

In the format provided by the authors and unedited.

# An interpretable mortality prediction model for COVID-19 patients

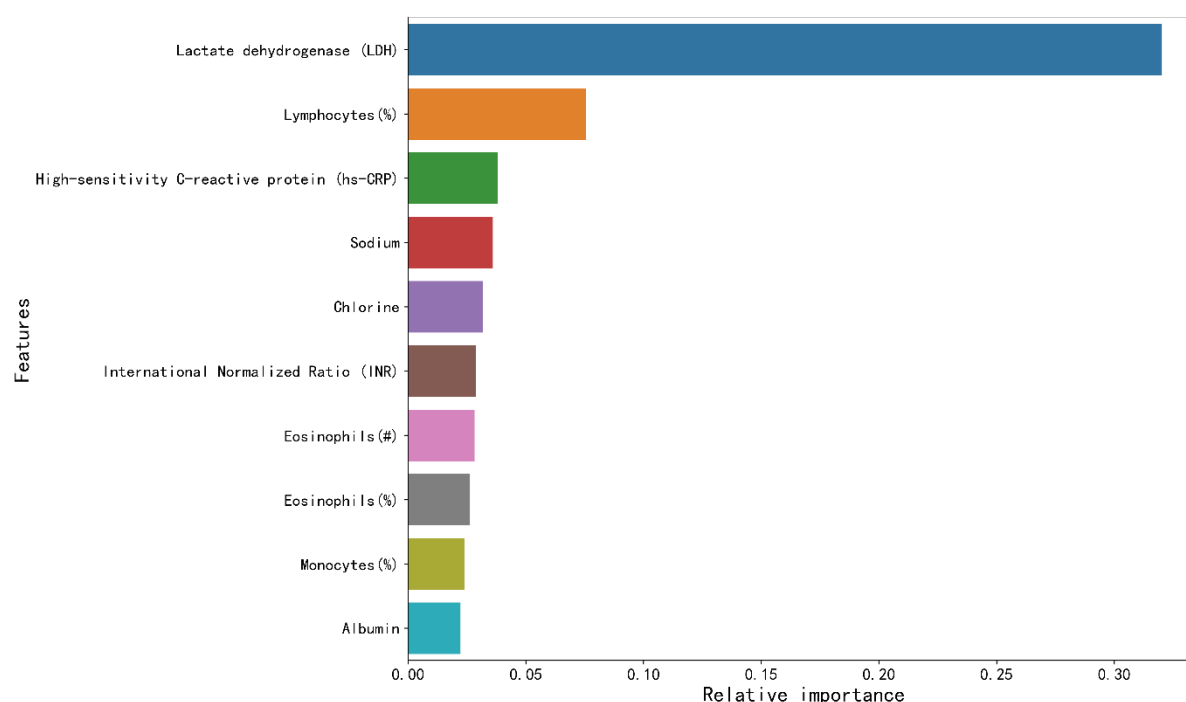
Li Yan<sup>1,10</sup>, Hai-Tao Zhang<sup>2,10</sup>, Jorge Goncalves<sup>3,4,10</sup>, Yang Xiao<sup>2</sup>, Maolin Wang<sup>2</sup>, Yuqi Guo<sup>2</sup>, Chuan Sun<sup>2</sup>, Xiuchuan Tang<sup>5</sup>, Liang Jing<sup>1</sup>, Mingyang Zhang<sup>2</sup>, Xiang Huang<sup>2</sup>, Ying Xiao<sup>2</sup>, Haosen Cao<sup>2</sup>, Yanyan Chen<sup>6</sup>, Tongxin Ren<sup>7</sup>, Fang Wang<sup>1</sup>, Yaru Xiao<sup>1</sup>, Sufang Huang<sup>1</sup>, Xi Tan<sup>8</sup>, Niannian Huang<sup>8</sup>, Bo Jiao<sup>8</sup>, Cheng Cheng<sup>2</sup>, Yong Zhang<sup>9</sup>, Ailin Luo<sup>8</sup>, Laurent Mombaerts<sup>3</sup>, Junyang Jin<sup>7</sup>, Zhiguo Cao<sup>2</sup>, Shusheng Li<sup>1</sup>✉, Hui Xu<sup>8</sup>✉ and Ye Yuan<sup>2</sup>✉

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

## Supplementary Information

In the Supplementary Information, we shall illustrate data analysis using a step-by-step procedure below:

**Step1.** Obtain the Top10 features using 375 samples with all features: Multi-tree XGBoost is trained with the parameters setting as the max depth with 4, the learning rate is equal 0.2, the tress number of estimators is set to 150, the value of the regularization parameter  $\alpha$  is set to 1, the 'subsample' and 'colsample\_bytree' both are set to 0.9 to prevent overfitting when there are many features and the sample size is not large.



Supplementary Figure 1: Top ten key clinical features that are ranked according to its importance in the Multi-tree XGBoost algorithm. We selected 100 random number seeds from 0 to 99, and then we split the data with 70% training data and 30% for validation with the different random number seeds and train a Multi-tree XGBoost model. After repeating these 100 times in order to ensure the stability of the features ranking, we can obtain the average importance of each feature.

**Step 2.** Reduce the number of features used:

Supplementary Algorithm 1:

---

**Algorithm 1** Feature selection

---

**Input:** The training set of  $\{X_{train} \in R^{n_1 \times m}, Y_{train} \in [0,1]^{n_1}\}$ , the validation set of  $\{X_{validation} \in R^{n_2 \times m}, Y_{validation} \in \{0,1\}^{n_2}\}$ , where  $n_1$  and  $n_2$  are the number of samples in the training set and validation set, respectively,  $m$  is the total number of features, and  $Y_{train}$ ,  $Y_{validation}$  are the true labels in the training set and the validation set, respectively. A list of features' name  $F_{all}$  is sorted by the importance in the Multi-tree XGBoost model.

**Output:** The selected features name list  $F_{selected-last}$ .

**Step 1 Initialization**

- 1)  $i = 0, AUC_{score-last} = 0, F_{selected} = [], F_{selected-last} = [], \text{threshold} = 0.5\%$

**Step 2 Feature selection**

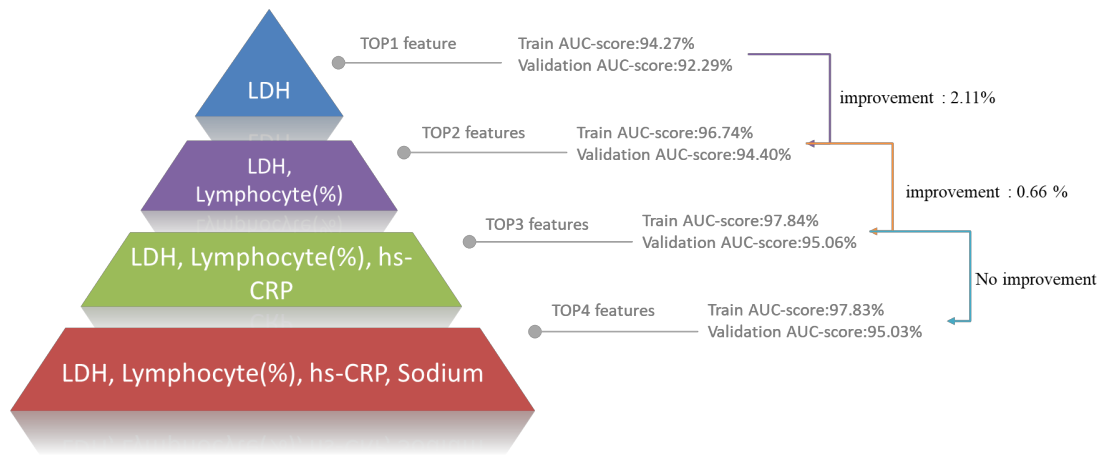
- 2)  $F_{selected-last} \leftarrow F_{selected}$
- 3) Add element  $F_{all}[i]$  to  $F_{selected}$
- 4)  $X_{train-selected}$  is the matrix formed by the corresponding columns of  $F_{selected}$  in  $X_{train}$ .
- 5)  $X_{validation-selected}$  is the matrix formed by the corresponding columns of  $F_{selected}$  in  $X_{validation}$ .

**Step 3 Training and Prediction**

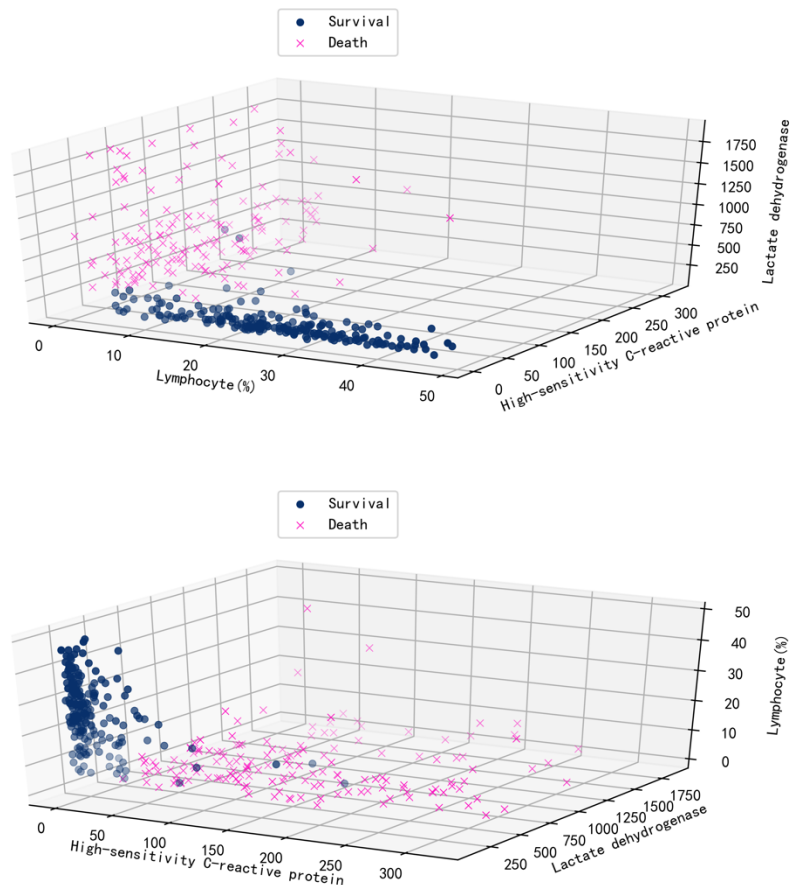
- 6) Fit the  $\{X_{train-selected}, Y_{train}\}$  with XGBoost and get the  $f^*(x)$
- 7) Predict the  $\tilde{Y}_{validation} \leftarrow f^*(X_{validation-selected})$ , where  $\tilde{Y}_{validation}$  is the predicted labels of validation set samples.

**Step 4 Calculating the f1-scores**

- 8)  $AUC_{score} \leftarrow AUC_{score}(Y_{validation}, \tilde{Y}_{validation})$
- 9) If  $(AUC_{score} - AUC_{score-last}) < \text{threshold}$
- 10) Then  $AUC_{score-best} = AUC_{score-last}$  and return  $F_{selected-last}$
- 11) Else  $i \leftarrow i + 1$  and return to Step 2
-



Supplementary Figure 2: Illustration of AUC scores using Supplementary Algorithm 1. Multi-tree XGBoost with 375 samples (Top-X features): It is trained with the parameters setting as the max depth with 4, the learning rate is equal 0.2, the value of the regularization parameter  $\alpha$  is set to 1.



Supplementary Figure 3: Visualization of 351 patients' outcome over three selected features.

**Step 3.** Reduce number of tree to 1, which leads to the Single-tree XGBoost algorithm. Because there are 24 samples with at least one of top-3 features missing. In order to

obtain a better decision rule, we have deleted these samples and obtained a new dataset with the 351 samples and 3 features.

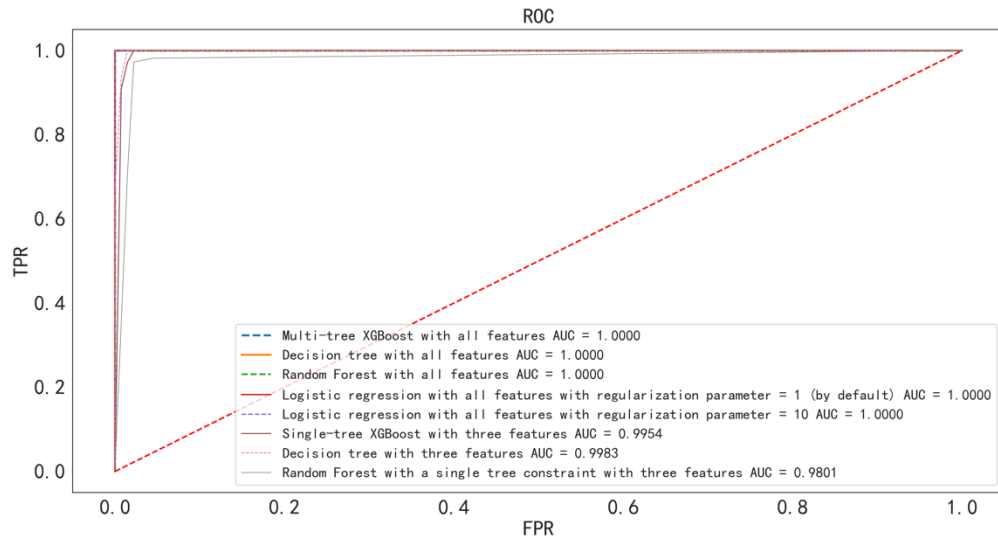
Single-tree XGBoost with 351 samples (Top3 features): It is trained with the parameter setting as the tress number of estimators is set to 1, the values of the two regularization parameters  $\alpha$  and  $\beta$  are both set to 0, the subsample and max features both are set to 1.

|                     | <b>Precision</b> | <b>Recall</b> | <b>F1-score</b> | <b>Support</b> |
|---------------------|------------------|---------------|-----------------|----------------|
| <b>Survival</b>     | 1.00             | 0.98          | 0.99            | 135            |
| <b>Death</b>        | 0.97             | 1.00          | 0.99            | 110            |
| <b>Accuracy</b>     |                  |               | 0.99            | 245            |
| <b>Macro avg</b>    | 0.99             | 0.99          | 0.99            | 245            |
| <b>Weighted avg</b> | 0.99             | 0.99          | 0.99            | 245            |

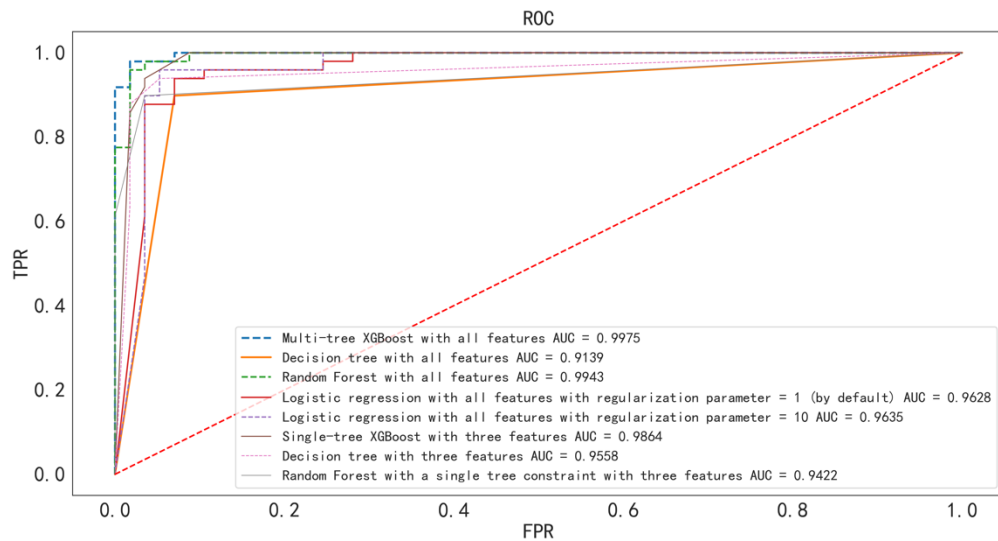
Supplementary Table 1: Performance of the proposed algorithm on training dataset for the Single-tree XGBoost algorithm.

|                     | <b>Precision</b> | <b>Recall</b> | <b>F1-score</b> | <b>Support</b> |
|---------------------|------------------|---------------|-----------------|----------------|
| <b>Survival</b>     | 0.95             | 0.96          | 0.96            | 57             |
| <b>Death</b>        | 0.96             | 0.94          | 0.95            | 49             |
| <b>Accuracy</b>     |                  |               | 0.95            | 106            |
| <b>Macro avg</b>    | 0.95             | 0.95          | 0.95            | 106            |
| <b>Weighted avg</b> | 0.95             | 0.95          | 0.95            | 106            |

Supplementary Table 2 : Performance of the proposed algorithm on validation dataset for the Single-tree XGBoost algorithm.



(A)



(B)

Supplementary Figure 4: The ROC curves (receiver operating characteristic curve) for training (A) and validation (B) datasets split from 351 patients is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings from 0 to 1. Where True Positive Rate:  $TPR = TP / (TP + FN)$  and False Positive Rate:  $FPR = FP / (FP + TN)$ . AUC is defined as the Area Under ROC curve, AUC belong to  $[0,1]$ . The value closer to 1, the better the model's classification ability.

| Model class   | Cross-validation AUC for training set (mean +- std) | Cross-validation AUC for validation set (mean +- std) |
|---|---|---|
| Multi-tree XGBoost with all features  | 100.00% +- 0.00%                                    | 99.74% +- 0.39%                                       |
| Single-tree XGBoost with three features   | 99.24%+- 0.28%                                      | 97.77% +- 1.82%                                       |
| Decision tree with all features   | 99.86% +- 0.18%                                     | 94.07%+- 2.95%  |
| Decision tree with three features   | 99.36%+- 0.36%                                      | 96.10% +- 2.59%                                       |
| Random Forest with all features   | 100.00% +- 0.00%                                    | 99.69% +- 0.37%                                       |
| Random Forest with a single tree constraint with three features                     | 98.21%+- 0.77%                                      | 96.63% +- 2.32%                                       |
| Logistic regression with all features with regularization parameter =1 (by default) | 100.00% +- 0.01%                                    | 96.95% +- 2.27%                                       |

Supplementary Table 3 : Comparison of the proposed algorithm on both training and validation datasets split from last measurements of 351 patients using three features with other machine learning method using 100-round 5-fold cross validation.

|            |          |                 |       |
|------------|----------|-----------------|-------|
| True Label | Survival | 94              | 3     |
|            | Death    | 0               | 13    |
|            |          | Survival        | Death |
|            |          | Predicted Label |       |

Supplementary Figure 5: Confusion matrix for the additional test dataset (110 patients' last measurements) using the Single-tree XGBoost algorithm.

| True Label      | Survival | Death |
|-----------------|----------|-------|
|                 | 459      | 70    |
| Death           | 15       | 365   |
| Predicted Label |          |       |

Supplementary Figure 6: Confusion matrices for all available blood samples for all 485 patients.

| True Level      | Survival | Death |
|-----------------|----------|-------|
|                 | 165      | 17    |
| Death           | 2        | 67    |
| Predicted Level |          |       |

Supplementary Figure 7: Confusion matrices for all available blood samples for all 110 patients in the external test dataset.