Development and External Evaluation of Predictions Models for Mortality of COVID-19 Patients using Machine Learning Method

Simin Li
Yidu Cloud Technology Inc.
Yulan Lin
Fujian Medical University
Tong Zhu
Yidu Cloud Technology Inc.
Mengjie Fan
Yidu Cloud Technology Inc.
Shicheng Xu
Yidu Cloud Technology Inc.
Weihao Qiu
Yidu Cloud Technology Inc.
Can Chen
Yidu Cloud Technology Inc.
Linfeng Li
Yidu Cloud Technology Inc.
Yao Wang
Yidu Cloud Techenology Inc.
Jun Yan
Yidu Cloud Technology Inc.
Justin Wong
Ministry of Health Brunei
Lin Naing
Universiti Brunei Darussalam
Minghuan Wang
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Chuan Qin
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Jia Song
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Shabei Xu
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology
Zheng Liu (zhengliuent@hotmail.com)
Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology

Research article

Keywords: coronavirus disease 2019, mortality, prediction, machine learning, China

Posted Date: July 24th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-38913/v1

Version of Record: A version of this preprint was published at Neural Computing and Applications on January 5th, 2021. See the published version at https://doi.org/10.1007/s00521-020-05592-1.

Abstract

Background To develop and evaluate the prognostic machine-learning model for mortality in patients with coronavirus disease 2019 (COVID-19).

Methods Clinical data of confirmed COVID-19 were retrospectively collected from Wuhan between 18th January and 29th March 2020. Gradient Boosting Decision Tree (GBDT), logistic regression (LR) model, and simplified LR with selected 5 features (LR-5) model were built to predict the mortality of COVID-19. 5-fold area under curve (AUC), accuracy, positive predictive value (PPV), and negative predictive value (NPV) were calculated and compared between models.

Results A total of 2,924 patients were included in the final analysis, 257(8.8%) of whom died during hospitalization and 2,667 (91.2%) survived. There were 21(0.7%) mild cases, 2,051(70.1%) moderate case, 779(26.6%) severe cases, and 73(2.5%) critically severe cases of COVID-19 on admission. T

he overall 5-fold AUC was observed highest in GBDT model (0.941), followed by LR (0.928) and LR-5 (0.913). The diagnostic accuracy were 0.889 in GBDT, 0.868 in LR and 0.887 in LR-5. GBDT model also showed the highest sensitivity (0.899) and speciality (0.889). The NPV of all three models exceeded 97%, while the PPV were relatively low in all models, 0.381 for LR, 0.402 for LR-5 and 0.432 for GBDT. In subgroups analysis with severe cases only, GBDT model also performed the best with a accuracy of 0.799 and 5-fold AUC (0.918).

Conclusion The finding revealed that mortality prediction performance of the GBDT was superior to the LR models in confirmed cases of COVID-19, regardless of disease severity.

Introduction

Coronavirus Disease 2019 (COVID-19) is a new form of respiratory disorder caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1]. According to WHO coronavirus disease situation reports [2], there has been nearly 2 million cases and thirteen thousand deaths as of 16th April 2020[2]. Patients with COVID-19 may develop acute respiratory distress syndrome and occasionally may progress to multiorgan failure[3]. Latest reports suggested that for COVID-19 infection the hospitalization rate is 20.7–31.4%, ICU admission rate is 4.9–11.5%[4] and the mortality is as high as 6.5% in confirmed cases[2]. The sharp increase of COVID-19 cases leads to a growing demand for medical equipment and intensive care unit admission. Clinical decision models for the prognosis of confirmed COVID-19 may support the clinician's decision-making, prioritize health-care resources and mitigate the burden on the healthcare system.

There are at least 37 models for disease progression prediction, although most of them were published in open access repositories, ahead of peer review, being developed by 24 march, 2020[6], and would be more now. Most prediction models for disease progression were single center studies with small sample sizes (26–577 cases) and developed with multivariable logistic regression [6–9], which led to an increased risk of overfitting.

Machine learning methods for predicting outcomes more accurately have been implemented in the medical field [10–12]. The COVID-19 related machine learning model mainly focused on the radiology diagnosis rather than the prediction of disease progression. With the technological advances of machine learning, such as high-volume information extraction from medical records and being not sensitive to missing data, we hypothesized a machine learning model can be used to predict fatal outcome of COVID-19 patients. Thus we conducted this study aiming to develop and evaluate the prognostic machine-learning model for mortality in patients with COVID-19.

Methods

Study Population and Data Sources

This retrospective study was conducted between 18th January 2020 and 29th March 2020 in Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology. A total of 3,057 patients diagnosed with COVID-19 during the study period. The medical records of those patients were accessed. The inclusion criteria were patients with laboratory confirmed COVID-19 and with definite outcomes (death or discharged). The exclusion criteria were as follows: patients still on hospitalization and did not develop the outcome by the end of the study period; patients lost to follow-up; or patients died within 24 hours after admission. Patients were discharged from hospital after both clinical recovery and detection of negative SARS-Cov-2 RNA twice in 24 hours apart.

The diagnosis of COVID-19 was based on the Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment (7th version)[13]. Four levels of disease severity for COVID-19 were defined by the guidance: mild, moderate, severe and critically ill. In this study, we classified the mild and moderate as non-severe cases, while rest two levels as severe cases. The primary outcome in this study was death during hospitalization.

Data collection

The medical records of all eligible patients were screened and data extraction was completed by research team. Demographic, clinical, laboratory, radiological characteristics and treatment and outcomes data were obtained with data collection forms from electronic medical reports.

Features Extraction and Selection

A total of 1,224 features were initially extracted from electronic medical records. Univariate chi-square and t-test were used to compare the distribution differences between survivor and non-survivor group. Eventually, 152 features with $p \leq 0.05$ were selected for further model development (see Supplementary Appendix A for list of features), including demographic variables (age and sex), comorbidities (hypertension, diabetes, heart disease, malignant tumor and etc.), initial vital signs (body temperature, systolic blood pressure, respiration rate, and heart rate), clinical symptoms(fever, cough, dyspnea, etc),blood gas analysis, routine blood test, biochemical examination, flow cytometry detection as well as cytokine profiles.

Machine Learning and External Validation

Figure 1 has illustrated the process of machine learning. Details on development of machine learning could be found in the Supplementary Appendix B. In brief, Gradient Boosting Decision Tree (GBDT), logistic regression (LR) model, and simplified LR (LR-5) model with 5 selected features were built. GBDT model was initially trained using all 152 features in training set, and only 83 features were retained in the final prediction model (selected 83 features were listed in Supplementary Appendix C). In order to make our LR model more user-friendly for clinicians, we developed a simplified LR model (LR-5) using only five features selected by stepwise regression. The five features in LR-5 model were serum lactic dehydrogenase (LDH), urea, leukomonocyte (%), age and SPO2. The formula used in the LR-5 could be found in Supplementary Appendix D. Finally, we also conducted an external validation test for LR-5 model using clinical data of 72 confirmed COVID-19 from Brunei.

Statistically analysis

Continuous variables were presented as median with interquartile range (IQR). Categorical variables were presented as n (%). χ^2 test and t-test were used to compare differences among non-survivors and survivors. All variables found to have a statistically significant association (two-tailed, p-value < 0.05). The prediction ability of different models was compared using 5-fold area under curve (AUC), positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity, accuracy, Yourden's index and threshold. In order to testify models' ability of death prediction based on disease severity, we also compared the performance of different models in two subgroups: non-severe (mild and moderate) group, severe (severe and critically severe) group. Each patients' data were transformed and contained 152 features, which was then randomly assigned to either training set (80%, n = 2,339) or testing set (20%, n = 585). Models were trained in the training set, and 5-fold areas were calculated based on testing set for further model comparisons.

Results

Baseline characteristics of patients

A total of 3,057 patients with COVID-19 were hospitalized in the study, 97 patients were excluded for lost to follow-up, 11 were still on hospitalization during study period, 25 patients died within 24 hours (Fig. 2). A total of 2,924 patients were eventually included in the final analysis, 257(8.8%) of whom died during hospitalization and 2,667 (91.2%) survived. There were 1,481 (50.6%) males, and the median age of the cohort was 59 years old. Approximately 43% patients had comorbidities, the most common disease was hypertension (29.6%), followed by cardiovascular disease (34.1%), diabetes (13.6%), coronary disease(7.1%), cerebrovascular disease (3.0%), malignancy (2.4%), COPD (1.2%). There were 21(0.7%) mild cases, 2,051(70.1%) moderate case, 779(26.6%) severe cases, and 73(2.5%) critically severe cases of COVID-19 on admission (Table 1). The death event was occurred in 0 mild cases, 95/1,956(4.86%) in moderate cases, 134/645(20.8%) in severe cases, and 28/45(62.2%) critically severe cases.

Table 1
Baseline characteristic of the patients on admission

Features	Total (N = 2924)	Survival (n = 2667)	Death (n = 257)	*P*	AUC
Age (years), median (IQR)	61.876(49.737–69.539)	60.703(48.381–68.692)	69.577(62.709–78.333)	< 0.001	0.718
Gender (%)					
Female	1443 (49.4)	1267 (47.5)	176 (68.5)	< 0.001	0.605
Male	1481 (50.6)	1400 (52.5)	81 (31.5)	< 0.001	0.605
Underling Comorbidity (%)					
Any	1263 (43.2)	1108 (41.5)	155 (60.3)	< 0.001	0.594
Cardiovasulcar disease	998.0 (34.1)	878.0 (32.9)	120.0 (46.7)		
Coronary disease	208.0 (7.1)	173.0 (6.5)	35.0 (13.6)	< 0.001	0.536
Hypertension	865.0 (29.6)	764.0 (28.6)	101.0 (39.3)	0.001	0.553
Cerebrovascular disease	87.0 (3.0)	70.0 (2.6)	17.0 (6.6)	0.001	0.520
COPD	35.0 (1.2)	27.0 (1.0)	8.0 (3.1)	0.009	0.511
Diabetes	397.0 (13.6)	358.0 (13.4)	39.0 (15.2)	0.445	0.509
Malignancy	70.0 (2.4)	53.0 (2.0)	17.0 (6.6)	< 0.001	0.523
Infectious disease	92.0 (3.1)	78.0 (2.9)	14.0 (5.4)	0.037	0.513
Tuberculosis	52.0 (1.8)	44.0 (1.6)	8.0 (3.1)	0.130	0.507
CKD	17.0 (0.6)	12.0 (0.4)	5.0 (1.9)	0.013	0.507
Hepatitis	45.0 (1.5)	40.0 (1.5)	5.0(1.9)	0.591	0.502
Severity of COVID-19 on admission (%)					
Mild	21 (0.7)	21 (0.8)	0 (0.0)	0.250	0.504
Moderate	2051 (70.1)	1956 (73.3)	95 (37.0)	< 0.001	0.682
Severe	779 (26.6)	645 (24.2)	134 (52.1)	< 0.001	0.640
Critical	73 (2.5)	45 (1.7)	28 (10.9)	< 0.001	0.546
Clinical manifestation (%)					
Fever	1964.0 (67.2)	1788.0 (67.0)	176.0 (68.5)	0.677	0.507
Cough	1510.0 (51.6)	1381.0 (51.8)	129.0 (50.2)	0.648	0.508
Pant	42.0 (1.4)	33.0 (1.2)	9.0 (3.5)	0.009	0.511
Dyspnea	962.0 (32.9)	844.0 (31.6)	118.0 (45.9)	< 0.001	0.571
Dizzy	63.0 (2.2)	48.0 (1.8)	15.0 (5.8)	< 0.001	0.520
Pharyngalgia	129.0 (4.4)	128.0 (4.8)	1.0 (0.4)	< 0.001	0.522
Temperature (℃)	36.8 (0.7)	36.8 (0.7)	37.0 (0.9)	< 0.001	0.585
Pulse (rates/min)	90.8 (22.0)	90.4 (20.0)	95.5 (27.5)	< 0.001	0.571
RR (rates/min)	23.5 (2.0)	23.4 (2.0)	25.2 (10.0)	< 0.001	0.682

Notes: Continuous variables were expressed as medians with interquartile range (IQRs)

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; COPD, chronic obstructive pulmonary disease; CKD, chronic kidney diseases; WBC, white blood cell count; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate (ESR); NLR, Neutrophil-to-Lymphocyte ratio; LDH, lactic dehydrogenase; eGFR, estimated glomerular filtration rate; HDL-C = high-density lipoprotein cholesterol; SBP = systolic blood pressure; RR, respiratory rate; DBP, diastolic blood pressure; BUN, blood urea nitrogen; AUC, area under curve.

Features	Total (N = 2924)	Survival (n = 2667)	Death (n = 257)	*P*	AUC
SBP (mmHg)	175.2 (24.0)	179.1 (23.0)	133.1 (26.0)	0.134	0.522
DBP (mmHg)	81.0 (17.0)	81.1 (16.0)	80.3 (17.0)	0.211	0.516
SPO_2 (%)	95.4 (3.0)	96.2 (2.0)	87.1 (15.0)	< 0.001	0.729
Laboratory test, median (IQR)				< 0.001	
WBC (× 10^9/L)	5.78(4.55–7.39)	5.69(4.49–7.145)	8.595(5.677–12.928)	< 0.001	0.721
Neutrophil (× 10^9/L)	3.73(2.67–5.28)	3.58(2.62–4.945)	7.465(4.5-11.622)	< 0.001	0.790
Lymphocyte (× 10^9/L)	1.22(0.81–1.68)	1.29(0.89–1.73)	0.585(0.42–0.8)	< 0.001	0.847
NLR	2.906(1.81–5.418)	2.69(1.756–4.57)	12.211(6.49-23.396)	< 0.001	0.883
Platelets (× 10^9/L)	222.0(170.0-284.0)	225.0(176.0-289.0)	152.0(112.0-222.0)	< 0.001	0.728
ESR (mm/h)	28.0(13.0–55.0)	27.0(12.0–54.0)	35.0(18.0–60.0)	0.008	0.562
LDH (U/L)	241.0(192.5–328.0)	233.0(189.0-305.0)	485.0(363.0-639.0)	< 0.001	0.876
CRP (mg/L)	10.2(1.6–55.9)	7.8(1.4–43.2)	103.7(59.85–162.4)	< 0.001	0.873
HDL-C (mmol/L)	0.96(0.79–1.2)	0.98(0.812–1.22)	0.76(0.55–0.92)	< 0.001	0.743
Procalcitonin (μg/L)	0.06(0.04–0.12)	0.06(0.04–0.09)	0.245(0.13–0.712)	< 0.001	0.870
Ferritin (ng/mL)	473.0(233.675–915.2)	421.7(213.7-792.35)	1436.8(771.75-2444.5)	< 0.001	0.826
Total bilirubin (μmol/L)	8.85(6.6–12.1)	8.6(6.4–11.7)	12.0(8.7–17.6)	< 0.001	0.692
ALT (U/L)	22.0(14.0–38.0)	22.0(14.0–37.0)	24.0(17.25-42.0)	0.001	0.562
AST (U/L)	25.0(18.0–36.0)	24.0(18.0–34.0)	41.0(29.0–58.0)	< 0.001	0.755
Prealbumin (g/L)	231.0(167.0-278.0)	236.0(178.0-279.0)	118.0(99.5-141.5)	< 0.001	0.843
Albumin (g/L)	36.7(32.6-40.85)	37.4(33.4–41.3)	31.3(28.2–34.2)	< 0.001	0.191
BUN (mmol/L)	4.5(3.5–5.8)	4.4(3.4–5.5)	8.3(5.5-12.775)	< 0.001	0.811
Creatinine (μmol/L)	68.0(56.0–83.0)	67.0(56.0–81.0)	86.5(67.0-110.75)	< 0.001	0.704
eGFR (ml/min)	93.4(79.3–104.0)	94.3(81.9-104.9)	73.2(48.7–90.6)	< 0.001	0.740
TNF-α (pg/ml)	8.1(6.5–10.5)	7.9(6.4–10.0)	11.45(9.025–18.975)	< 0.001	0.760
IL-2R (pg/ml)	405.0(281.0-649.0)	381.0(277.0-581.0)	1096.5(726.75–1717.0)	< 0.001	0.881
IL-6 (pg/ml)	6.03(2.76-22.525)	5.025(2.63-18.362)	59.69(23.16–122.0)	< 0.001	0.887
IL-8 (pg/ml)	10.9(7.6-18.075)	10.4(7.325–16.65)	23.95(13.55–52.35)	< 0.001	0.785
IL-10 (pg/ml)	8.6(6.3–13.4)	7.9(6.1–11.6)	14.6(9.525–25.5)	< 0.001	0.748

Notes: Continuous variables were expressed as medians with interquartile range (IQRs)

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; COPD, chronic obstructive pulmonary disease; CKD, chronic kidney diseases; WBC, white blood cell count; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate (ESR); NLR, Neutrophil-to-Lymphocyte ratio; LDH, lactic dehydrogenase; eGFR, estimated glomerular filtration rate; HDL-C = high-density lipoprotein cholesterol; SBP = systolic blood pressure; RR, respiratory rate; DBP, diastolic blood pressure; BUN, blood urea nitrogen; AUC, area under curve.

Comparisons of baseline between survivors and non-survivors

Table 1 presents the comparison of the baseline characteristics between survivors and non-survivors. Compared to survivors, non-survivors were older (69.577[62.709–78.333] vs. 60.703[48.381–68.692], p < 0.001), and more likely to be female (68.5% vs. 47.5%, p < 0.001). Comorbidities were more common in non-survivors, with 60.3% in non-survivors and 41.5% in survivors (p < 0.001). Specifically, the cardiovascular diseases (46.7%), Chronic obstructive pulmonary disease (COPD) (3.1%) and cancer (6.6%) were prominent in non-

survivors. Lower lymphocyte (0.7[0.4] vs. 1.4[0.8], p < 0.001) and higher neutrophils (8.6[7.1] vs. 4.1[2.3], p < 0.001) and NLR (17.7 [16.9] vs. 4.1[2.8], p < 0.001) level were found in non-survivors than survivors. LDH, hs-CRP, and pro-inflammatory cytokines as such IL-6, TNF-α, IL-10 were higher in non-survivors than survivors.

Comparisons of different models in the full cohort

The top ten features with the highest predictive accuracy in the models are shown in Table 2. Three models were finally developed and tested with 5 fold cross validation (Table 3). LR model comprised 152 features and GBDT models had 83 features. We then simplified LR model as LR-5 comprised the top 5 common clinical indices. The overall 5-fold AUC of LR, LR-5 and GBDT models were 0.928, 0.913 and 0.941, respectively, among which, GBDT models has the largest AUC. Similarly, estimated AUC on testing set was also highest in GBDT model (0.939), followed by LR (0.928) and LR-5 (0.915) (Fig. 3A). The diagnostic accuracy were 0.889 in GBDT, 0.868 in LR and 0.887 in LR-5. GBDT model also obtained the highest sensitivity (0.899) and specificity (0.889). The NPV of all three models exceeded 97%, while the PPV were not high in all models, with 0.381 for LR, 0.402 for LR-5 and 0.432 for GBDT.

Table 2
Top ten features with highest predictive ability.

Feature No.	Feature added	P value of coef	AUC on train	AUC on test
1.0	LDH	< 0.001	0.840	0.876
2.0	BUN	< 0.001	0.882	0.877
3.0	Lymphocyte (%)	< 0.001	0.895	0.903
4.0	Age	< 0.001	0.903	0.911
5.0	SPO_2	< 0.001	0.915	0.917
6.0	Platelets	< 0.001	0.923	0.925
7.0	CRP	< 0.001	0.930	0.921
8.0	IL-10	0.001	0.932	0.930
9.0	HDL-C	0.005	0.934	0.932
10.0	SaO_2	0.005	0.935	0.931
Abbreviations: LDH, lactic dehydrogenase; BUN, blood urea nitrogen; CRP, C-reactive protein; HDL-C = high-density lipoprotein cholesterol;				
AUC, area under curve.				

Table 3
Prediction accuracy of different models in different cohort.

	LR model			LR-5 model			GBDT model		
No. of included feature	152			5			83		
	Total	Non-severe	Severe	Total	Non-severe	Severe	Total	Non-severe	Severe
Total (death)	2924(257)	2072 (95)	852(162)	2924(257)	2072(95)	852(162)	2924(257)	2072(95)	852(162)
Threshold	0.110	0.110	0.110	0.140	0.140	0.140	0.090	0.090	0.090
5-fold AUC	0.928	0.924	0.891	0.913	0.895	0.887	0.941	0.932	0.918
AUC on testing set	0.928	0.946	0.855	0.915	0.902	0.864	0.939	0.940	0.897
AUC on training set	0.937	0.931	0.913	0.913	0.897	0.888	0.997	0.997	0.997
Sensitivity (95%CI)	0.878	0.933	0.714	0.898	0.952	0.711	0.899	0.940	0.774
Specificity (95% CI)	0.769	0.714	0.806	0.771	0.588	0.871	0.788	0.619	0.903
Accuracy	0.868	0.922	0.732	0.887	0.938	0.743	0.889	0.924	0.799
Positive predictive value	0.381	0.357	0.397	0.402	0.333	0.435	0.432	0.351	0.483
Negative predictive value	0.975	0.984	0.941	0.978	0.983	0.956	0.978	0.979	0.972

Abbreviations: AUC, area under curve.

Performance of models in COVID-19 patients with different disease severity

As patients with mild or moderate COVID-19 are not hospitalized due to the scarcity of medical resources in most countries, we tried to test models under different clinical scenarios. Table 3 also shows the performance result of models stratified by disease severity. All models performed excellent in non-severe cases with accuracy of 0.922, 0.938, and 0.924 in LR, LR-5 and GBDT models, respectively. LR model however had the highest AUC on testing set (Fig. 3B). In severe cases, the accuracy of LR model for predicting mortality was the lowest (0.732), followed by LR-5 model (0.743). The GBDT model performed the best in severe cases with a accuracy of 0.799. GBDT also showed the highest 5-fold AUC (0.918) as well as highest AUC on testing set (0.897) in severe cases (Fig. 3C). The NPV remained high in both severe and non-severe cases. The PPV of GBDT model for predicting death was even greater in severe cases(0.483) than over all cohort(0.432), indicating an excellent ability in early identification of patients with poor outcome.

Discussion

In this study, we applied machine-learning algorithms to develop prognostic models for predicting mortality in confirmed cases of COVID-19. All models performed well in overall population. Particularly, prediction performance of the GBDT was superior to LR models in the subgroup of severe COVID-19. Furthermore, we developed a simplified LR-5 models with 5 indices as a convenient tool for clinical doctors that showed an acceptable AUC and accuracy.

The demographic and clinical characteristics of this cohort were representative. Most of the risk factors found in non-survivors have been reported in previous study[14–16]. The top ten features in the models included LDH, BUN, lymphocyte count, age, SPO2, platelets, CRP, IL-10, HDL-C, and SaO2, most of which have been repeatedly documented in literature [6,17,18]. These variables reflected different aspects of the characteristics of COVID-19, for example, the respiratory failure (SpO2 and SaO2), the renal dysfunction (BUN). Notably the indicators of the systemic inflammation (LDH, CRP, IL-10, Platelets) comprised almost half of top ten features. Systemic inflammation has been reported in severe COVID-19[19]. The cytokine storm may play a crucial role in the development of respiratory failure and consequently organ failure[20,21]. Higher cytokine level (IL-2R, IL-6, IL-10, and TNF-a) has been found in non-survivor group patients in this study, which was consistent with previous studies[21,22]. Moreover, one of the top ten features in the machine learning models was IL-10, which is a

cytokine with potent anti-inflammatory properties that can induce T cell exhaustion[23,24]. This might partially contribute to the lymphopenia in severe COVID-19.

The models in this study were derived from real-world data with comprehensive details, thus the selection bias was limited and the results were more representative than other models. All of the three models performed well with AUC of 0.911–0.943 and NPVs exceeded 97%. However, the PPVs were relatively low, which were consistent with all the other prediction models reported in literature. The major reason for this could be the dynamic change of the disease. All the models in this study as well as in the literature were derived from baseline data collected on admission, where highly heterogeneity exited. A dynamic model could have better performance.

Compared with LR models, GBDT performed better in mortality prediction in both full cohort and subgroup of different severity. GBDT is not sensitive to missing data, therefore can serve as a good tool for early detection of potential critical patients and optimize the medical resource allocation. In contrast, LR model has superiority on high-speed calculation and provides result handy for interpretation, which might be more user-friendly in clinics. However, this LR full model included 161 features and the application could be cumbersome for daily clinical practice, especially when the healthcare systems were confronting severe human resource shortage. As a simplified model, the LR-5 model incorporating only 5 common variables with an excellent PPV and satisfying accuracy could be recommended as a simple tool for clinical use.

We also conducted an external validation test using data from Brunei. During 29th Feb and 29th March 2020, a total of 72 confirmed cases of COVID-19 were followed, among whom 2 died (Supplemental Appendix E). Based on LR-5 model, patients' data of leukomonocyte (%), urea, age and SPO2 were collected for analysis, while data on LDH were unavailable. LDH was then filled using the median value that estimated from training set (median = 239 U/L). As a result, leukomonocyte (%) turned to show the highest AUC (0.917), followed by urea (0.867), age (0.826), and SPO2 (0.704) (data not shown). As a prediction tool, LR-5 model showed a strong ability in death prediction with a very high AUC of 0.97. However, it shall be noted that selection bias due to small sample size could never be eliminated and further external validation study using larger sample size should provide warranty.

There were several limitations in this study. Firstly, we only used 5-fold cross validation rather than external validation due to the lack of external data. Second, only the Chinese patients were included, the generalisability and implementation of these models across different settings and populations remains unknown.

In conclusion, three models were developed in this study. GBDT models performed the best in different severity. LR-5 is a simple tool for routine care.

Declarations

Ethics approval and content to participate

The study was approved by the institutional committee board of Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, and Universiti Brunei Darussalam Research Ethics Committee. All cases were anonymous in the final analysis. The requirement for informed consent was waived due to the retrospective nature of the analyses.

Data and material availability

Data could be requested to corresponding author upon reasonable request.

Consent for publication

NA

Competing interests

We declare no competing interests.

Authors' contributions

SL, YL and TZ plan the study. SL, TZ, SX obtained the data and performed the statistical analysis and data summarization. SL and YL drafted the manuscript. All authors discussed the results and contributed to the final manuscript.

References

1. El Zowalaty ME, Järhult JD. From SARS to COVID-19: A previously unknown SARS- related coronavirus (SARS-CoV-2) of pandemic potential infecting humans - Call for a One Health approach. One health (Amsterdam Netherlands). 2020;9:100124.
2. WHO Coronavirus disease situation reports, 16. April 2020, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
3. Yang F, Shi S, Zhu J, Shi J, Dai K, Chen X. Analysis of 92 deceased patients with COVID-19. *Journal of medical virology.* 2020.
4. Severe Outcomes Among Patients with Coronavirus. Disease 2019 (COVID-19) — United States, February 12–March 16, 2020. MMWR Morb Mortal Wkly Rep 2020;69:343–346.. 2018.
5. Lippi G, Mattiuzzi C, Sanchis-Gomar F, Henry BM. Clinical and demographic characteristics of patients dying from COVID-19 in Italy versus China. *Journal of medical virology.* 2020.
6. Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. 2020;369:m1328.
7. Ji D, Zhang D, Xu J, et al. Prediction for Progression Risk in Patients with COVID-19 Pneumonia: the CALL Score. *Clin Infect Dis.* 2020.
8. Tan L, Wang Q, Zhang D, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal transduction targeted therapy. 2020;5:33.
9. Shi Y, Yu X, Zhao H, Wang H, Zhao R, Sheng J. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit Care (London England). 2020;24(1):108.
10. Mähringer-Kunz A, Wagner F, Hahn F, et al. Predicting survival after transarterial chemoembolization for hepatocellular carcinoma using a neural network: A Pilot Study. Liver international: official journal of the International Association for the Study of the Liver. 2020;40(3):694–703.
11. Cui C, Wang S, Zhou J, et al. Machine Learning Analysis of Image Data Based on Detailed MR Image Reports for Nasopharyngeal Carcinoma Prognosis. BioMed research international. 2020;2020:8068913.
12. Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine Learning in Oncology: A Clinical Appraisal. *Cancer letters.* 2020.
13. The Chinese Clinical Guidance for COVID-19. Pneumonia Diagnosis and Treatment(7th version). http://kjfymeetingchinaorg/msite/news/show/cn/3337html.
14. Zhao Q, Meng M, Kumar R, et al. The impact of COPD and smoking history on the severity of Covid-19: A systemic review and meta-analysis. *Journal of medical virology.* 2020.
15. Guan W-j, Ni Z-y, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine.* 2020.
16. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395(10223):497–506.
17. Zhu J, Ji P, Pang J, et al. Clinical characteristics of 3,062 COVID-19 patients: a meta-analysis. *Journal of medical virology.* 2020.
18. Lagunas-Rangel FA. Neutrophil-to-lymphocyte ratio and lymphocyte-to-C-reactive protein ratio in patients with severe coronavirus disease 2019 (COVID-19): A meta-analysis. *Journal of medical virology.* 2020.
19. Qin C, Zhou L, Hu Z, et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. *Clin Infect Dis.* 2020.
20. Henderson LA, Canna SW, Schulert GS, et al. On the alert for cytokine storm: Immunopathology in COVID-19. *Arthritis & rheumatology (Hoboken, NJ).* 2020.
21. Gao Y, Li T, Han M, et al. Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *Journal of medical virology.* 2020.
22. Li X, Xu S, Yu M, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. *The Journal of allergy and clinical immunology.* 2020.
23. Iyer SS, Cheng G. Role of interleukin 10 transcriptional regulation in inflammation and autoimmune disease. Crit Rev Immunol. 2012;32(1):23–63.
24. Brooks DG, Trifilo MJ, Edelmann KH, Teyton L, McGavern DB, Oldstone MB. Interleukin-10 determines viral clearance or persistence in vivo. Nature medicine. 2006;12(11):1301–9.

Figures

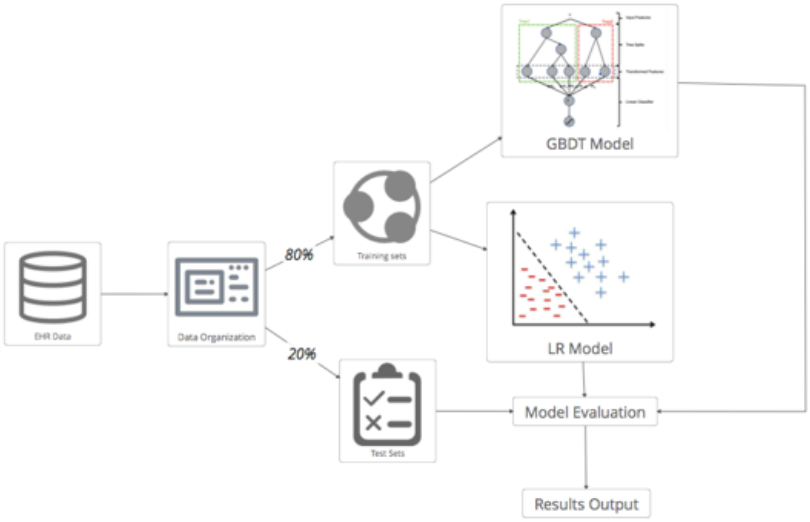

Figure 1

Machine learning process

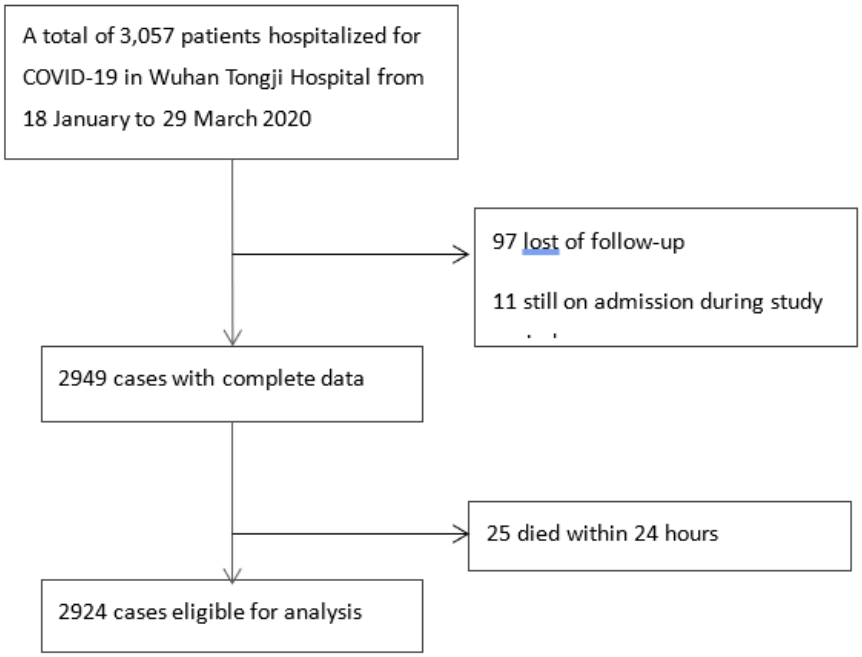

Figure 2

Recruitment process of study population

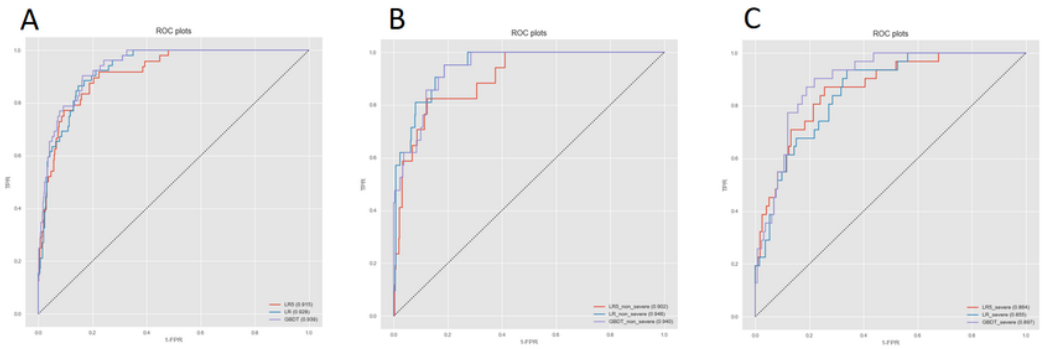

Figure 3

Comparisons of receiver operating characteristic (ROC) curve estimated based on testing set for LR, LR-5, GMBT models in full cohort and subgroups of non-severe and severe patients. A. ROC curve of GBMT, LR, and LR-5 models in testing set B. ROC curve of GBMT, LR, and LR-5 models among non-severe cases in testing set C. ROC curve of GBMT, LR, and LR-5 models among severe cases in testing set

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalAppendix.docx