ORIGINAL PAPER

Deep Learning–Based Time-to-Death Prediction Model for COVID-19 Patients Using Clinical Data and Chest Radiographs

Toshimasa Matsumoto[2] · Shannon Leigh Walston[2] · Michael Walston[1] · Daijiro Kabata[3] · Yukio Miki[2] · Masatsugu Shiba[1,3] · Daiju Ueda[1,2]

Abstract

Accurate estimation of mortality and time to death at admission for COVID-19 patients is important and several deep learning models have been created for this task. However, there are currently no prognostic models which use end-to-end deep learning to predict time to event for admitted COVID-19 patients using chest radiographs and clinical data. We retrospectively implemented a new artificial intelligence model combining DeepSurv (a multiple-perceptron implementation of the Cox proportional hazards model) and a convolutional neural network (CNN) using 1356 COVID-19 inpatients. For comparison, we also prepared DeepSurv only with clinical data, DeepSurv only with images (CNNSurv), and Cox proportional hazards models. Clinical data and chest radiographs at admission were used to estimate patient outcome (death or discharge) and duration to the outcome. The Harrel's concordance index (c-index) of the DeepSurv with CNN model was 0.82 (0.75–0.88) and this was significantly higher than the DeepSurv only with clinical data model (c-index = 0.77 (0.69–0.84), $p = 0.011$), CNNSurv (c-index = 0.70 (0.63–0.79), $p = 0.001$), and the Cox proportional hazards model (c-index = 0.71 (0.63–0.79), $p = 0.001$). These results suggest that the time-to-event prognosis model became more accurate when chest radiographs and clinical data were used together.

Keywords Deep learning · Chest radiography · Artificial intelligence · COVID-19 · Prognosis

Introduction

As of November 2021, there are 250 million confirmed cases of COVID-19 worldwide, with more than 5 million deaths. The number of new cases is still increasing daily (https://covid19.who.int/). Therefore, it is essential for healthcare providers to efficiently triage patients with COVID-19. Predicting disease severity and progression in COVID-19 patients is important, as early intervention has been shown to reduce mortality [1, 2].

The Cox proportional hazards model, which can contrast variables associated with event and time to event, is a frequently applied analysis in medical research [3]. The model provides us not just the outcome (i.e., deceased or not) but also the time to event, which is more helpful for clinical practice. There are several studies which estimate the prognosis of COVID-19 patients using this model [4–6]. These include models that predict the time to death [4], the severity of illness [5], and the length of hospital stay [6] for patients. In these studies, the Cox proportional hazards model showed high performance but it has a limitation. It assumes linearity rather than performing nonlinear analysis which could better reflect actual clinical characteristics [3]. For example, BMI is a known nonlinear risk factor for COVID-19 admission and death [7]. Therefore, there is a need for a better solution that focuses on nonlinear variables.

In recent years, deep learning has been attracting attention in the medical field [8, 9]. With deep learning, it is possible to extract the complex linear and nonlinear relationship between clinical characteristics and individual prognosis. Integrating deep learning into a Cox proportional hazards model has led to the development of the deep

✉ Daiju Ueda
ai.labo.ocu@gmail.com

1   Smart Life Science Lab, Center for Health Science Innovation, Osaka Metropolitan University, 1-4-3 Asahi-machi, Abeno-ku, Osaka 545-8585, Japan

2   Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka Metropolitan University, 1-4-3 Asahi-machi, Abeno-ku, Osaka 545-8585, Japan

3   Department of Medical Statistics, Graduate School of Medicine, Osaka Metropolitan University, 1-4-3 Asahi-machi, Abeno-ku, Osaka 545-8585, Japan

learning survival neural network (DeepSurv) [10]. This has been shown to perform as well or better than other survival analysis methods on survival data with linear and nonlinear covariates. The advantage of time-to-death estimation is that it can provide more information than the conventional binary classification task. Conventional binary classification does not estimate how many days until a patient is at increased risk of death. A time-to-death model, on the other hand, can estimate the risk of death over time from data at a fixed point in time (at the time of admission in this model).

Medical images are known to be useful for prognostication in COVID-19. For example, the usefulness of chest radiographs [11–14] and chest CT [15, 16] was reported. Although CT is three dimensional and highly sensitive, chest radiographs may be more useful in the COVID-19 pandemic because they are relatively quick, low cost, portable, and accessible. Some reports show chest radiography for COVID-19 patients is indicative of the risk of hospitalization, duration of hospitalization, and risk of serious outcomes [12–14].

We hypothesized that we could build a better prognostic model by using chest radiographs with clinical data. Since DeepSurv does not have a mechanism to handle images, we developed a new model that integrated a convolutional neural network (CNN) which is one of the deep learning fields into DeepSurv. This allows us to handle both clinical data and images at once for prognosis estimation. So far, no study has developed an end-to-end deep learning model to predict time to event which combines clinical data and whole images as inputs. Using this newly created model, we predicted the mortality and time to death of patients hospitalized with COVID-19. Additionally, we scored the importance of the images compared to various clinical data.

Methods

Study Design

At first, we integrated a CNN into DeepSurv (DeepSurv with CNN model). Then, we developed and tested the model to estimate time to death for patients hospitalized with COVID-19. After developing the DeepSurv with CNN model, we compared the importance relative to other clinical data and visualized the region of interest of the radiographs. For comparison, we also developed the model with only the CNN component (CNNSurv model), with only the clinical variables (DeepSurv-only model), and a Cox proportional hazards model. An overview of our study is shown in Fig. 1. Chest radiographs were collected from the Stony Brook University COVID-19 dataset [17] in The Cancer Imaging Archive [18]. Since this dataset is open source, there is no need for review by the ethics board. We have created this article in compliance with the STARD statements [19].

Study Patients and Ground Truth Labeling

This dataset was acquired at Stony Brook University from patients who tested positive by PCR for COVID-19. Since this dataset was consecutively extracted from

Fig. 1 Overview of the prognostic models. We developed four prognostic models: a Cox proportional hazards model using only clinical data at the time of admission, a DeepSurv model using only clinical data at the time of admission, a DeepSurv with CNN model using clinical data at the time of admission and chest radiographs, and a CNNSurv-only model using chest radiographs
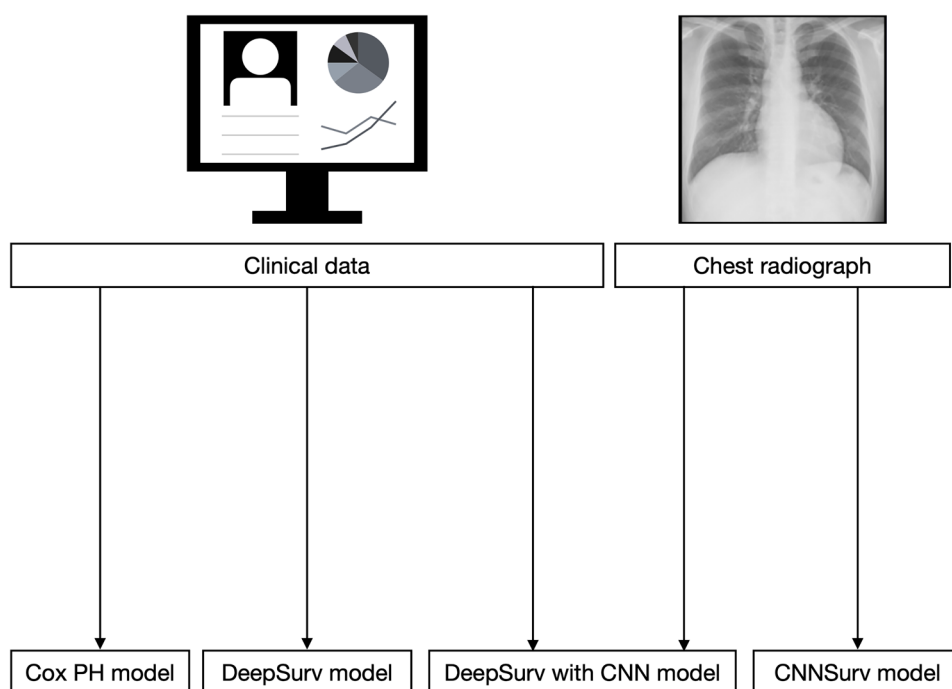
the electronic medical records, it is representative of the population at that center. The dataset consists of pre- and post-admission images (Xp, CT, MR, etc.) and a csv listing test results and patient information. All imaging from the pre- and post-admission periods are available; those closest to the time of admission were extracted for this study. As for the latter records, only data at the time of admission were available, and none of the subsequent data during hospitalization were available. In this data set, anticoagulant therapy is used as a therapeutic intervention. However, the data show no significant difference between those patients who did and did not take anticoagulant therapy [20–22], so the impact of therapeutic interventions taken prior to hospitalization on survival is likely to be minimal. Clinical data includes medical history, blood tests, and vital signs. This dataset included 1384 COVID-19 patients. We extracted the one radiograph taken closest to the time of admission. All radiographs were taken in anterior–posterior view. A total of 1356 patients were used for this study after excluding 28 patients who did not have a chest radiograph at admission. The eligibility flowchart is shown in Fig. 2. Clinical data and chest radiographs at admission were extracted as explanatory variables. As the ground truth, patient outcome (death or discharge) and duration until the outcome were extracted and used as objective variables. Detailed demographics are shown in Table 1.

Clinical Data Selection

The objective was to create a model that could predict patient prognosis with data available at the time of hospitalization. We chose variables which have been shown to be risk factors for severe COVID-19 [23–25]. Clinical data includes gender, age, smoking history, BMI, and medical history (hypertension, diabetes, chronic heart disease, chronic renal failure, chronic lung disease, and malignancy). Additionally, vital signs (heart rate, systolic blood pressure, respiratory rate, and blood oxygen saturation) and laboratory data (white blood cell count, sodium, potassium, c-reactive protein, aspartate aminotransferase, alanine aminotransferase, urea nitrogen, creatinine, lactate, brain natriuretic peptide, and d-dimer) were used. Mean vital signs and laboratory test results for each dataset are available in the Online Resource, Table 1.
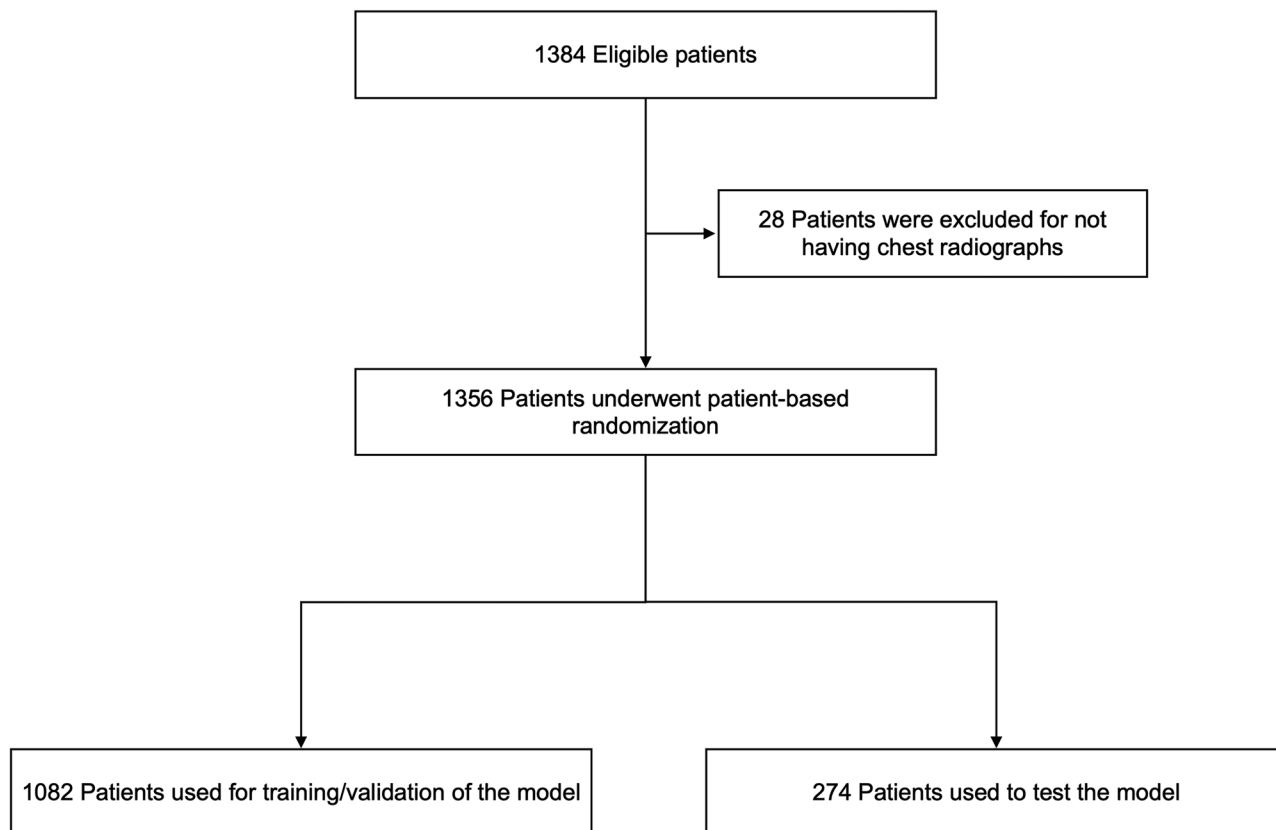
Fig. 2 Eligibility diagram

Table 1 Demographics

	Training/ validation dataset	Test dataset
Total no. of patients	1082	274
Male	621	159
Female	461	115
Age		
18–59	69	18
60–74	277	69
75–90	208	51
Period between admission and radiography (mean ± SD)	1 ± 1 day	1 ± 1 day
Smoking history	224	58
Body mass index, mean ± std	29.4 ± 6.0	29.2 ± 5.4
Disease history		
Hypertension	394	95
Diabetes	221	53
Chronic heart disease	151	42
Chronic kidney disease	65	16
Chronic lung disease	160	44
Malignancy	79	14
Outcomes		
Death	141	39
Discharge	941	235
Ventilation	175	38
ICU admission	215	45

Data Partition

All patients were randomly divided into training and test datasets at a ratio of 4:1. Definition of training and test datasets are shown in the Online Resource, Methods a. Since the partition was performed on a patient basis, there was no overlap of images or patients among the respective datasets. The training dataset included 1082 patients and the test dataset included 274 patients.

Image Processing

All chest radiographs in each dataset were resized to three sizes (256, 320, and 512 pixels). First, the longer side was downscaled while maintaining the aspect ratio. Second, the shorter side of the radiographs was padded black.

Model Implementation

We combined a CNN into DeepSurv [10]. Specifically, we concatenated the output of the CNN to the fully connected layer of DeepSurv to create an end-to-end deep learning model. This model is composed of both CNN and MLP structures. During forward propagation, the output of the CNN calculated from a chest radiograph is concatenated with clinical data and then they are passed to the MLP. The loss value is calculated on the output values of the MLP; in other words, it is calculated on both the radiograph and tabular data. The weights in both the CNN and MLP are then simultaneously updated. In each training session, the model took both the images and clinical data as input, predicted the outcome (death or discharge), and then both DeepSurv and the CNN in the model were simultaneously trained by back propagation. The CNN was developed using ResNet [26], DenseNet [27], and EfficientNet [28] architectures in the PyTorch framework [29]. It was trained from scratch with the training dataset using fivefold cross validation and independently tested with the test dataset. All images were augmented using random rotation, random shift, brightness shifts, and horizontally flipped. Detailed processes for development of the deep learning model are shown in the Online Resource, Methods b; machine environments are shown in the Online Resource, Methods c; an outline of the model is shown in the Online Resource, Fig. 1; and the source code is available online (https://github.com/deepsurv-cnn/).

Additionally, we prepared a CNNSurv model, DeepSurv-only model, and a Cox proportional hazards model for comparison. For the CNNSurv model, chest radiographs were used to estimate patients' prognosis. For the DeepSurv-only and the Cox proportional hazards models, clinical data were used. The CNNSurv model and the DeepSurv-only model were trained from scratch with the training dataset using fivefold cross validation and independently tested with the test dataset. As for the Cox proportional hazards model, principal component analysis was applied and used thirteen variables due to the large number of explanatory variables to prevent overfitting. Then the Cox proportional hazards model was independently evaluated with the test dataset.

Table 2 Results of each model

	C-index (95% CI)	Brier score (95% CI)	*p* value
Cox proportional hazards model	0.71 (0.63–0.79)	0.26 (0.20–0.32)	0.001
Deepsurv model	0.77 (0.69–0.84)	0.20 (0.13–0.27)	0.011
CNNsurv model	0.70 (0.63–0.79)	0.21 (0.19–0.23)	0.001
Deepsurv with CNN model	0.82 (0.75–0.88)	0.20 (0.13–0.27)	ref

CNN convolutional neural network

Importance Values and Saliency Maps

Importance values for each explanatory variables including chest radiographs were calculated using permutation importance with scikit-learn version 1.1.1 [30]. Permutation feature importance is a model inspection technique that is especially useful for nonlinear or opaque estimators. The permutation feature importance is defined as the decrease in a model score when a single feature value is randomly shuffled. This procedure breaks the relationship between the feature and the target; thus, the drop in the model score is indicative of how much the model depends on the feature.

A saliency map was generated for each chest radiograph to visualize the focus of the deep learning model as it estimated patient prognosis. A classification activation map was applied to create class-discriminative visualization of the chest radiograph [31]. A detailed explanation of the saliency map generation model is shown in the Online Resource, Fig. 2, and the source code is available online (https://github.com/deepsurv-cnn/).

Statistical Analysis

To evaluate the performance of the prognosis prediction models, we applied Harrell's concordance index (*c*-index) [32] of right-censored data and the brier score [33]. The *c*-index of the models compared progression information (death or discharged, and duration) with the rank of the predicted risk score. In addition, the Kaplan–Meier method was used to stratify patients into high- and low-risk subgroups according to the median progression risk score. Stratification performance was assessed using the log-rank test based on the predicted risk score of the stratified subgroups [34]. Time-dependent area under the curve (AUC) was calculated based on the predicted results of the DeepSurv with CNN model. Different prediction models were compared using binomial tests to show the difference in performance. A *p*-value less than 0.05 was considered significant. All analyses were performed using R (version 4.0.0.) and Python 3.8.1.

Results

Model Development

The models were each independently developed using the training dataset applied for 100 training epochs using fivefold cross validation. The final hyperparameters for the DeepSurv with CNN, CNNSurv, and DeepSurv-only models were the Adam optimizer (learning ratio = 0.001), a chest radiograph size of 256 pixels, a batch size of 64, and DenseNet. The cumulative contribution using the principal

component analysis was 0.97 for the Cox proportional hazards model.

Model Evaluation

The Cox proportional hazards model had a *c*-index of 0.71 (0.63–0.79) and a brier score of 0.26 (0.20–0.32), the DeepSurv-only model had a *c*-index of 0.77 (0.69–0.84) and a brier score of 0.20 (0.13–0.27), and the CNNSurv model had a *c*-index of 0.70 (0.63−0.79) and a brier score of 0.21 (0.19−0.23), and the DeepSurv with CNN model had a *c*-index of 0.82 (0.75–0.88) and a brier score of 0.20 (0.13–0.27). The *c*-index of the DeepSurv with CNN model was significantly higher (*p*-values were 0.001 compared to the Cox proportional hazards model, 0.001 compared to the CNNSurv model and 0.011 compared to the DeepSurv-only model) than the other models (Table 2).

Kaplan–Meier curves for risk stratification are shown in Fig. 3. As shown, the Cox proportional hazards model, DeepSurv model, CNNSurv model, and DeepSurv with CNN model were discriminative in stratifying patients into high-risk and low-risk subgroups with *p*-values of 0.01, < 0.005, and < 0.005. Time-dependent AUC was over 0.8 throughout the first week (Fig. 4).

Importance Values and Saliency Maps

The importance values showed that age was the most important factor, followed by being male. Images were in the top five—the highest of all the examinations and laboratory tests done in the hospital (Fig. 5). As for the saliency maps, the hottest region was on the area of infiltration (Online Resource 1).

Discussion

In this study, we developed a deep learning-based model to predict mortality and time to event by integrating clinical data and imaging information of COVID-19 patients. To our knowledge, this is the first study to develop an end-to-end deep learning model to predict time to event which combines clinical data and whole images as inputs. The results showed that the *c*-index of the DeepSurv with CNN model was 0.82 (0.75–0.88) in the test dataset, which enabled correct stratification of COVID-19 patients. This model performed higher than the Cox proportional hazards model, CNNSurv-only model, and the DeepSurv-only model (*p*-value < 0.05). The time-dependent AUC shows excellent performance throughout the first week.

Predicting disease severity and progression in COVID-19 patients is important, as early intervention has been shown to reduce mortality [1, 2]. In COVID-19, for example, being
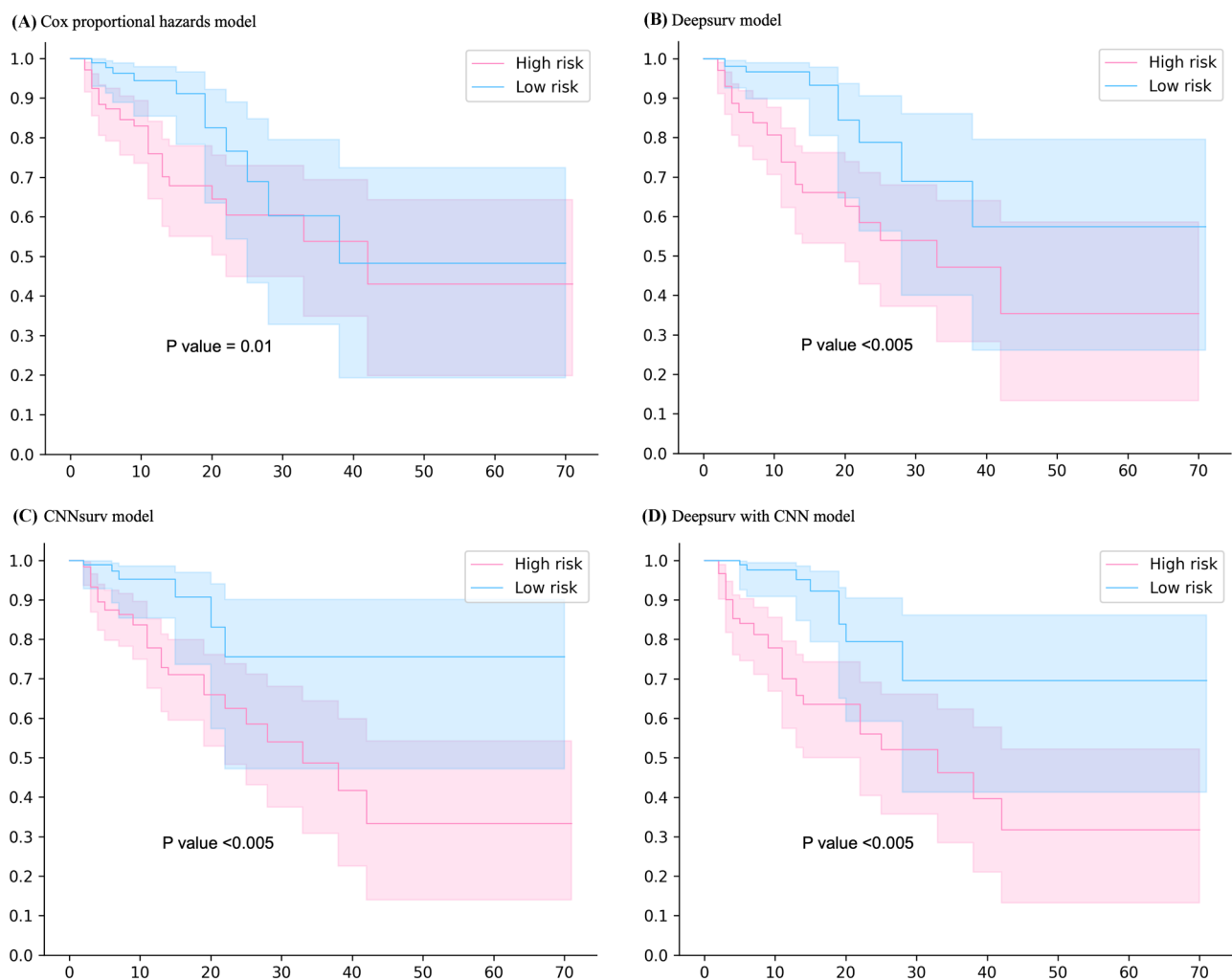
Fig. 3 Kaplan–Meier plots. The high-risk and low-risk patients from each model were divided based on the median model output value. This plot shows the ground truth survival of these patients, and the shaded area represents the accuracy of the prediction

male, advanced age, diabetes, and chronic respiratory disease are risk factors [23–25]. Chest radiography is important as a versatile imaging modality that has shown promise in aiding diagnosis and prognosis during the COVID-19 pandemic [11–14]. By merging chest radiography information with known risk factors, our model showed higher performance for estimating COVID-19 prognosis. There are some differences between our study and previous studies [35–44]. First of all, our model is a time-to-death predictive model with image and clinical data, which allows us to estimate how likely it is that death will occur in the days following hospitalization, rather than only a binary classification. In this respect, it differs from many previous studies [35–43]. On the other hand, one study showed a CNN and a random survival forest-based model that predicts death or discharge, with time to event for COVID-19 patients [44]. This study is similar in concept to ours. Although the implementation is well designed, the training of the CNN and the random

survival forest was performed separately, while our model is trained simultaneously. Training simultaneously allows the model to represent more complex relationships between images and other explanatory variables. There is no study to implement a model which can predict time to event using end-to-end deep learning. Moreover, there has been no research comparing the importance of imaging among these factors. Here, we perform this comparison using permutation importance [45]. The top 10 results showed that in addition to chest radiographs, age, gender, medical history (chronic heart disease, chronic lung disease), oxygen saturation, and blood tests (C-reactive protein, lactate, creatinine) were important. The importance of age, gender, and medical history (chronic heart disease and chronic lung disease) have been reported in previous studies [23–25]. It also makes sense that oxygen saturation is important because it is an indicator of the severity of pneumonia. C-reactive protein and lactate represent the severity of the inflammatory

**Fig. 4** Time-dependent AUC
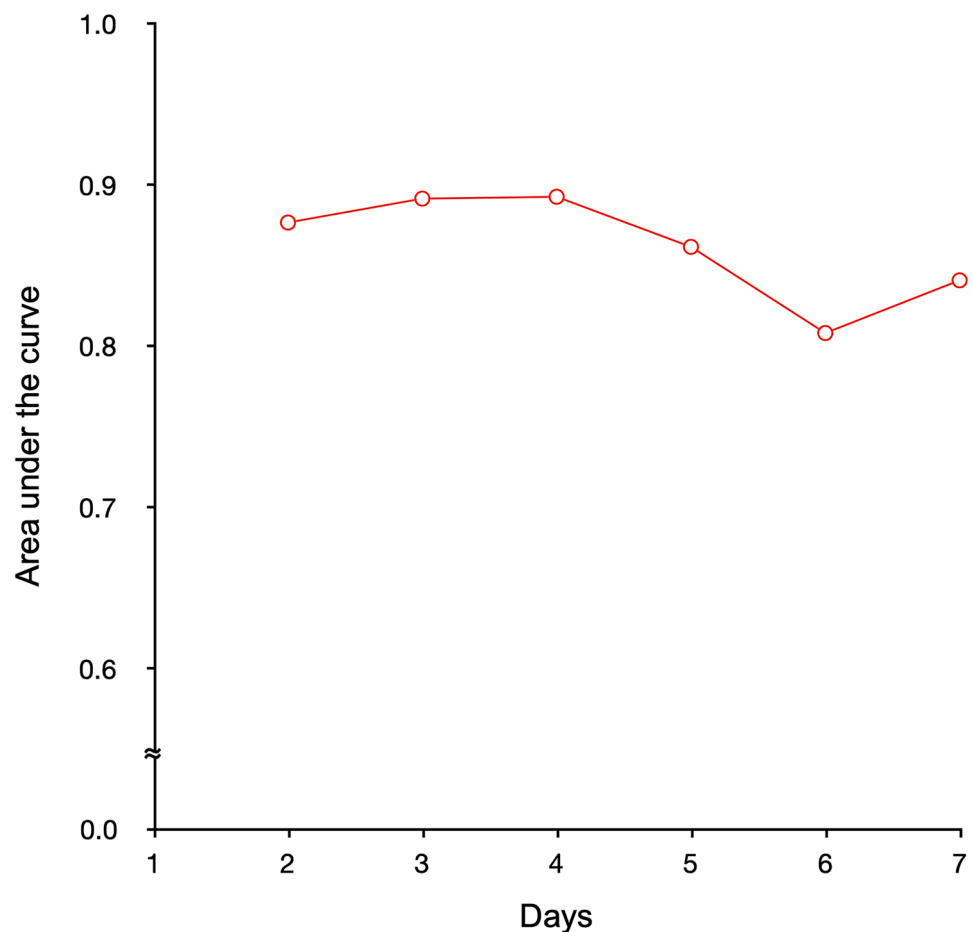
response, and creatinine is a value indicating renal function. All of these are well-known indicators of severity [46, 47]. Chest radiographs contain information such as age, gender, and oxygen saturation, which are covariates, and permutation importance is known to be lower when covariates are present. Even under these unfavorable conditions, the fact that the image is ranked in the top five means that the image is of outstanding importance.

DeepSurv, which applies deep learning to the Cox proportional hazards model, is gradually being introduced to the field of medicine [43, 48–50]. For example, it has been applied to head and neck cancer [48], oral cancer [43], lung cancers [49], and brain metastasis [50] to create more accurate and personalized prognostic models. However, the explanatory variables which can be used in the model are tabular data and not images. Until recently, it has been difficult to integrate images into a prognostic model. We can overcome this difficulty by the evolution of CNNs, starting with the neocognitron [51], and advances in machine power. Our DeepSurv with CNN model shows the best performance and may predict prognosis more accurately than the conventional Cox hazards proportional hazards model [3] or the DeepSurv-only model [10], which use risk factors other than imaging.

The model presented here has implications for other diseases which also currently rely on tabular clinical data to determine patient prognosis. For example, one of the most famous models for stratification of patient prognosis is the TNM staging of cancer patients [52]. Information about the tumor itself is aggregated into T, which most commonly uses only the diameter of the tumor. The malignancy of the tumor may be defined by the shape, volume, and internal properties of the tumor margins, but these are not taken into account in TNM staging. If imaging information can be used directly for stratification, as in our model, more individualized and accurate prognosis prediction will be possible.

Our model does not require high machine power, and the radiographs we handle are 256×256-pixel images which is much smaller than chest radiographs in digital imaging and communications in medicine format. Therefore, the model can be implemented into daily practice using any system with a central processing unit [53]. However, systems in hospitals are not simple, and in most cases, multiple systems coexist and cooperate with each other. Therefore, clinical implementation of this model may require additional investment in medical technology, such as an image extraction system from picture archiving and communication systems,

Fig. 5 Permutation importance. These values show the relative importance of each of the variables included in the models. These values have been sorted from greatest impact to least impact for ease of reading. The bar for chest radiograph images has been highlighted in pink
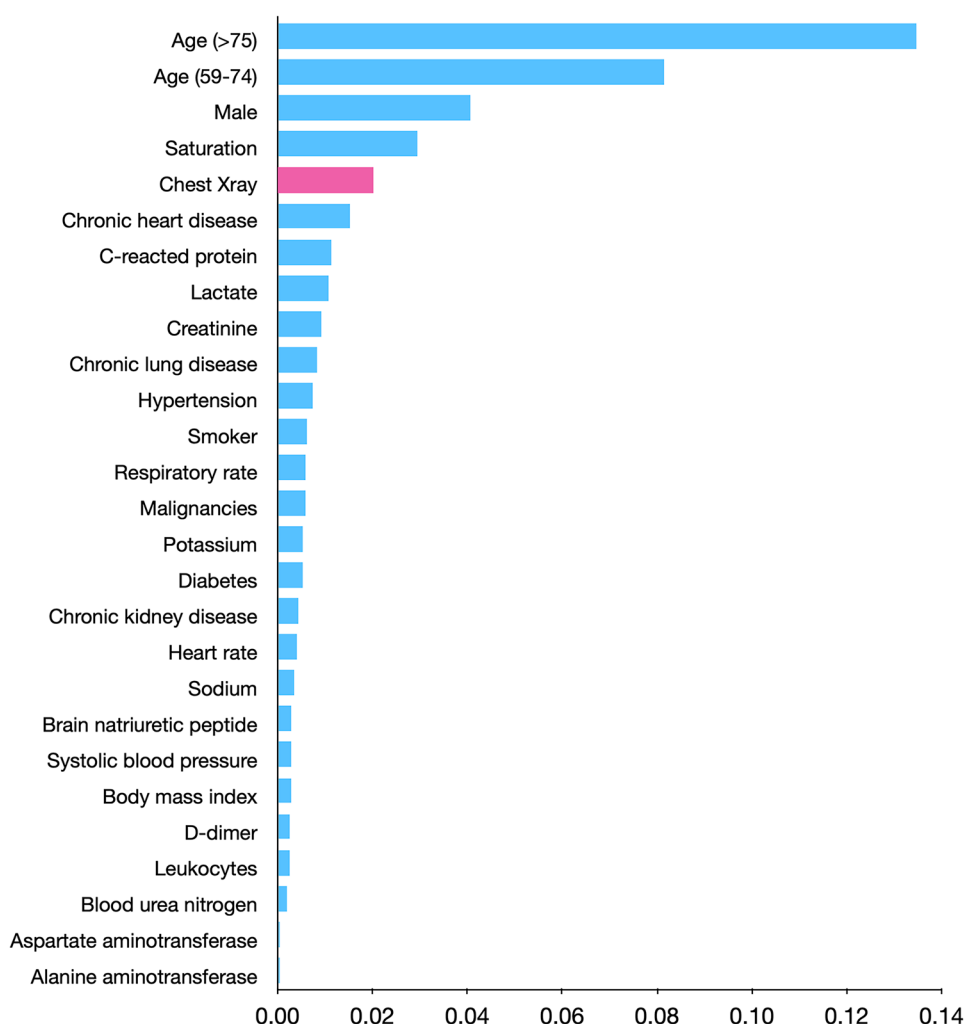
a computer analysis system for the images, and a system to provide the results to the physician.

This study had several limitations. The data in this study were collected from a single center. Further validation with a test dataset acquired in another institution is needed to show robustness of the model. In addition, this was a retrospective study and should be reviewed prospectively. In the clinical application of this model, it is best to retrain or fine tune it with data taken more recently because of the data set shift problem [54].

In clinical practice, patients admitted with COVID-19 have a chest radiograph taken as routine clinical practice. Our model was able to predict patient survival with high performance by using conventional tests taken upon admission and patient information. It also revealed the importance of the images themselves compared to these tests. Predicting patient prognosis allows healthcare providers to perform appropriate triage and management, and optimize the use of resources. Application of this model may not only support patients but also the hospital

systems which have struggled throughout this pandemic to maintain supplies. We plan to validate this model using a multicenter dataset and develop an even more comprehensive model which includes other pneumonias.

Declarations

References

1. Goyal DK, Mansab F, Iqbal A and Bhatti S. Early intervention likely improves mortality in COVID-19 infection. Clin Med (Lond). 2020;20(3):248-250.
2. Sun Q, Qiu H, Huang M and Yang Y. Lower mortality of COVID-19 by early recognition and intervention: experience from Jiangsu Province. Annals of Intensive Care. 2020;10(1):33.
3. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological). 1972;34(2):187–220.
4. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, Vargas-Vázquez A, González-Díaz A, Márquez-Salinas A, Fermín-Martínez CA, Naveja JJ and Aguilar-Salinas CA. Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. J Clin Endocrinol Metab. 2020;105(8).
5. Ji D, Zhang D, Xu J, Chen Z, Yang T, Zhao P, Chen G, Cheng G, Wang Y, Bi J, Tan L, Lau G and Qin E. Prediction for progression risk in patients with COVID-19 pneumonia: The CALL Score. Clin Infect Dis. 2020;71(6):1393-1399.
6. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, Wang M, Qiu X, Li H, Yu H, Gong W, Bai Y, Li L, Zhu Y, Wang L and Tian J. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. European Respiratory Journal. 2020:2000775.
7. Gao M, Piernas C, Astbury NM, Hippisley-Cox J, O'Rahilly S, Aveyard P and Jebb SA. Associations between body-mass index and COVID-19 severity in 6·9 million people in England: a prospective, community-based, cohort study. The Lancet Diabetes & Endocrinology. 2021;9(6):350-359.
8. Hinton G. Deep learning—a technology with the potential to transform health care. JAMA. 2018;320(11):1101-1102.
9. LeCun Y, Bengio Y and Hinton G. Deep learning. Nature. 2015;521(7553):436-444.
10. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T and Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology. 2018;18(1):24.
11. Chau T-N, Lee P-O, Choi K-W, Lee C-M, Ma K-F, Tsang T-Y, Tso Y-K, Chiu M-C, Tong W-L, Yu W-C and Lai S-T. Value of initial chest radiographs for predicting clinical outcomes in patients with severe acute respiratory syndrome. Am J Med. 2004;117(4):249-254.
12. Cozzi D, Albanesi M, Cavigli E, Moroni C, Bindi A, Luvarà S, Lucarini S, Busoni S, Mazzoni LN and Miele V. Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. Radiol Med. 2020;125(8):730-737.
13. Kim HW, Capaccione KM, Li G, Luk L, Widemon RS, Rahman O, Beylergil V, Mitchell R, D'Souza BM, Leb JS, Dumeer S, Bentley-Hibbert S, Liu M, Jambawalikar S, Austin JHM and Salvatore M. The role of initial chest X-ray in triaging patients with suspected COVID-19 during the pandemic. Emerg Radiol. 2020;27(6):617-621.
14. Toussie D, Voutsinas N, Finkelstein M, Cedillo MA, Manna S, Maron SZ, Jacobi A, Chung M, Bernheim A, Eber C, Concepcion J, Fayad ZA and Gupta YS. Clinical and chest radiography features determine patient outcomes in young and middle-aged adults with COVID-19. Radiology. 2020;297(1):E197-E206.
15. Bai HX, Hsieh B, Xiong Z, Halsey K, Choi JW, Tran TML, Pan I, Shi L-B, Wang D-C, Mei J, Jiang X-L, Zeng Q-H, Egglin TK, Hu P-F, Agarwal S, Xie F-F, Li S, Healey T, Atalay MK and Liao W-H. Performance of radiologists in differentiating COVID-19 from Non-COVID-19 viral pneumonia at chest CT. Radiology. 2020;296(2):E46-E54.
16. Yang W, Sirajuddin A, Zhang X, Liu G, Teng Z, Zhao S and Lu M. The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). Eur Radiol. 2020;30(9):4874-4882.
17. Saltz J, Saltz M, Prasanna P, Moffitt R, Hajagos J, Bremer E, Balsamo J and Kurc T. (2021). Stony brook university COVID-19 positive cases. https://doi.org/10.7937/TCIA.BBAG-2923
18. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L and Prior F. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging. 2013;26(6):1045-1057.
19. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HCW, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA and Cohen JF. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ : British Medical Journal. 2015;351:h5527.
20. Sholzberg M, Tang GH, Rahhal H, AlHamzah M, Kreuziger LB, Áinle FN, Alomran F, Alayed K, Alsheef M, AlSumait F, Pompilio CE, Sperlich C, Tangri S, Tang T, Jaksa P, Suryanarayan D, Almarshoodi M, Castellucci LA, James PD, Lillicrap D, Carrier M, Beckett A, Colovos C, Jayakar J, Arsenault MP, Wu C, Doyon K, Andreou ER, Dounaevskaia V, Tseng EK, Lim G, Fralick M, Middeldorp S, Lee AYY, Zuo F, da Costa BR, Thorpe KE, Negri EM, Cushman M and Jüni P. Effectiveness of therapeutic heparin versus prophylactic heparin on death, mechanical ventilation, or intensive care unit admission in moderately ill patients with covid-19 admitted to hospital: RAPID randomised clinical trial. Bmj. 2021;375:n2400.
21. Spyropoulos AC, Goldin M, Giannis D, Diab W, Wang J, Khanijo S, Mignatti A, Gianos E, Cohen M, Sharifova G, Lund JM, Tafur A, Lewis PA, Cohoon KP, Rahman H, Sison CP, Lesser ML, Ochani K, Agrawal N, Hsia J, Anderson VE, Bonaca M, Halperin JL and Weitz JI. Efficacy and safety of therapeutic-dose heparin vs standard prophylactic or intermediate-dose heparins for thromboprophylaxis in high-risk hospitalized patients with COVID-19: The HEP-COVID randomized clinical trial. JAMA Intern Med. 2021;181(12):1612-1620.
22. Goligher EC, Bradbury CA, McVerry BJ, Lawler PR, Berger JS, Gong MN, Carrier M, Reynolds HR, Kumar A, Turgeon AF, Kornblith LZ, Kahn SR, Marshall JC, Kim KS, Houston BL, Derde LPG, Cushman M, Tritschler T, Angus DC, Godoy LC, McQuilten Z, Kirwan BA, Farkouh ME, Brooks MM, Lewis RJ, Berry LR, Lorenzi E, Gordon AC, Ahuja T, Al-Beidh F, Annane D, Arabi YM, Aryal D, Baumann Kreuziger L, Beane A, Bhimani Z, Bihari S, Billett HH, Bond L, Bonten M, Brunkhorst F, Buxton M, Buzgau A, Castellucci LA, Chekuri S, Chen JT, Cheng AC, Chkhikvadze T, Coiffard B, Contreras A, Costantini TW, de Brouwer S, Detry MA, Duggal A, Džavík V, Effron MB, Eng HF, Escobedo J, Estcourt LJ, Everett BM, Fergusson DA, Fitzgerald M, Fowler RA, Froess JD, Fu Z, Galanaud JP, Galen BT, Gandotra S, Girard TD, Goodman AL, Goossens H, Green C, Greenstein YY, Gross PL, Haniffa R, Hegde SM, Hendrickson CM, Higgins AM, Hindenburg AA, Hope AA, Horowitz JM, Horvat CM, Huang DT, Hudock K, Hunt BJ, Husain M, Hyzy RC, Jacobson JR, Jayakumar D, Keller NM, Khan A, Kim Y, Kindzelski A, King AJ, Knudson MM, Kornblith AE, Kutcher ME, Laffan MA, Lamontagne F, Le Gal G, Leeper CM, Leifer ES, Lim G, Gallego Lima F, Linstrum K, Litton E, Lopez-Sendon J, Lother SA, Marten N, Saud Marinez A, Martinez M, Mateos Garcia E, Mavromichalis S, McAuley DF, McDonald EG, McGlothlin A, McGuinness SP, Middeldorp S, Montgomery SK, Mouncey PR, Murthy S, Nair GB, Nair R, Nichol AD, Nicolau JC, Nunez-Garcia B, Park JJ, Park PK, Parke RL, Parker JC, Parnia S, Paul JD, Pompilio M, Quigley JG, Rosenson RS, Rost NS, Rowan K, Santos FO, Santos M, Santos MO, Satterwhite L, Saunders CT, Schreiber J, Schutgens REG, Seymour CW, Siegal DM, Silva DG,

Jr., Singhal AB, Slutsky AS, Solvason D, Stanworth SJ, Turner AM, van Bentum-Puijk W, van de Veerdonk FL, van Diepen S, Vazquez-Grande G, Wahid L, Wareham V, Widmer RJ, Wilson JG, Yuriditsky E, Zhong Y, Berry SM, McArthur CJ, Neal MD, Hochman JS, Webb SA and Zarychanski R. Therapeutic anticoagulation with heparin in critically Ill patients with Covid-19. N Engl J Med. 2021;385(9):777-789.

23. Barry M, Alotaibi M, Almohaya A, Aldrees A, AlHijji A, Althabit N, Alhasani S, Akkielah L, AlRajhi A, Nouh T, Temsah MH and Al-Tawfiq JA. Factors associated with poor outcomes among hospitalized patients with COVID-19: Experience from a MERS-CoV referral hospital. J Infect Public Health. 2021;14(11):1658-1665.

24. Izcovich A, Ragusa MA, Tortosa F, Lavena Marzio MA, Agnoletti C, Bengolea A, Ceirano A, Espinosa F, Saavedra E, Sanguine V, Tassara A, Cid C, Catalano HN, Agarwal A, Foroutan F and Rada G. Prognostic factors for severity and mortality in patients infected with COVID-19: A systematic review. PLoS One. 2020;15(11):e0241955.

25. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, Curtis HJ, Mehrkar A, Evans D, Inglesby P, Cockburn J, McDonald HI, MacKenna B, Tomlinson L, Douglas IJ, Rentsch CT, Mathur R, Wong AYS, Grieve R, Harrison D, Forbes H, Schultze A, Croker R, Parry J, Hester F, Harper S, Perera R, Evans SJW, Smeeth L and Goldacre B. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020;584(7821):430-436.

26. He K, Zhang X, Ren S and Sun J, editors. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.

27. Huang G, Liu Z, Van Der Maaten L and Weinberger KQ, editors. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.

28. Tan M and Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning; 2019: PMLR.

29. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N and Antiga L. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems. 2019;32:8026-8037.

30. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.

31. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D and Batra D, editors. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision; 2017.

32. Harrell FE, Jr., Lee KL and Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-387.

33. Brier GW. Verification of forecasts expressed in terms of probability. Monthly Weather Review. 1950;78(1):1-3.

34. Heagerty PJ and Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005;61(1):92-105.

35. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, Bonten MMJ, Dahly DL, Damen JA, Debray TPA, de Jong VMT, De Vos M, Dhiman P, Haller MC, Harhay MO, Henckaerts L, Heus P, Kammer M, Kreuzberger N, Lohmann A, Luijken K, Ma J, Martin GP, McLernon DJ, Andaur Navarro CL, Reitsma JB, Sergeant JC, Shi C, Skoetz N, Smits LJM, Snell KIE, Sperrin M, Spijker R, Steyerberg EW, Takada T, Tzoulaki I, van Kuijk SMJ, van Bussel BCT, van der Horst ICC, van Royen FS, Verbakel JY, Wallisch C, Wilkinson J, Wolff R, Hooft L, Moons KGM and van Smeden M. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328.

36. Xu M, Ouyang L, Han L, Sun K, Yu T, Li Q, Tian H, Safarnejad L, Zhang H, Gao Y, Bao FS, Chen Y, Robinson P, Ge Y, Zhu B,

Liu J and Chen S. Accurately differentiating between patients with COVID-19, patients with other viral infections, and healthy individuals: Multimodal late fusion learning approach. J Med Internet Res. 2021;23(1):e25535.

37. Ho TT, Park J, Kim T, Park B, Lee J, Kim JY, Kim KB, Choi S, Kim YH, Lim J-K and Choi S. Deep learning models for predicting severe progression in COVID-19-infected patients: Retrospective study. JMIR Med Inform. 2021;9(1):e24973.

38. Fang C, Bai S, Chen Q, Zhou Y, Xia L, Qin L, Gong S, Xie X, Zhou C, Tu D, Zhang C, Liu X, Chen W, Bai X and Torr PHS. Deep learning for predicting COVID-19 malignant progression. Med Image Anal. 2021;72:102096.

39. Soda P, D'Amico NC, Tessadori J, Valbusa G, Guarrasi V, Bortolotto C, Akbar MU, Sicilia R, Cordelli E, Fazzini D, Cellina M, Oliva G, Callea G, Panella S, Cariati M, Cozzi D, Miele V, Stellato E, Carrafiello G, Castorani G, Simeone A, Preda L, Iannello G, Del Bue A, Tedoldi F, Alí M, Sona D and Papa S. AIforCOVID: Predicting the clinical outcomes in patients with COVID-19 applying AI to chest-X-rays. An Italian multicentre study. Med Image Anal. 2021;74:102216.

40. Kwon YJF, Toussie D, Finkelstein M, Cedillo MA, Maron SZ, Manna S, Voutsinas N, Eber C, Jacobi A, Bernheim A, Gupta YS, Chung MS, Fayad ZA, Glicksberg BS, Oermann EK and Costa AB. Combining initial radiographs and clinical variables improves deep learning prognostication in patients with COVID-19 from the emergency department. Radiology. Artificial intelligence. 2020;3(2):e200098-e200098.

41. Chieregato M, Frangiamore F, Morassi M, Baresi C, Nici S, Bassetti C, Bnà C and Galelli M. A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data. Sci Rep. 2022;12(1):4329.

42. Khozeimeh F, Sharifrazi D, Izadi NH, Joloudari JH, Shoeibi A, Alizadehsani R, Gorriz JM, Hussain S, Sani ZA, Moosaei H, Khosravi A, Nahavandi S and Islam SMS. Combining a convolutional neural network with autoencoders to predict the survival chance of COVID-19 patients. Scientific Reports. 2021;11(1):15343.

43. Kim DW, Lee S, Kwon S, Nam W, Cha I-H and Kim HJ. Deep learning-based survival prediction of oral cancer patients. Scientific Reports. 2019;9(1):6994.

44. Jiao Z, Choi JW, Halsey K, Tran TML, Hsieh B, Wang D, Eweje F, Wang R, Chang K, Wu J, Collins SA, Yi TY, Delworth AT, Liu T, Healey TT, Lu S, Wang J, Feng X, Atalay MK, Yang L, Feldman M, Zhang PJL, Liao W-H, Fan Y and Bai HX. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. The Lancet Digital Health. 2021;3(5):e286-e294.

45. Lundberg SM and Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA: Curran Associates Inc.; 2017. p. 4768–4777.

46. Kermali M, Khalsa RK, Pillai K, Ismail Z and Harky A. The role of biomarkers in diagnosis of COVID-19 - A systematic review. Life sciences. 2020;254:117788-117788.

47. Lambden S, Laterre PF, Levy MM and Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. Critical care (London, England). 2019;23(1):374-374.

48. Howard FM, Kochanny S, Koshy M, Spiotto M and Pearson AT. Machine learning-guided adjuvant treatment of head and neck cancer. JAMA Netw Open. 2020;3(11):e2025881.

49. She Y, Jin Z, Wu J, Deng J, Zhang L, Su H, Jiang G, Liu H, Xie D, Cao N, Ren Y and Chen C. Development and validation of a deep learning model for non–small cell lung cancer survival. JAMA Network Open. 2020;3(6):e205842-e205842.

50. Bice N, Kirby N, Bahr T, Rasmussen K, Saenz D, Wagner T, Papanikolaou N and Fakhreddine M. Deep learning-based

survival analysis for brain metastasis patients with the national cancer database. Journal of Applied Clinical Medical Physics. 2020;21(9):187-192.

51. Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics. 1980;36(4):193-202.

52. Brierley JD, Gospodarowicz MK and Wittekind C. TNM classification of malignant tumours: John Wiley & Sons; 2017.

53. Dreyer KJ and Geis JR. When machines think: Radiology's next frontier. Radiology. 2017;285(3):713-718.

54. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, Kohane IS and Saria S. The clinician and dataset shift in artificial intelligence. New England Journal of Medicine. 2021;385(3):283-286.