



OPEN

U-survival for prognostic prediction of disease progression and mortality of patients with COVID-19

Janne J. Näppi^{1,3}, Tomoki Uemura^{1,2,3}, Chinatsu Watari¹, Toru Hironaka¹, Tohru Kamiya² & Hiroyuki Yoshida^{1✉}

The rapid increase of patients with coronavirus disease 2019 (COVID-19) has introduced major challenges to healthcare services worldwide. Therefore, fast and accurate clinical assessment of COVID-19 progression and mortality is vital for the management of COVID-19 patients. We developed an automated image-based survival prediction model, called U-survival, which combines deep learning of chest CT images with the established survival analysis methodology of an elastic-net Cox survival model. In an evaluation of 383 COVID-19 positive patients from two hospitals, the prognostic bootstrap prediction performance of U-survival was significantly higher ($P < 0.0001$) than those of existing laboratory and image-based reference predictors both for COVID-19 progression (maximum concordance index: 91.6% [95% confidence interval 91.5, 91.7]) and for mortality (88.7% [88.6, 88.9]), and the separation between the Kaplan–Meier survival curves of patients stratified into low- and high-risk groups was largest for U-survival ($P < 3 \times 10^{-14}$). The results indicate that U-survival can be used to provide automated and objective prognostic predictions for the management of COVID-19 patients.

The rapid increase in the number of patients who have the coronavirus disease 2019 (COVID-19) has introduced major challenges for healthcare services worldwide. According to the World Health Organization (WHO), the first nine months of 2020 saw more than 34 million COVID-19 infections and more than 1 million deaths worldwide¹, and these numbers are still increasing rapidly. Therefore, a fast and accurate clinical assessment of disease progression and mortality of patients with COVID-19 is vital for logistic planning and for management of the patients.

Recently, the WHO published specific recommendations about the use of chest imaging for the management of COVID-19 patients². Chest imaging can help clinicians to decide whether to admit or discharge patients with mild COVID-19 symptoms, whether to admit patients with moderate-to-severe COVID-19 symptoms to a regular ward or an intensive care unit (ICU), and to provide information about therapeutic management of hospitalized patients with moderate-to-severe COVID-19 symptoms².

Chest computed tomography (CT) is the most sensitive chest imaging method for COVID-19^{3–5}; therefore, several image-based prognostic predictors have been reported for chest CT. Some of these predictors are based on quantification of the radiologist's visual assessment of CT images^{6–8}. A semi-quantitative total severity score^{9,10}, which characterizes the extent of lobar involvement based on the bilateral multiple lobular ground-glass opacity and consolidation of COVID-19 on chest CT, has been shown to be predictive of patients' mortality⁷. The extent of lung parenchyma, which can be quantified visually or by the use of image processing software for determining the volumetric size of the well-aerated lung parenchyma, has been shown to be predictive of the admission to an ICU or death⁸.

Several computer-assisted prognostic predictors have also been reported. The general approach is to extract features from CT images and to subject those features to prognostic prediction. For performing the feature extraction, deep learning^{11–17} or semi-automated image processing software^{6,8,18,19} is used for segmentation of the complete lung regions, or for segmentation of infected regions within the lungs such as ground-glass opacities, semi-consolidation, and consolidation. In studies in which complete lung regions are extracted, the segmented

¹3D Imaging Research, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, 25 New Chardon Street, Suite 400C, Boston, MA 02114, USA. ²Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Kitakyushu, Japan. ³These authors contributed equally: Janne J. Näppi and Tomoki Uemura. ✉email: yoshida.hiro@mgh.harvard.edu

Variables	
Age (years)	67 [58.2, 78]
Gender, n (%)	
Male	128 (59.8)
Female	86 (40.2)
ICU admissions, n (%)	109 (50.9)
Deaths, n (%)	46 (21.4)
Laboratory tests^a	
Lactic dehydrogenase (U/l)	324.0 [238.5, 438.0]
C-reactive protein (mg/l)	74.3 [35.4, 148.0]
Lymphocytes (%)	13.9 [8.6, 21.2]
Total severity score	8 [5, 13]
Percentage of well-aerated lung parenchyma ^b (%)	69.2 [55.6, 82.1]

Table 1. Demographics, clinical characteristics, laboratory information, and reference predictors for the patient cohort. Continuous variables are expressed in terms of the median [interquartile range: Q1, Q3]. ^aFor a sub-cohort of 175 patients. ^bFor a sub-cohort of 209 patients.

regions were characterized by use of radiomics with a large number of features^{12,20}, whereas the segmented infected regions are typically characterized by use of a small number of well-understood manually defined features such as the relative size of a thresholded region within the complete segmented region^{8,11,13–16,18,19}. In a majority of studies, the subsequent prognostic prediction has been based on logistic regression, which is largely limited to binary predictions^{8,18–22}. In studies that made use of established survival analysis methodologies, the survival analysis was typically based on Cox proportional hazards regression and/or Kaplan–Meier survival curves^{6,12,16}.

These previous studies on prognostic predictors for COVID-19 on chest CT had various limitations. Visual or semi-automated quantification of CT images tends to be subjective, and it can have great inter- and intra-observer variability. Only few of the automated image-based survival prediction models made use of established survival analysis methodologies, and in most studies, the analysis was limited to selected regions of interest rather than considering the complete lung region. Although some of these studies made use of deep learning, the role of deep learning has been limited to the segmentation of regions rather than using deep learning for generating new prognostic predictors²³. Yet another limitation has been the use of manually crafted or mathematically defined features that are not necessarily ideal for the construction of optimal image-based prognostic predictors.

In this study, we present an automated image-based survival prediction model, called U-survival, which integrates the image information extracted by deep learning (U-Net) directly into a Cox proportional hazards model with an elastic-net penalty (elastic-net Cox model) for performing the prognostic prediction of patients with COVID-19. After training the U-Net to perform semantic segmentation of the lung tissue patterns of chest CT images, we subject the bottleneck section of the U-Net to an elastic-net Cox model that automatically selects a sparse subset of features to build an optimal survival model for the input data. See the subsection of “[The U-survival model](#)” under Methods for details of the design of the model.

Our approach is inspired by radiomics in the sense that we use an elastic-net penalty to construct a deep radiomic signature for survival analysis from a large number of features that are extracted from the images internally by the U-Net. Thus, we use deep learning as an integral part of the survival model, rather than as a segmentation tool. We show that the prognostic performance of the resulting U-survival model can exceed that of existing laboratory²⁴ as well as visual^{6,10,25} and quantitative⁸ image-based prognostic predictors in predicting the disease progression and mortality for patients with COVID-19. We also demonstrate how the integration of an established traditional survival analysis methodology into the model makes it possible to obtain additional information, such as survival curves or risk stratification, which was not available with the previously proposed prediction models based on logistic regression. Thus, the U-survival model can be used for providing an automated and objective survival analysis with high accuracy for COVID-19 patients. Such a model could be used for logistic planning, clinical decision making, and management of COVID-19 patients.

Results

Summary of the dataset. We retrospectively identified 383 patients who had been confirmed as being COVID-19 positive between March 1 and June 28, 2020, at the Massachusetts General Hospital or the Brigham and Women’s Hospital (Boston, MA, USA); they were followed up until July 28, 2020. After the application of our exclusion criteria, 214 of these patients were included in the study. Table 1 shows the demographics, clinical characteristics, laboratory information, and three of the reference predictors for the 214 COVID-19 patients. Some of the information was available only for subcohorts, as indicated in Table 1.

Prognostic predictions. Bootstrap-based internal validation was performed on the U-survival model as well as on the existing prognostic predictors (hereafter called reference predictors) of (1) a combination of the laboratory tests of lactic dehydrogenase, lymphocyte, and C-reactive protein^{24,26}, (2) 4D-curvature features of

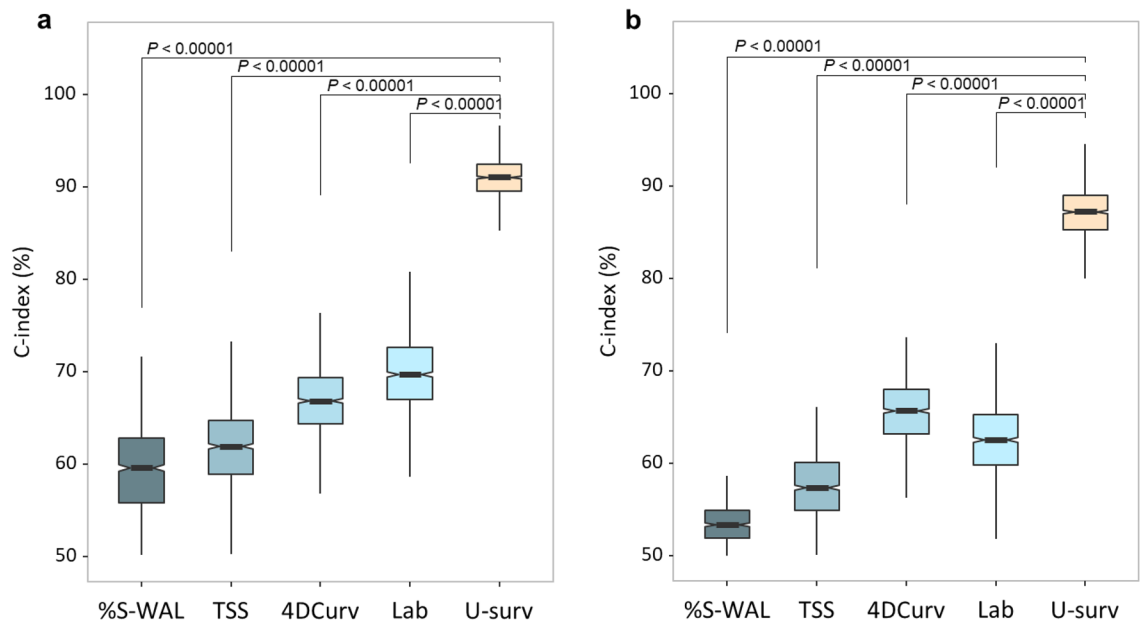


Figure 1. Performance of U-survival and the reference predictors in the prediction of COVID-19 progression and mortality for the subcohorts of patients for whom all the predictors were available. (a,b) Notched boxplots, in which the middle lines and the notches show the means and 95% confidence intervals, of the C-index values of U-survival and the reference predictors in the prediction of disease progression (a), where the outcome is defined as ICU admission or death, and in the prediction of mortality (b), where the outcome is defined as death. The C-index values were obtained by per-patient bootstrapping with 1000 replicates. The P values were obtained by use of a bootstrap hypothesis test. %S-WAL percentage of well-aerated lung parenchyma, TSS total severity score, 4DCurv 4D-curvature, Lab laboratory tests, U-surv U-survival.

the CT images (abbreviated as 4DCurv)²⁷, (3) the total severity score (TSS)^{10,28}, and (4) the percentage of well-aerated lung parenchyma (%S-WAL)⁸.

Figure 1 shows the performance of the prognostic prediction, as measured by the concordance index (C-index)²⁹, of U-survival and the reference predictors in the prediction of COVID-19 (a) progression, where the outcome was defined as ICU admission or death, and (b) mortality, where the outcome was defined as death, for the subcohorts of patients for whom all of the predictors were available. The subcohort of the progression analysis in (a) included 108 patients, of whom 38 were admitted to the ICU ($n = 26$) or who expired ($n = 12$), whereas that of the mortality analysis in (b) included 173 patients, of whom 40 expired.

For the prediction of progression, U-survival yielded an estimated C-index value of 90.9% [95% confidence interval (CI) 90.8, 91.0], whereas those of the reference predictors of laboratory tests, 4DCurv, TSS, and %S-WAL were 69.8% [CI 69.5, 70.1], 66.6% [CI 66.4, 66.9], 61.8% [CI 61.5, 62.0], and 59.0% [CI 58.7, 59.3], respectively. For the prediction of mortality, U-survival showed a C-index value of 87.1% [CI 86.9, 87.3], whereas those of the reference predictors of laboratory tests, 4DCurv, TSS, and %S-WAL, were 62.5% [CI 62.3, 62.8], 65.5% [CI 65.3, 65.7], 57.0% [CI 56.8, 57.2], and 53.3% [CI 53.1, 53.4], respectively. These results indicate that U-survival outperformed the reference predictors by a large margin. The prediction performance of U-survival was statistically significantly higher than those of the reference predictors both for COVID-19 progression and for mortality ($P < 0.00001$).

Table 2 shows the performance of U-survival and the reference predictors in the prediction of the COVID-19 progression (left) and mortality (right) for the subcohorts of patients, in which all available patients (second and fourth columns) were used for each predictor. In each type of prediction and for each predictor, the C-index values were increased by 0.7–4.5% from those shown in Fig. 1. Similar to the trend shown in Fig. 1, U-survival yielded C-index value of 91.6% [CI: 91.5, 91.7] for progression and 88.7% [88.6, 88.9] for mortality, which statistically significantly ($P < 0.00001$) outperformed the other predictors by a large margin.

Risk stratification. Figure 2 shows the Kaplan–Meier survival curves of 173 patients stratified into low- and high-risk groups based on the mortality prediction results shown in Fig. 1b. Figure 2 indicates that the difference between the survival curves for the two groups was statistically significant for U-survival and the 4D-curvature, with log-rank P values of 3×10^{-14} and 7×10^{-4} , respectively. The difference between the survival curves was not statistically significant ($P > 0.01$) for the other predictors. Visually, it is evident that the difference between the two curves was largest with U-survival, indicating that U-survival is more effective than any of the other predictors for mortality risk stratification of COVID-19 patients.

Similarly, the survival curves of 108 patients stratified into low- and high-risk groups based on the progression prediction results in Fig. 1a show that the difference between the low- and high-risk groups was largest with U-survival (Supplementary Fig. S1).

Predictors	Progression			Mortality		
	Patients n	C-index % [95% CI]		Patients n	C-index % [95% CI]	
%S-WAL	138	56.2 [56.0, 56.4]	$P < 10^{-4}$ $P < 10^{-4}$ $P < 10^{-4}$ $P < 10^{-4}$	209	55.9 [55.7, 56.1]	$P < 10^{-4}$ $P < 10^{-4}$ $P < 10^{-4}$ $P < 10^{-4}$
Total severity score	141	62.9 [62.6, 63.1]		214	70.1 [69.9, 70.3]	
4D-curvature	138	70.6 [70.4, 70.8]		209	60.4 [60.2, 60.7]	
Laboratory tests	109	69.1 [68.9, 69.4]		175	63.3 [63.0, 63.5]	
U-survival	141	91.6 [91.5, 91.7]		214	88.7 [88.6, 88.9]	

Table 2. Performance of U-survival and the reference predictors in the prediction of the outcomes of COVID-19 for the subcohorts where all of the available patients were used for each predictor. The table shows C-index values and their 95% confidence intervals (CI) obtained by per-patient bootstrapping with 1000 replicates in the prediction of disease progression (left) and mortality (right). “Patients” columns show the number of patients used for each predictor. The P values were obtained by application of a bootstrap hypothesis test. %S-WAL percentage of well-aerated lung parenchyma.

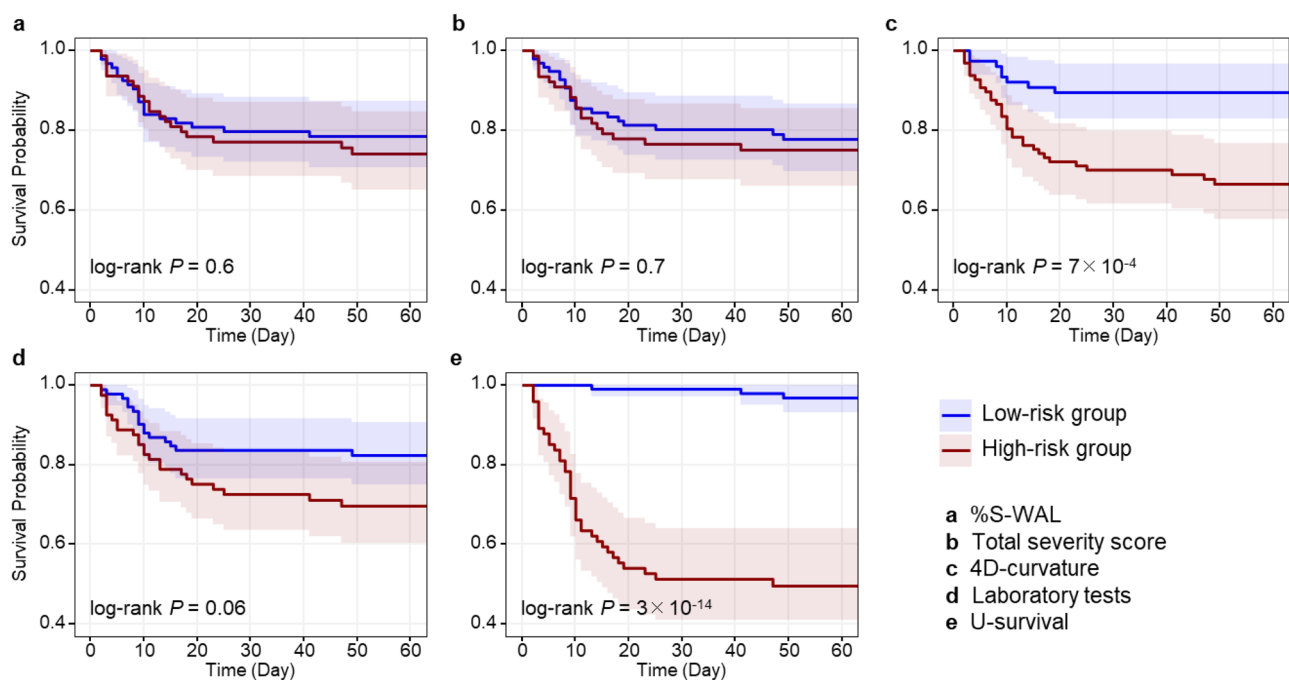


Figure 2. Kaplan–Meier survival curves of COVID-19 patients stratified into low- and high-risk groups based on the mortality predictions in Fig. 1b. The estimated survival curves for the low- and high-risk groups are shown in blue and red, respectively, with shaded areas representing the 95% confidence interval. The P values were obtained by application of the log-rank test to the two survival curves. %S-WAL percentage of well-aerated lung parenchyma.

Discussion

Although a quantitative assessment of CT images has been reported to be a predictor of disease progression and mortality for COVID-19 patients, most of the studies were based on logistic regression that is limited to the observation of binary outcomes at a given time point³⁰. The use of traditional survival analysis methodologies enables the calculation of survival data as a function of time, which is desirable for the analysis of time-to-event information for applications such as the comparison of survival distributions among various risk groups^{30,31}. The automated image-based U-survival model that we developed in this study is based on Cox regression analysis; therefore, it provides a continuous image-based prognostic prediction that enables us to acquire the proportional hazard for predicting the survival probability at any time point and to estimate survival curves for individual COVID-19 patients.

Our approach uses deep learning and chest CT images as an integral part of the survival prediction. Our results that greatly expand upon our previously reported methods and preliminary results³² show that the resulting U-survival model can substantially outperform existing laboratory tests and image-based quantitative scoring methods in the prediction of the disease progression and mortality of COVID-19 patients. An advantage of deep learning is that it is designed to learn optimal features for solving problems based on the available training data,

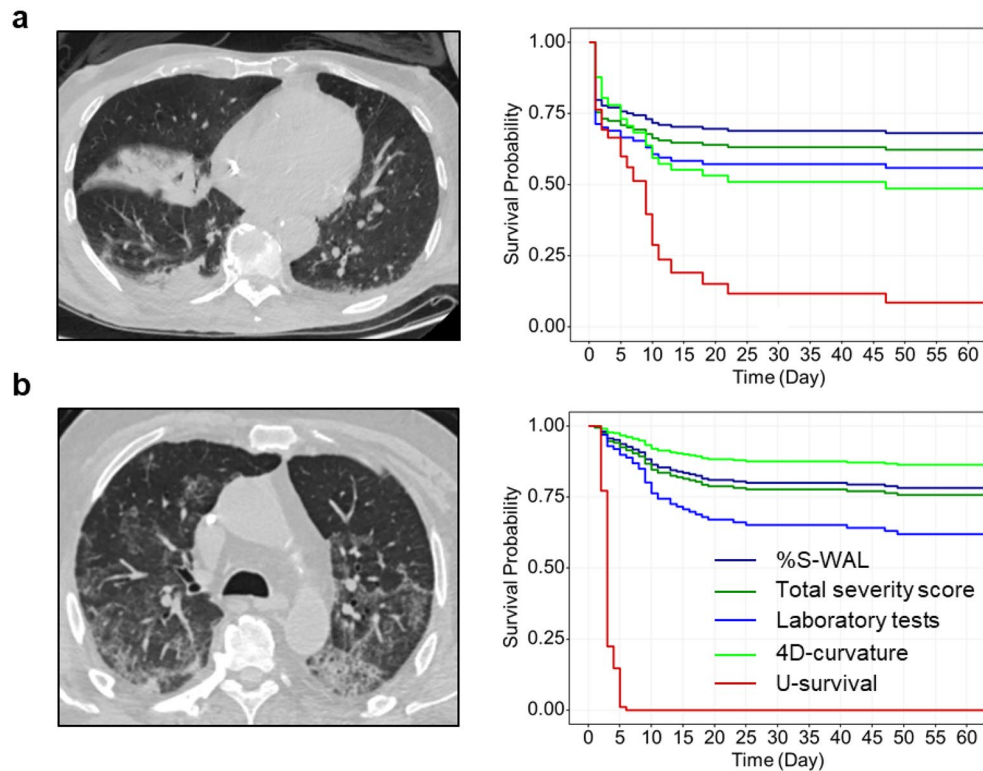


Figure 3. Representative chest CT images (left) and the corresponding per-patient survival curves (right) estimated by use of the predictors included in this study. (a) A 71-year-old male who was admitted to the ICU nine days after chest CT examination. The laboratory test results were dehydrogenase (LDH) 192 U/l, C-reactive protein (CRP) 126 mg/l, and lymphocytes 11.1%. The total severity score was 8, and the percentage of well-aerated lung parenchyma (%S-WAL) was 81.2%. The median survival times estimated by U-survival and 4DCurv were 8 and 47 days, respectively, whereas no median survival time was estimated by the other predictors. (b) A 64-year old male who expired three days after a chest CT examination. The laboratory test results were LDH 276 U/l, CRP 157 mg/l, and lymphocytes 4.6%. The total severity score was 12, and %S-WAL was 64.4%. The median survival time estimated by the U-survival was 3 days, whereas no median survival time was estimated by the other predictors.

whereas traditional manually crafted or mathematically defined image-based features are not necessarily ideal, or they are highly challenging to design for solving complex clinical problems such as the prognostic prediction investigated in this study. It should be noted that, in previous image-based prediction models for COVID-19, deep learning has been used only for the segmentation of images rather than for the extraction of prognostic features from the CT images.

In this study, we performed a systematic performance validation including the use of established performance metrics and the bootstrap method to demonstrate the generalizability of our results^{35,36}. Our experiments were designed based on the international expert consensus guideline of “Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)”³⁵, which provides specific recommendations regarding appropriate scientific reporting of studies involved with the development and validation of prediction models. In particular, the bootstrap method, which we used as a primary method for internal validation in this study, provides a mechanism that accounts for overfitting and uncertainty in the model development, thereby providing realistic optimism-corrected estimates of the generalization performance of the developed models^{35,36}. A further advantage of bootstrapping is that the effects of predictor selection strategies on the model building, and thus the extent of model overfitting and optimism, can be quantified by repeating the predictor selection process in each bootstrap sample³⁵. It has been demonstrated that the sample-based average of bootstrapping is a nearly unbiased estimate of the expected value of the optimism that would be observed in external validation^{36–38}.

Figure 3 shows two examples of the estimation of survival curves for individual COVID-19 patients by use of the prognostic predictors included in this study. For both patients, the median survival times obtained from the survival curves calculated by U-survival, as indicated by the red curve in each plot, are more accurate estimates of the observed survival time than are those of the reference predictors, demonstrating the potential usefulness of the U-survival model for the management of COVID-19 patients.

This was a retrospective study; therefore, our CT protocol was representative of routine clinical practice rather than being a carefully designed uniform study protocol. In particular, although non-contrast CT is considered sufficient for the evaluation of COVID-19 cases³³, 57% of our cases originated from contrast-enhanced CT. In practice, contrast-enhanced CT may be used for the evaluation of COVID-19 patients who may have a clinically

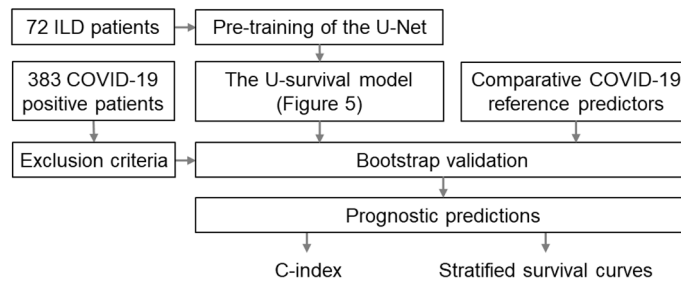


Figure 4. Flowchart of the key developments and evaluations. The U-survival model and the reference predictors are validated by use of bootstrapping with COVID-19 patients. The prognostic predictions are compared by use of the C-index and stratified survival curves. *ILD* Interstitial lung disease.

worsening cardiorespiratory status or suspicion of a pulmonary embolism, or for avoiding redundant imaging studies³³. Regardless, our experiments indicated that the prognostic performance of U-survival did not vary meaningfully between non-contrast and contrast-enhanced CT (Supplementary Fig. S2).

The U-Net that we used in this study had been trained with an independent external training dataset of interstitial lung diseases for the segmentation of chest CT images into distinct lung tissue patterns. Previously, we had successfully used the pre-trained U-Net for the analysis of chest CT images of patients with idiopathic pulmonary fibrosis³⁴. The use of an independent external training dataset improved the generalizability of our results, and it enabled us to maximize the number of independent COVID-19 test cases available for this study. In the future, access to a large number of CT images of COVID-19 patients could enable us to construct an independent COVID-19 dataset for the training of the U-Net, and this could yield an even higher performance of U-survival than what was reported in this study.

Our imaging study included some unique clinical information such as laboratory test results for the patients. Although CT imaging databases of COVID-19 patients are currently being made available in public imaging repositories, it is not clear whether clinical information such as laboratory tests will be included in these repositories. Thus, it may not be possible to validate our study in a larger external patient cohort any time soon.

Our results are based on the use of a 2D U-Net. In the future, it may be beneficial to use a 3D U-Net instead. However, at present, effective use of 3D deep learning is constrained by the memory limitations of currently available graphics processing units (GPUs) and by the acquisition of clinical chest CT studies at an anisotropic image resolution. Because of the GPU memory limitations, it is not possible to fit a full high-resolution chest CT volume and a sophisticated 3D deep learning model into a same GPU without the use of countermeasures such as subsampling of the image volume, which can have a detrimental effect on the performance of deep learning³⁹. Because of the anisotropic image acquisition, many chest CT studies produce a stack of 2D image slices rather than a true 3D volume, and this makes volumetric analysis less meaningful than is an independent analysis of each image slice. Also, the use of 3D volumes would reduce the amount of training data in comparison to the use of 2D images. Nevertheless, in the long run, the use of a 3D U-Net with a high-capacity GPU and a large number of isotropic chest CT volumes could yield an even higher performance of U-survival than that reported in this study.

In summary, the limitations of this study as elaborated above include that this was a retrospective study and that we did not perform an external validation with an independent study population. Potential future directions include an external validation study with a large number of prospective cases and human observers from more than two institutions, an investigation of the potential benefit of a 3D U-Net in the U-survival model, and ultimately an introduction of the U-survival model to a clinical routine.

Conclusion

We showed that deep learning of chest CT images can be used as an integral part of an automated image-based survival prediction model based on traditional survival analysis methodology. This makes it possible to obtain complete survival information that was not available with previously proposed prediction models. In our evaluation of 383 COVID-19 positive patients from two hospitals, the U-survival model significantly ($P < 0.0001$) outperformed existing laboratory tests and image-based visual and quantitative predictors in the prediction of the disease progression and mortality of COVID-19 patients and in the separation between the Kaplan–Meier survival curves of patients stratified into low- and high-risk groups ($P < 3 \times 10^{-14}$). The results indicate that the U-survival model can be used to provide automated and objective prognostic predictions for the management of COVID-19 patients.

Methods

Figure 4 shows a flowchart of the key developments and evaluation steps described below.

Study cohort. The study was reviewed and approved by the Mass General Brigham (MGB) institutional review board (IRB). All procedures involving human participants were performed in accordance with the ethical

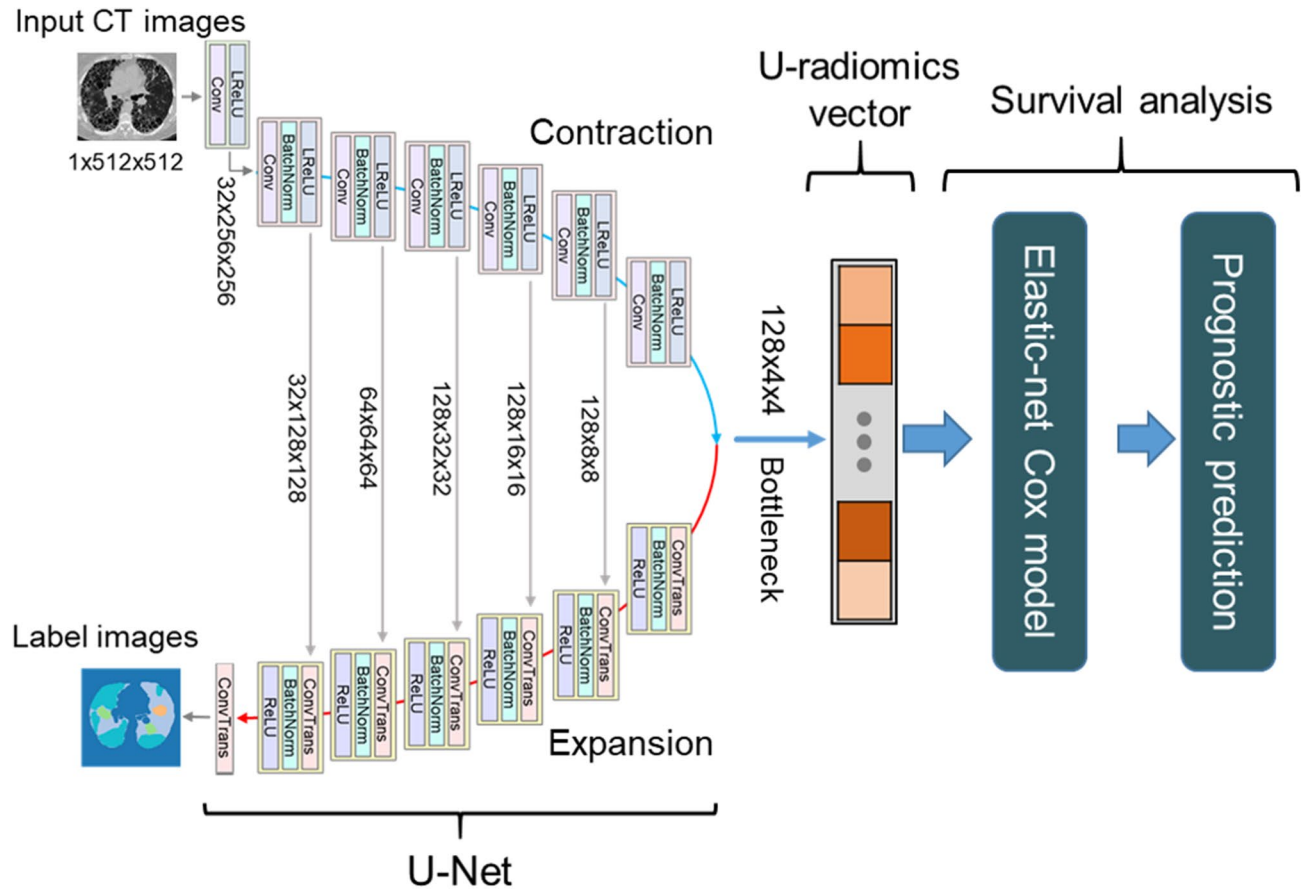


Figure 5. Schematic diagram of the U-survival model. The U-radiomics vector derived from the bottleneck section of the U-Net is subjected to an elastic-net Cox model for predicting the disease progression and mortality of a patient. *Conv* convolutional layer (kernel 4, stride: 2), *ConvTrans* transposed convolution (kernel 4, stride: 2), *LReLU* leaky rectified linear unit, *BatchNorm* batch normalization.

standards of the IRB and with the 1964 Declaration of Helsinki and its later amendments. The informed consent of the patients was waived for this study by the MGB IRB.

We retrospectively identified 383 COVID-19 positive patients between March 1 and June 28, 2020, from the medical records of the Massachusetts General Hospital and the Brigham and Women’s Hospital; they were followed up until July 28, 2020. The patients who were included in the study (1) were at least 18 years old, (2) had been diagnosed as COVID-19 positive based on a positive result for severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) by reverse transcriptase-polymerase chain reaction (RT-PCR) with samples obtained from the nasopharynx, oropharynx, or lower respiratory tract, and (3) had a high-resolution chest CT examination. The resulting cohort consisted of 302 patients. After excluding the patients whose CT examinations had been performed for diseases other than COVID-19, we established a cohort of a total of 214 COVID-19 patients.

The chest CT images were acquired by use of single-phase low-dose acquisition with a multi-channel CT scanner (Canon/Toshiba Aquilion ONE, GE Discovery CT750 HD and Revolution CT/Frontier, Siemens SOMATOM Definition AS/AS+/Edge/Flash, SOMATOM Force, Biograph 64, or Sensation 64) by use of 80–140 kVp voltage, auto tube-current modulation, a pitch of 0.3–1.6, and a slice thickness of 0.625–2.0 mm. The CT images were reconstructed by use of a neutral or medium sharp reconstruction kernel.

For the patients included in this study, the mean and standard deviation of the number of days between the CT image acquisition and admission to ICU were 2.2 ± 4.1 ($n = 36$), whereas those between the CT image acquisition and death were 11.4 ± 10.9 ($n = 46$).

A subcohort of 141 patients who had a CT examination before ICU admission or who were not admitted to the ICU, was used for the prediction analysis of COVID-19 progression. For these patients, the survival time was defined as the number of days from the baseline CT image acquisition to ICU admission or death (for uncensored patients) or otherwise set to the most recent follow-up date (for censored patients). For the analysis of mortality, the entire cohort of 214 patients was used, where the survival time was defined as the number of days from the baseline CT image acquisition to death (for uncensored patients) or as the most recent follow-up date (for censored patients).

The U-survival model. The U-survival model has two basic components: a convolutional neural network (U-Net) and an elastic-net Cox model (Fig. 5). The U-Net has three principal layer sections: contraction, bot-

tleneck, and expansion⁴⁰. The output of the bottleneck section that is located between the contraction and expansion sections is essentially a compressed low-dimensional representation of the input data to the U-Net.

The U-Net of this study was pre-trained, as described in our previous study³⁴, for performing semantic segmentation of the lung region of axial CT images into five distinct lung tissue patterns (ground-glass opacity, reticulation, consolidation, honeycombing, and a normal healthy lung pattern) by use of the chest CT images of 72 patients who had interstitial lung diseases and corresponding tissue-pattern-labeled images (“label images” in Fig. 5). After such training, the output of the bottleneck section of the U-Net, called U-radiomics vector, contains information regarding the interpretation of the lung tissue patterns of chest CT images^{41–43}.

To use this information for survival analysis, we subject the U-radiomics vector to an elastic-net Cox model⁴⁴. This model automatically selects a sparse subset of the features of the U-radiomics vector to build an optimal survival model for the input data.

We can formulate the construction of the U-survival model as follows: Let $\{(y_n, \mathbf{x}_n, \delta_n) | n = 1, \dots, N\}$ denote time-to-event data for patient n , where $\mathbf{x}_n = (x_n^m | m = 1, \dots, M)$ is the U-radiomics vector from the U-Net. If $\delta_n = 1$, y_n is the observed survival time of an outcome such as disease progression or death for patient n . If $\delta_n = 0$, y_n is a right-censoring time. Moreover, let $\{t_i | i = 1, \dots, K\}$ be an ascending list of the event times, and let $j(i)$ denote the mapping of the event time t_i to the index of the patient at risk at that time. The construction of the Cox model involves the estimation of a hazard vector $e^\beta = (e^{\beta_m} | m = 1, \dots, M)$ by maximization of the partial likelihood $L(\beta) = \prod_{i=1}^K \left(e^{x_{j(i)} \cdot \beta} / \sum_{j \in R_i} e^{x_j \cdot \beta} \right)$, where R_i is the set of indices for patients at risk at time t_i ⁴⁵.

In cases where the number of features in the U-radiomics vector is much larger than the number of patients ($M \gg N$), the above maximum-likelihood estimation can lead to a degeneration of β . The elastic-net Cox model obviates this problem by use of an elastic-net regularization term.

$$P_\alpha(\beta) = \alpha \sum_{i=1}^M |\beta_i| + (1 - \alpha) \sum_{i=1}^M \beta_i^2 \quad (0 \leq \alpha \leq 1) \quad (1)$$

where β is estimated by $\hat{\beta} = \operatorname{argmax}_\beta (\log L(\beta) - P_\alpha(\beta))$. The first term of Eq. (1) is the $L1$ or Lasso regularization that is known effectively to obviate the degeneration and to provide a sparse representation of β ⁴⁶. The second term of Eq. (1) is the $L2$ or ridge regularization, which shrinks all β_m of the hazard vector to zero to obviate the degeneration, but also provides highly correlated predictors with an equal weight. When these two terms are mixed by use of the parameter α , the elastic-net Cox model allows the U-survival model to perform simultaneous feature selection and optimal estimation of the hazard vector e^β .

Application of U-survival to CT images. To apply the U-survival model to prognostic prediction, we first apply the pre-trained U-Net to the axial CT images of a COVID-19 patient in our study cohort (see Supplementary Fig. S3). For each image, we identify the corresponding U-radiomics vector as the expanded output of the bottleneck section of the U-Net. These per-image U-radiomics vectors of the patient are then combined into a single per-patient U-radiomics vector by use of the median value of each feature of the per-image U-radiomics vectors across the axial CT images of the patient. Finally, the per-patient U-radiomics vector is subjected to the elastic-net Cox model for building a survival model for predicting the prognosis of the patient.

Implementation and parameters of the U-survival model. The CT images of this study and those of the external training dataset for the U-Net³⁴ were pre-processed for the U-Net by clipping of the intensity values to a Hounsfield unit (HU) range of -1024 to 1024 and by mapping of these values linearly to the range of -1 to $+1$.

The free hyper-parameters of the U-survival model were optimized by use of a grid search algorithm. Specifically, the output of the bottleneck section of the U-Net was configured to be $c \times n \times n = 128 \times 4 \times 4$, where c and n represent the number of channels and the size of the output, respectively (Fig. 5). Each element of the output was post-processed by z-score normalization across the patients, and the resulting z-score normalized output was converted to a U-radiomics vector with a length of 2048. This configuration of the bottleneck section yielded 1–9% higher performance than did other possible configurations of the bottleneck section (see Supplementary Fig. S4). Similarly, we set the mixing parameter α in Eq. (1) to take advantage of both lasso and ridge regularization terms for optimizing the performance of the U-survival model (see Supplementary Fig. S5). For the training, we used Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The dropout ratio was set to 0.5, batch size was 64, and the learning rate was 2.0×10^{-4} .

Computational environment. The U-survival model was implemented by use of our custom-made Linux-based computational server that was equipped with $10 \times V100S$ GPUs (NVIDIA Corporation, Santa Clara, CA, USA) with 32 GB memory and 12-core 3.6 GHz Xeon Gold 6256 CPUs (Intel Corporation, Santa Clara, CA, USA) with 512 GB memory. The U-Net was implemented by use of PyTorch 1.5⁴⁷. The elastic-net Cox model was implemented by use of glmnet 4.0⁴⁴.

Evaluation methods. The performance of the prognostic prediction of the U-survival model was assessed by use of the C-index²⁹. The C-index is similar to the area under the receiver operating characteristic (ROC) curve (AUC) that evaluates the classification performance for binary outcomes, except that the C-index is used for estimating the concordance between predicted and observed outcomes in the presence of censoring. The C-index performs the estimation of concordance based on usable pairs, in which one patient is known to have

an outcome before the other patient, who may have an outcome later or who may be censored. The C-index has a value range of 0–100%, where 50% indicates random prediction and 100% indicates perfect prediction.

To obtain optimism-corrected, generalizable estimates of the C-index values and their confidence intervals, we used per-patient bootstrapping with 1000 replications^{35,38} for the evaluation of the U-survival model. The bootstrap method enables the use of all of the available data, provides a mechanism for accounting for overfitting and uncertainty in the model development, and generates a nearly unbiased estimate of what would be observed in external validation^{35,36}.

The performance of the U-survival model in risk stratification was evaluated by use of the Kaplan–Meier estimator⁴⁸. We first calculated the optimism-corrected predictions by use of the above per-patient bootstrapping process⁴⁹. Then, we generated the Kaplan–Meier survival curves that show the survival probability estimates over time by use of the Kaplan–Meier estimator. The survival curves are stratified into low- and high-risk groups by use of the mean of the bootstrap predictions. The difference between the survival curves between the risk groups was evaluated by use of the log-rank test, where a *P* value of less than 0.01 was considered to indicate a statistically significant difference.

The statistical analyses of the evaluation results were conducted by use of the R 4.0.2 software environment⁵⁰.

Comparative reference predictors. For reference, we compared the prediction performance of the U-survival model with those of several previously reported COVID-19 predictors, including (1) a combination of the laboratory tests of lactic dehydrogenase, lymphocytes, and C-reactive protein^{24,26}, (2) a visual image-based CT assessment in terms of the total severity score^{10,28}, and (3) a quantitative image-based CT assessment in terms of the software-based percentage of well-aerated lung parenchyma (%S-WAL)⁸. In addition, we included (4) a quantitative image-based CT assessment based on the 4D-curvature features of the CT images (4DCurv)²⁷, which we previously showed to be effective in the prognostic prediction of the mortality of patients with interstitial lung diseases^{51,52}.

The total severity score was obtained by a pulmonologist with 20-year experience (C.W.) based on a protocol reported in previously published studies^{9,10}. Each of the five lung lobes was assessed for the degree of acute lung involvement, which was classified as none (0%) with a lobe score of 0, minimal (1–25%) with a lobe score of 1, mild (26–50%) with a lobe score of 2, moderate (51–75%) with a lobe score of 3, or severe (76–100%) with a lobe score of 4. An overall total severity score was calculated by summing of the five lobe scores (range of possible total score: 0–20). The distribution of the total severity score in the patient cohort is shown in Table 1.

The %S-WAL was calculated based on a previously published study⁸ and image processing software²⁷. The software extracted the complete lung region automatically from the CT images, after which the %S-WAL was calculated as the relative volume of the well-aerated 3D lung region determined by the density interval of -950 HU and -700 HU to the volumetric size of the complete segmented 3D lung region. The distribution of the %S-WAL in the study patient cohort is shown in Table 1.

The 4DCurv characterizes each point of the complete lung region by use of hyper-curvature features of the chest CT images, including the principal curvatures, curvedness, bright/dark sheets, cylinders, blobs, and curvature scales⁵¹. The 4DCurv was calculated by use of previously published image processing software²⁷ that generated 363 hyper-curvature features per patient⁵¹.

We calculated the prognostic predictions obtained from these predictors by subjecting them to the elastic-net Cox model and by applying per-patient bootstrapping with 1000 replications to the model. The C-index values obtained from the predictions were compared with those of U-survival by use of a bootstrap hypothesis test³⁸, and a *P* value of less than 0.01 was considered to indicate that the difference was statistically significant.

The performances of the reference predictors in risk stratification were evaluated by use of the Kaplan–Meier survival curves that were stratified into low- and high-risk groups by the mean of the bootstrap predictions of the predictors. The difference between the survival curves for the risk groups was evaluated by use of the log-rank test, and the *P* values obtained from the test were compared with that of the U-survival model.

Data availability

The datasets used and analyzed during the current study are available from the corresponding author upon request.

Received: 9 October 2020; Accepted: 13 April 2021

Published online: 29 April 2021

References

1. WHO. Coronavirus Disease (COVID-19). *Weekly Epidemiological Update* (2020). <https://doi.org/10.1097/jcn.0000000000000710>.
2. Akl, E. A. *et al.* Use of chest imaging in the diagnosis and management of COVID-19: a WHO rapid advice guide. *Radiology* **298**, E63–E69 (2021).
3. Mei, X. *et al.* Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. *Nat. Med.* **26**, 1224–1228 (2020).
4. Harmon, S. A. *et al.* Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat. Commun.* **11**, 4080 (2020).
5. Rubin, G. D. *et al.* The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner Society. *Radiology* **296**, 172–180 (2020).
6. Francone, M. *et al.* Chest CT score in COVID-19 patients: Correlation with disease severity and short-term prognosis. *Eur. Radiol.* **30**, 6808–6817 (2020).
7. Li, K. *et al.* Predictors of fatality including radiographic findings in adults with COVID-19. *Respir. Res.* **21**, 146 (2020).
8. Colombi, D. *et al.* Well-aerated lung on admitting Chest CT to predict adverse outcome in COVID-19 pneumonia. *Radiology* **296**, E86–E96 (2020).

9. Bernheim, A. *et al.* Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology* **295**, 200463 (2020).
10. Lyu, P., Liu, X., Zhang, R., Shi, L. & Gao, J. The performance of chest CT in evaluating the clinical severity of COVID-19 pneumonia: identifying critical cases based on CT characteristics. *Invest. Radiol.* **55**, 412–421 (2020).
11. Liu, F. *et al.* CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. *Theranostics* **10**, 5613–5622 (2020).
12. Wu, Q. *et al.* Radiomics analysis of computed tomography helps predict poor prognostic outcome in COVID-19. *Theranostics* **10**, 7231–7244 (2020).
13. Huang, L. *et al.* Serial quantitative chest CT assessment of COVID-19: deep-learning approach. *Radiol. Cardiothorac. Imaging* **2**, e200075 (2020).
14. Wang, Y. *et al.* Quantitative analysis of chest CT imaging findings with the risk of ARDS in COVID-19 patients: A preliminary study. *Ann. Transl. Med.* **8**, 594–594 (2020).
15. Yu, Q. *et al.* Multicenter cohort study demonstrates more consolidation in upper lungs on initial CT increases the risk of adverse clinical outcome in COVID-19 patients. *Theranostics* **10**, 5641–5648 (2020).
16. Zhang, K. *et al.* Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell* **181**, 1423–1433 (2020).
17. Bai, X. *et al.* Predicting COVID-19 malignant progression with AI techniques. *medRxiv* (2020). <https://doi.org/10.1101/2020.03.20.20037325>.
18. Lanza, E. *et al.* Quantitative chest CT analysis in COVID-19 to predict the need for oxygenation support and intubation. *Eur. Radiol.* **30**, 6770–6778 (2020).
19. Matos, J. *et al.* Evaluation of novel coronavirus disease (COVID-19) using quantitative lung CT and clinical data: Prediction of short-term outcome. *Eur. Radiol. Exp.* **4**, 39 (2020).
20. Homayounieh, F. *et al.* CT radiomics, radiologists and clinical information in predicting outcome of patients with COVID-19 pneumonia. *Radiol. Cardiothorac. Imaging* **2**, e200322 (2020).
21. Li, M. D. *et al.* Automated assessment of COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese neural networks. *medRxiv* (2020). <https://doi.org/10.1101/2020.05.20.20108159>.
22. Xiao, L. *et al.* Development and validation of a deep learning-based model using computed tomography imaging for predicting disease severity of coronavirus disease 2019. *Front. Bioeng. Biotechnol.* **8**, 898 (2020).
23. Kundu, S., Elhalawani, H., Gichoya, J. W. & Kahn, C. E. How might AI and chest imaging help unravel COVID-19's mysteries?. *Radiol. Artif. Intell.* **2**, e200053 (2020).
24. Yan, L. *et al.* An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.* **2**, 283–288 (2020).
25. Li, K. *et al.* The clinical and chest CT features associated with severe and critical COVID-19 pneumonia. *Invest. Radiol.* **55**, 327–331 (2020).
26. Ji, D. *et al.* Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clin. Infect. Dis.* **71**, 1393–1399 (2020).
27. Kawata, Y., Kubo, H., Niki, N., Ohmatsu, H. & Moriyama, N. A study of three-dimensional curvatures and curvatures of four-dimensional hypersurface for analyzing pulmonary nodules on high-resolution CT images. *Syst. Comput. Japan* **36**, 16–29 (2005).
28. Li, K. *et al.* CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur. Radiol.* **30**, 4407–4416 (2020).
29. Harrell, F. E., Lee, K. L. & Mark, D. B. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* **15**, 361–387 (1996).
30. George, B., Seals, S. & Aban, I. Survival analysis and regression models. *J. Nucl. Cardiol.* **21**, 686–694 (2014).
31. Cui, L. *et al.* A deep learning-based framework for lung cancer survival analysis with biomarker interpretation. *BMC Bioinformatics* **21**, 112 (2020).
32. Uemura, T., Näppi, J. J., Kamiya, T. & Yoshida, H. U-radiomics for predicting survival of patients with COVID-19. in *Proc. SPIE Medical Imaging* 1160110 (2021). <https://doi.org/10.1117/12.2581907>.
33. Kalra, M. K., Homayounieh, F., Arru, C., Holmberg, O. & Vassileva, J. Chest CT practice and protocols for COVID-19 from radiation dose management perspective. *Eur. Radiol.* **30**, 6554–6560 (2020).
34. Uemura, T. *et al.* U-radiomics for predicting survival of patients with idiopathic pulmonary fibrosis. in *Proc. SPIE Medical Imaging* 113140T (2020). <https://doi.org/10.1117/12.2551273>.
35. Moons, K. G. M. *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Ann. Intern. Med.* **162**, W1–W73 (2015).
36. Smith, G. C. S., Seaman, S. R., Wood, A. M., Royston, P. & White, I. R. Correcting for optimistic prediction in small data sets. *Am. J. Epidemiol.* **180**, 318–324 (2014).
37. Harrell, F. Jr., Lee, K. & Mark, D. Tutorial in biostatistics: multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing error. *Stat. Med.* **15**, 361–387 (1996).
38. Efron, B. & Tibshirani, R. J. *An introduction to the bootstrap. Monographs on Statistics and Applied Probability* vol. 57 (Chapman & Hall, 1993).
39. Sabottke, C. F. & Spieler, B. M. The effect of image resolution on deep learning in radiography. *Radiol. Artif. Intell.* **2**, (2020).
40. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
41. Shi, F. *et al.* Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. *IEEE Rev. Biomed. Eng.* **14**, 4–15 (2021).
42. Ker, J., Wang, L., Rao, J. & Lim, T. Deep learning applications in medical image analysis. *IEEE Access* **6**, 9375–9379 (2017).
43. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
44. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Softw.* **39**, 1–13 (2011).
45. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc.* **34**, 187–220 (1972).
46. Tibshirani, R. Regression shrinkage and selection via the lasso: a retrospective. *J. R. Stat. Soc. Ser. B (Statistical Methodol.)* **73**, 273–282 (2011).
47. Paszke, A. *et al.* PyTorch: an imperative style, high-performance deep learning library. in *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)* (2019).
48. Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958).
49. Bello, G. A. *et al.* Deep learning cardiac motion analysis for human survival prediction. *Nat. Mach. Intell.* **1**, 95–104 (2019).
50. R Core Team. R: A Language and Environment for Statistical Computing. (2020).
51. Watari, C. *et al.* Radiomic texture-curvature (RTC) features for precision medicine of patients with rheumatoid arthritis-associated interstitial lung disease. in *Proc. SPIE Medical Imaging* (eds. Zhang, J. & Chen, P.-H.) vol. 10579 105791N (SPIE, 2018).
52. Uemura, T. *et al.* U-radiomics combined with hyper-curvature features for predicting survival of patients with idiopathic pulmonary fibrosis. *Int. J. Comput. Assist. Radiol. Surg.* **15**, S119–S120 (2020).

Acknowledgements

The authors acknowledge that this work did not receive any funding.

Author contributions

J.N. and T.U. contributed to methodology, software development, experiment, formal analysis, and writing the manuscript. C.W. contributed to clinical data collection and formal analysis. T.H. contributed to data collection and software developments. T.K. provided technical consultation. H.Y. conceptualized and supervised the study, carried out formal analyses, and reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-88591-z>.

Correspondence and requests for materials should be addressed to H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021