



An interpretable mortality prediction model for COVID-19 patients

Li Yan^{1,10}, Hai-Tao Zhang^{2,10}, Jorge Goncalves^{3,4,10}, Yang Xiao², Maolin Wang², Yuqi Guo², Chuan Sun², Xiuchuan Tang⁵, Liang Jing¹, Mingyang Zhang², Xiang Huang², Ying Xiao², Haosen Cao², Yanyan Chen⁶, Tongxin Ren⁷, Fang Wang¹, Yaru Xiao¹, Sufang Huang¹, Xi Tan⁸, Niannian Huang⁸, Bo Jiao⁸, Cheng Cheng², Yong Zhang⁹, Ailin Luo⁸, Laurent Mombaerts³, Junyang Jin⁷, Zhiguo Cao¹⁰, Shusheng Li¹✉, Hui Xu⁸✉ and Ye Yuan²✉

The sudden increase in COVID-19 cases is putting high pressure on healthcare services worldwide. At this stage, fast, accurate and early clinical assessment of the disease severity is vital. To support decision making and logistical planning in healthcare systems, this study leverages a database of blood samples from 485 infected patients in the region of Wuhan, China, to identify crucial predictive biomarkers of disease mortality. For this purpose, machine learning tools selected three biomarkers that predict the mortality of individual patients more than 10 days in advance with more than 90% accuracy: lactic dehydrogenase (LDH), lymphocyte and high-sensitivity C-reactive protein (hs-CRP). In particular, relatively high levels of LDH alone seem to play a crucial role in distinguishing the vast majority of cases that require immediate medical attention. This finding is consistent with current medical knowledge that high LDH levels are associated with tissue breakdown occurring in various diseases, including pulmonary disorders such as pneumonia. Overall, this Article suggests a simple and operable decision rule to quickly predict patients at the highest risk, allowing them to be prioritized and potentially reducing the mortality rate.

Outbreaks of the COVID-19 epidemic have been causing worldwide health concerns since December 2019. The virus causes fever, cough, fatigue and mild to severe respiratory complications, which, if very severe, can lead to patient death. On 6 March, there were 98,192 cumulated cases of infection across the world and 3,045 deaths had been reported¹. On 11 March, the virus outbreak was declared a pandemic by the World Health Organization². So far, it has been reported that 13.8–19.1% of COVID-19-infected patients in Wuhan, China, became severely ill^{3–5}. Furthermore, recent reports have exposed an astonishing case fatality rate of 61.5% for critical cases, increasing sharply with age and for patients with underlying comorbidities⁶. The severity of cases is putting great pressure on medical services, leading to a shortage of intensive care resources.

Unfortunately, there is no currently available prognostic biomarker to distinguish patients that require immediate medical attention and to estimate their associated mortality rate. The capacity to identify cases that are at imminent risk of death has thus become an urgent yet challenging necessity. Under these circumstances, we retrospectively analysed the blood samples of 485 patients from the region of Wuhan, China, to identify robust and meaningful markers of mortality risk. A mathematical modelling approach based on state-of-the-art interpretable machine learning algorithms was devised to identify the most discriminative biomarkers of patient mortality. The problem was formulated as a classification task,

where the inputs included basic information, symptoms, blood samples and the results of laboratory tests, including liver function, kidney function, coagulation function, electrolytes and inflammatory factors, taken from originally general, severe and critical patients (Table 1), as well as their associated outcomes corresponding to either survival or death at the end of the examination period. Through optimization, this classifier aims to reveal the most crucial biomarkers distinguishing patients at imminent risk, thereby relieving clinical burden and potentially reducing the mortality rate.

Medical records were collected by using standard case report forms that included epidemiological, demographic, clinical, laboratory and mortality outcome information (Table 2 and Supplementary Data 1). The clinical outcomes were followed up to 24 February 2020. The study was approved by the Tongji Hospital Ethics Committee.

Data resources

The medical information of all patients collected between 10 January and 18 February 2020 were used for model development. Data originating from pregnant and breast-feeding women, patients younger than 18 years and recordings with data material less than 80% complete were excluded from subsequent analysis. For 375 patients, fever was the most common initial symptom (49.9%), followed by cough (13.9%), fatigue (3.7%) and dyspnoea (2.1%). The age distribution of the patients was 58.83 ± 16.46 years, and 59.7%

¹Department of Emergency, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ²School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. ³Luxembourg Centre for System Biomedicine, Luxembourg, Luxembourg. ⁴Department of Plant Sciences, University of Cambridge, Cambridge, UK. ⁵School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan, China. ⁶Department of Information Management, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ⁷Huazhong University of Science and Technology – Wuxi Research Institute, Wuhan, China. ⁸Department of Anesthesiology, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China. ⁹School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, China. ¹⁰These authors contributed equally: Li Yan, Hai-Tao Zhang, Jorge Goncalves. ✉e-mail: shushengli16@sina.com; sophia_wh@hotmail.com; yue@hust.edu.cn

Table 1 | Criteria for assessment of disease severity upon hospital admission

| Diagnosis criteria | | |
|--|---|---|
| One epidemiological history | Any two clinical manifestations | |
| Lived in Wuhan within 14 days before onset; had contact with patients with fever and respiratory symptoms from Wuhan within 14 days before onset; had contact with COVID-19 patients (positive for COVID-19 nucleic acid) within 14 days before onset; or part of a familial cluster of onsets | Fever and/or respiratory symptoms; normal or decreased total white blood cell count or decreased lymphocyte count during early stage of onset; typical imaging features | |
| | One of the following aetiological evidences: | |
| | SARS-CoV-2 nucleic acid is positive in respiratory or blood samples detected by RT-PCR; virus sequence detected in respiratory or blood samples share high homology with the known sequence of SARS-CoV-2 | |
| Patient's classification | | |
| General | Severe | Critical |
| At least one epidemiological history + at least two clinical manifestations + one aetiological evidence of COVID infection | Evidence of COVID infection + one of the following aetiological evidences: | Evidence of COVID infection + one of the three following conditions: |
| Patients with pneumonia signs in CT scans are defined as general | RR interval ≥ 30 b.p.m.; SpO ₂ $\leq 93\%$ at rest. | Shock; need mechanical ventilation; admitted into ICU because of MODS |
| RT-PCR, polymerase chain reaction with reverse transcription; SpO ₂ , saturation of peripheral oxygen; ICU, intensive care unit; MODS, multiple organ dysfunction syndrome. | | |

RT-PCR, polymerase chain reaction with reverse transcription; SpO₂, saturation of peripheral oxygen; ICU, intensive care unit; MODS, multiple organ dysfunction syndrome.

were male. The epidemiological history included Wuhan residents (37.9%), familial cluster (6.4%) and health workers (1.9%). The laboratory results are shown in Table 2. Of the 375 cases included in the subsequent analysis, 201 recovered from COVID-19 and were discharged from the hospital, while the remaining 174 died> deceased. Following this, 110 newly discharged or deceased patients between 19 February 2020 and 24 February 2020 were enrolled for analysis as an external test dataset.

The minimal, maximal and median follow-up times (from admission to hospital to death or discharge) for all 485 (375 + 110) patients are 0 days 02:01:58 (hours: minutes: seconds), 35 days 04:05:54 and 11 days 04:15:36, respectively. The high mortality rate seen in our study was related to the fact that Tongji Hospital admitted a higher rate of severe and critical cases in Wuhan. A patient's severity was empirically assessed by medical doctors according to the criteria in Table 1 only at admission⁷. Figure 1 summarizes the outcome of patients in three different classes.

Development of a machine learning model

Most patients had multiple blood samples taken throughout their stay in hospital. However, the model training and testing uses only the data from the final sample as inputs to the model to assess the crucial biomarkers of disease severity, distinguish patients that require immediate medical assistance and accurately match corresponding features to each label. Nevertheless, the model can be applied to all other blood samples and the predictive potential of the identified biomarkers estimated (see Estimation of the prediction horizon section). Missing data were '−1' padded. The model

Table 2 | Epidemiological, demographic, clinical, laboratory and mortality outcome information collected from medical records

| Characteristics | Overall |
|--|-------------------------|
| Age, mean (s.d.) (years) | 58.83 (16.46) |
| Gender, <i>n</i> (%) | |
| Male | 224 (59.7) |
| Female | 151 (40.3) |
| Epidemiological history, <i>n</i> (%) | |
| Wuhan residents | 142 (37.9) |
| Contact with confirmed or suspected patients | 2 (0.5) |
| Familial cluster | 24 (6.4) |
| Health worker | 7 (1.9) |
| Contact with Huanan Seafood Market | 2 (0.5) |
| Undefined contact history | 198 (52.8) |
| Symptoms on onset, <i>n</i> (%) | |
| Fever | 187 (49.9) |
| Cough | 52 (13.9) |
| Fatigue | 14 (3.7) |
| Dyspnoea | 8 (2.1) |
| Chest distress | 7 (1.9) |
| Muscular soreness | 2 (0.5) |
| Outcomes, <i>n</i> (%) | |
| Survival rate | 201 (53.6) |
| Mortality rate | 174 (46.4) |
| Laboratory test (patient's last measurements) | |
| LDH, median (range, Q1–Q3) (U l ^{−1}) | 273.50 (199.00, 617.75) |
| Lymphocytes, median (range, Q1–Q3) (%) | 14.35 (4.17, 27.53) |
| High-sensitivity C-reactive protein (mg l ^{−1}) | 26.3 (2.0, 99.10) |
| Sodium median (range, Q1–Q3) (mmol l ^{−1}) | 140.7 (138.3, 143.3) |
| Chlorine median (range, Q1–Q3) (mmol l ^{−1}) | 102.3 (99.53, 105.58) |
| International normalized ratio (range, Q1–Q3) | 1.10 (1.02, 1.30) |
| Eosinophils, median (range, Q1–Q3) (×10 ⁹ l ^{−1}) | 0.02 (0.00, 0.09) |
| Eosinophils, median (range, Q1–Q3) (%) | 0.25 (0.00, 1.50) |
| Monocytes, median (range, Q1–Q3) (%) | 6.25 (2.98, 8.90) |
| Albumin, mean (s.d.) (g l ^{−1}) | 32.67 (6.31) |

Data were first tested for normality. The Kolmogorov–Smirnov test was used to test whether a single sample was from a particular distribution, then this one-sample Kolmogorov–Smirnov test checked the normality of the data. A test level of $\alpha = 0.05$ and $P < 0.05$ indicate that a sample does not fit a normal distribution. LDH, lactic dehydrogenase. The continuity variables of normal distributions are described by the mean (s.d.) and the continuity variables of non-normal distributions by the median and quartiles.

output corresponds to patient mortality. Patients that survived were assigned to class 0 and those that died to class 1.

The performance models were evaluated by assessing the classification accuracy (ratio of true predictions over all predictions), the precision, sensitivity/recall and F1 scores (defined below):

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (1)$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (2)$$

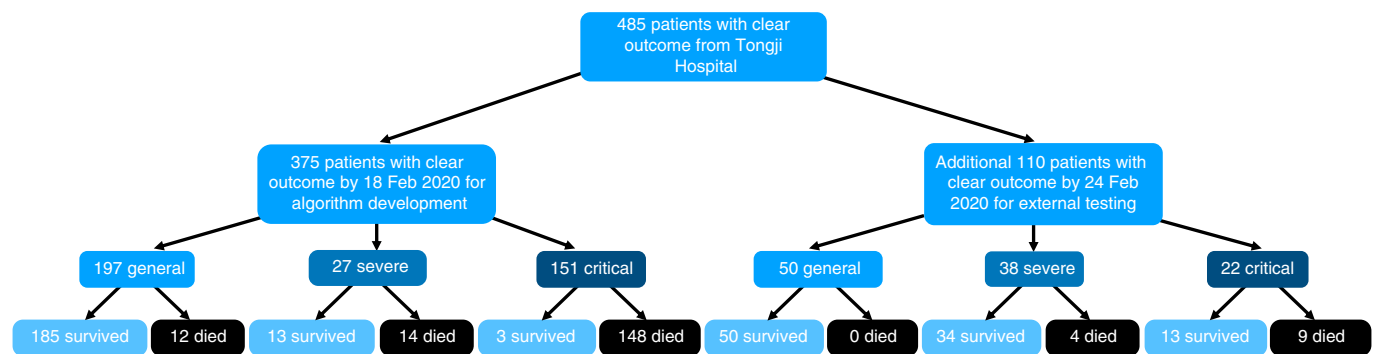


Fig. 1 | A flowchart of patient enrolment. Originally, 375 patients with a definite outcome before 18 February 2020 were used for model development, then an additional 110 patients with a definite outcome between 19 February 2020 and 24 February 2020 were used as an external test dataset.

$$F1_i = \frac{2 \times \text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$\text{Macro averages(score)} = \frac{1}{C} \sum_i \text{score}_i \quad (5)$$

$$\text{Weighted averages(score)} = \frac{1}{N} \sum_i N_i \cdot \text{score}_i \quad (6)$$

$\text{score} \in \{\text{Precision}, \text{Recall}, F1\}$

where $i \in C$ represents the class, N is the number of all samples, C is the number of all classes, N_i is the number of samples, TN_i , in class i , TP_i , FP_i and FN_i stand for true positive, true negative, false positive and false negative rates for class i , respectively. In total, 75 features were considered.

This study uses a supervised XGBoost classifier⁸ as the predictor model. XGBoost is a high-performance machine learning algorithm that benefits from great interpretability potential due to its recursive tree-based decision system. In contrast, internal model mechanisms of black-box modelling strategies are typically difficult to interpret. The importance of each individual feature in XGBoost is determined by its accumulated use in each decision step in trees. This computes a metric characterizing the relative importance of each feature, which is particularly valuable to estimate features that are the most discriminative of model outcomes, especially when they are related to meaningful clinical parameters.

XGBoost was originally trained with the following default parameter settings: maximum depth equal to 4, learning rate equal to 0.2, number of tree estimators set to 150, value of the regularization parameter α set to 1 and 'subsample' and 'colsample_bytree' both set to 0.9 to prevent overfitting for cases with many features and small sample size⁸. We refer to it as the 'Multi-tree XGBoost algorithm'.

Feature importance for an operable decision tree

To evaluate the markers of imminent mortality risk, we assessed the contribution of each patient parameter to decisions of the algorithm. Features were ranked by Multi-tree XGBoost according to their importance (Supplementary Figs. 1 and 2 and Supplementary algorithm 1). The performances of the model showed no improvement in area under the curve (AUC) scores when the number of top features increased to four. Hence, the number of key features was set

Table 3 | Performances of the Multi-tree XGBoost classification in discriminating between mortality outcomes using 100-round fivefold cross-validation using Supplementary algorithm 1

| Features | AUC score for training sets (%) | AUC score for validation sets (%) |
|-------------------------|---------------------------------|-----------------------------------|
| LDH | 94.27 ± 0.82 | 92.29 ± 2.62 |
| LDH, lymphocyte | 96.74 ± 0.45 | 94.40 ± 2.31 |
| LDH, lymphocyte, hs-CRP | 97.84 ± 0.37 | 95.06 ± 2.21 |

Data presented as mean ± s.d.

to the following three: lactic dehydrogenase (LDH), lymphocytes and high-sensitivity C-reactive protein (hs-CRP).

Table 3 summarizes the performances of the Multi-tree XGBoost model. The results show that the model is able to accurately identify the outcome of patients, regardless of their original diagnosis upon hospital admission. Notably, the performance of the external test set (detailed below) is similar to that of the training and validation sets, which suggests that the model captures the key biomarkers of patient mortality. The set of selected features is represented graphically for each patient in Supplementary Fig. 3, demonstrating a clear separability. Table 3 further emphasizes the importance of LDH as a crucial biomarker for patient mortality rate.

Development of a clinically operable decision tree

Following previous findings on the importance of LDH, lymphocytes and hs-CRP, we aimed to construct a simplified and clinically operable decision model. XGBoost algorithms are based on recursive decision tree building from past residuals and can identify those trees that contribute the most to the decision of the predictive model. Decision trees are simple classifiers consisting of sequences of binary decisions organized hierarchically. Hence, if the accuracy of a tree remains high, reducing the complexity of the model to such a structure has the potential to reveal a clinically portable decision algorithm. In the following, we refer to the latter as an 'interpretable model' or 'single-tree XGBoost'.

There were 24 patients with incomplete measurements for at least one of the three principal biomarkers in their last blood samples, leaving 351 patients to identify a single-tree XGBoost model. To identify the model, XGBoost was re-trained with the same parameters as described above, except for the following: number of tree estimators set to 1, values of the regularization parameters α and β both set to 0, and the subsample and max features both set to 1 as overfitting issues have been avoided based on previous modelling⁸. The interpretable decision tree was obtained by a random split of

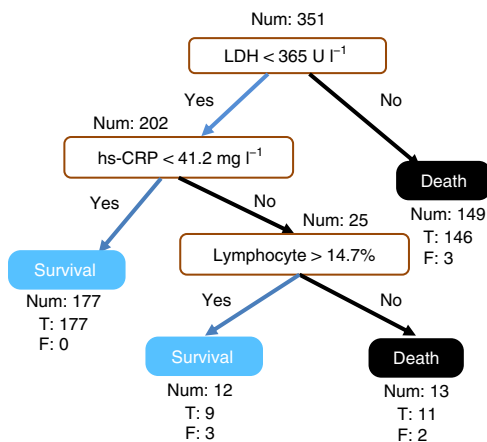


Fig. 2 | A decision rule using three key features and their thresholds in absolute value. Num, the number of patients in a class; T, the number of correctly classified; F, the number of misclassified patients.

the 351 patients to training and validation datasets in the ratio 7:3. The resulting tree structure and performances are shown, respectively, in Fig. 2 and Supplementary Tables 1 and 2.

In addition, the performances of the interpretable model were estimated for the external test set on the latest blood samples of 110 patients, which were not part of the training or validation of the Single-tree XGBoost model (Table 4). The associated confusion matrix is presented in Supplementary Fig. 5, which shows 100% survival prediction accuracy and 81% mortality prediction accuracy. Overall, the scores for survival and death prediction, accuracy, macro and weighted averages are consistently over 0.90.

Finally, for benchmark purposes, the performances of the interpretable model were compared with other standard methods such as random forest and logistic regression⁹. The receiver operating characteristic curves and AUC scores are shown in Supplementary Table 3 and Supplementary Fig. 4.

Estimation of the prediction horizon

Most patients had multiple blood samples taken throughout their hospital stay. In total, there were 909 blood samples with complete measurements of these three features for all 485 patients used for training and validation, and 251 blood samples with complete measurements of these three features for the 110 patients in the external test set. The predictive potential of our model was evaluated on all blood tests for all 485 patients and 110 patients in the external test dataset (Fig. 3 and Supplementary Figs. 6 and 7). On average, the accuracy of our algorithm was 90%, further showing that the model could be applied to any blood sample, including those that were taken far ahead of the day of primary clinical outcome. On average, the model could predict the outcome of all true positive patients at about 10 days (11 days for patients in the external test set) in advance of outcome using all their blood samples (Fig. 3b,c). The model can even predict 18 days in advance with a cumulative accuracy above 90% (Fig. 3d,e). The accuracy of the prediction increases closer to the patient's outcome. This prediction horizon analysis suggests that, where a patient's condition deteriorates, the clinical route is able to give an early warning to clinicians a few days in advance.

Discussion

The significance of our work is twofold. First, it goes beyond providing high-risk factors⁴. It provides a simple and intuitive clinical test to precisely and quickly quantify the risk of death. For example, a routine sequential respiratory support therapy for patients with SpO₂ below 93% comprises intranasal catheterization of oxygen, oxygen supply through a mask, high-flow oxygen supply through a

Table 4 | Performance of the proposed interpretable model on the external test dataset

| | Precision | Recall | F1 score | Support |
|--------------|-----------|--------|----------|---------|
| Survival | 1.00 | 0.97 | 0.98 | 97 |
| Death | 0.81 | 1.00 | 0.90 | 13 |
| Accuracy | | | 0.97 | 110 |
| Macro avg | 0.91 | 0.98 | 0.94 | 110 |
| Weighted avg | 0.98 | 0.97 | 0.97 | 110 |

nasal catheter, non-invasive ventilation support, invasive ventilation support and extracorporeal membrane oxygenation. Predicting that for some patients this sequential oxygen therapy leads to unsatisfactory therapeutic effects could preempt physicians to pursue different approaches. The goal is for the model to identify high-risk patients before irreversible consequences occur. Second, the three key features, LDH, lymphocytes and hs-CRP, can be easily collected in any hospital. In crowded hospitals, and with shortages of medical resources, this simple model can help to quickly prioritize patients, especially during a pandemic when limited healthcare resources have to be allocated¹⁰.

The increase of LDH reflects tissue/cell destruction and is regarded as a common sign of tissue/cell damage. Serum LDH has been identified as an important biomarker for the activity and severity of idiopathic pulmonary fibrosis¹¹. In patients with severe pulmonary interstitial disease, the increase of LDH is significant and is one of the most important prognostic markers of lung injury¹¹. For critically ill patients with COVID-19, the rise in LDH level indicates an increase of the activity and extent of lung injury.

The increase of hs-CRP, an important marker for poor prognosis in acute respiratory distress syndrome^{12,13}, reflects a persistent state of inflammation¹⁴. The result of this persistent inflammatory response is large grey-white lesions in the lungs of patients with COVID-19 (seen in autopsy)¹⁵. In tissue sections, a large amount of sticky secretion is also seen overflowing from the alveoli¹⁵.

Finally, our results also suggest that lymphocytes may serve as a potential therapeutic target. This hypothesis is supported by the results of clinical studies^{4,16}. Lymphopenia is a common feature in patients with COVID-19 and might be a critical factor associated with disease severity and mortality¹⁷. Injured alveolar epithelial cells could induce the infiltration of lymphocytes, leading to persistent lymphopenia, as was seen in SARS-CoV-2 and MERS-CoV (they share similar alveolar penetrating and antigen presenting cell (APC) impairing pathways)^{18,19}. A biopsy study has provided strong evidence of substantially reduced counts of peripheral CD4 and CD8 T cells, while their status was hyperactivated²⁰. Also, Jing and colleagues have reported that the lymphopenia is mainly related to the decrease in CD4 and CD8 T cells²¹. It is thus likely that lymphocytes play distinct roles in COVID-19, which deserves further investigation.

This study has room for further improvement, which is left for future work. First, given that the proposed machine learning method is purely data-driven, our model may vary if starting from different datasets. As more data become available, the whole procedure can easily be repeated to obtain more accurate models. This is a single-centred, retrospective study, which provides a preliminary assessment of the clinical course and outcome of patients. We look forward to subsequent large-sample and multi-centred studies. Second, although we had a pool of more than 70 clinical features, our modelling principle is a trade-off between having a minimal number of features and the capacity of good prediction, therefore avoiding overfitting. Finally, this study strikes a balance between model interpretability and improved accuracy. Although clinical settings tend to prefer interpretable models, it is possible that a black-box model may lead to improved performance.

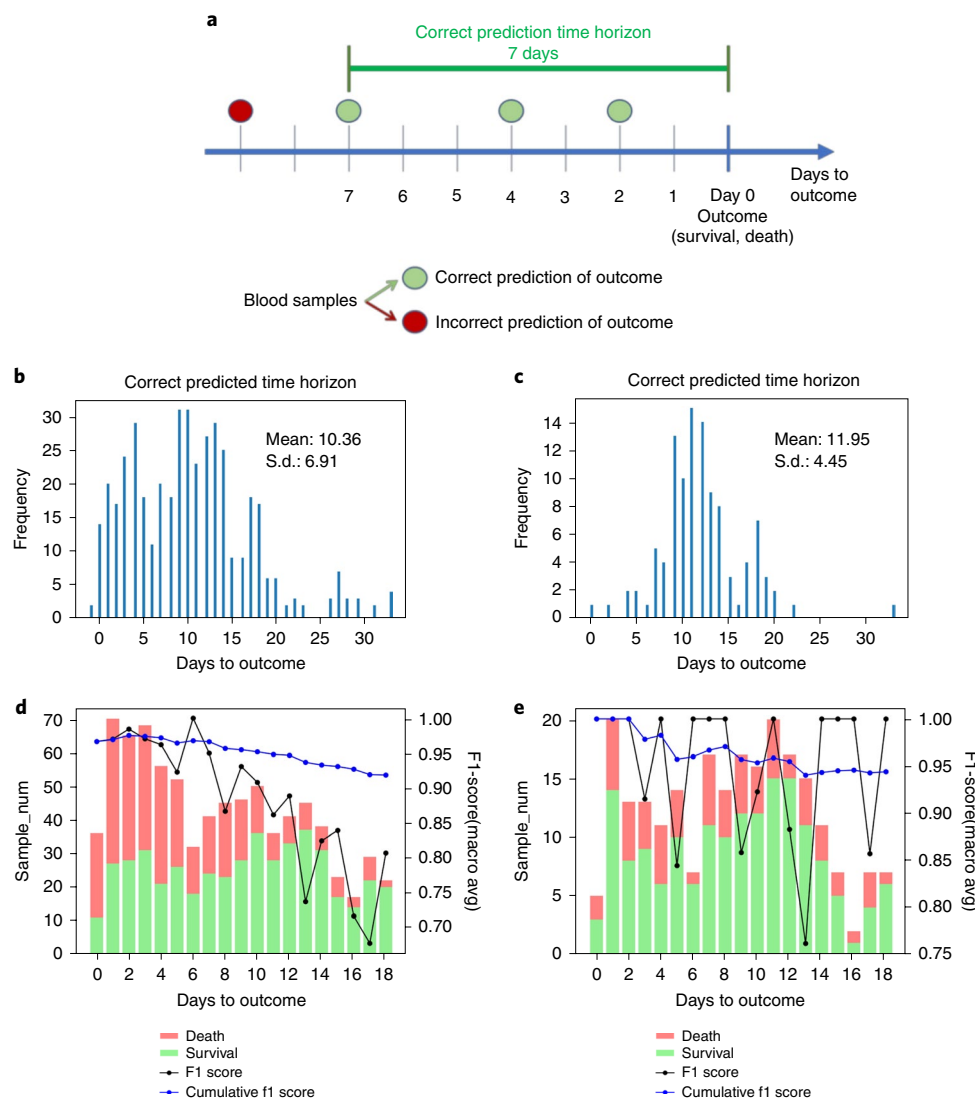


Fig. 3 | Estimation of the prediction horizon of the decision rule with three features. **a**, Illustration of the concept of the correct prediction time horizon. **b**, Histogram of the maximum correct prediction time horizons for all 485 patients with true positive prediction. Note that there are two patients with negative days, as their only blood sample results arrived one day after their clinical outcome. **c**, Histogram of the maximum correct prediction time horizons for 110 patients in the external test set with true positive prediction. **d**, The predictive performance (F1 score and cumulative F1 score) evaluated with respect to the day of outcome for all 485 patients. **e**, The predictive performance (F1 score and cumulative F1 score) evaluated with respect to the day of outcome for the 110 patients in the external test set.

Conclusion

In summary, this study has identified three indicators (LDH, hs-CRP and lymphocytes), together with a clinical route (Fig. 2), for COVID-19 prognostic prediction. We have developed an XGBoost machine learning-based model that can predict the mortality rates of patients more than 10 days in advance with more than 90% accuracy, enabling detection, early intervention and potentially a reduction of mortality in patients with COVID-19.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this Article.

Data and code availability

Data are available in the Supplementary Information. The code implementation is available at https://github.io/HAIRLAB/Pre_Surv_COVID_19 under an MIT licence (<https://doi.org/10.5281/zenodo.3758806>).

Received: 15 March 2020; Accepted: 29 April 2020;

Published online: 14 May 2020

References

1. World Health Organization. *Coronavirus Disease 2019 (COVID-19) Situation Report 46*, 6 March 2020 (2020); <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf>
2. World Health Organization. *Coronavirus Disease 2019 (COVID-19) Situation Report 68*, 28 March 2020 (2020); <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200328-sitrep-68-covid-19.pdf>
3. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
4. Chen, N. et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**, 507–513 (2020).
5. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Zhonghua Liu Xing Bing Xue Za Zhi* **41**, 145–151 (2020).
6. Yang, X. et al. Clinical course and outcomes of critically ill patients with SARS CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Resp. Med.* **8**, 475–481 (2020).

7. *Diagnosis and Treatment of Pneumonia Infected by the New Novel Coronavirus* (the trial fifth edition) Medical Letter from the National Health Office (National Health Commission of the People's Republic of China, 2020).
8. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
9. Lundberg, S. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
10. Truog, R. D., Mitchell, C. & Dalley, G. Q. The toughest triage—allocating ventilators in a pandemic. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMp2005689> (2020).
11. Kishaba, T., Tamaki, H., Shimaoka, Y., Fukuyama, H. & Yamashiro, S. Staging of acute exacerbation in patients with idiopathic pulmonary fibrosis. *Lung* **192**, 141–149 (2014).
12. Ridker, P. M. et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N. Engl. J. Med.* **359**, 2195–2207 (2008).
13. Sharma, S. K. et al. Aetiology, outcomes & predictors of mortality in acute respiratory distress syndrome from a tertiary care centre in North India. *Indian J. Med. Res.* **143**, 782–792 (2016).
14. Bajwa, E. K. et al. Plasma C-reactive protein levels are associated with improved outcome in ARDS. *Chest* **136**, 471–480 (2009).
15. Liu, X. et al. A general report on the systematic anatomy of COVID-19. *J. Forensic Med.* **36**, 1–3 (2020).
16. Wang, D. et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020).
17. Chan, J. F. et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514–523 (2020).
18. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
19. Ge, X. Y. et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013).
20. Xu, Z. et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Resp. Med.* **8**, 420–422 (2020).
21. Liu, J. et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EbioMedicine* **55**, 102763 (2020).

Acknowledgements

We would like to dedicate this paper to those who have devoted their lives to the battle with coronavirus.

Author contributions

Y.Y. conceptualized the idea. Y.Y., H.-T.Z. and L.Y. initialized, conceived and supervised the project. L.Y., H.X. and S.L. collected data. Y.Y., M.W., Y.G. and C.S. discovered key features and the clinical route. L.Y., H.-T.Z., Yang Xiao, L.M., H.X., J.G. and Y.Y. drafted the manuscript. All authors provided critical review of the manuscript and approved the final draft for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42256-020-0180-7>.

Correspondence and requests for materials should be addressed to S.L., H.X. or Y.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.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- ☐ ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- ☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- ☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- ☒ ☐ A description of all covariates tested
- ☒ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- ☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- ☐ ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- ☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- ☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- ☒ ☐ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Microsoft excel

Data analysis

Python

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data is in the supplement

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- ☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|-----------------|---------------------|
| Sample size | 485 patients |
| Data exclusions | no data is excluded |
| Replication | no replication |
| Randomization | no randomization |
| Blinding | no blinding |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |

Methods

| | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Human research participants

Policy information about [studies involving human research participants](#)

| | |
|----------------------------|--|
| Population characteristics | We considered the medical information of all COVID-19 patients collected between 10 January and 24 February 2020 from Tongji Hospital. Data originating from pregnant and breast-feeding women, patients younger than 18 years old, and recordings without at least 80% of complete data materials, were excluded from subsequent analysis. There are a total number of 485 patients with the definite outcome involved in this study. |
| Recruitment | Since this is a retrospective study, there was no specific recruitment process involved. |
| Ethics oversight | The study was approved by the Ethics Committee of Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.