Outcome prediction model and prognostic biomarkers for COVID-19 patients in Vietnam

Nguyen, Hien Thi Thu; Le-Quy, Vang; Ho, Son Van; Thomsen, Jakob Holm Dalsgaard; Stoico, Malene Pontoppidan; Tong, Hoang Vang; Nguyen, Nhat-Linh; Krarup, Henrik Bygum; Nguyen, Son Hong; Tran, Viet Quoc; Nguyen, Toan Linh; Dinh-Xuan, Anh-Tuan

Published in:
ERJ Open Research

DOI (link to publication from Publisher):
10.1183/23120541.00481-2022

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Outcome prediction model and prognostic biomarkers for COVID-19 patients in Vietnam

Hien Thi Thu Nguyen [1,2], Vang Le-Quy [2,3,8], Son Van Ho [4,8], Jakob Holm Dalsgaard Thomsen [1,5], Malene Pontoppidan Stoico [1,5], Hoang Van Tong [6], Nhat-Linh Nguyen [2], Henrik Bygum Krarup [1,5], Son Hong Nguyen [4], Viet Quoc Tran [4], Linh Toan Nguyen [6,9] and Anh Tuan Dinh-Xuan [2,7,9]

[1]Department of Molecular Diagnostics, Aalborg University Hospital, Aalborg, Denmark. [2]AVSE Global Medical Translational Research Network, Paris, France. [3]Novodan ApS, Aalborg, Denmark. [4]Military Hospital 175, Ho Chi Minh City, Vietnam. [5]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark. [6]Department of Pathophysiology, Vietnam Military Medical University, Hanoi, Vietnam. [7]Department of Respiratory Medicine and Physiology, Hôpital Cochin, Paris, France. [8]These authors contributed equally to this work. [9]Joint senior authors.

Corresponding author: Anh Tuan Dinh-Xuan (anh-tuan.dinh-xuan@aphp.fr)

Shareable abstract (@ERSpublications)
Outcome prediction model and prognostic biomarkers for COVID-19 patients in Vietnam https://bit.ly/3BKyf7n

Cite this article as: Nguyen HTT, Le-Quy V, Ho SV, *et al*. Outcome prediction model and prognostic biomarkers for COVID-19 patients in Vietnam. *ERJ Open Res* 2023; 9: 00481-2022 [DOI: 10.1183/23120541.00481-2022].

Received: 23 Sept 2022
Accepted: 11 Dec 2022

Abstract

Background Accurate prognosis is important either after acute infection or during long-term follow-up of patients infected by severe acute respiratory syndrome coronavirus 2. This study aims to predict coronavirus disease 2019 (COVID-19) severity based on clinical and biological indicators, and to identify biomarkers for prognostic assessment.

Methods We included 261 Vietnamese COVID-19 patients, who were classified into moderate and severe groups. Disease severity prediction based on biomarkers and clinical parameters was performed by applying machine learning and statistical methods using the combination of clinical and biological data.

Results The random forest model could predict with 97% accuracy the likelihood of COVID-19 patients who subsequently worsened to the severe condition. The most important indicators were interleukin (IL)-6, ferritin and D-dimer. The model could still predict with 92% accuracy after removing IL-6 from the analysis to generalise the applicability of the model to hospitals with limited capacity for IL-6 testing. The five most effective indicators were C-reactive protein (CRP), D-dimer, IL-6, ferritin and dyspnoea. Two different sets of biomarkers (D-dimer, IL-6 and ferritin, and CRP, D-dimer and IL-6) are applicable for the assessment of disease severity and prognosis. The two biomarker sets were further tested through machine learning algorithms and relatively validated on two Danish COVID-19 patient groups (n=32 and n=100). The results indicated that various biomarker sets combined with clinical data can be used for detection of the potential to develop the severe condition.

Conclusion This study provided a simple and reliable model using two different sets of biomarkers to assess disease severity and predict clinical outcomes in COVID-19 patients in Vietnam.

Introduction

The pandemic caused by the novel strain of coronavirus (severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)) since December 2019 is still spreading over the world [1] with new variants causing at least six or seven successive waves of infections to date [2]. Although vaccines for coronavirus disease 2019 (COVID-19) are available and vaccination has been administered in most parts of the world, the number of COVID-19 cases is still increasing rapidly and numerous people are still losing their lives due to the disease [3]. The course of COVID-19 disease ranges from mild self-limiting symptoms to severe and lethal complications such as acute respiratory distress syndrome (ARDS) and multi-organ failure [1, 3, 4]. SARS-CoV-2 virus, especially delta and omicron variants, is likely to spread very quickly, and the mortality rate due to SARS-CoV-2 is estimated at 2–3% and mainly in patients with certain disease

backgrounds [4]. When infected with SARS-CoV-2, the viral particles enter the host cells *via* angiotensin-converting enzyme 2 receptors. Subsequently, the host immune system responds to SARS-CoV-2 infection through cascades of inflammation, cytokine storms and activation of coagulation. These host responses may eventually lead to severe complications and death [5, 6]. Thus, looking for laboratory parameters that could help forecast the progression of COVID-19 toward severity and/or mortality is highly needed especially in at-risk patients and many studies have been conducted to serve this purpose. A large number of biomarkers such as inflammatory factors (interleukins (ILs), C-reactive protein (CRP), procalcitonin (PCT) and lactate dehydrogenase (LDH)), coagulative parameters (D-dimer and fibrinogen) and blood counts have been identified for the management of COVID-19 patients [7]. Of these, levels of IL-6, CRP, D-dimer, PCT and ferritin have been highlighted as potential prognostic and treatment response biomarkers for clinical practice [7, 8].

The cytokine storm has been described in COVID-19 patients and is characterised by a massive secretion of cytokines including IL-1, IL-6 and tumour necrosis factor-α, which are associated with disease severity and death from COVID-19 [9]. IL-6 is secreted by a variety of immune cells in response to infection and plays a central role in regulating the outcome of infectious diseases including SARS-CoV-2 infection [10]. IL-6 is important for regulating B- and T-cell responses, and for coordinating innate and adaptive immunity [11]. IL-6 is able to regulate cell survival, proliferation and differentiation, and may direct both pro-inflammatory and anti-inflammatory outcomes in infections [12]. IL-6 is strongly induced and serum IL-6 levels are significantly elevated in inflammatory and infectious conditions. IL-6 is also observed to be massively increased in a majority of COVID-19 patients [13], and is closely correlated with severity and bilateral interstitial lung involvement [14]. Thus, IL-6 has been identified as a vital biomarker of disease severity and predictor of mortality.

D-dimer, a product of fibrin degradation during fibrinolysis, is found in the circulatory system and is a marker for coagulation disorders [15]. Recently, increased D-dimer levels have been commonly observed in patients with COVID-19 and identified as a vital prognostic biomarker for disease severity and mortality [16, 17]. CRP levels appear as one of the best markers for the prediction of development of severe disease among non-severe COVID-19 cases [7]. Although the contribution of ferritin levels in COVID-19 patient stratification is still debated, a study identified ferritin as an independent risk factor for severity in COVID-19 patients [18]. In addition, a meta-analysis of 25 studies with 5350 patients showed that elevated serum CRP, D-dimer and serum ferritin levels were associated with an increased risk of severe COVID-19 but not with mortality in COVID-19 patients [8].

The outcome of SARS-CoV-2 infection is largely dependent on viral virulence and variability as well as the host's genetic background [19]. Although serum levels of D-dimer, IL-6, CRP and ferritin have been well documented as biomarkers for the prediction of disease progression and mortality, and guide the management of COVID-19 patients in different world populations, there is no study describing the prognostic values of combinations of these biomarkers with various clinical parameters in COVID-19 patients living in low- and middle-income countries (LMICs) such as Vietnam. The current study aims to analyse the serum levels of these biomarkers and their association with clinical outcomes in Vietnamese COVID-19 patients. The results could provide more evidence of the diagnostic and prognostic value of these biomarkers for the management of COVID-19 patients hospitalised both in general and specialised medical centres.

Methods

Ethical approval

The study was conducted in accordance with the principles of the Declaration of Helsinki. The study was approved by Military Hospital 175 (Ho Chi Minh City, Vietnam) according to decision 4889 BV/NCKH signed on 31 December 2021 and the North Denmark Region Committee on Health Research Ethics (protocol N-20200031).

Patient selection and data collection

A retrospective study was performed on 261 Vietnamese patients with SARS-CoV-2 infection enrolled at the COVID-19 Treatment Centre, Military Hospital 175, from 26 July to 22 November 2021. These 261 Vietnamese patients were confirmed positive with SARS-CoV-2 infection by real-time reverse transcriptase (RT)-PCR and were classified into two groups (mild/moderate and severe/critical) based on medical history, clinical symptoms and results of laboratory tests (figure 1). The classification was based on the World Health Organization classification (WHO/2019-nCoV/clinical/2021.2) and adapted by the Vietnam Ministry of Health according to decision 4689/QD-BYT on the issuance of "Guidelines for the diagnosis and treatment of COVID-19" in Vietnam. The "severe" group included 200 patients who were diagnosed

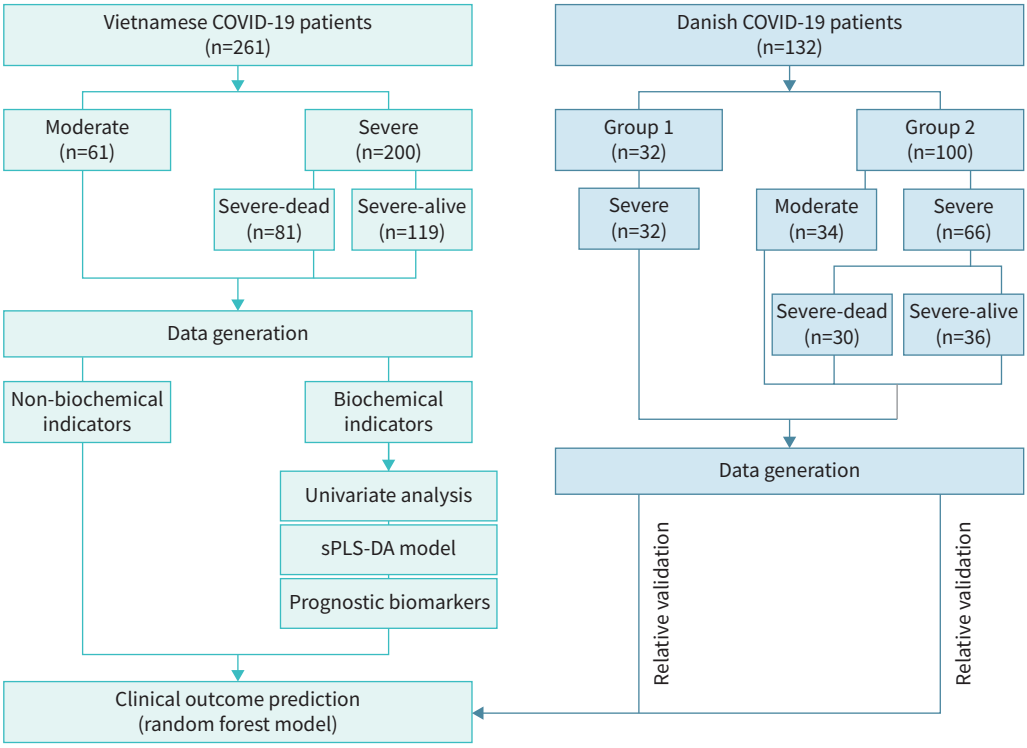


FIGURE 1 Flowchart of the study. sPLS-DA: spare partial least squares discriminant analysis.

with severe/critical conditions, and were further classified into two subgroups based on mortality status ("severe-alive" and "severe-dead)". The "moderate" group included 61 patients who were diagnosed with mild/moderate conditions.

Clinical symptoms classification

The clinical symptoms of severe disease include signs and symptoms of pneumonia accompanied by one of the following signs: breathing >25 breaths·min^{-1}, accessory respiratory muscle contraction, peripheral oxygen saturation (S_{pO_2}) <94% when breathing room air, circulatory symptoms such as fast or slow heart rate, nervous symptoms such as restlessness or languor and fatigue. Other paraclinical manifestations include: chest radiography and chest computed tomography (CT) scan showing lung lesions in >50% area of the lungs, levels of arterial blood gas partial pressure of oxygen (P_{aO_2})/fraction of inspired oxygen (F_{IO_2}) of 200–300, and detection of multiple sonographic B-lines by ultrasound. Patients are considered to have the critical condition if clinical symptoms include rapid breathing >30 breaths·min^{-1} or slow breathing <10 breaths·min^{-1}, signs of severe respiratory failure with laboured breathing, abnormal breathing, decreased consciousness or coma, fast or possibly slow heart rate, low blood pressure, or oliguria or anuria. Other paraclinical manifestations include: chest radiography or chest CT scan showing >50% lung damage, P_{aO_2}/F_{IO_2} <200, respiratory acidosis, blood lactate >2 mmol·L^{-1} and detection of multiple sonographic B-lines by ultrasound.

Comorbidities were diagnosed based on medical history, clinical examination and results of laboratory tests, regardless of age and gender. All related epidemiological and demographic information, clinical symptoms and laboratory parameters were collected for further analyses from medical records.

All clinical data were extracted from the hospital database. Details of the laboratory tests are described in the supplementary material.

Relative validation cohort

A relative validation cohort was chosen from 132 Danish patients with SARS-CoV-2 infection enrolled at the COVID-19 Treatment Centre, Aalborg University Hospital (Aalborg, Denmark). These Danish COVID-19 patients were divided into two groups (figure 1). Group 1 included 32 patients who were

confirmed positive with SARS-CoV-2 infection by RT-PCR from 12 April 2020 to 5 May 2020, and were classified into a severe/hospitalised ("severe") subgroup based on medical history, clinical symptoms and results of laboratory tests according to the WHO classification (WHO/2019-nCoV/clinical/2021.2).

Group 2 recruited 100 Danish COVID-19 patients from 25 March 2020 to 2 February 2021, divided into two subgroups. The first subgroup included 66 patients who were diagnosed with severe/critical conditions ("severe") and were further classified into two subgroups based on mortality status ("severe-alive" and "severe-dead"). The second subgroup included 34 patients who were diagnosed with mild/moderate ("moderate") conditions.

Patients from each validation group included several follow-up samples. All clinical data were extracted from the LABKA database that holds information on blood samples from all hospital visits, including emergency departments, outpatient consultations and admissions to the hospital.

Statistical analysis

Statistical evaluation, biomarker prognostic design and testing

We used the modified MetaboAnalystR 2.0 R package (https://github.com/biocyberman/MetaboAnalystR/tree/dev) and R package for all univariate and multivariate statistical analysis, biomarker design and validation.

Spare partial least squares discriminant analysis (sPLS-DA) was conducted to identify the biochemical marker contributing to discrimination between groups. For evaluation of the classification performance of the sPLS-DA model, a five-fold cross-validation was chosen. With this model, the error rate has been calculated to evaluate the performances.

A classical univariate receiver operating characteristic (ROC) curve analysis was performed to assess the strength of association between individual metabolites and stages of COVID-19 disease. The ROC curve summarises the sensitivity and specificity of that single feature to accurately classify data, which can then be used to compare the overall accuracy of different biomarkers.

Clinical outcome prediction

The random forest algorithm [20] was used to develop classification models and predict outcomes for the patients. Data is divided randomly into a training set and a test set, with 80% of patients in the training set and 20% in the test set. Hyperparameters (*i.e.* max_depth, max_features and n_estimators) were tuned by using cross-validation search. The best set of parameters is chosen by setting the "scoring" parameter to the combination of "accuracy", "F1" and "precision" score. The combination of hyperparameters that resulted in the highest score of the scoring criteria was selected and used for training the model. The test dataset was then used to test prediction accuracy. The result was then summarised in a confusion matrix. Features (*i.e.* columns of data) that have the most influence on the outcome are sorted as produced by the importance list.

Results

Patient baseline, non-biochemical indicator outcome groups comparison

A total of 261 Vietnamese patients with confirmed COVID-19 were eligible for the final analysis, of which 61 were stratified to the moderate group, 81 were stratified to the severe group and dead (severe-dead) and 119 were stratified to the severe group who eventually survived (severe-alive) (figure 1). The median age was 55 years (interquartile range 20 years; range 19–98 years) and 125 (47.9%) were male. Of the 261 patients, 213 (81.6%) had ⩾1 coexisting medical conditions. Hypertension (104 (39.8%)), diabetes (77 (29.5%)), heart failure (28 (10.7%)) and overweight (56 (21.5%)) were the most common coexisting conditions. The most common symptoms at onset of illness were fever (95 (36.4%)), fatigue (47 (18%)), dry cough (130 (49.8%)), exhausted (33 (12.6%)) and dyspnoea (162 (62.1%)) (table 1). The number of comorbid diseases in severe patients (n=200) was significantly higher than that in moderate patients, particularly for hypertension (102 (50.2%) *versus* 2 (3.3%)), diabetes (75 (37.5%) *versus* 2 (3.3%)), heart failure (28 (14.0%) *versus* 0 (0.0%)), high risk (127 (63.5%) *versus* 13 (21.3%)) and overweight (56 (28.0%) *versus* 0 (0.0%)). Compared with the moderate group, severe patient groups were more likely to report ARDS (130 (65.0%) *versus* 0 (0.0%)). There was a much higher number of ARDS in the severe-dead group compared with the severe-alive group (72 (88.9%) *versus* 58 (48.7%)). The severe patients were more likely to report dyspnoea (80.5% *versus* 2%, respectively).

There were no significant differences in age and sex of patients who belong to the moderate group and patients who belong to the severe group.

TABLE 1 Study characteristics, comorbidities and presenting symptom-related variables among Vietnamese COVID-19 patients

Characteristics	Total (n=261)	Moderate (n=61)	Severe (n=200)	
			Severe-alive (n=119)	Severe-dead (n=81)
Age (years)	55±15	57±17	52±14	58±13
Male	125 (48)	26 (43)	68 (57)	31 (38)
Comorbidities				
ARDS	130 (50)	0 (0)	58 (49)	72 (89)
CART	14 (5)	0 (0)	7 (6)	7 (9)
Diabetes	77 (30)	2 (3)	39 (33)	36 (44)
ECMO	7 (3)	0 (0)	7 (6)	0 (0)
Heart failure	28 (11)	0 (0)	15 (13)	13 (16)
High risk[#]	140 (54)	13 (21)	73 (61)	54 (67)
Hypertension	104 (40)	2 (3)	59 (50)	43 (53)
Postpartum	6 (2)	0 (0)	5 (4)	1 (1)
Clinical symptoms				
Dry cough	130 (50)	2 (3)	82 (69)	46 (57)
Dyspnoea	162 (62)	1 (2)	96 (81)	65 (80)
Exhausted	33 (13)	0 (0)	15 (13)	18 (22)
Fatigue	47 (18)	1 (2)	28 (24)	18 (22)
Fever	95 (36)	0 (0)	61 (51)	34 (42)

Data are presented as mean±SD or n (%). ARDS: acute respiratory distress syndrome; CRRT: continuous renal replacement therapy; ECMO: extracorporeal membrane oxygenation. [#]: people with health conditions that affect their immune system such as lung disease.

Biochemical profile variance from COVID-19 patients in different groups

We included all 12 common biochemical markers from COVID-19 samples for the analysis (figure 2). A total of eight markers were significantly different between at least one pair of groups. Higher levels of alanine transaminase (ALT), IL-6, D-dimer, CRP and ferritin were observed in the severe groups than in the moderate group ($p<0.05$ for all). In addition, patients from the severe-dead group showed lower levels of albumin and platelets compared with the severe-alive and moderate groups.

Through sPLS-DA, we could differentiate between moderate *versus* severe-dead groups and between moderate *versus* severe-alive groups with an overall cross-validation accuracy of 82% and 79%, respectively (figure 3a and b). sPLS-DA cannot distinguish severe-alive from severe-dead based on the 12 biochemical markers (data not shown).

We discovered that CRP, D-dimer, IL-6, ferritin and white blood cells were contributing biochemical markers to separate moderate from severe-dead groups based on a combination of a p-value of <0.05 and a fold change value. ROC curve analysis for individual biochemical markers demonstrated that D-dimer, IL-6, CRP and urea had the highest area under the curve (AUC) of 0.95, 0.93, 0.83 and 0.70, respectively (figure 4b).

The volcano plot analysis showed that CRP, D-dimer, IL-6 and ferritin were also top biochemical markers changing between moderate and severe-alive groups (data not shown). Among them D-dimer, IL-6 and ferritin showed the highest AUC values (figure 4a).

Clinical outcome prediction and marker selection

Based on the random forest analysis, IL-6 was the most important predictor of critical illness in patients with COVID-19 pneumonia, followed by ferritin, dyspnoea, D-dimer and CRP (figure 5a). 10 variables with the most significant values identified by random forest (IL-6, ferritin, dyspnoea, D-dimer, CRP, cough, fever, hypertension, platelets and ALT) were then used for machine learning models to predict critically ill patients with COVID-19 pneumonia. The model could predict with 96% accuracy the likelihood of COVID-19 patients worsening to the severe condition. The random forest model achieved a sensitivity of 92%, specificity of 100% and F1 score of 0.96% (figure 5b). When the model went deeper for analysing the subgroups of severe patients, the random forest model could predict with 87% and 82% accuracy for the severe-alive and severe-dead group, respectively (figure 5c). The model could still predict with 92% accuracy for severe patients after removing IL-6 from the analysis to generalise the applicability

https://doi.org/10.1183/23120541.00481-2022

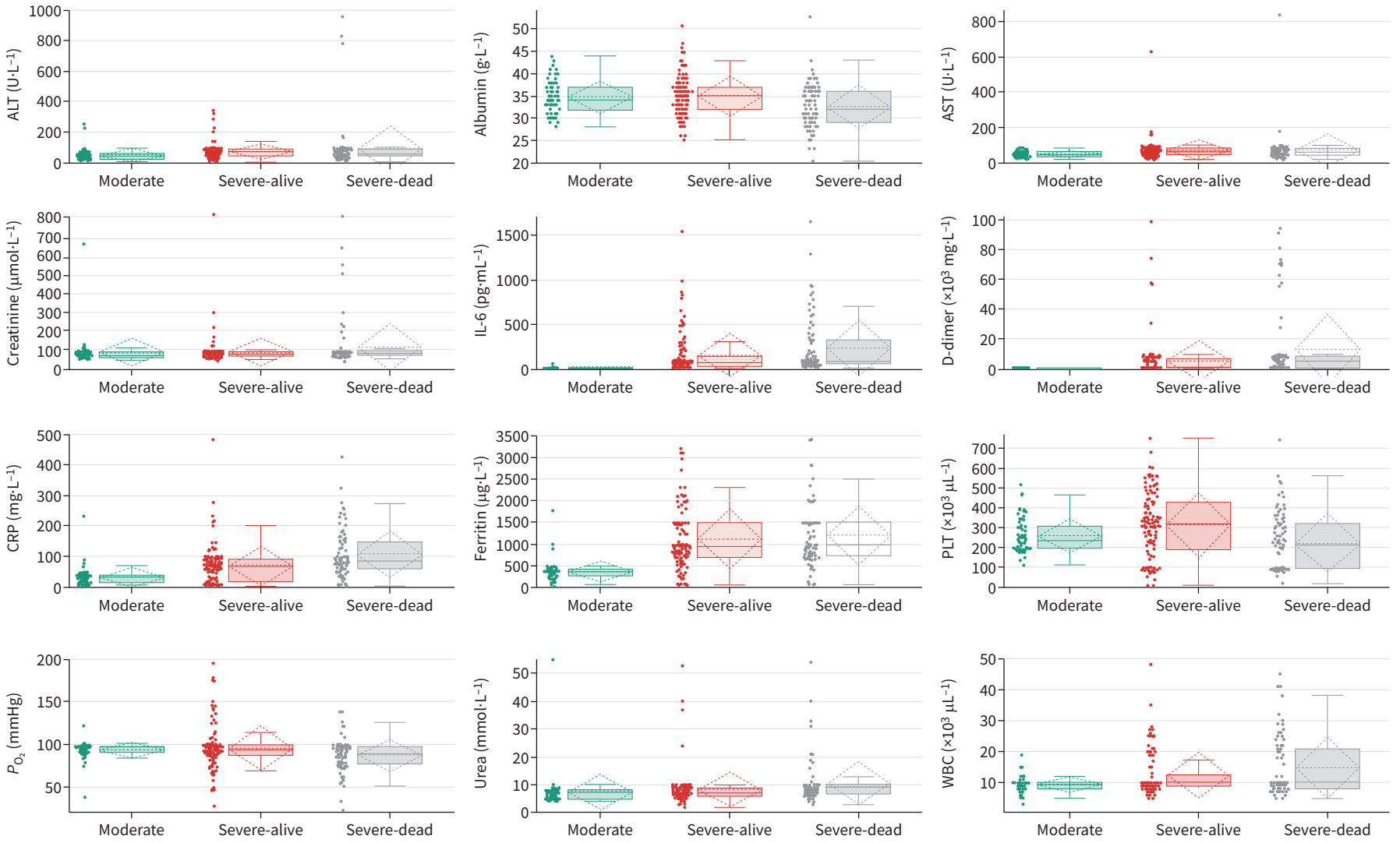


FIGURE 2 Biochemical profile variance from COVID-19 patients with different disease severity. Data for each biomarker are presented as dot blots and box-and-whisker plots. Boxes show median and interquartile range; whiskers show maximum–minimum values; dashed diamonds show standard deviation. ALT: alanine transaminase; AST: aspartate transaminase; IL: interleukin; CRP: C-reactive protein; PLT: platelets; P_{O_2}: partial pressure of oxygen; WBC: white blood cells.

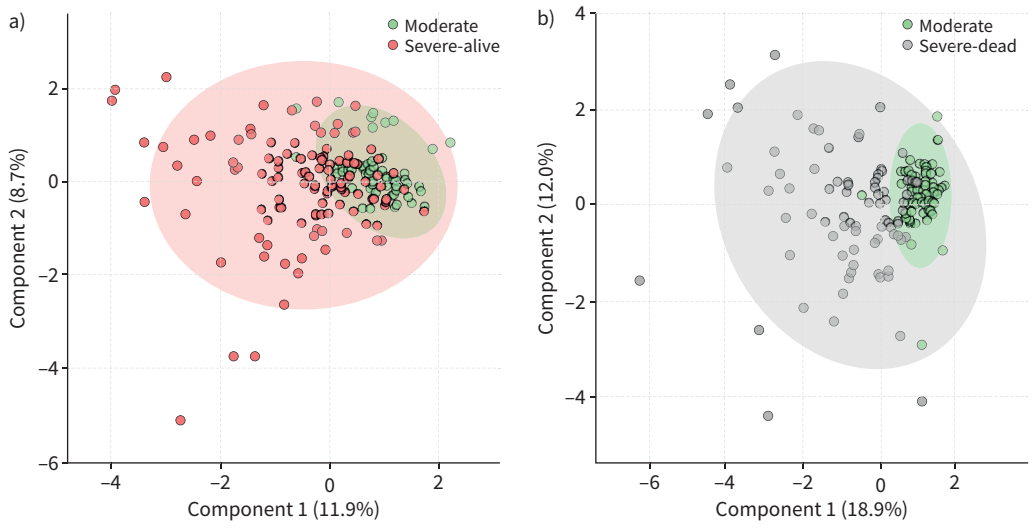


FIGURE 3 Spare partial least squares score plots of a) moderate *versus* severe-alive groups and b) moderate *versus* severe-dead groups.

of the model to hospital settings with limited capacity for IL-6 testing. However, the prediction model without IL-6 could not predict so well with the subgroups of severe patients; the accuracy of each group was 79% and 72% for the severe-alive and severe-dead group, respectively (data not shown).

Biomarker analysis using classical univariate ROC curve analysis also showed that four biomarkers that are the most important predictors of critical illness with the random forest model (CRP, D-dimer, IL-6 and ferritin) were related to the potential severe condition. Among those, a set of three biomarkers (D-dimer, IL-6 and ferritin) was linked to the likelihood of recovery in patients with severe COVID-19. Another set of four biomarkers (CRP, D-dimer, IL-6 and urea) could differentiate moderate from severe-dead patients. Therefore, these four variables (D-dimer, CRP, ferritin and IL-6) were further used for machine learning models. The random forest model could predict with 93% accuracy the likelihood of COVID-19 patients worsening to the severe condition, with a sensitivity of 92%, specificity of 95% and F1 score of 88% (figure 6a).

The model could still predict with 86% accuracy after removing IL-6 from the analysis to generalise the applicability of the model to hospital settings with limited capacity for IL-6 testing (figure 6b).

When CRP was not included in the model due to its low value of importance, the random forest model could predict with 98% accuracy the likelihood of COVID-19 patients worsening to the severe condition. This model with only three factors (D-dimer, ferritin and IL-6) could achieve a sensitivity of 100%, specificity of 95% and F1 score of 92% (figure 6c). We should emphasise that the four factors used in the random forest model could not give high accuracy prediction for deeper group separation of severe patients.

Relative biomarker validation

Two Danish COVID-19 patient groups (Group 1: n=32 (median age 80 years, range 54–84 years) and Group 2: n=100 (median age 79 years, range 30–96 years)) were used to validate the potential of CRP, D-dimer, ferritin and IL-6 markers for predicting the severe condition. In the first Danish COVID-19 patient group, when using all four markers (CRP, D-dimer, ferritin and IL-6) for predicting outcome, the random forest model could predict the severe patient group with 84.8% accuracy (figure 7a). Significantly higher levels of IL-6 were observed in the hospitalised COVID-19 patient subgroups than in the none COVID-19 group with an AUC of 0.82 (figure 7b). A higher prediction value (88% accuracy) was observed when only three factors (D-dimer, ferritin and IL-6) were used (figure 7c). However, when IL-6 was removed from the model, the accuracy of model prediction for the severe group was only 60% (figure 7d). A similar result was seen in the second Danish COVID-19 patient cohort. 76.5% accuracy was obtained from the prediction model for the severe group when using only three markers (CRP, D-dimer and ferritin) (data not shown).

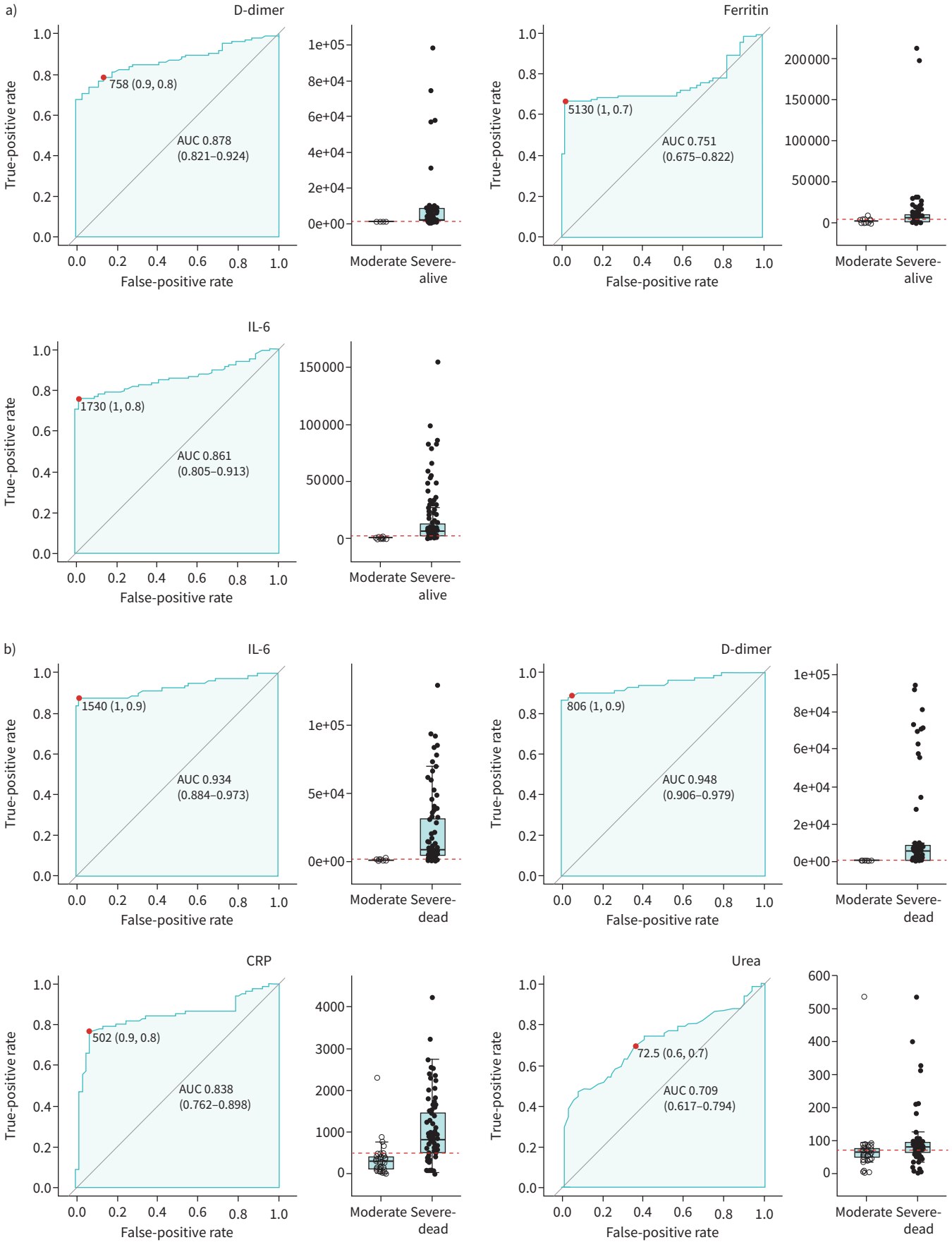

a)
D-dimer
758 (0.9, 0.8)
AUC 0.878
(0.821–0.924)
True-positive rate
False-positive rate
Moderate
Severe-alive
Ferritin
5130 (1, 0.7)
AUC 0.751
(0.675–0.822)
IL-6
1730 (1, 0.8)
AUC 0.861
(0.805–0.913)
b)
IL-6
1540 (1, 0.9)
AUC 0.934
(0.884–0.973)
Severe-dead
D-dimer
806 (1, 0.9)
AUC 0.948
(0.906–0.979)
CRP
502 (0.9, 0.8)
AUC 0.838
(0.762–0.898)
Urea
72.5 (0.6, 0.7)
AUC 0.709
(0.617–0.794)

FIGURE 4 Biomarkers of early COVID-19 severity diagnosis. Receiver operating characteristic (ROC) curve analysis of individual biomarkers for distinguishing a) moderate and severe-alive groups and b) moderate and severe-dead groups. The computed 95% confidence interval for individual biomarkers is highlighted in the shaded background over the ROC curve. The area under the ROC curve (AUC) is shown in red to highlight the diagnostic potential of the corresponding biomarker. The box-and-whisker plots shown on the right side of each ROC curve plot revealed significantly different levels of these biomarkers in the two compared groups. IL: interleukin; CRP: C-reactive protein.

The results indicated the model based on four markers, *i.e.* CRP, D-dimer, ferritin and IL-6, had the potential to predict the likelihood of the severe condition in patients with COVID-19.

Discussions

Outcome prediction is of high interest for many purposes, and even more during the COVID-19 pandemic, especially in LMICs with overall limited medical resources such as Vietnam. It is useful for the planning of resources and infrastructure in efforts to monitor and control the epidemic. It is a reference for doctors to understand risk factors. For patients, however, the information should be handled with care due to the personal and psychological impacts.

This study is the first of its kind to build outcome prediction models and prognostic biomarkers for COVID-19 patients in Vietnam based on common biomarkers. In particular, four biological markers, *i.e.* CRP, D-dimer, ferritin and IL-6, had the potential to predict the likelihood of severity in patients with COVID-19 with high accuracy (up to 96%). Our analytical model could still predict with 92% accuracy after removing IL-6 from the analysis to generalise the applicability of the model to hospitals with limited or no capacity for IL-6 testing.

In the present study, patients who died were more likely to be middle aged and with a pre-existing comorbidity. Our results are different from those published from large cohorts of patients hospitalised with COVID-19 in the UK, USA and Denmark. The median age of patients who died in our study was 58 years, which is much younger compared with patients included in studies from the UK (median 80 years) [21] and Denmark (median 81 years) [22]. Another remarkable difference relates to our findings of a greater proportion of deaths among female patients with pre-existing comorbidity. Differences are most likely due to a variety of factors, including regional differences concerning demographics, healthcare systems, preparedness and knowledge of COVID-19 at the time of the pandemic outbreak.

The pandemic caused by SARS-CoV-2 is still spreading over the world and has led to a tremendous burden on healthcare systems in developing countries such as Vietnam, where resources are scarce, including medical equipment and human resources. A simple and accurate prediction model based on critical clinical parameters to classify patients and allocate scarce resources efficiently is needed. This is especially important to identify vulnerable patients at risk of severe illness that may rapidly deteriorate, requiring timely more aggressive support to reduce mortality. Hence, our study results may be applied as

a)

Factor	Importance
IL-6	17.36
Ferritin	13.40
Dyspnoea	9.87
D-dimer	9.35
CRP	4.90
Cough	3.93
Fever	2.06
Hypertension	1.75
PLT	1.59
ALT	1.59

b)

Metric	Value
Sensitivity	0.92
Specificity	1.00
PPV	1.00
NPV	0.98
Precision	1.00
Recall	0.92
F1	0.96
Prevalence	0.24
Detection rate	0.22
Detection prevalence	0.22
Balanced accuracy	0.96

c)

Metric	Moderate	Severe-alive	Severe-dead
Sensitivity	1.00	0.91	0.69
Specificity	1.00	0.82	0.94
PPV	1.00	0.81	0.85
NPV	1.00	0.92	0.87
Precision	1.00	0.81	0.85
Recall	1.00	0.91	0.69
F1	1.00	0.86	0.76
Prevalence	0.24	0.45	0.31
Detection rate	0.24	0.41	0.22
Detection prevalence	0.24	0.51	0.25
Balanced accuracy	1.00	0.87	0.82

FIGURE 5 Prediction model metrics using 10 factors: interleukin (IL)-6, ferritin, dyspnoea, D-dimer, C-reactive protein (CRP), cough, fever, hypertension, platelets (PLT) and alanine transaminase (ALT). a) The 10 most important predictors of critical illness in patients with COVID-19 pneumonia identified by random forest analysis. b, c) Prediction model metrics using 10 factors for b) two groups (moderate and severe) and c) three groups (moderate, severe-alive and severe-dead). PPV: positive predictive value; NPV: negative predictive value.

a)

Factor	Importance
IL-6	30.99
Ferritin	24.61
D-dimer	13.27
CRP	6.16

	Moderate	Severe
Moderate	11	2
Severe	1	38

Metric	Value
Sensitivity	0.92
Specificity	0.95
PPV	0.85
NPV	0.97
Precision	0.85
Recall	0.92
F1	0.88
Prevalence	0.23
Detection rate	0.21
Detection prevalence	0.25
Balanced accuracy	0.93

b)

Factor	Importance
Ferritin	42.23
D-dimer	24.54
CRP	8.27

	Moderate	Severe
Moderate	9	1
Severe	3	39

Metric	Value
Sensitivity	0.75
Specificity	0.98
PPV	0.90
NPV	0.93
Precision	0.90
Recall	0.75
F1	0.82
Prevalence	0.23
Detection rate	0.17
Detection prevalence	0.19
Balanced accuracy	0.86

c)

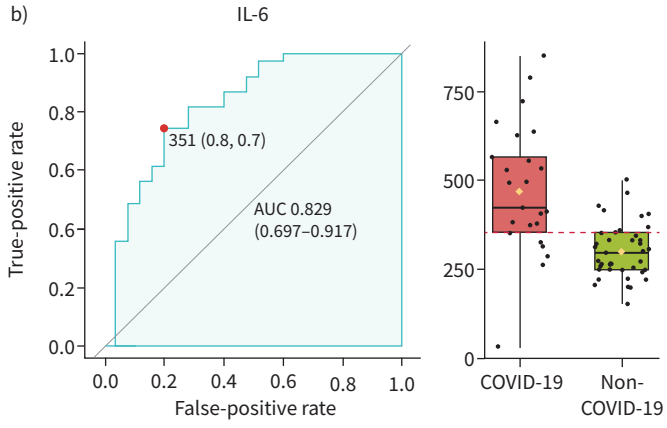
Factor	Importance
IL-6	37.57
Ferritin	27.24
D-dimer	10.23

	Moderate	Severe
Moderate	12	2
Severe	0	38

Metric	Value
Sensitivity	1.00
Specificity	0.95
PPV	0.86
NPV	1.00
Precision	0.86
Recall	1.00
F1	0.92
Prevalence	0.23
Detection rate	0.23
Detection prevalence	0.27
Balanced accuracy	0.98

FIGURE 6 a) Prediction model metrics using four factors (D-dimer, CRP, ferritin and interleukin (IL)-6), b) removal of IL-6 from analysis to generalise the applicability of the model to hospital settings and c) prediction model metrics using three factors (IL-6, D-dimer and ferritin). PPV: positive predictive value; NPV: negative predictive value.

helpful biomarkers building a simple prediction model in the early management of high-risk COVID-19 patients to improve prognosis and mortality rates. Despite the good results, the accuracy of the prediction model obtained from the validation cohort was not as high as the accuracy obtained from the investigation



a)

	Moderate	Severe
Moderate	0	19
Severe	0	106

c)

	Moderate	Severe
Moderate	0	15
Severe	0	110

d)

	Moderate	Severe
Moderate	0	50
Severe	0	75

FIGURE 7 a) Confusion matrix: four factors (C-reactive protein (CRP), D-dimer, ferritin and interleukin (IL)-6)). b) Receiver operating characteristic curve (ROC) for IL-6: hospitalised COVID-19 patient group *versus* non-COVID-19 patient group. The computed 95% confidence interval for IL-6 is highlighted in the shaded background over the ROC curve. The area under the ROC curve (AUC) is shown in red to highlight the diagnostic potential of IL-6. The box-and-whisker plots shown on the right side of the ROC curve plot revealed significantly higher levels of IL-6 in the COVID-19 group compared with the non-COVID-19 group. c) Confusion matrix: three factors (D-dimer, ferritin and IL-6). d) Confusion matrix: three factors (CRP, D-dimer and ferritin).

cohort (84.8% *versus* 96%). However, the results are sufficient and helpful in developing policies for prevention and responses to reduce critical adverse and poor outcomes of COVID-19.

Although the sPLS-DA model cannot differentiate severe-alive from severe-dead based on 12 biochemical markers, alternative biomarker sets can still be used to detect patients who will potentially develop the severe condition. The markers CRP, D-dimer, IL-6 and urea were associated with higher risk of death, while the combination of D-dimer, IL-6 and ferritin was linked to the likelihood of recovery in patients with severe COVID-19. Our study findings are consistent with those from same-size cohorts of COVID-19 patients hospitalised in China and Italy [23], and data reported in large meta-analyses and review papers [24, 25]. Ghahramani *et al*. [26] completed a meta-analysis in Asian populations looking at laboratory features of severe *versus* non-severe COVID-19 patients and identified a range of abnormal biomarkers that were elevated in severe patients, including high levels of neutrophils, ALT, aspartate transaminase, total bilirubin, blood urea nitrogen, creatinine, erythrocyte sedimentation rate, CRP, PCT, LDH, fibrinogen, prothrombin time, D-dimer, glucose and neutrophil-to-lymphocyte ratio, as well as reduced lymphocytes, monocytes and eosinophils, haemoglobin, platelets, albumin, serum sodium, and lymphocyte-to-CRP ratio, leukocyte-to-CRP ratio and leukocyte-to-IL-6 ratio. We were able to expand on these findings, reporting that several laboratory parameters including elevated CRP, D-dimer, ferritin, IL-6 and ALT were associated with the severe group. Albumin, platelets and urea were the only biomarkers associated with increased risk of patient death. Although this association did not reach statistical significance as the AUC was <0.7, these biomarkers might help in early triage selecting COVID-19 patients with high risk of death in whom more extensive investigations and closer monitoring would be indicated. Abnormal values of biomarkers reflecting organ damage, including urea and D-dimer, were associated with a higher risk of mortality. Coagulopathy in severe COVID-19 is known to be associated with high morbidity and mortality [27]. COVID-19 coagulopathy is generally characterised by increased D-dimer (a fibrin degradation product released when plasma cleaves cross-linked fibrin), increased LDH and mild thrombocytopenia [28]. Elevated D-dimer may in turn reflect respiratory failure among patients with COVID-19 related to microvascular thrombotic processes [29]. High levels of urea observed in this study indicate the impact of COVID-19 on renal functions. Blood urea nitrogen has been considered to be an independent risk factor for mortality in COVID-19 patients [30]. COVID-19 patients with higher risk of mortality had higher levels of CRP and IL-6, consistent with the known association between inflammatory pathways leading to the cytokine storm resulting from severe infection and causing fatal end-organ damage [28].

Strengths and limitations

The main strength of this study is the multicentre nature, with a discovery cohort from Vietnamese patients and a relative validation cohort from Danish patients. There are some limitations to this study that need to be considered when interpreting the results. The main limitation of the study is its observational design and the number of patients was relatively small to achieve a subgroup analysis. The timing of blood draws is also a possible confounding factor. Some patients from the Vietnamese cohort may have had laboratory biomarker assessment only at admission, but over time it may have provided more information about the prognosis of variation in D-dimer, ferritin and IL-6. Moreover, various confounding conditions may have affected the levels of laboratory biomarkers, including nutritional status, age, comorbidities and insufficiently documented medications.

Conclusions

Our study results demonstrated that CRP, D-dimer, ferritin and IL-6 were significantly correlated with the prognosis of severe COVID-19 patients. Furthermore, this study provided a simple and reliable model to predict clinical outcomes for COVID-19 patients in Vietnam. Two different sets of biomarkers are applicable for the assessment of disease severity and prognosis.

Provenance: Submitted article, peer reviewed.

Conflict of interest: No conflicts of interest exist for any of the authors.

References

1 Lu R, Zhao X, Li J, *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2020; 395: 565–574.
2 Callaway E. What omicron's BA.4 and BA.5 variants mean for the pandemic. *Nature* 2022; 606: 848–849.
3 Eyre DW, Taylor D, Purver M, *et al.* Effect of Covid-19 vaccination on transmission of alpha and delta variants. *N Engl J Med* 2022; 386: 744–756.

4 Gautret P, Million M, Jarrot PA, *et al.* Natural history of COVID-19 and therapeutic options. *Expert Rev Clin Immunol* 2020; 16: 1159–1184.
5 Tang N, Li D, Wang X, *et al.* Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost* 2020; 18: 844–847.
6 Berlin DA, Gulick RM, Martinez FJ. Severe Covid-19. *N Engl J Med* 2020; 383: 2451–2460.
7 Bivona G, Agnello L, Ciaccio M. Biomarkers for prognosis and treatment response in covid-19 patients. *Ann Lab Med* 2021; 41: 540–548.
8 Huang I, Pranata R, Lim MA, *et al.* C-reactive protein, procalcitonin, D-dimer, and ferritin in severe coronavirus disease-2019: a meta-analysis. *Ther Adv Respir Dis* 2020; 14: 1753466620937175.
9 Hu B, Huang S, Yin L. The cytokine storm and COVID-19. *J Med Virol* 2021; 93: 250–256.
10 McGonagle D, Sharif K, Regan AO, *et al.* The role of cytokines including IL-6 in COVID-19 induced pneumonia and MAS-like disease. *Autoimmun Rev* 2020; 19: 102537.
11 Heinrich PC, Castell JV, Andus T. Interleukin-6 and the acute phase response. *Biochem J* 1990; 265: 621–636.
12 Hunter CA, Jones SA. IL-6 as a keystone cytokine in health and disease. *Nat Immunol* 2015; 16: 448–457.
13 Chen R, Sang L, Jiang M, *et al.* Longitudinal hematologic and immunologic variations associated with the progression of COVID-19 patients in China. *J Allergy Clin Immunol* 2020; 146: 89–100.
14 Liu T, Zhang J, Yang Y, *et al.* The role of interleukin-6 in monitoring severe case of coronavirus disease 2019. *EMBO Mol Med* 2020; 12: e12421.
15 Linkins LA, Takach Lapner S. Review of D-dimer testing: good, bad, and ugly. *Int J Lab Hematol* 2017; 39: Suppl. 1, 98–103.
16 Rostami M, Mansouritorghabeh H. D-dimer level in COVID-19 infection: a systematic review. *Expert Rev Hematol* 2020; 13: 1265–1275.
17 Ponti G, Maccaferri M, Ruini C, *et al.* Biomarkers associated with COVID-19 disease progression. *Crit Rev Clin Lab Sci* 2020; 57: 389–399.
18 Lin Z, Long F, Yang Y, *et al.* Serum ferritin as an independent risk factor for severity in COVID-19 patients. *J Infect* 2020; 81: 647–679.
19 Schoggins J. Defective viral RNA sensing linked to severe COVID-19. *Science* 2021; 374: 534–535.
20 Breiman L. Random forests. *Mach Learn* 2001; 45: 5–32.
21 Martins-Filho PR, Tavares CSS, Santos VS. Factors associated with mortality in patients with COVID-19. A quantitative evidence synthesis of clinical and laboratory data. *Eur J Intern Med* 2020; 76: 97–99.
22 Hodges G, Pallisgaard J, Schjerning Olsen AM, *et al.* Association between biomarkers and COVID-19 severity and mortality: a nationwide Danish cohort study. *BMJ Open* 2020; 10: e041295.
23 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* 2020; 369: m1328.
24 Malik P, Patel U, Mehta D, *et al.* Biomarkers and outcomes of COVID-19 hospitalisations: systematic review and meta-analysis. *BMJ Evid Based Med* 2021; 26: 107–108.
25 Kermali M, Khalsa RK, Pillai K, *et al.* The role of biomarkers in diagnosis of COVID-19 – a systematic review. *Life Sci* 2020; 254: 117788.
26 Ghahramani S, Tabrizi R, Lankarani KB, *et al.* Laboratory features of severe vs. non-severe COVID-19 patients in Asian populations: a systematic review and meta-analysis. *Eur J Med Res* 2020; 25: 30.
27 Zhou F, Yu T, Du R, *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020; 395: 1054–1062.
28 Wang D, Hu B, Hu C, *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020; 323: 1061–1069.
29 Magro C, Mulvey JJ, Berlin D, *et al.* Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: a report of five cases. *Transl Res* 2020; 220: 1–13.
30 Cheng Y, Luo R, Wang K, *et al.* Kidney disease is associated with in-hospital death of patients with COVID-19. *Kidney Int* 2020; 97: 829–838.