

Time Windows Voting Classifier for COVID-19 Mortality Prediction

Background.

The ability to predict COVID-19 patients' level of severity (death or survival) enables clinicians to prioritise treatment. Recently, using three blood biomarkers, an interpretable machine learning model was developed to predict the mortality of COVID-19 patients. The method was reported to be suffering from performance stability because the identified biomarkers are not consistent predictors over an extended duration.

Methods.

To sustain performance, the proposed method partitioned data into three different time windows. For each window, an end-classifier, a mid-classifier and a front-classifier were designed respectively using the XGboost single tree approach. These time window classifiers were integrated into a majority vote classifier and tested with an isolated test data set.

Results.

The voting classifier strengthens the overall performance of 90% cumulative accuracy from a 14 days window to a 21 days prediction window.

Conclusions.

An additional 7 days of prediction window can have a considerable impact on a patient's chance of survival. This study validated the feasibility of the time window voting classifier and further support the selection of biomarkers features set for the early prognosis of patients with a higher risk of mortality.

Keywords.

Decision tree classifier, XGboost, Biomarkers, voting classifier, COVID-19

Time Windows Voting Classifier for COVID-19 Mortality Prediction

Tiong-Thye Goh¹, MengJun Liu²

²Corresponding author

¹Wellington School of Business and Government, Victoria University of Wellington, New Zealand.

²School of Education, Hubei University

No.368# You Yi Road, Wuchang District, Wuhan, Hubei, PRC China.

Email: lmj_wlu@163.com

There is an urgent need to assess the likelihood of a COVID-19 patient dying from the disease and to support the decisions for prioritising treatment and resource planning in the hospital. Recently Yan et al. (2020) and his team used machine learning tools to identify three biomarkers – lactic dehydrogenase, lymphocyte and high-sensitivity C-reactive protein – reported to predict the severity of the disease in individual patients more than 10 days in advance with more than 90% accuracy. Their training set consists of 375 cases of which 201 (53.6%) survived and 174 (46.4%) died. Their algorithm was developed using the last complete blood test data. However, a recent study has suggested that the performance degraded over a longer predictive window and suffered from a high level of fluctuation (Huang et al., 2020). The assumption that the same biomarkers are consistent predictors over an extended duration is unrealistic. Biomarkers in a different time frame reveal different stages of illness as each stage is characterised by specific biochemical alterations (Marcello & Luisa, 2020). Using only the last complete blood test data suffers from information loss and lower predictive performance over an extended time window. To sustain the performance, this study proposes to use all the available data to minimise information loss by designing three classifiers based on samples over three different time windows. Subsequently, the classifiers are integrated into a majority vote classifier to optimise performance (Lam & Suen, 1997).

MATERIALS AND METHODS

Data Collection

For this study, we used the data in Yan et al. (2020). The medical data of all patients were collected from hospitals in Wuhan, China between 10 January and 18 February 2020 using standard case report forms that included epidemiological, demographic, clinical, laboratory and mortality outcome information (Yan et al., 2020). Only the laboratory, mortality outcome and days to outcome fields were used for modelling, validating and testing. The data consists of 375 patients with 201 (53.6%) discharged and 174 (46.4%) deceased. In total 75 blood sample features were sparingly recorded across the data set. The minimal and maximal follow-up times (from admission to hospital to death or discharge) for the patients are 0 to 35 days.

Laboratory Methods

Statistical Analysis

A parametric test (t test) and nonparametric test (Mann-Whitney U test) were used for continuous variables with or without normal distribution, respectively. Missing data was handled by XGBoost during model development. A two-sided α of less than 0.05 was considered statistically significant. Statistical analyses were computed using SPSS (version 24).

Classifier Development

The original blood sample data set was merged into daily data for each patient. The data were partitioned into 3 sub-sample sets of 7 day windows. Data set 0 consists of the final 7 days of data with 237 samples for training and validation and 101 samples for testing. Data set 1 consists of samples from 8 to 14 days with 146 samples for training and validation and 62 samples for testing, and data set 2 consists of samples beyond 14 days with 72 samples for training and validation and 31 samples for testing. For each data set, we used the XGBoost method to select the important features for the respective window and used the important features to generate a single decision tree. This process created three classifiers each representing a specific time window. Finally, we combined the three classifiers into a majority vote classifier to establish the final prediction outcome. For features extraction, XGBoost parameters setting were maximum depth equal to 4, learning rate equal to 0.2, number of tree estimators set to 150, regularisation parameter α set to 1 and 'subsample' and 'colsample_bytree' both set to 0.9. To generate the single tree, we set the maximum depth to 2 and tree estimator to 1. Figure 1 depicts the classifier development process.

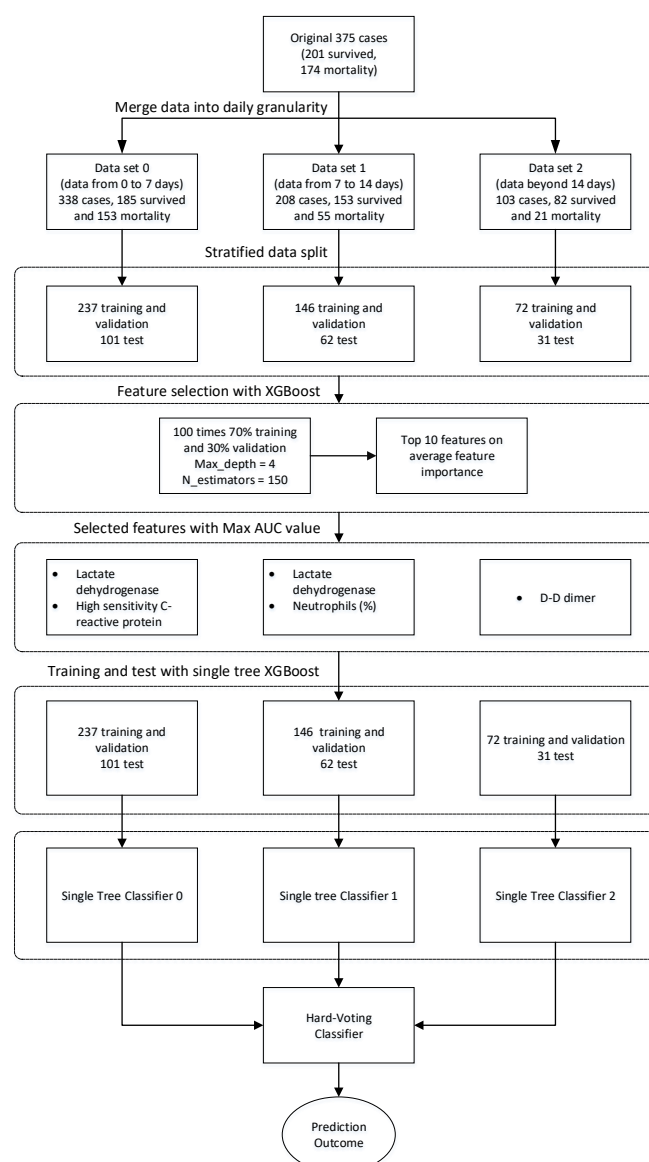


Figure 1. Data processing and classifier development procedures

RESULTS

Data Characteristics of COVID-19 Patients

Table 1 depicts the differences between features in the 3 data sets and patients' outcomes. For the outcome=1 group, 13 features are significantly different ($p < 0.05$) between the 3 data sets, whereas the outcome=0 group 33 features are significantly different ($p < 0.05$) between the 3 data sets. This suggests that for the outcome=0 group, biomarkers are changing over the 3 time windows. When comparing patient outcomes between corresponding data sets, a significant majority of the features were different between the survival and the mortality group ($p_{dso} < 0.05$, $p_{ds1} < 0.05$, $p_{ds2} < 0.05$). These differences provide the basis for building machine learning classifiers. However, the difference statistics are not sufficient to provide information on which features are most important to provide diagnostics and predictive analysis.

Prognostic Factors of Severe COVID-19

The XGBoost algorithm is used to select the important features for each of the data sets. For each data set we built a single tree decision classifier for each time window. Table 2 depicts the selected prognostic features and the rules for each decision tree classifier. The biomarkers selected by the algorithm are linked to serious outcomes in patients (Ponti et al., 2020). Lactate dehydrogenase measures tissue damage associated with a wide range of disorders, including liver disease and interstitial lung disease, and C-reactive protein is a marker of inflammation or infection-reduced lung function. Lymphopenia is a common feature in patients with COVID-19 and might be a critical factor associated with disease severity and mortality (Zhao et al., 2020). Neutrophils are associated with infection, increasing the inflammation and haemorrhagic lesions in the lungs of infected patients (Hemmat et al., 2020). Neutrophils may be responsible for mortality in these severe coronavirus cases (Tomar et al., 2020). Biomolecules such as D-dimer indicate a breakdown product of blood clots. These features were identified as relevant to complications associated with COVID-19 and were dramatically increased in patients that died versus those that recovered. A study by Zhou et al. (2020) showed that D-dimer greater than 1 $\mu\text{g/mL}$ could help clinicians to identify patients with a poor prognosis about 20 days in advance. D-dimer values are even higher in patients with severe COVID-19 than in those with milder forms and therefore, D-dimer measurement may be associated with evolution toward worse clinical outcomes (Lippi & Favaloro, 2020).

Table 1. Data Sets Characteristics

Features	OUTCOME= 1				OUTCOME=0				P0	P_DS0	P_DS1	P_DS2
	DS0	DS1	DS2	P1	DS0	DS1	DS2					
Gender	1.29(0.45)	1.24(0.43)	1.29(0.46)	0.803	1.52(0.5)	1.52(0.5)	1.51(0.5)	0.978	0.000	0.000	0.000	0.073
Age	69.19(12.24)	69.57(9.33)	68.1(9.08)	0.879	50.6(15.21)	49.32(14.78)	51.09(13.95)	0.621	0.000	0.000	0.000	0.000
Days	2.62(1.43)	9.67(1.58)	22.19(4.56)	0.000	3.4(1.39)	10.91(1.65)	18.34(3.47)	0.000	0.000	0.000	0.000	0.000
(%)lymphocyte	6.34(5.87)	6.6(5.03)	6.86(5.18)	0.904	25.75(10.4)	24.32(10.83)	23.09(10.29)	0.163	0.000	0.000	0.000	0.000
Activation of partial thromboplastin time	43.1(10.33)	39.69(8.08)	41.67(5.6)	0.169	38.29(6.4)	39.62(6.02)	41.16(4.71)	0.029	0.000	0.961	0.707	0.000
Alkaline phosphatase	101.97(48.9)	83.95(40.89)	82.34(37.43)	0.023	66.36(49.55)	67.9(27.34)	62.65(17.56)	0.607	0.000	0.002	0.001	0.001
Amino-terminal brain natriuretic peptide precursor(NT-proBNP)	5221.81(9469.79)	1978.36(2513.65)	1096.04(1076.89)	0.020	2634.95(10934.24)	118.76(142.93)	140.31(200.56)	0.068	0.138	0.000	0.000	0.000
Corrected calcium	2.34(0.13)	2.37(0.11)	2.35(0.12)	0.224	2.36(0.12)	2.37(0.11)	2.32(0.1)	0.021	0.140	0.674	0.374	0.374
D-D dimer	12.62(8.72)	9.66(9.12)	12.82(8.51)	0.117	0.84(0.77)	1.52(3.28)	1.23(2.77)	0.217	0.000	0.000	0.000	0.000
Direct bilirubin	15.55(33.14)	7.78(5.03)	10.4(15.96)	0.204	3.87(1.84)	4.2(2.28)	3.83(1.53)	0.268	0.000	0.000	0.000	0.000
ESR	39.75(28.91)	40.31(23.96)	34.06(23.46)	0.704	30.5(26.85)	26.61(17.67)	32.98(22.23)	0.162	0.087	0.001	0.859	0.859
Eosinophil count	0.02(0.04)	0.01(0.03)	0.01(0.03)	0.648	0.08(0.06)	0.05(0.07)	0.02(0.04)	0.000	0.000	0.000	0.000	0.181
Fibrin degradation products	90.94(60.25)	75.9(68.98)	75.63(51.04)	0.426	4.42(1.2)	9.87(21.61)	5.16(5.5)	0.157	0.000	0.000	0.000	0.000
HBsAg	19.82(65.73)	13.52(57.31)	1.78(7.07)	0.526	0.01(0.01)	5.01(32.42)	0.49(2.37)	0.408	0.136	0.347	0.230	0.230
HCO3-	21(4.14)	21.41(4.21)	21.72(4.15)	0.684	25.62(2.93)	24.45(2.67)	23.3(2.11)	0.000	0.000	0.000	0.000	0.017
HCV antibody quantification	0.16(0.33)	0.13(0.27)	0.11(0.19)	0.787	0.1(0.09)	0.07(0.08)	0.09(0.12)	0.491	0.336	0.118	0.551	0.551
HIV antibody quantification	0.1(0.05)	0.1(0.05)	0.08(0.01)	0.079	0.13(0.05)	0.1(0.03)	0.1(0.02)	0.000	0.056	0.785	0.005	0.005
High sensitivity C-reactive protein	129.72(70.61)	129.06(73.16)	118.11(71.81)	0.784	10.3(27.88)	24.77(32.92)	43.98(47.13)	0.000	0.000	0.000	0.000	0.000
Hypersensitive cardiac troponinI	1652.89(6284.83)	2031.74(7360.49)	920.1(2709.46)	0.810	20.79(87.94)	6.32(7.11)	7.44(10.42)	0.242	0.066	0.023	0.028	0.028
Interleukin 10	27.28(94)	13.4(8.53)	18.11(14.37)	0.787	5.69(3.64)	6.64(3.73)	9.02(5.78)	0.002	0.167	0.000	0.000	0.001
Interleukin 1β	7.8(11.31)	6.18(4.87)	5.05(0.17)	0.600	6.29(7.41)	6.66(5.76)	5.76(2.66)	0.711	0.467	0.754	0.360	0.360
Interleukin 2 receptor	1332.8(813.75)	1699.82(1540.21)	1193.88(839.25)	0.311	592(617.17)	625.19(302.53)	599.42(255.41)	0.905	0.000	0.000	0.000	0.000
Interleukin 6	255.85(726.72)	108.06(142.72)	294.61(726.4)	0.676	8.64(21.48)	18.29(27.49)	25.88(26.94)	0.011	0.034	0.000	0.000	0.015
Interleukin 8	107.38(381.44)	62(65.65)	345.42(970.5)	0.221	9.88(8.72)	18.77(28.69)	28.73(49.01)	0.042	0.124	0.000	0.000	0.031
International standard ratio	1.54(0.8)	1.24(0.3)	1.34(0.29)	0.022	1.03(0.07)	1.05(0.08)	1.06(0.1)	0.020	0.000	0.000	0.000	0.000
Lactate dehydrogenase	732.57(367.83)	620.98(346.66)	635.91(304.89)	0.116	210.93(68.48)	259.07(88.63)	281.79(99.56)	0.000	0.000	0.000	0.000	0.000
Mean platelet volume	11.47(1.16)	11.23(0.88)	11.41(0.86)	0.409	10.46(0.82)	10.51(0.83)	10.6(0.87)	0.465	0.000	0.000	0.000	0.000
PH value	6.57(0.73)	6.3(0.74)	6.37(0.71)	0.215	6.24(0.6)	6.53(0.49)	6.36(0.54)	0.028	0.003	0.132	0.988	0.988
PLT distribution width	14.44(3.32)	13.6(2.27)	14.16(2.5)	0.247	11.97(1.95)	12.07(1.86)	12.19(1.97)	0.685	0.000	0.000	0.000	0.000
Platelet count	137.82(89.78)	149.86(68.49)	131.44(63.9)	0.590	251.1(78.84)	245.89(97.34)	201.44(68.29)	0.000	0.000	0.000	0.000	0.000
Prothrombin activity	65.37(19.38)	77.67(18.93)	70.25(14.03)	0.001	96.9(11.16)	94.15(11.61)	92.73(13.64)	0.083	0.000	0.000	0.000	0.000
Prothrombin time	18.6(8.28)	15.71(2.87)	16.62(2.75)	0.034	13.48(0.76)	13.78(0.82)	13.94(1)	0.003	0.000	0.000	0.000	0.000
Quantification of Treponema pallidum antibodies	0.27(1.43)	0.1(0.21)	0.05(0.04)	0.638	0.08(0.04)	0.05(0.03)	0.12(0.51)	0.470	0.500	0.105	0.588	0.588
RBC distribution width SD	44.47(7.91)	42.48(6.1)	45.24(4.72)	0.181	40.25(3.53)	39.68(3.37)	39.51(2.95)	0.202	0.000	0.000	0.000	0.000
Red blood cell count	9.11(23.24)	12.38(18.68)	7.78(7.52)	0.569	6.27(17.96)	4.86(2.65)	9.52(30.5)	0.198	0.225	0.000	0.797	0.797
Red blood cell distribution width	13.74(2.09)	13.04(2.06)	13.64(1.34)	0.103	12.48(0.97)	12.3(0.92)	12.33(1.05)	0.245	0.000	0.001	0.000	0.001
Serum chloride	104.2(8.87)	100.38(6.96)	108.39(9.52)	0.001	100.99(3.86)	100.7(3.48)	100.62(3.28)	0.689	0.000	0.688	0.000	0.000
Serum potassium	4.52(0.79)	4.63(0.71)	4.6(0.65)	0.667	4.43(0.49)	4.22(0.49)	4.06(0.45)	0.000	0.234	0.000	0.000	0.000
Thrombin time	18.33(3.94)	22(24.11)	19.58(5.67)	0.276	16.55(1.25)	16.33(1.31)	16.55(1.22)	0.469	0.000	0.034	0.001	0.001
Total bilirubin	23.95(41.77)	14.65(6.69)	16.99(22.25)	0.229	9.85(5.17)	9.98(5.03)	8.7(3.57)	0.144	0.000	0.000	0.000	0.002
Total cholesterol	3.3(0.89)	3.56(0.91)	3.42(0.69)	0.205	4.31(0.96)	3.85(0.81)	3.63(0.71)	0.000	0.000	0.038	0.232	0.232
Tumor necrosis factorα	15.71(14.46)	15.76(15.09)	18.69(19.58)	0.822	7.17(4.91)	8.08(2.84)	9.37(3.12)	0.021	0.001	0.000	0.003	0.003
Urea	12.58(8.89)	11.9(8.02)	15.42(12.87)	0.323	5.37(4.97)	4.35(1.63)	4.19(1.84)	0.015	0.000	0.000	0.000	0.000
Uric acid	296.63(170.75)	322.9(168.09)	304.29(174.57)	0.640	278.34(103.11)	248.02(92.55)	244.2(68.93)	0.006	0.267	0.000	0.000	0.017
White blood cell count	14.04(9.84)	15.11(28.95)	13.55(10.28)	0.905	14.94(56.81)	8.83(28.05)	14.34(45.58)	0.484	0.853	0.172	0.937	0.937
albumin	28.58(4.02)	29.68(3.89)	28.91(5.69)	0.275	36.93(3.91)	35.83(4.31)	36.35(4.4)	0.083	0.000	0.000	0.000	0.000
antithrombin	78.2(17.49)	86.98(17.01)	84.96(17.93)	0.043	98.34(13.21)	95.46(14.02)	89.85(10.12)	0.016	0.000	0.026	0.225	0.225
aspartate aminotransferase	74.56(129.3)	56.02(105.55)	47.54(28.82)	0.451	23.43(17.52)	25.7(14.11)	28.06(13.53)	0.093	0.000	0.001	0.000	0.001
basophil count(%)	0.02(0.02)	0.02(0.02)	0.01(0.02)	0.087	0.02(0.02)	0.01(0.01)	0.01(0.01)	0.000	0.512	0.219	0.050	0.050
basophil(%)	0.16(0.17)	0.13(0.11)	0.1(0.09)	0.134	0.36(0.28)	0.23(0.2)	0.18(0.23)	0.000	0.000	0.000	0.120	0.120
calcium	2(0.12)	2.04(0.11)	2.01(0.13)	0.143	2.2(0.11)	2.17(0.1)	2.14(0.1)	0.000	0.000	0.000	0.000	0.000
creatinine	116.67(78.62)	128.75(111.55)	156.8(180.72)	0.215	99.09(183.06)	69(22.88)	69.93(20.01)	0.069	0.289	0.000	0.000	0.000
eGFR	69.62(28.35)	64.47(27.92)	66.49(28.55)	0.517	95.86(26.07)	99.13(23.16)	97.8(18.75)	0.507	0.000	0.000	0.000	0.000
eosinophils(%)	0.16(0.4)	0.1(0.22)	0.08(0.18)	0.474	1.47(1.28)	1.05(1.44)	0.46(0.7)	0.000	0.000	0.000	0.000	0.016
ferritin	2602.4(3470.67)	2221.81(1642.86)	5745.61(14095.98)	0.211	631.14(426.3)	617.84(567.11)	666.1(634.7)	0.898	0.000	0.000	0.000	0.017
fibrinogen	4.31(2)	4.74(2.36)	3.8(1.57)	0.254	4.23(1.33)	4.77(1.26)	4.53(1.08)	0.023	0.757	0.936	0.034	0.034
globulin	34.68(5.56)	36.14(4.87)	32.8(4.41)	0.045	31.32(5.27)	32.52(4.19)	32.99(3.3)	0.013	0.000	0.000	0.823	0.823
glucose	10.32(4.59)	10.71(5.62)	11.26(4.58)	0.677	6.36(3.12)	6.9(3.13)	7.51(2.73)	0.032	0.000	0.000	0.000	0.000
glutamic-pyruvic transaminase	50.64(110.29)	36.04(57.22)	31.81(17.62)	0.504	32.71(26.74)	30.54(26.84)	24.27(17.29)	0.053	0.051	0.377	0.081	0.081
hematocrit	36.42(6.12)	37.71(5.44)	35.52(6.15)	0.273	37.08(4.37)	37.38(4.21)	36.89(4.05)	0.690	0.277	0.662	0.223	0.223
hemoglobin	124.77(20.93)	130.68(17.69)	114.93(22.83)	0.011	127.23(15.48)	125.71(19.81)	124.27(17.59)	0.453	0.241	0.116	0.046	0.046
indirect bilirubin	8.39(9.8)	6.95(3.11)	6.56(6.97)	0.440	6.09(3.5)	5.87(3.29)	4.87(2.42)	0.023	0.008	0.047	0.076	0.076
lymphocyte count	0.6(0.34)	0.53(0.28)	0.5(0.19)	0.208	1.73(3.3)	1.25(0.53)	1.09(0.44)	0.054	0.000	0.000	0.000	0.000
mean corpuscular hemoglobin	31.29(3.18)	31.37(2.24)	30.32(2.79)	0.342	30.9(2.35)	30.82(2.75)	30.29(2.77)	0.205	0.228	0.204	0.961	0.961
mean corpuscular hemoglobin concentration	345.01(18.47)	347.3(15.63)	331.43(15.09)	0.002	343.21(13.86)	344.82(16.63)	342.01(12.8)	0.378	0.334	0.355	0.002	0.002
mean corpuscular volume	90.65(7.1)	90.33(5.04)	91.53(7.96)	0.788	90(5.1)	89.31(5.36)	88.47(6.12)	0.115	0.355	0.241	0.059	0.059
monocytes count	0.46(0.27)	0.41(0.26)	0.37(0.28)	0.193	0.75(2.87)	0.45(0.18)	0.39(0.17)	0.265	0.243	0.247	0.600	0.600
monocytes(%)	4.4(4.13)	4.45(4.3)	3.91(2.75)	0.847	8.59(2.99)	8.3(2.96)	8.07(3.08)	0.413	0.000	0.000	0.000	0.000
neutrophils count	11.83(5.53)	9.88(6.05)	9.77(5.08)	0.053	4.41(2.69)	3.95(2.15)	3.67(1.82)	0.051	0.000	0.000	0.000	0.000
neutrophils(%)	88.94(9.54)	88.72(7.52)	89.04(7.28)	0.985	63.73(12.92)	66.1(12.93)	68.21(12.12)	0.031	0.000	0.000	0.000	0.000
platelet large cell ratio	36.22(9.03)	34.34(6.91)	35.67(6.72)	0.402	28.22(6.64)	28.63(6.74)	29.36(6.95)	0.471	0.000	0.000	0.000	0.000
procalcitonin	1.48(3.09)	1.51(6.64)	4.27(8.96)	0.068	0.13(0.49)	0.06(0.13)	0.07(0.09)	0.369	0.002	0.057	0.000	0.000
serum sodium	141.91(8.24)	139.25(7.6)	146.93(10.14)	0.002	139.92(2.89)	139.67(2.93)	138.45(					

Table 2 The selected prognostic features and the rules for each decision tree classifier.

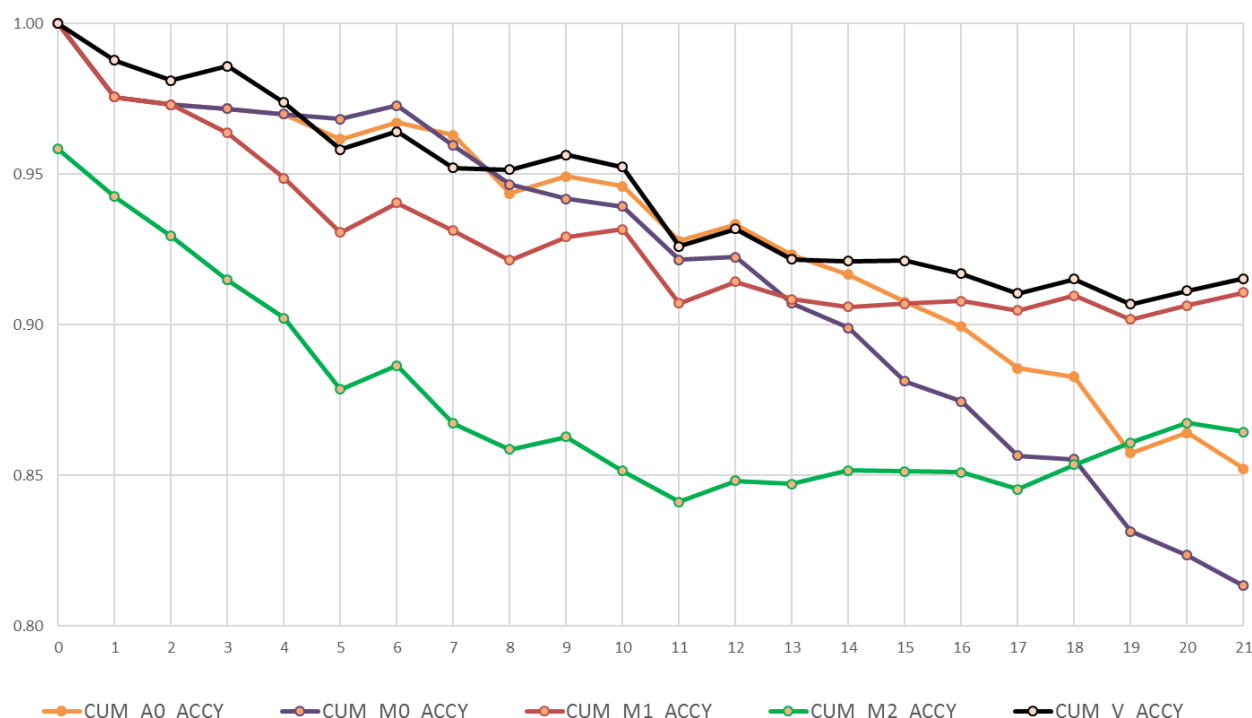
Classifier	Features Selected	Decision Rules
0	LDH: Lactate dehydrogenase HCP: High sensitivity C-reactive protein	=IF(AND([@[LDH]] <340, [@[HCP]] <55.233),0,1)
1	LDH: Lactate dehydrogenase NEU: Neutrophils (%)	=IF(AND([@[NEU]] <88.25), ([@[LDH]] <393.5),0,1)
2	DMR: D-D dimer	=IF([@[DMR]] <2.1),0,1)

Validation of the Predictive Accuracy of the Voting Classifier

Table 3 depicts the performance of the 3 classifiers for the training, validation and the testing phases. Overall classifier 0's average performance is better than those of classifiers 1 and 2. However when we break it down into different time windows, classifier 0's performance degraded at time >14 days. This is the typical when using a single classifier to estimate an extended period. At time>14 days, classifier 2 performed much better than classifier 0. For all the time windows, the voting classifier was either equal or better than the best classifiers. Figure 2 shows the performance over 21 days. The voting classifier performed better than the original classifier by 5% and at above 90% accuracy.

Table 3 The performance of the 3 classifiers for the training, validation and the testing phases.

		Survival Vs Mortality		
Model	N	AUC (95% CI)	Sensitivity	Specificity (%)
Model 0				
Training	166	0.990(.980-1.0)	100.0%	97.2%
Validation	71	0.965(.917-1.0)	94.7%	100%
Testing	101	0.97	90.9%	93.2%
Model 1				
Training	102	0.977(.970-.993)	98.6%	96.2%
Validation	44	0.812(.687-.917)	90.6%	58.3%
Testing	62	0.863	89.1%	68.8%
Model 2				
Training	50	0.945(.937-.955)	100.0%	90.9%
Validation	22	0.814(.612-1.00)	94.1%	75.0%
Testing	31	0.826	95.8%	66.7%



CUM_A0_ACCY: Cumulative accuracy for A0 (Yan); CUM_M0_ACCY: Cumulative accuracy for classifier 0; CUM_M1_ACCY: Cumulative accuracy for classifier 1; CUM_M2_ACCY: Cumulative accuracy for classifier 2

Figure 2. Comparing cumulative accuracy performance of classifiers over days (all available data)

Table 4. Comparison of classifiers performance based on different time window (all available data)

	Average Classifiers Accuracy (Mean/ SD)				
Time Window	A0	M0	M1	M2	V
0-7 days	0.963	0.960	0.931	0.867	0.952
	0.028	0.042	0.060	0.082	0.055
8-14 days	0.864	0.830	0.877	0.834	0.886
	0.108	0.091	0.128	0.085	0.131
15-21 days	0.714	0.630	0.921	0.892	0.903
	0.198	0.150	0.094	0.106	0.105

A0: Classifier based on Yan et al.; M0: Classifier 0; M1 Classifier 1; M2: Classifier 2 V: Voting Classifier

DISCUSSION

The purpose of the study is to enhance the classifier performance of Yan et al. (2020) by using all the available data instead of just the last set of complete blood data. Overall, the proposed voting classifier performs better than the original classifier from the 15 to 21 days window with 5% higher cumulative accuracy (Figure 2), 20% average accuracy (Table 4) or overall, 3.5% higher AUC (Table S).

Classifier 0's performance which is based on the ending segment of the data is compatible with the Yan et al. results as shown in Table 4 and Figure 3, especially for the 0 to 7 days window. When evaluated with a separate test data the AUCs are .888, .905, .873, and .937 for classifier 0, classifier 1, classifier 2 and the voting classifier respectively as shown in Table S. Overall the voting classifier performed as designed, especially for the 15 to 21 days window, to obtain a longer time window for the severe patients to be identified.

Figure 2 demonstrates a similar decrease in performance with distant data which is also depicted by Yan et al. (2020) in their daily performance plots. Model 1, which uses the middle segment of data, appears to be more stable than models 0 and 2. The difference in performance in each different time segment suggests that using only ending data for training may not be the best option. As can be seen from the analysis, different time segments can decrease performance due to different relevant feature indicators. When patients are admitted, the ability to establish their illness timeframe and severity improves their chance of survival.

The hard-voting classifier leverages the strength of each classifier (Lam & Suen, 1997). Figure 2 demonstrates that the overall cumulative accuracy improved from 85% (original classifier from Yan et al.) to 92% at day 21 which suggests that the simple voting classifier is a suitable approach to identify high risk patients.

Our study has several strengths. First, the decision tool is based on only 4 features, which are relatively inexpensive and easily obtained from routine blood tests. The biomarkers are linked to different stages of illness progression for poor prognosis patients. Compared to Yan et al. (2020), the proposed time window voting classifier is easy to implement and the performance is more consistent over a 21 day period. The additional test data evaluation ensures the study is rigorous.

Our study comes with certain limitations. First, the samples are retrospective and consist of 375 cases with 201 (53.6%) who survived and 174 (46.4%) who died. While the small sample size limits generalisation, it does provide a direction for further analysis when larger data sets become available. Second, a more comprehensive analysis is needed to establish the stability of the algorithm. Our method may have reduced the instability, but we do need further evidence in support. Third, the data distribution with respect to mortality would affect the performance of the algorithm which requires further validation that takes hospital context into consideration (Wang et al., 2020).

Supplementary Data

Notes

Author contributions.

Financial support.

Potential conflicts of interest. The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

Supplement materials

Table S. 100 runs of single tree XGboost classifiers accuracy on test data (Mean /SD)

	Classifier 0	Classifier 1	Classifier 2	Voting
All Test Data	0.849 0.115	0.895 0.050	0.858 0.061	0.910 0.050
Test--DS0	0.969 0.016	0.913 0.035	0.852 0.040	0.948 0.025
Test-DS1	0.861 0.038	0.865 0.040	0.810 0.047	0.882 0.038
Test-DS2	0.719 0.080	0.907 0.058	0.914 0.043	0.900 0.056

References

- Hemmat, N., Derakhshani, A., Bannazadeh Baghi, H., Silvestris, N., Baradaran, B., & De Summa, S. (2020, 2020-June-09). Neutrophils, Crucial, or Harmful Immune Cells Involved in Coronavirus Infection: A Bioinformatics Study [Original Research]. *Frontiers in Genetics*, 11(641). <https://doi.org/10.3389/fgene.2020.00641>
- Huang, C., Long, X., Zhan, Z., & van den Heuvel, E. (2020). Model stability of COVID-19 mortality prediction with biomarkers. *medRxiv*, 2020.2007.2029.20161323. <https://doi.org/10.1101/2020.07.29.20161323>
- Lam, L., & Suen, S. (1997). Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(5), 553-568.
- Lippi, G., & Favaloro, E. J. (2020, May). D-dimer is Associated with Severity of Coronavirus Disease 2019: A Pooled Analysis. *Thromb Haemost*, 120(5), 876-878. <https://doi.org/10.1055/s-0040-1709650>
- Marcello, C., & Luisa, A. (2020, 24 Jun. 2020). Biochemical biomarkers alterations in Coronavirus Disease 2019 (COVID-19). *Diagnosis*(0), 000010151520200057. <https://doi.org/https://doi.org/10.1515/dx-2020-0057>
- Ponti, G., Maccaferri, M., Ruini, C., Tomasi, A., & Ozben, T. (2020). Biomarkers associated with COVID-19 disease progression. *Critical reviews in clinical laboratory sciences*, 1-11. <https://doi.org/10.1080/10408363.2020.1770685>
- Tomar, B., Anders, H.-J., Desai, J., & Mulay, S. R. (2020). Neutrophils and Neutrophil Extracellular Traps Drive Necroinflammation in COVID-19. *Cells*, 9(6), 1383. <https://doi.org/10.3390/cells9061383>
- Wang, C., Deng, C., & Wang, S. (2020, 2020/08/01/). Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognition Letters*, 136, 190-197. <https://doi.org/https://doi.org/10.1016/j.patrec.2020.05.035>
- Yan, L., Zhang, H.-T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., & Zhang, M. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence*, 1-6.
- Zhao, Q., Meng, M., Kumar, R., Wu, Y., Huang, J., Deng, Y., Weng, Z., & Yang, L. (2020, 2020/07/01/). Lymphopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A systemic review and meta-analysis. *International Journal of Infectious Diseases*, 96, 131-135. <https://doi.org/https://doi.org/10.1016/j.ijid.2020.04.086>
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., & Cao, B. (2020, 2020/03/28/). Clinical course

and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229), 1054-1062.
[https://doi.org/https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/https://doi.org/10.1016/S0140-6736(20)30566-3)