

Supplementary Material

An externally validated fully automated deep learning algorithm to classify COVID-19 and other pneumonias on chest CT

Akshayaa Vaidyanathan ^{1,2,†}, Julien Guiot ^{3,†}, Fadila Zerka ^{1,2}, Flore Belmans ¹, Ingrid Van Peuflück ¹, Louis Deprez ⁴, Denis Danthine ⁴, Gregory Canivet ⁵, Philippe Lambin ², Sean Walsh ¹, Mariaelena Occchipinti ¹, Paul Meunier ³, Wim Vos ¹, Pierre Lovinfosse ⁶ & Ralph T.H. Leijenaar ¹

¹Radiomics (Oncoradiomics SA), Liège, Belgium

²The D-Lab, Department of Precision Medicine, Department of Radiology & Nuclear Medicine, GROW–School for Oncology, Maastricht University, Maastricht, The Netherlands

³Department of Pneumology, University Hospital of Liège, Liège, Belgium

⁴Department of Radiology, University Hospital of Liège, Liège, Belgium

⁵Department of Computer Applications, University Hospital of Liège, Liège, Belgium

⁶Department of Nuclear Medicine and Oncological Imaging, University Hospital of Liège, Liège, Belgium

* Correspondence:

Akshayaa Vaidyanathan,

akshayaa.vaidyanathan@radiomics.bio

Radiomics, Clos Chanmurly 13, 4000 Liège

[†]These authors have contributed equally to this work and share first authorship

Lung Segmentation

The lungs were segmented as a single structure using RadiomiX (Oncoradiomics SA, Liège, Belgium) based on convolutional neural networks by combining 3D and 2D architectures. Details on both these architectures are given below.

3D lung segmentation

This model architecture consists of a 3D U-Net (1) with residual blocks (2) in the encoder part of the network. Publicly available data from the cancer imaging archive (3) was used to train and validate the model. The specific dataset contains CT scans of 422 confirmed non-small cell lung cancer cases, along with manual segmentations of the left and right lungs. The segmentations were performed by an experienced radiologist and these segmentations were used as a reference standard. The data was randomly partitioned into a training set (n = 322), a tuning set (n = 50), and a test set (n = 50). In order to generate homogeneous CT volumes as input for the model, the following pre-processing steps were performed. All the volumes were resized to 160 x 160 x 448 along the x, y and z axis and image intensities were clipped at a window width of 1500 HU and a window level of -600 HU (i.e., a standard lung CT window level settings).

The following data augmentations were performed to avoid overfitting (4) on the training dataset:

- Flipping in different directions: up and down, left and right

- Randomly resampling volumes to varying voxel sizes and resize to dimensions (160 x 160 x 448)
- Rotating (10-30 degrees) onto the left or right direction
- Reversing the sequence of axial slices

The model was trained with the pre-processed volumes and their corresponding reference labels, using Jaccard loss (5) as an objective function. Here, the loss is calculated in a mini batch of two images per iteration. The network was trained for 10 epochs and at the end of each epoch the Jaccard loss was calculated on the model's predictions to ensure validation loss convergence.

2D lung segmentation

The 2D lung segmentation model architecture is based on a 2D Feature Pyramid Network (6) adapted with ResNext blocks (7) in the encoder. The model was trained and validated on the following datasets,

- 1) Publicly available dataset with 888 CT scans and the corresponding reference annotations for lungs available from LUNA16 challenge (8)
- 2) Publicly available data from the cancer imaging archive (3) containing CT scans of 422 confirmed non-small cell lung cancer cases, along with manual segmentations of the left and right lungs. The segmentations were performed by an experienced radiologist and these segmentations were used as a reference standard.

The network was trained with the 2D axial slices clipped at a window width of 1500 HU and a window level of -600 HU and with their corresponding reference labels. The network's weights were updated by using the Adam optimizer at an initial learning rate of 1e-5 (9). The model was trained using customized Jaccard loss (5) as an objective function where the loss is calculated in a mini batch of 8 images per iteration. The network was trained for 5 epochs and at the end of each epoch, the Jaccard loss was calculated on the model's predictions to ensure validation loss convergence.

The predicted segmentations of each architecture (i.e., the segmentation output from both the 3D and the 2D segmentation models) were ensembled and the intersection constitutes the final total lung segmentation which is used for extraction of radiomics features. The deep learning-based lung segmentation achieved a mean Dice similarity coefficient score of 0.92 across the publicly available datasets which indicates adequate precision (i.e. no significant over or under segmentation).

Lung abnormalities segmentation

The segmentation model is based on 2D U-Net combined with Res Next as encoder and deep supervision and was trained on axial unenhanced chest CT scans of 199 COVID-19 patients coming from three different centres in three different countries (10). The model's performance was evaluated on an external test set of 50 COVID-19 patients coming from several different centres in Moscow, Russia (11). All datasets are open source, freely available online. An automatic in-house lung segmentation model (see above 1 Lung segmentation) was used to crop the lung region from the CT volumes. Axial slices with no segmented lung regions were removed from the volumes.

Different sets of 48 consecutive axial slices with an overlap of 10 slices between one set and the other (extracted from the whole volume) were used to train the model. Each set contains at least one slice with lung abnormalities. Each data point containing the consecutive axial slices was pre-processed in following ways to obtain a three channel input to the model:

- The first channel contains slices with intensities clipped at lung window level settings (W:1500 HU, L:-600 HU) with lungs and the abnormalities cropped.
- The second channel contains the slices with original intensities with lungs and abnormalities cropped.
- The third channel contains slices with intensities clipped at Mediastinal window level settings (W:350 HU, L:50 HU) with the region containing the lungs cropped. A rectangular crop was obtained with x_{min} = minimum x value for which lungs or lung abnormalities pixels are present, x_{max} = maximum x value for which lungs or lung abnormalities pixels are present and y_{min} = minimum y value for which lungs or lung abnormalities pixels are present, y_{max} = maximum y value for which lungs or lung abnormalities pixels are present. Fig.1 report an example of the input for the three channels.

The automatic deep learning segmentation algorithm achieved good performances (mean DSC 0.6 ± 0.1) on the external test set.

Figures and Tables

Table S1. Summary of imaging parameters for the training and external validation datasets

	Training set	External validation set
Manufacturer		
SIEMENES	60 %	100 %
GE MEDICAL SYSTEM	40 %	-
PHILIPS	< 1 %	-
Kernel		
B30f	-	7 %
B41s	-	45 %
D40s	-	48 %
B30s	2 %	-
Br32f	17 %	-
Br59f	12 %	-
Tr20f	13 %	-
LUNG	11 %	-
STANDARD	31 %	-
Others	14 %	-
Pixel spacing (mm)		
< 0.5	1 %	-
0.5 to 0.6	7 %	11 %
0.6 to 0.7	21 %	48 %
0.7 to 0.8	16 %	38 %
> 0.8	55 %	3 %
Slice thickness (mm)		
2	100 %	100 %

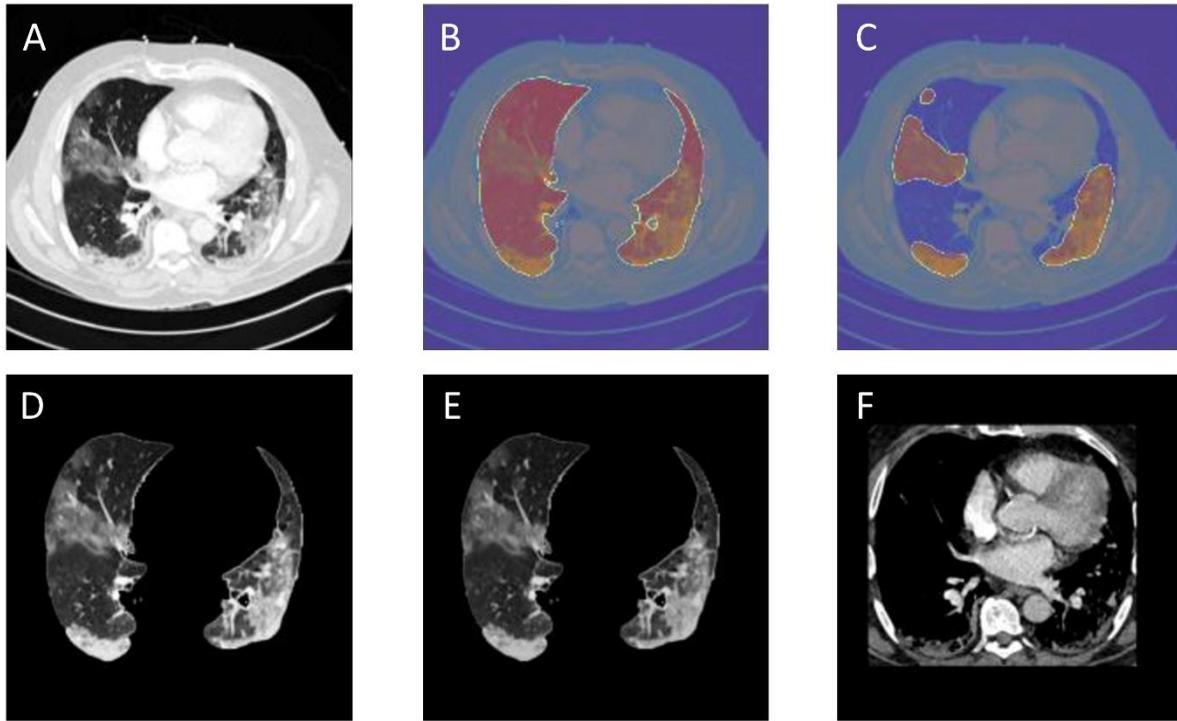


Fig. S1 Lungs plus abnormalities segmentation on a slice from a COVID-19 patient. A) Original axial slice from case with COVID-19 label; B) Lung segmentation obtained on the same slice; C) Ground Glass Opacities segmented by the lung abnormalities model. Three channel input obtained from the same slice, D) Channel 1; E) Channel 2; F) Channel 3.

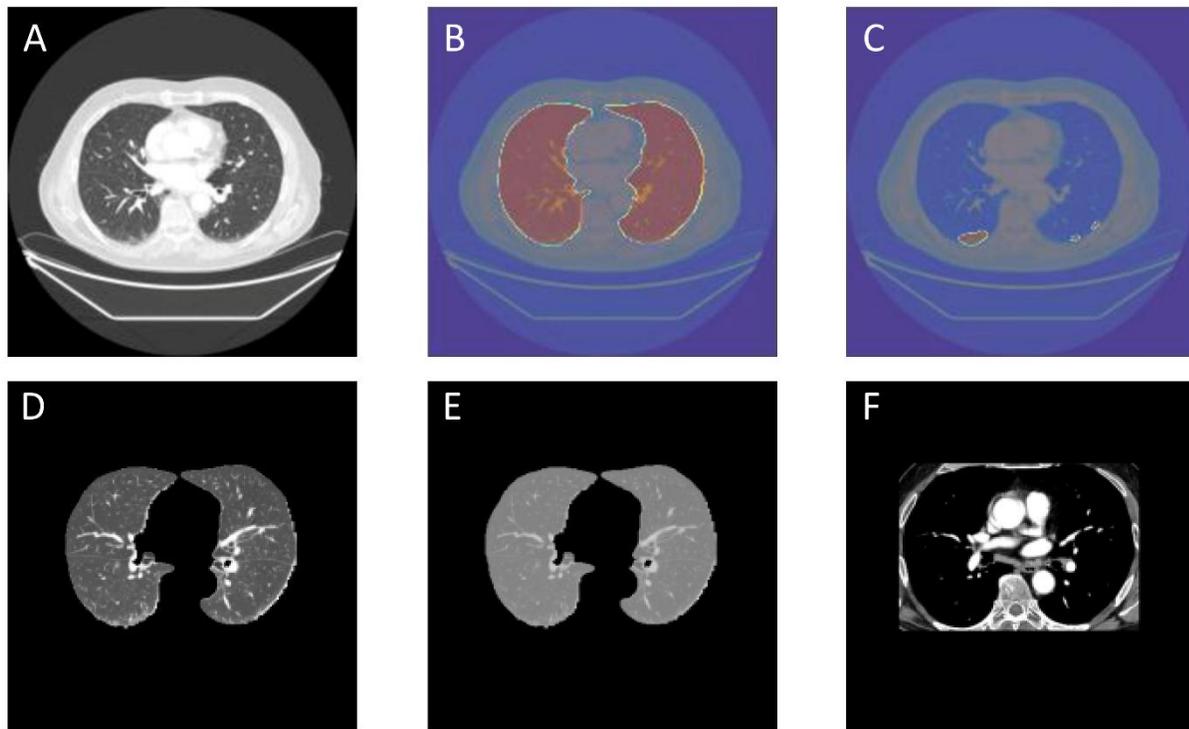


Fig. S2 Lungs and abnormalities segmentation on a slice from a No Infection patient. A) Original axial slice from case with No Infection label; B) Lung segmentation obtained on the same slice; C) Aspecific abnormalities segmented by the lung abnormalities model; Three channel input obtained from the same slice, D) Channel 1; E) Channel 2; F) Channel 3.

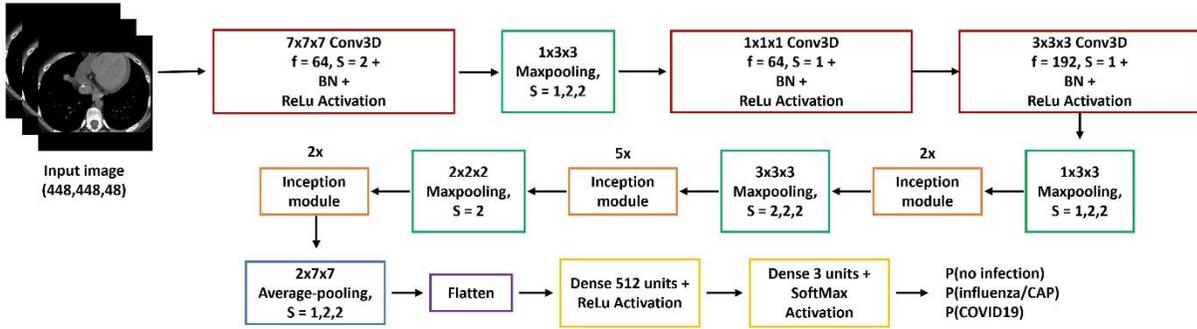


Fig. S3 3D CNN model network. Inception module architecture is based on the implementation described in [13]. Convolution blocks (red); Maxpooling blocks (green); Inception modules (orange); Average pooling layer (blue); Flatten layer (purple); Fully connected layers (yellow).

Patient Information

Patient ID: PATIENT1

Scan ID: Scan 1

Scan date: DD/MM/YYYY

Diagnosis: Influenza/CAP

Probability: [N: 0.02, Inf/CAP: 0.69, COVID: 0.29]

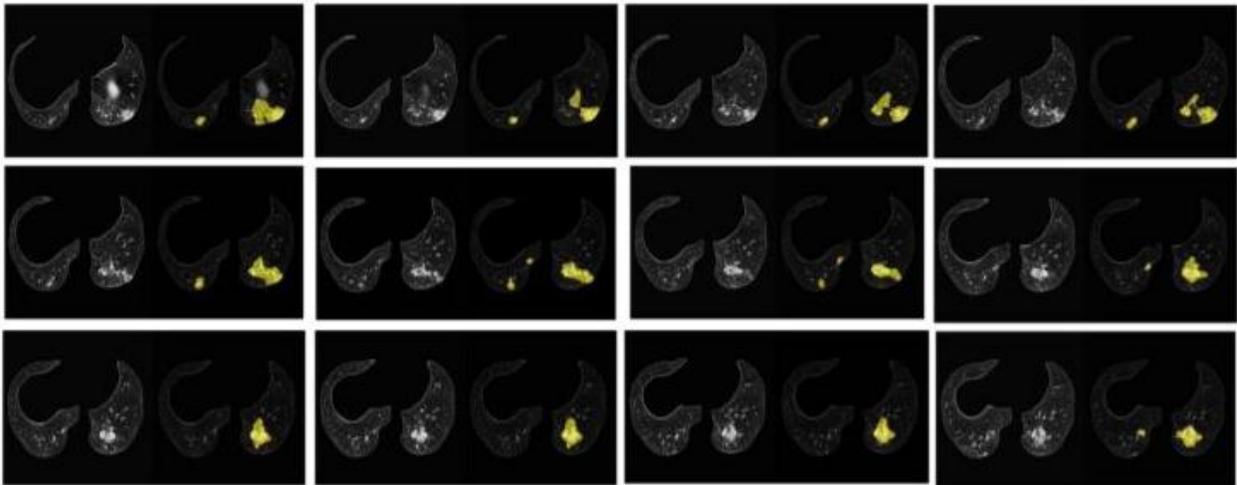


Fig. S4 Example of clinical summary report for Influenza/CAP patient.

Patient Information

Patient ID: PATIENT1

Scan ID: Scan 1

Scan date: DD/MM/YYYY

Diagnosis: COVID-19

Probability: [N: 0.02, Inf/CAP: 0.10, COVID: 0.87]

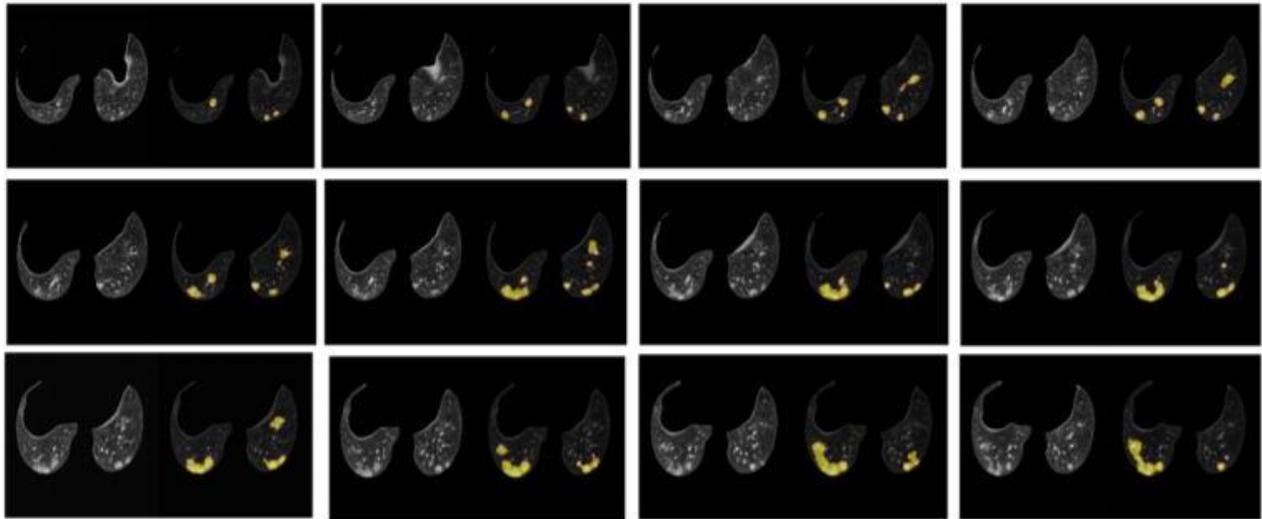


Fig. S5 Example of clinical summary report for COVID-19 patient.

TRIPOD checklist

		Development [D]	External validation [V]	Combined Development & External validation [D+V]
Y=yes; N=no; R=referenced; NA=not applicable				
Title and abstract				
1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.			0
i	The words developing/development, validation/validating, incremental/added value (or synonyms) are reported in the title	Y	Y	Y
ii	The words prediction, risk prediction, prediction model, risk models, prognostic models, prognostic indices, risk scores (or synonyms) are reported in the title	N	N	N
iii	The target population is reported in the title	N	N	N
iv	The outcome to be predicted is reported in the title	Y	Y	Y
2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.			0
i	The objectives are reported in the abstract	Y	Y	Y
ii	Sources of data are reported in the abstract <i>E.g. Prospective cohort, registry data, RCT data.</i>	Y	Y	Y
iii	The setting is reported in the abstract <i>E.g. Primary care, secondary care, general population, adult care, or paediatric care. The setting should be reported for both the development and validation datasets, if applicable.</i>	Y	Y	Y
iv	A general definition of the study participants is reported in the abstract <i>E.g. patients with suspicion of certain disease, patients with a specific disease, or general eligibility criteria.</i>	Y	Y	Y
v	The overall sample size is reported in the abstract	Y	Y	Y

vi	The number of events (or % outcome together with overall sample size) is reported in the abstract <i>If a continuous outcome was studied, score Not applicable (NA).</i>	Y	Y	Y
vii	Predictors included in the final model are reported in the abstract. For validation studies of well-known models, at least the name/acronym of the validated model is reported <i>Broad descriptions are sufficient, e.g. ‘all information from patient history and physical examination’. Check in the main text whether all predictors of the final model are indeed reported in the abstract.</i>	N	N	N
viii	The outcome is reported in the abstract	Y	Y	Y
ix	Statistical methods are described in the abstract <i>For model development, at least the type of statistical model should be reported. For validation studies a quote like “model’s discrimination and calibration was assessed” is considered adequate. If done, methods of updating should be reported.</i>	N	N	N
x	Results for model discrimination are reported in the abstract <i>This should be reported separately for development and validation if a study includes both development and validation.</i>	Y	Y	Y
xi	Results for model calibration are reported in the abstract <i>This should be reported separately for development and validation if a study includes both development and validation.</i>	N	N	N
xii	Conclusions are reported in the abstract <i>In publications addressing both model development and validation, there is no need for separate conclusions for both; one conclusion is sufficient.</i>	Y	Y	Y
3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.			1
i	The background and rationale are presented	Y	Y	Y
ii	Reference to existing models is included (or stated that there are no existing models)	Y	Y	Y

3b	Specify the objectives, including whether the study describes the development or validation of the model or both.			1
i	It is stated whether the study describes development and/or validation and/or incremental (added) value	Y	Y	Y
Methods				
4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.			1
i	The study design/source of data is described <i>E.g. Prospectively designed, existing cohort, existing RCT, registry/medical records, case control, case series. This needs to be explicitly reported; reference to this information in another article alone is insufficient.</i>	Y	Y	Y
4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.			1
i	The starting date of accrual is reported	Y	Y	Y
ii	The end date of accrual is reported	Y	Y	Y
iii	The length of follow-up <u>and</u> prediction horizon/time frame are reported, if applicable <i>E.g. "Patients were followed from baseline for 10 years" and "10-year prediction of..."; notably for prognostic studies with long term follow-up. If this is not applicable for an article (i.e. diagnostic study or no follow-up), then score Not applicable (NA).</i>	NA	NA	NA
5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.			1
i	The study setting is reported (e.g. primary care, secondary care, general population) <i>E.g.: 'surgery for endometrial cancer patients' is considered to be enough information about the study setting.</i>	R	R	R
ii	The number of centres involved is reported <i>If the number is not reported explicitly, but can be concluded from the name of the centre/centres, or if clearly a single centre study, score Yes.</i>	Y	Y	Y

iii	The geographical location (at least country) of centres involved is reported <i>If no geographical location is specified, but the location can be concluded from the name of the centre(s), score Yes.</i>	Y	Y	Y
5b	Describe eligibility criteria for participants.			0
i	In-/exclusion criteria are stated <i>These should explicitly be stated. Reasons for exclusion only described in a patient flow is not sufficient.</i>	N	N	N
5c	Give details of treatments received, if relevant. <i>(i.e. notably for prognostic studies with long term follow-up)</i>			Not applicable
i	Details of any treatments received are described <i>This item is notably for prognostic modelling studies and is about treatment at baseline or during follow-up. The 'if relevant' judgment of treatment requires clinical knowledge and interpretation. If you are certain that treatment was not relevant, e.g. in some diagnostic model studies, score Not applicable.</i>	NA	NA	NA
6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.			1
i	The outcome definition is clearly presented <i>This should be reported separately for development and validation if a publication includes both.</i>	Y	Y	Y
ii	It is described how outcome was assessed (including all elements of any composite, for example CVD [e.g. MI, HF, stroke]).	R	R	R
iii	It is described when the outcome was assessed (time point(s) since T0)	R	R	R
6b	Report any actions to blind assessment of the outcome to be predicted.			0
i	Actions to blind assessment of outcome to be predicted are reported <i>If it is clearly a non-issue (e.g. all-cause mortality or an outcome not requiring interpretation), score Yes. In all other instances, an explicit mention is expected.</i>	N	N	N
7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.			0

i	All predictors are reported <i>For development, “all predictors” refers to all predictors that potentially could have been included in the ‘final’ model (including those considered in any univariable analyses). For validation, “all predictors” means the predictors in the model being evaluated.</i>	N	N	N
ii	Predictor definitions are clearly presented	Y	Y	Y
iii	It is clearly described how the predictors were measured	Y	Y	Y
iv	It is clearly described when the predictors were measured	N	N	N
7b	Report any actions to blind assessment of predictors for the outcome and other predictors.			0
i	It is clearly described whether predictor assessments were blinded for outcome <i>For predictors for which it is clearly a non-issue (e.g. automatic blood pressure measurement, age, sex) and for instances where the predictors were clearly assessed before outcome assessment, score Yes. For all other predictors an explicit mention is expected.</i>	N	N	N
ii	It is clearly described whether predictor assessments were blinded for the other predictors	N	N	N
8	Explain how the study size was arrived at.			1
i	It is explained how the study size was arrived at <i>Is there any mention of sample size, e.g. whether this was done on statistical grounds or practical/logistical grounds (e.g. an existing study cohort or data set of a RCT was used)?</i>	Y	Y	Y
9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.			0

i	<p>The method for handling missing data (predictors and outcome) is mentioned</p> <p><i>E.g. Complete case (explicit mention that individuals with missing values have been excluded), single imputation, multiple imputation, mean/median imputation.</i></p> <p><i>If there is no missing data, there should be an explicit mention that there is no missing data for all predictors and outcome. If so, score Yes.</i></p> <p><i>If it is unclear whether there is missing data (from e.g. the reported methods or results), score No.</i></p> <p><i>If it is clear there is missing data, but the method for handling missing data is unclear, score No.</i></p>	N	N	N
ii	<p>If missing data were imputed, details of the software used are given</p> <p><i>When under 9i explicit mentioning of no missing data, complete case analysis or no imputation applied, score Not applicable.</i></p>	NA	NA	NA
iii	<p>If missing data were imputed, a description of which variables were included in the imputation procedure is given</p> <p><i>When under 9i explicit mentioning of no missing data, complete case analysis or no imputation applied, score Not applicable.</i></p>	NA	NA	NA
iv	<p>If multiple imputation was used, the number of imputations is reported</p> <p><i>When under 9i explicit mentioning of no missing data, complete case analysis or no imputation applied, score Not applicable.</i></p>	NA	NA	NA
10a	Describe how predictors were handled in the analyses.			1
i	<p>For continuous predictors it is described whether they were modelled as linear, nonlinear (type of transformation specified) or categorized</p> <p><i>A general statement is sufficient, no need to describe this for each predictor separately.</i></p> <p><i>If no continuous predictors were reported, score Not applicable.</i></p>	NA	Not applicable	NA
ii	<p>For categorical or categorized predictors, the cut-points were reported</p> <p><i>If no categorical or categorized predictors were reported, score Not applicable.</i></p>	Y	Not applicable	Y

iii	For categorized predictors the method to choose the cut-points was clearly described <i>If no categorized predictors, score Not applicable.</i>	NA	Not applicable	NA
10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.			0
i	The type of statistical model is reported <i>E.g. Logistic, Cox, other regression model (e.g. Weibull, ordinal), other statistical modelling (e.g. neural network)</i>	Y	Not applicable	Y
ii	The approach used for predictor selection <u>before</u> modelling is described <i>'Before modelling' means before any univariable or multivariable analysis of predictor-outcome associations.</i> <i>If no predictor selection before modelling is done, score Not applicable.</i> <i>If it is unclear whether predictor selection before modelling is done, score No.</i> <i>If it is clear there was predictor selection before modelling but the method was not described, score No.</i>	NA	Not applicable	NA
iii	The approach used for predictor selection <u>during</u> modelling is described <i>E.g. Univariable analysis, stepwise selection, bootstrap, Lasso.</i> <i>'During modelling' includes both univariable or multivariable analysis of predictor-outcome associations.</i> <i>If no predictor selection during modelling is done (so-called full model approach), score Not applicable.</i> <i>If it is unclear whether predictor selection during modelling is done, score No.</i> <i>If it is clear there was predictor selection during modelling but the method was not described, score No.</i>	N	Not applicable	N
iv	Testing of interaction terms is described <i>If it is explicitly mentioned that interaction terms were not addressed in the prediction model, score Yes.</i> <i>If interaction terms were included in the prediction model, but the testing is not described, score No.</i>	N	Not applicable	N
v	Testing of the proportionality of hazards in survival models is described <i>If no proportional hazard model is used, score Not</i>	N	Not applicable	N

	<i>applicable.</i>			
vi	Internal validation is reported <i>E.g. Bootstrapping, cross validation, split sample. If the use of internal validation is clearly a non-issue (e.g. in case of very large data sets), score Yes. For all other situations an explicit mention is expected.</i>	N	Not applicable	N
10c	For validation, describe how the predictions were calculated.			0
i.	It is described how predictions for individuals (in the validation set) were obtained from the model being validated <i>E.g. Using the original reported model coefficients with or without the intercept, and/or using updated or refitted model coefficients, or using a nomogram, spreadsheet or web calculator.</i>	Not applicable	N	N
10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models. <i>These should be described in methods section of the paper (item 16 addresses the reporting of the results for model performance).</i>			0
i	Measures for model discrimination are described <i>E.g. C-index / area under the ROC curve.</i>	Y	Y	Y
ii	Measures for model calibration are described <i>E.g. calibration plot, calibration slope or intercept, calibration table, Hosmer Lemeshow test, O/E ratio.</i>	N	N	N
iii	Other performance measures are described <i>E.g. R2, Brier score, predictive values, sensitivity, specificity, AUC difference, decision curve analysis, net reclassification improvement, integrated discrimination improvement, AIC.</i>	Y	Y	Y
10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.			Not applicable
i	A description of model-updating is given <i>E.g. Intercept recalibration, regression coefficient recalibration, refitting the whole model, adding a new predictor If updating was done, it should be clear which updating method was applied to score Yes. If it is not explicitly mentioned that updating was</i>	Not applicable	NA	NA

	<i>applied in the study, score this item as 'Not applicable'.</i>			
11	Provide details on how risk groups were created, if done. <i>If risk groups were not created, score this item as Yes.</i>			0
i	If risk groups were created, risk group boundaries (risk thresholds) are specified <i>Score this item separately for development and validation if a study includes both development and validation.</i> <i>If risk groups were not created, score this item as not applicable.</i>	N	N	N
12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome and predictors.			0
i	Differences or similarities in <u>definitions</u> with the development study are described <i>Mentioning of any differences in all four (setting, eligibility criteria, predictors and outcome) is required to score Yes.</i> <i>If it is explicitly mentioned that there were no differences in setting, eligibility criteria, predictors and outcomes, score Yes.</i>	Not applicable	N	N
Results				
13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.			0
i	The flow of participants is reported	Y	Y	Y
ii	The number of participants with and without the outcome are reported <i>If outcomes are continuous, score Not applicable.</i>	N	N	N
iii	A summary of follow-up time is presented <i>This notably applies to prognosis studies and diagnostic studies with follow-up as diagnostic outcome.</i> <i>If this is not applicable for an article (i.e. diagnostic</i>	NA	NA	NA

	<i>study or no follow-up), then score Not applicable.</i>			
13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.			0
i	Basic demographics are reported	Y	Y	Y
ii	Summary information is provided for all predictors included in the final developed/validated model	Y	Y	Y
iii	The number of participants with missing data for predictors is reported	N	N	N
iv	The number of participants with missing data for the outcome is reported	N	N	N
13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).			0
i	Demographic characteristics (at least age and gender) of the validation study participants are reported along with those of the original development study	Not applicable	Y	Y
ii	Distributions of predictors in the model of the validation study participants are reported along with those of the original development study	Not applicable	Y	Y
iii	Outcomes of the validation study participants are reported along with those of the original development study	Not applicable	N	N
14a	Specify the number of participants and outcome events in each analysis.			1
i	The number of participants in each analysis (e.g. in the analysis of each model if more than one model is developed) is specified	Y	Not applicable	Y
ii	The number of outcome events in each analysis is specified (e.g. in the analysis of each model if more than one model is developed) <i>If outcomes are continuous, score Not applicable.</i>	NA	Not applicable	NA
14b	If done, report the unadjusted association between each candidate predictor and outcome.			Not applicable

i	The unadjusted associations between each predictor and outcome are reported <i>If any univariable analysis is mentioned in the methods but not in the results, score No. If nothing on univariable analysis (in methods or results) is reported, score this item as Not applicable.</i>	NA	Not applicable	NA
15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).			0
i	The regression coefficient (or a derivative such as hazard ratio, odds ratio, risk ratio) for each predictor in the model is reported	N	Not applicable	N
ii	The intercept or the cumulative baseline hazard (or baseline survival) for at least one time point is reported	N	Not applicable	N
15b	Explain how to use the prediction model.			1
i	An explanation (e.g. a simplified scoring rule, chart, nomogram of the model, reference to online calculator, or worked example) is provided to explain how to use the model for individualised predictions.	Y	Not applicable	Y
16	Report performance measures (with confidence intervals) for the prediction model. <i>These should be described in results section of the paper (item 10 addresses the reporting of the methods for model performance).</i>			0
i	A discrimination measure is presented <i>E.g. C-index / area under the ROC curve.</i>	Y	Y	Y
ii	The confidence interval (or standard error) of the discrimination measure is presented	N	Y	N
iii	Measures for model calibration are described <i>E.g. calibration plot, calibration slope or intercept, calibration table, Hosmer Lemeshow test, O/E ratio.</i>	N	N	N
iv	Other model performance measures are presented <i>E.g. R2, Brier score, predictive values, sensitivity, specificity, AUC difference, decision curve analysis, net reclassification improvement, integrated discrimination improvement, AIC.</i>	Y	Y	Y

17	If done, report the results from any model updating (i.e., model specification, model performance, recalibration). <i>If updating was not done, score this TRIPOD item as 'Not applicable'.</i>			Not applicable
0	Model updating was done <i>If "No", then answer 17i-17v with "Not applicable"</i>	Not applicable	N	N
i	The updated regression coefficients for each predictor in the model are reported <i>If model updating was described as 'not needed', score Yes.</i>	Not applicable	NA	NA
ii	The updated intercept or cumulative baseline hazard or baseline survival (for at least one time point) is reported <i>If model updating was described as 'not needed', score Yes.</i>	Not applicable	NA	NA
iii	The discrimination of the updated model is reported	Not applicable	NA	NA
iv	The confidence interval (or standard error) of the discrimination measure of the updated model is reported	Not applicable	NA	NA
v	The calibration of the updated model is reported	Not applicable	NA	NA
Discussion				
18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).			1
i	Limitations of the study are discussed <i>Stating any limitation is sufficient.</i>	Y	Y	Y
19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.			1
i	Comparison of results to reported performance in development studies and/or other validation studies is given	Not applicable	Y	Y
19b	Give an overall interpretation of the results considering objectives, limitations, results from similar studies and other relevant evidence.			1

i	An overall interpretation of the results is given	Y	Y	Y
20	Discuss the potential clinical use of the model and implications for future research.			1
i	The potential clinical use is discussed <i>E.g. an explicit description of the context in which the prediction model is to be used (e.g. to identify high risk groups to help direct treatment, or to triage patients for referral to subsequent care).</i>	Y	Y	Y
ii	Implications for future research are discussed <i>E.g. a description of what the next stage of investigation of the prediction model should be, such as "We suggest further external validation".</i>	Y	Y	Y
Other information				
21	Provide information about the availability of supplementary resources, such as study protocol, web calculator, and data sets.			
i	Information about supplementary resources is provided	Y	Y	Y
22	Give the source of funding and the role of the funders for the present study.			1
i	The source of funding is reported or there is explicit mention that there was no external funding involved	Y	Y	Y
ii	The role of funders is reported or there is explicit mention that there was no external funding	Y	Y	Y

Number of applicable TRIPOD items			32
Number of TRIPOD items adhered			15
OVERALL adherence to TRIPOD			47%

References

- Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* doi:10.1007/978-3-319-46723-8_49
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE Computer Society)*, 770–778. doi:10.1109/CVPR.2016.90

3. Prior F, Smith K, Sharma A, Kirby J, Tarbox L, Clark K, Bennett W, Nolan T, Freymann J. The public cancer radiology imaging collections of The Cancer Imaging Archive. *Sci data* (2017) **4**:170124. doi:10.1038/sdata.2017.124
4. Caruana R, Lawrence S, Giles L. Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping. in *Advances in Neural Information Processing Systems*
5. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, Blaschko M. Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory & Practice. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* (2019) **11765 LNCS**:92–100. doi:10.1007/978-3-030-32245-8_11
6. Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Institute of Electrical and Electronics Engineers Inc.), 936–944. doi:10.1109/CVPR.2017.106
7. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated residual transformations for deep neural networks. in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* (Institute of Electrical and Electronics Engineers Inc.), 5987–5995. doi:10.1109/CVPR.2017.634
8. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard C van den, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal* (2016) **42**:1–13. doi:10.1016/j.media.2017.06.015
9. Kingma DP, Ba JL. Adam: A method for stochastic optimization. in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*
10. An P, Xu S, Harmon S, Turkbey E, Sanford T, Amalou A, Kassin M, Varble N, Blain M, Anderson V, et al. CT Images in Covid-19 [Data set]. *Cancer Imaging Arch* (2020) doi:https://doi.org/10.7937/tcia.2020.gqry-nc81
11. Morozov SP, Andreychenko AE, Pavlov NA, Vladzimirskyy A V, Ledikhova N V, Gombolevskiy VA, Blokhin IA, Gelezhe PB, Gonchar A V, Chernina Y. MosMedData: Chest CT Scans With COVID-19 Related Findings Dataset. (2020) Available at: <http://europepmc.org/abstract/PPR/PPR268614>