	
CURB65	0.812 (0.736 - 0.888)	0.700 (0.592 - 0.808)	
COVID-GRAM	0.781 (0.691 - 0.872)	0.789 (0.688 - 0.891)	
MIT analytics	0.853 (0.786 - 0.920)	0.644 (0.532 - 0.756)	
<b>Death or ICU</b>			
Radiologist report	0.713 (0.628 - 0.799)	0.736 (0.639 - 0.833)	0.720 (0.603 - 0.835)
AI-severity	0.767 (0.685 - 0.849)	0.825 (0.748 - 0.902)	0.749 (0.632 - 0.857)
AI-segment	0.676 (0.583 - 0.769)	0.767 (0.675 - 0.858)	0.724 (0.594 - 0.820)
Clinical and bio	0.768 (0.687 - 0.850)	0.812 (0.732 - 0.892)	0.781 (0.679 - 0.876)
ScanCov (Trimodal Radiologist report)	0.781 (0.703 - 0.860)	0.830 (0.755 - 0.905)	0.794 (0.716 - 0.892)
Trimodal AI-severity	0.806 (0.732 - 0.880)	0.844 (0.771 - 0.917)	0.805 (0.721 - 0.896)
Trimodal AI-segment	0.764 (0.684 - 0.844)	0.848 (0.776 - 0.920)	0.802 (0.728 - 0.902)
CALL	0.584 (0.488 - 0.680)	0.592 (0.488 - 0.695)	
Yan et al. (2020)	0.723 (0.641 - 0.806)	0.670 (0.576 - 0.763)	
Colombi et al (2020)	0.682 (0.592 - 0.772)	0.588 (0.486 - 0.689)	
CURB65	0.688 (0.598 - 0.778)	0.709 (0.618 - 0.800)	
COVID-GRAM	0.718 (0.632 - 0.803)	0.702 (0.600 - 0.805)	
MIT Analytics	0.712 (0.623 - 0.800)	0.656 (0.558 - 0.753)	

**Supp Table 3: AUC values for the different models on the different sets.** Each model was trained on 646 patients from KB. Results are reported on the validation set from KB (150 patients) and the external validation set from IGR (135 patients), as well as on the training set using 5 fold cross validation stratified by outcome and age (CV KB).

Variable	AUC on KB validation set	AUC on IGR validation set
Age > 60	0.884 (0.828 - 0.940)	0.786 (0.710 - 0.862)
Sex	0.933 (0.892 - 0.975)	0.893 (0.838 - 0.947)
Oxygen saturation > 90	0.761 (0.681 - 0.840)	0.782 (0.676 - 0.888)
Disease extent > 2	0.926 (0.887 - 0.965)	0.881 (0.819 - 0.943)
Crazy paving	0.775 (0.700 - 0.851)	0.725 (0.637 - 0.812)
Condensation	0.6365 (0.534 - 0.737)	0.675 (0.583 - 0.767)
GGO	0.800 (0.655 - 0.944)	0.583 (0.475 - 0.690)

**Supp Table 4: AI-severity model performances on other classification tasks than severity prediction.** AUC scores are reported on both KB and IGR validation sets when re-training the AI-severity model to predict a few clinical and radiological variables we have selected. We considered the feature vector of AI-severity obtained when

Models	Variables included									
ScanCovIA	Oxygen saturation	Disease extent	Age	Sex	Platelet	Urea				
Tri AI-severity	AI-severity	Oxygen saturation	Urea	Sex	Platelet	Age	LDH	Diastolic pressure	Hypertension	Neutrophil
Tri AI-segment	Oxygen saturation	Consolidation AI	Age	Sex	Platelet	GGO AI	Urea	LDH	Crazy paving AI	Dyspnea
Clinical and bio (C & B)	Oxygen saturation	Age	Sex	LDH	Platelet	Chronic kidney disease	Dyspnea	Hypertension	Neutrophil	Urea

**Supp Table 5: Names of the variables included in the 4 different models.**

Variable	Coding/unit	Transformation	Coefficient
Oxygen saturation	%	$-\log(1 + 100 - X)$	-0.745
Disease extent	0 to 5 scale	None	0.611
Age		None	0.025
Sex	1 for male 0 for woman	None	0.545
Platelet	G/L	$\log(0.001 + X)$	-0.838
Urea	mmol/L	$\log(0.001 + X)$	0.608

**Supp Table 6: Coefficients, transformation, and units to compute the ScanCov score.** Disease extent values can be: 0 (no extent of disease) / 1 (<10%) / 2 (10-24%) / 3 (25-49%) / 4 (50-74%) / 5 >75%.

Variable	Correlation	Lower C.I.	Upper C.I
LDH	0.52	0.45	0.58
CRP	0.45	0.39	0.51
Oxygen saturation	-0.43	-0.49	-0.37
Ferritin	0.3	0.18	0.4
Monocyte	-0.23	-0.29	-0.15
Respiratory rate	0.18	0.1	0.25
Weight	0.16	0.07	0.24
Neutrophil	0.15	0.08	0.22
Symptoms duration before examination	0.12	0.05	0.19
BMI	0.12	0.02	0.22
Height	0.12	0.02	0.21
Total bilirubin	0.1	0.02	0.17
Leucocytes	0.09	0.02	0.17
Lymphocyte	-0.09	-0.16	-0.01
Cardiac frequency	0.08	0.01	0.15
Urea	0.07	-0.01	0.14
Conjugated bilirubin	0.07	-0.21	0.34
Creatine kinase	0.06	-0.02	0.14
Haemoglobin	0.05	-0.02	0.12
Body temperature	0.05	-0.02	0.12
Platelet	0.03	-0.04	0.11
Systolic pressure	-0.03	-0.1	0.04
Age	0.02	-0.05	0.09
Diastolic pressure	-0.01	-0.08	0.06

**Supp Table 7: Correlation of clinical and biological variables with a radiologist quantification of disease extent.** Correlation was computed using 817 patients from the KB hospital. Variables are sorted in decreasing order when considering the squared correlation value for ranking.