Critical Care Explorations
June 2025 • Volume 7 • Number 6 • e1262

ORIGINAL CLINICAL REPORT

OPEN

Machine Learning Accurately Predicts Need for Critical Care Support in Patients Admitted to Hospital for Community-Acquired Pneumonia

OBJECTIVES: Hospitalized community-acquired pneumonia (CAP) patients are admitted for ventilation, vasopressors, and renal replacement therapy (RRT). This study aimed to develop a machine learning (ML) model that predicts the need for such interventions and compare its accuracy to that of logistic regression (LR).

DESIGN: This retrospective observational study trained separate models using random-forest classifier (RFC), support vector machines (SVMs), Extreme Gradient Boosting (XGBoost), and multilayer perceptron (MLP) to predict three endpoints: eventual use of invasive ventilation, vasopressors, and RRT during hospitalization. RFC-based models were overall most accurate in a derivation COVID-19 CAP cohort and were validated in one COVID-19 CAP and two non-COVID-19 CAP cohorts.

SETTING: This study is part of the Community-Acquired Pneumonia: Toward InnoVAtive Treatment (CAPTIVATE) Research program.

PATIENTS: Two thousand four hundred twenty COVID-19 and 1909 non-COVID-19 CAP patients over 18 years old hospitalized and not needing invasive ventilation, vasopressors, and RRT on the day of admission were included.

INTERVENTIONS: None.

MEASUREMENTS AND MAIN RESULTS: Performance was evaluated with area under the receiver operating characteristic curve (AUROC) and accuracy. RFCs performed better than XGBoost, SVM, and MLP models. For comparison, we evaluated LR models in the same cohorts. AUROC was very high ranging from 0.74 to 0.95 in predicting ventilation, vasopressors, and RRT use in our derivation and validation cohorts. ML used and variables such as FIO_2, Glasgow Coma Scale, and mean arterial pressure to predict ventilator, vasopressor use, creatinine, and potassium to predict RRT use. LR was less accurate than ML, with AUROC ranging 0.66 to 0.8.

CONCLUSIONS: A ML algorithm more accurately predicts need of invasive ventilation, vasopressors, or RRT in hospitalized non-COVID-19 CAP and COVID-19 patients than regression models and could augment clinician judgment for triage and care of hospitalized CAP patients.

KEYWORDS: community-acquired pneumonia; COVID-19; machine learning; renal replacement therapy; vasopressors; ventilation

George S. Chen, BSc[1]
Terry Lee, PhD[2]
Jennifer L.Y. Tsang, MD[3,4]
Alexandra Binnie, MD[5,6,7]
Anne McCarthy, MD[8]
Juthaporn Cowan, MD[8]
Patrick Archambault, MD[9]
Francois Lellouche, MD[10,11]
Alexis F. Turgeon, MD, MSc[10,11]
Jennifer Yoon, MD[12]
Francois Lamontagne, MD[13]
Allison McGeer, MD[14]
Josh Douglas, MD[15]
Peter Daley, MD[16]
Robert Fowler, MD[17]
David M. Maslove, MD[18]
Brent W. Winston, MD[19]
Todd C. Lee, MD[20]
Karen C. Tran, MD[21]
Matthew P. Cheng, MD[20]
Donald C. Vinh, MD[20]
John H. Boyd, MD[22,23]
Keith R. Walley, MD[22,23]
Joel Singer, PhD[2]
John C. Marshall, MD[24]
James A. Russell, MD[22,23]
for the Community-Acquired Pneumonia: Toward InnoVAtive Treatment (CAPTIVATE) Investigators

DOI: 10.1097/CCE.0000000000001262

Community-acquired pneumonia (CAP) causes 4 million deaths/yr (7% of global deaths). Severely ill patients are admitted to hospital for oxygen, IV antibiotics, adjunctive care (e.g., corticosteroids [1]), ventilation, vasopressors, and renal replacement therapy (RRT) (2).

Usually, it is clear when ICU admission is needed. However, many noncritically ill patients are admitted to a ward but later deteriorate and require urgent resuscitation and ICU admission. Deterioration may be rapid and unpredictable.

KEY POINTS

Question: Can machine learning (ML) accurately predict eventual need for ventilation, vasopressor, and renal replacement therapy in hospitalized COVID-19 and non-COVID-19 community-acquired pneumonia (CAP) patients on day of admission?

Findings: In this retrospective observational study, a random-forest classifier outperforms regression models in predicting eventual need for critical care intervention in hospitalized COVID-19 and CAP patients. Area under the receiver operating characteristic curve was very high (0.74–0.95) in predicting these interventions across both derivation and validation cohorts.

Meanings: An ML algorithm accurately predicts need for critical care intervention in hospitalized CAP and COVID-19 patients and could augment clinician judgment for triage and care of hospitalized CAP patients.

Prior studies identified predictors of deterioration (3–12) in COVID-19 CAP and non-COVID-19 CAP but are not in clinical use. Furthermore, the enormous amount of clinical and laboratory data makes it difficult for the human mind to predict critical care interventions. We reasoned that machine learning, a so-called artificial intelligence clinician, could accurately predict need for invasive ventilation, vasopressors, and RRT in CAP patients who are admitted to wards.

Accordingly, our hypothesis was that machine learning predicts later (i.e., after the day of admission) need for invasive ventilation, vasopressors, and RRT in patients hospitalized for CAP who do not need invasive ventilation, vasopressors, and RRT on the day of admission. We tested this hypothesis in: 1) derivation in acute COVID-19 CAP, 2) validation in acute COVID-19 CAP, and 3) additional validation in acute non-COVID-19 CAP. For comparison, we compared machine learning results with logistic regression models in the same cohorts.

METHODS

This is an observational study nested within the Community-Acquired Pneumonia: Toward InnoVAtive Treatment (CAPTIVATE) Research program—a multicenter, pan-Canadian cohort study of hospitalized patients older than 18 years old with acute CAP (13–19). In this nested observational study, we included patients enrolled between March 2018 and December 2023 in 16 Canadian hospitals.

Ethics

This study was approved by the Providence Health Care and University of British Columbia Human Research Committee (Approval number H20-00600) and by each of the contributing clinical sites on March 5, 2020. Anonymized clinical data and use of discarded plasma from clinical blood tests were deemed low risk and informed consent was deemed not necessary for this research. The ethics committee study title is Angiotensin Receptor Blockers Corona (ARBs CORONA). This study was approved by the institutional review board of each participating site (**Supplement Table 15**, https://links.lww.com/CCX/B512). All the sites in France are approved via the approval for Dr. Asfar at Angers as coordinating site for France. Procedures were followed in accordance with the Helsinki Declaration of 1975.

Inclusion and Exclusion Criteria

COVID-19 Patients. Inclusion criteria were individuals over 18 years old with confirmed COVID-19 infection by local hospital or provincial laboratories clinically approved laboratory severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) testing who were admitted to hospital for acute COVID-19. Acute COVID-19 was defined based on best evidence at the time (20–24) and site investigators judged that the admitting illness was consistent with a clinical presentation of acute COVID-19. We excluded acute COVID-19 readmissions, emergency department admissions without hospitalization, and those admitted to hospital with positive SARS-CoV-2 test but whose acute illness was not due to acute COVID-19.

The definition of CAP was defined prospectively in the protocol and was followed by each site. There was a training period when the coordinating center reviewed with each site the first two to three cases to be sure inclusion and exclusion criteria and key variables were recorded accurately and consistently. In addition, the coordinating center adjudicated and

reviewed each case as reported in the e-Case Report Form and validated that inclusion and exclusion were appropriate.

Non-COVID-19 CAP Patients. Non-COVID-19 CAP patients were included if they were admitted to the hospital for acute non-COVID-19 CAP and were older than 18 years old and had: one of fever, chills, leukocytosis, and leukopenia; one of: cough, sputum, and dyspnea; and new infiltrates on chest radiograph consistent with CAP. Acute non-COVID-19 CAP readmissions and emergency department visits without hospitalization were excluded.

Baseline Characteristics

Baseline characteristics included age, sex, and underlying comorbidities; vital signs (heart rate, respiratory rate, temperature, blood pressure, arterial oxygen saturation [Sao_2]); laboratory evaluation of organ function (serum creatinine, alanine transaminase, aspartate transaminase [AST], bilirubin, platelet count, WBC count, Glasgow Coma Scale [GCS] score); and use of vasopressors, invasive ventilation, and RRT.

Acute COVID-19 severity was based on a modified version of the Coronavirus Clinical Characterisation Consortium (4C) Mortality Score (25): age, sex, comorbidities, respiratory rate, oxygen saturation, GCS, and urea and C-reactive protein. GCS and C-reactive protein were excluded from our calculation of a modified 4C Mortality Score because data were not consistently captured for GCS and not at all for C-reactive protein. Definition of comorbidities were as predefined in CAPTIVATE (26) rather than by the Charlson Comorbidity Index. Urea was not captured in CAPTIVATE so serum creatinine level was used to measure renal function as follows: normal: less than 110 μmol/L; moderate: 110–220 μmol/L; and more than moderate: greater than 220 μmol/L.

Outcomes

The co-primary outcomes were use of invasive mechanical ventilation, vasopressors, and RRT after the day of admission at any later time during hospitalization.

Data Sources

Patients were identified prospectively at the sites and data were collected by CAPTIVATE Research Coordinators on electronic Case Report Forms. Baseline data was data within 24 hours of admission. Quebec sites were unable to recruit patents in wave 3 due to research coordinator shortages. Random samples of 15% of the records were reviewed for accuracy by the data monitoring team. There were few concerns regarding the quality of data and any missing data was requested and included in the database.

Statistical Analyses

No formal sample size calculation was performed as this was a substudy of ARBs CORONA I. The initial planned sample size of ARBs CORONA I was 497 (27); enrollment was later continued in this open cohort as research funding increased. Baseline characteristics were compared using chi-square test, Fisher exact test (when > 20% of cells have expected cell counts < 5 or any expected cell count is < 1), analysis of variance, or Kruskal-Wallis test. Analyses were conducted in Python using scikit-learn for metrics, pandas for correlations, and scipy/statsmodels for statistical tests. Figures were generated with Microsoft Excel (Microsoft Corporation, Redmond, WA) and Matplotlib (28).

Machine Learning: Prediction Endpoints and Cohorts

Machine learning models were trained to predict the need for vasopressor, ventilator, and RRT after hospital admission (day 0) in each cohort. We used a derivation cohort of COVID-19 patients (n = 1210) and three validation cohorts: a second COVID-19 cohort (n = 1210) and two non-COVID-19 CAP cohorts (n = 1184, n = 725) (**Supplement Table 1**, https://links.lww.com/CCX/B512). The derivation cohort included the first 1210 COVID-19 patients, while the COVID-19 validation cohort comprised 1210 patients that were subsequently enrolled. The two non-COVID-19 CAP cohorts were similarly non-COVID-19 CAP patients enrolled early and later.

Data Preprocessing

Data preprocessing included column and row processing, imputation, normalization, class balancing, feature selection, and feature engineering (**Supplement Fig. 1** and **Supplement Text File**, https://links.lww.com/CCX/B512).

Feature Selection. We removed text-based, nonmedical (e.g., patient identification), sparse (> 75% missing), zero variance, and ICU admission features. We grouped similar features (**Supplement Table 3**, https://links.lww.com/CCX/B512). All features collected after day of admission were removed. Accordingly, any repeated measurements (e.g., laboratory work, vital signs, etc.) taken after admission were also excluded from model training (Supplement Fig. 1 and Supplement Text File, https://links.lww.com/CCX/B512).

Patient Selection. COVID-19 negative patients and patients with unknown outcomes (n = 95) were excluded (Supplement Fig. 1, https://links.lww.com/CCX/B512). For RRT prediction, an additional 12 patients with unknown RRT status were excluded (**Fig. 1**). Ward deaths without ICU admission were classified as experiencing ventilation, vasopressor, and RRT outcomes, assuming that critical care interventions would have been beneficial.

Our cohorts included data on patients with chronic kidney disease (CKD), but not end-stage renal disease (ESRD). CKD patients were included in the prediction of ventilation, vasopressors, and RRT because the minority of CKD patients are in ESRD and on dialysis at hospital admission for COVID-19 or non-COVID-19 CAP.

Data Normalization and Imputation. Features were normalized with MinMaxScaler from scikit-learn. Missing troponin and D-dimer values were imputed as the mean normal values respectively, as we assume that missingness indicated these tests were not ordered due to low clinical suspicion for elevation. Additionally, both troponin and D-dimer are highly right-skewed. All other missing values were imputed using K-Nearest Neighbors (Supplement Fig. 1, https://links.lww.com/CCX/B512) and KNNImputer from scikit-learn (29) (**Supplement Table 4**, https://links.lww.com/CCX/B512).

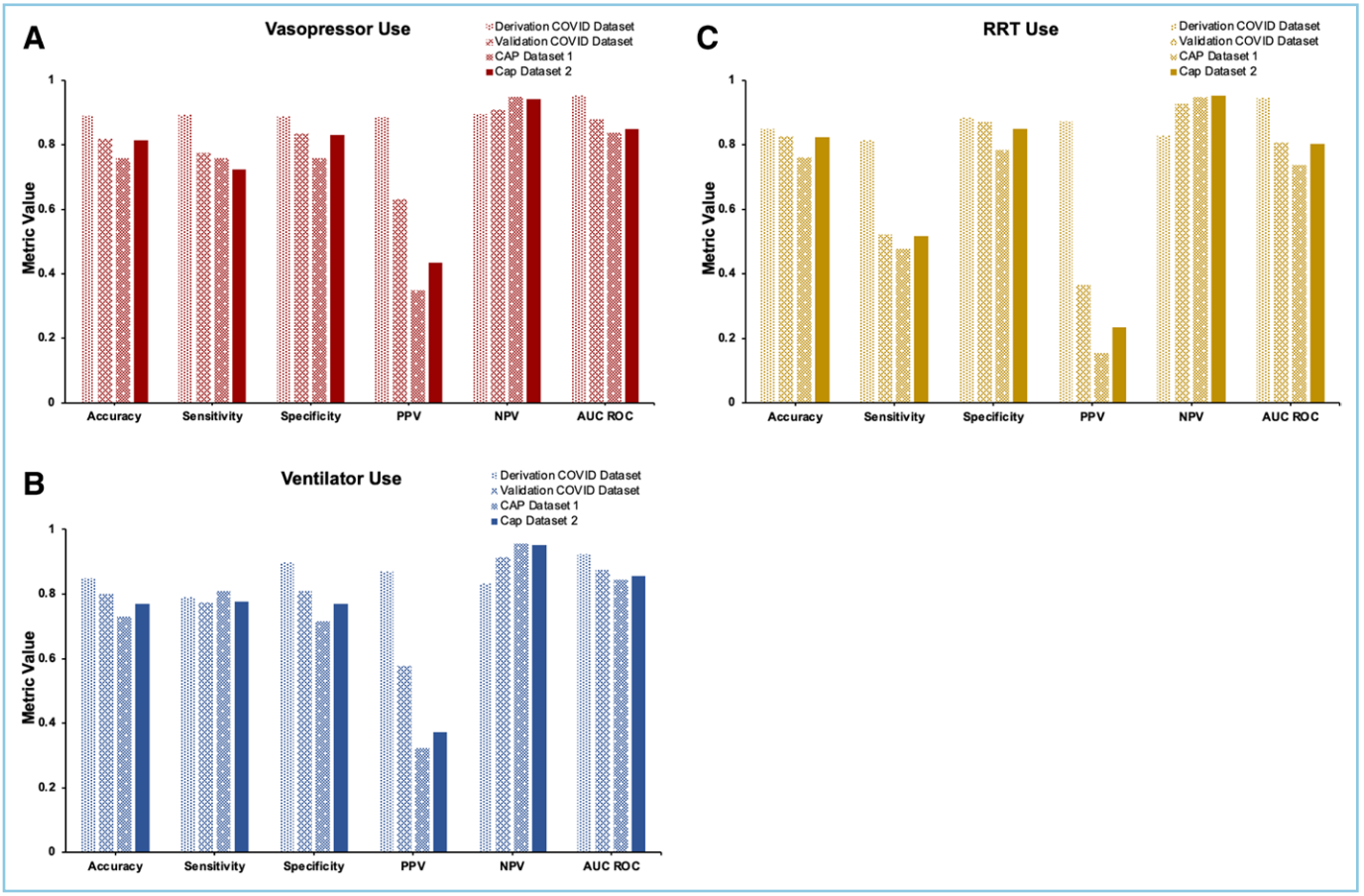

Figure 1. Performance metrics for models predicting critical care intervention needs for COVID-19 and community-acquired pneumonia (CAP) patients. Accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUC ROC) metrics for our model that predicts vasopressor use (**A**), ventilator use (**B**), and renal replacement therapy (RRT) use (**C**). Results are shown for model performance against the COVID-19 cohort used to train the models, an unseen COVID-19 cohort, and a CAP cohort. The derivation cohort has been class-balanced, whereas other cohorts have not.

Class Balancing. Class balancing was achieved through a combination of oversampling and undersampling. This enabled models to better learn from the minority outcomes and reduces bias toward the majority outcome. Class balancing was done as shown in Supplement Text File (https://links.lww.com/CCX/B512).

Model Training

The derivation cohort (n = 1210) was split 80/20 for training (n = 968) and testing (n = 242), respectively. We trained models (Supplement Text File, https://links.lww.com/CCX/B512) using random-forest classifier (RFC), support vector machine, XGBoost, and multilayer perceptron models. We performed hyperparameter optimization with GridSearchCV from scikit-learn (29) (**Supplement Table 5**, https://links.lww.com/CCX/B512) for our final models (**Supplement Table 6**, https://links.lww.com/CCX/B512).

External Validation and Model Generalizability. A validation cohort of COVID-19 pneumonia patients (n = 1210) was used for external validation. These patients were admitted later than patients from the derivation COVID-19 cohort and served as a temporal validation (i.e., to determine whether the model was accurate in later COVID-19 patients who could differ by wave, treatment, and vaccine status, e.g.). We validated model generalizability in two validation cohorts of non-COVID-19 CAP patients (n = 1184, n = 725). All cohorts were preprocessed identically and as in Supplement Text File (https://links.lww.com/CCX/B512). Validation cohorts were left unbalanced to reflect the models' performance in real-world clinical conditions, where the natural distribution of outcomes in the unbalanced dataset determines performance.

To build generalizability and reproducibility confidence, we did additional randomized testing of our COVID-19 and non-COVID-19 CAP cohorts. We randomized all COVID-19 patients into derivation and validation cohorts to create "new" COVID-19 cohorts and similarly randomized all non-COVID-19 CAP patients into two validation cohorts to create "new" non-COVID-19 CAP cohorts as follows. Original COVID cohorts (derivation and validation) were merged and randomly split into new derivation and validation cohorts. A new RFC model was trained on this new derivation cohort and tested with the new validation cohort. The same process was performed with the two original CAP validation cohorts.

Misclassification Analysis. McNemar's test was used to determine whether there were significant differences in how often the models made false-positive and false-negative errors with the models' predictions. The Mann-Whitney U test assessed model confidence in making correct vs. incorrect predictions and differences in feature distributions between cohorts.

Logistic Regression Models. Logistic regression is one of the most widely used methods to create prediction models. We applied the most used strategy in the literature to build such a model and compared logistic regression accuracy to machine learning accuracy. The univariate relationship between baseline characteristics and outcomes (vasopressor use, ventilator use, and RRT use) was first examined by chi-square test, Fisher exact test, t test, or Wilcoxon rank-sum test as appropriate. Given the large number of available features, univariate analysis was first performed to select the candidate variables for the prediction model. Factors with p value of less than 0.2 in the univariate analysis were included in the stepwise regression as candidate variables (bidirectional selection with null model as the starting model) with Akaike Information Criterion (AIC) to select the set of variables to be included in the prediction model. The optimal classification threshold for the prediction model was the point closest to the top-left part of the plot with perfect sensitivity or specificity.

Code Access

Code for this project is available on GitHub: https://github.com/ubcgchen/AI-COVID-CAP.

RESULTS

The derivation COVID-19 patients differed from the COVID-19 validation patient cohort (Supplement Table 1, https://links.lww.com/CCX/B512). Derivation COVID-19 cohort patients had more and different comorbidities (e.g., more chronic cardiac, kidney, rheumatologic and hematologic disease, and malignancy, but less dementia), were admitted to ICU less frequently, and had lower rates of invasive ventilation and vasopressors at admission day than validation COVID-19 cohort patients.

Of the machine learning models tested, RFC-based models overall had the highest accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve (AUROC) (**Fig. 2**; and **Supplement Table 2**, https://links.lww.com/CCX/B512). For each prediction, our RFC model outputs a probability score between 0 and 1 that indicates the confidence that the patient will require ventilation, vasopressor(s), or RRT, whereas the complement of this score reflects the model's confidence that the patient will not require these interventions.

To compare results of prediction across the derivation and validation cohorts, we describe accuracy and AUROC (other statistics are in Figs. 1 and 2; and **Supplement Tables 7** and **8**, https://links.lww.com/CCX/B512). Our RFC model demonstrated accuracy of 0.89 (95% CI, 0.867–0.874; range, 0.823–0.906), 0.85 (95% CI, 0.849–0.856; range, 0.810–0.891), and 0.85 (95% CI, 0.872–0.881; range, 0.806–0.917) and AUROC of 0.95 (95% CI, 0.936–0.941; range, 0.902–0.963), 0.92 (95% CI, 0.929–0.934; range, 0.886–0.958), and 0.94 (95% CI, 0.954–0.959; range, 0.924–0.982) for vasopressor, ventilator, and RRT use, respectively (Figs. 1 and 2; and Supplement Tables 7 and 8, https://links.lww.com/CCX/B512) in our derivation COVID-19 cohort.

In our validation COVID-19 cohort, accuracy scores were 0.82, 0.80, and 0.83 and AUROC scores were 0.88, 0.88, and 0.81 for vasopressor, ventilator, and RRT use, respectively. Model accuracy was stratified by whether patients received the intervention on the day of admission. For vasopressor use, accuracy was 0.97 for patients who received the intervention on day 0 and 0.81 for all other patients. For ventilator use, accuracy was 0.91 and 0.79, respectively, and for RRT use, accuracy was 0.73 and 0.83, respectively.

For the first non-COVID-19 CAP cohort, accuracy scores were 0.76, 0.73, and 0.76 and AUROC scores were 0.84, 0.84, and 0.74 for vasopressor, ventilator, and RRT use, respectively. Stratified by day of admission intervention, accuracy for vasopressor use was 0.84 for patients who received the intervention on day 0 and 0.75 for those who did not. For ventilator use, accuracy was 0.93 and 0.72, respectively, and for RRT use, accuracy was 0.86 and 0.76, respectively.

Finally, for the second non-COVID-19 CAP cohort, accuracy scores were 0.81, 0.77, and 0.82 and AUROC scores were 0.85, 0.86, and 0.80 for vasopressor, ventilator, and RRT use, respectively. Stratified by day of admission intervention, accuracy for vasopressor use was 0.81 for both patients who received the intervention on day 0 and those who did not. For ventilator use, accuracy was 0.96 and 0.76, respectively, and for RRT use, accuracy was 1.00 and 0.82, respectively.

We repeated the analysis after randomly creating new cohorts. Retraining RFC models with a random half of our COVID-19 data and validating with the other COVID-19 half continued to show high accuracy (0.81–0.92) and AUROC (0.78–0.98) across derivation and validation cohorts. Similarly, within our non-COVID-19 CAP, the randomized halves had high accuracy (0.86–0.92) and AUROC (0.79–0.97) (**Supplement Table 9**, https://links.lww.com/CCX/B512).

Features That Were Important Predictors of Ventilation, Vasopressor, and RRT

Certain features were more important in predicting ventilation, vasopressors, and RRT (**Supplement Fig. 2** and **Supplement Tables 10** and **11**, https://links.lww.com/CCX/B512). Features that most predicted vasopressor use included Fio_2 (r = 0.41), mean arterial pressure (MAP; r = –0.35), and GCS (r = –0.31); vasopressor use patients had higher bilirubin, respiratory rate, Fio_2, and AST; and lower MAP and Sao_2 compared with those never on vasopressors (Mann-Whitney $p = 0.04$, 2.38×10^{-19}, 1.89×10^{-23}, 0.004, 7.44×10^{-12}, 1.43×10^{-24}, respectively; **Supplement Fig. 3*A***, https://links.lww.com/CCX/B512).

The features that predicted ventilator use were also Fio_2, MAP, and GCS (r = 0.44, –0.32, and –0.44, respectively). Patients who required ventilation had significantly higher Fio_2, AST, respiratory rate, and WBC count, and lower MAP and Sao_2 than those who did not ($p = 2.10 \times 10^{-25}$, 0.002, 5.06×10^{-18}, 3.49×10^{-14}, 3.21×10^{-10}, 8.00×10^{-29}, respectively; **Supplement Fig. 3*B***, https://links.lww.com/CCX/B512).

GCS, serum creatinine, and troponin were most predictive of RRT use (r = –0.21, 0.34, and –0.003, respectively; Supplement Tables 8 and 9, https://links.lww.com/CCX/B512). In patients who required RRT, day 0 creatinine, potassium, respiratory rate, and troponin were significantly higher while hemoglobin and Sao_2 were significantly lower than patients not ever on RRT ($p = 5.65 \times 10^{-17}$, 5.86×10^{-6}, 0.022, 1.53×10^{-6},

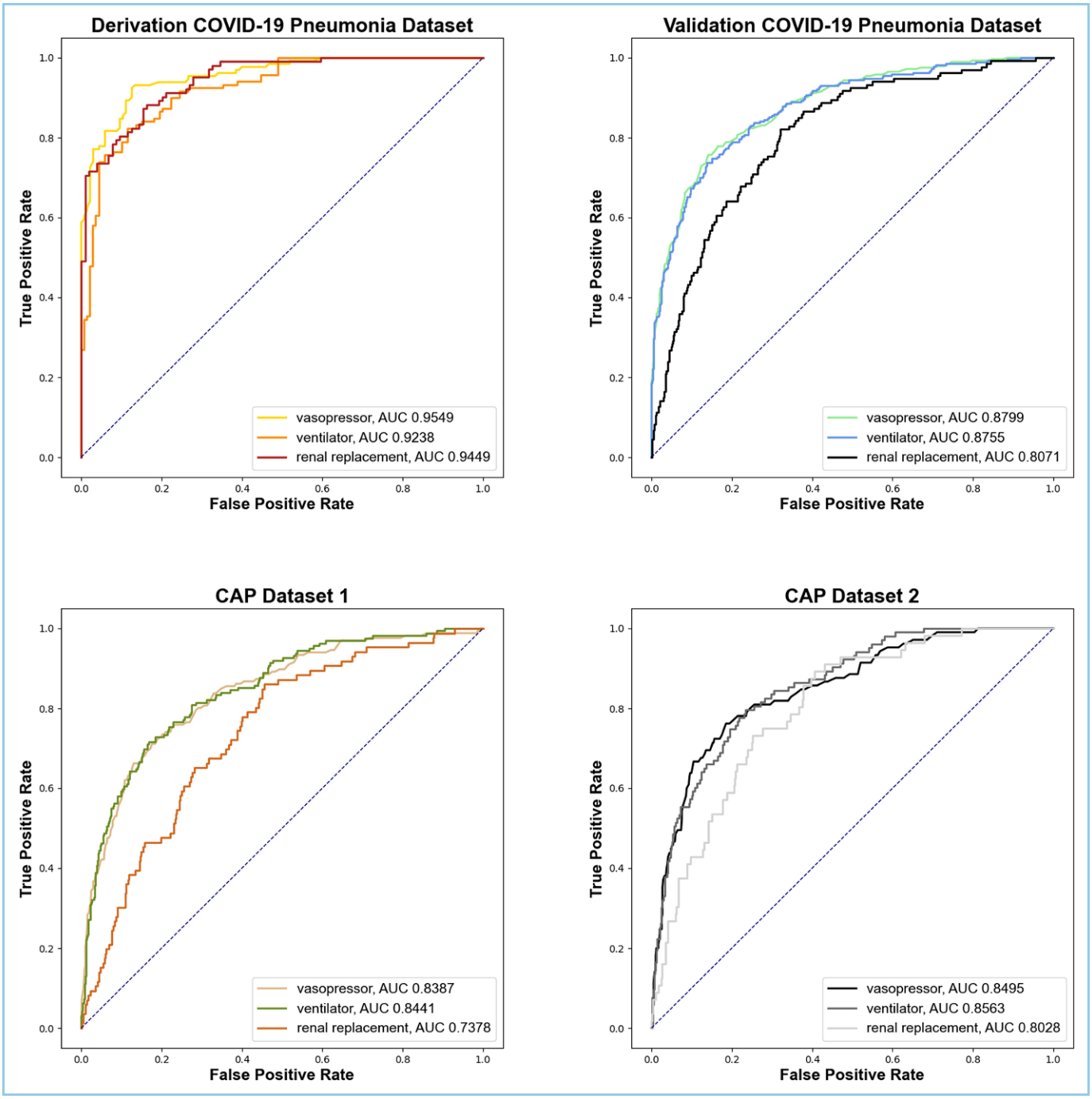

Figure 2. Area under the curve (AUC) for vasopressor, ventilator, and renal replacement use for our original COVID-19 cohort, unseen COVID-19 cohort, the first community-acquired pneumonia (CAP) cohort, and the second CAP cohort.

2.15×10^{-9}, 0.025, respectively; **Supplement Fig. 3*C***, https://links.lww.com/CCX/B512).

Misclassification Analysis

Our models demonstrated significantly different confidence levels in correctly vs. incorrectly classified patients for all endpoints in the derivation COVID-19 dataset ($p = 3.80 \times 10^{-13}$, 1.24×10^{-11}, and 7.95×10^{-5} for vasopressor, ventilator, and RRT). The Mann-Whitney *U* test confirms a statistical difference between the confidence scores; combining it with visual inspection suggests that the model is significantly more confident in its correct predictions than its incorrect ones. (Supplement Fig. 4 and Supplement Text File, https://links.lww.com/CCX/B512).

Logistic Regression Model Prediction of Ventilation, Vasopressors, and RRT

As a standard widely used to assess accuracy of prediction, we used a logistic regression model and logistic regression was less accurate (e.g., AUROC for invasive ventilation, vasopressors, and RRT: COVID-19 derivation: 0.77, 0.74, and 0.8; COVID-19 validation: 0.66, 0.65, and 0.77; non-COVID-19 CAP first cohort: 0.78, 0.75, and 0.71; and non-COVID-19 CAP second cohort: 0.73, 0.71, and 0.73, respectively) (**Table 1**; and Supplement Text File, https://links.lww.com/CCX/B512) than machine learning.

Randomizing the COVID-19 and non-COVID-19 CAP cohorts also showed lower accuracy of logistic regression (AUROC: ventilation, vasopressors, and RRT: randomized COVID-19 derivation: 0.76, 0.76, and 0.83; randomized COVID-19 validation: 0.71, 0.71, and 0.73; randomized non-COVID-19 CAP derivation: 0.78, 0.76, and 0.80; and randomized non-COVID-19 CAP validation: 0.77, 0.78, and 0.76, respectively) (**Supplement Table 14**, https://links.lww.com/CCX/B512) than machine learning.

DISCUSSION

In patients who were admitted to hospital for COVID-19 or non-COVID-19 CAP, a machine learning algorithm more accurately (all above 80%) predicted use of invasive ventilation, vasopressor, and RRT in patients not on this support on hospital admission day compared with conventional logistic regression.

A machine learning algorithm such as we propose could be used clinically in future by several teams: by emergency department staff for triage, by ward staff for evaluation of risk of ICU admission, and by Critical Care Outreach teams for evaluation of CAP patients admitted to medical wards from emergency

TABLE 1.
Metrics by Endpoint in Derivation Set COVID-19, Validation Set COVID-19, and Two Non-COVID-19 Community-Acquired Pneumonia Datasets for the Logistic Regression Prediction Model Based on Stepwise Logistic Regression With Akaike Information Criterion

Endpoint	Metric	Original COVID-19	Unseen COVID-19	CAP 1	CAP 2
Vasopressor use	Accuracy	0.698	0.601	0.698	0.609
	Sensitivity	0.682	0.671	0.689	0.727
	Specificity	0.703	0.588	0.699	0.602
	PPV	0.440	0.236	0.201	0.096
	NPV	0.866	0.904	0.953	0.097
	AUC ROC	0.741	0.648	0.752	0.709
Ventilator use	Accuracy	0.726	0.635	0.697	0.685
	Sensitivity	0.670	0.646	0.745	0.710
	Specificity	0.746	0.633	0.692	0.684
	PPV	0.494	0.230	0.203	0.108
	NPV	0.859	0.914	0.963	0.978
	AUC ROC	0.771	0.660	0.776	0.728
Renal replacement therapy use	Accuracy	0.705	0.676	0.661	0.680
	Sensitivity	0.778	0.741	0.692	0.600
	Specificity	0.702	0.674	0.660	0.681
	PPV	0.102	0.066	0.048	0.015
	NPV	0.986	0.988	0.989	0.995
	AUC ROC	0.802	0.765	0.708	0.608

AUC ROC = area under the receiver operating characteristic curve, CAP = community-acquired pneumonia, NPV = negative predictive value, PPV = positive predictive value.

department who are at high risk of need for ventilation, vasopressors, or RRT. Of note, the accuracy of prediction of use of invasive ventilation, vasopressors, and RRT was very high, consistent, and very similar in different types of CAP, that is, COVID-19 (caused by a single virus, i.e., SARS-CoV-2-induced CAP) and non-COVID-19 CAP (caused by a variety of bacteria and viruses). As expected, the machine learning accuracy was highest in the derivation COVID-19 cohort (> 0.85) but remained very high (0.80–0.85) in a COVID-19 and two non-COVID-19 CAP validation cohorts.

A logistic regression model was less accurate than machine learning in predicting critical care support in two COVID-19 and two non-COVID-19 CAP cohorts. Reasons could include different independent variables and weights/odds ratios used in machine learning vs. logistic regression (**Supplement Tables 12, 13**, and 16, https://links.lww.com/CCX/B512), a stronger iterative process in machine learning, and the low frequency of some events such as RRT that could have adversely affected logistic regression more than machine learning. Also, we chose the most widely used strategy for variable selection when building the logistic prediction model, univariate analysis, and stepwise selection based on AIC. Other more advanced variable selection strategies are available, such as genetic algorithm and Least Absolute Shrinkage and Selection Operator, which might improve accuracy so future work is required to explore these options.

The machine learning algorithm identified clinically logical features for prediction of critical care support but combined them in the model in a way that could augment human clinician prediction by including many more and many seemingly less important features in its prediction. Interestingly, the features for prediction using only hospital day of admission were similar for invasive ventilation, vasopressor, and RRT use.

Our study is novel; it is the first to use machine learning to predict: 1) invasive ventilation, vasopressor, and RRT use and 2) both COVID-19 and non-COVID-19 CAP. A few studies compared machine learning in COVID-19 and non-COVID-19 CAP (e.g., COVID-19 CAP vs. influenza CAP [8]). Other machine learning techniques were used to predict COVID-19 CAP mortality (3, 5, 6, 11), COVID-19 mortality and severity (7), RRT (12), and ventilation use in COVID-19 (4, 10). Previous studies sometimes had higher sample sizes, but we used many more admission day variables (i.e., features) for the machine learning than prior studies.

Our study strengths included consistent recruitment of both non-COVID-19 CAP and COVID-19 wave patients into our multicenter pan-Canadian open cohort (13–15, 17–19) and very detailed phenotyping using a standardized case report form. The main reason for hospital and ICU admission for CAP is for oxygen, IV antibiotics (2), corticosteroids (1, 30, 31), and invasive ventilation, vasopressor, or RRT so we evaluated the prediction of use of invasive ventilation, vasopressor, or RRT. Our model is robust and continued to perform well when retrained with a randomized half of the COVID-19 and randomized validation COVID-19 and non-COVID-19 CAP cohorts. Finally, we compared machine learning to logistic regression prediction in the same cohorts.

We acknowledge that standard logistic regression is a significant limitation of the study.

We acknowledge that this is a potentially inferior model development method to compare with machine learning approaches, such that the latter performed better. A recent study demonstrated that using advanced variable selection strategies could lead to similarly accurate prediction models as machine learning approaches (32). Limitations also include that we did not record advanced directives that could have modified decisions about ventilator, vasopressor, or RRT use. In this observational study, we could not determine causation, but add evidence regarding use of machine learning to accurately predict use of invasive ventilation, vasopressor, or RRT in COVID-19 and in non-COVID-19 CAP. The event rates of invasive ventilation, vasopressors, and RRT were lower in the validation than derivation cohorts, and this may have adversely affected accuracy and AUROC of the validation cohorts. Comorbidities were identified at time of admission, but the duration of comorbidities was not recorded nor was treatment and level of control of comorbidities all of which could alter use of ventilation, vasopressors, or RRT. We did not record COVID-19 therapies (e.g., anticoagulants and immunomodulatory drugs) that could have altered invasive ventilation, vasopressor, or RRT use. We did not evaluate capture machine learning performance in subgroups, such as by ethnicity due to the data sparsity: over 60% of patients did not have ethnicity recorded in the derivation cohort,

and greater than 70% in the validation COVID and non-COVID CAP cohorts. Most ethnicities had less than 50 data points available (Supplement Table 1, https://links.lww.com/CCX/B512). We educated the sites to not include patients who came from a healthcare facility or nursing home, but it is possible that we included such patients so may have included some patients with healthcare-acquired pneumonia. Finally, we did not record caregiver predictions of use of ventilator, vasopressor, or RRT for comparison with machine learning and logistic regression predictions.

CONCLUSIONS

A machine learning algorithm predicts use of invasive ventilation, vasopressors, or RRT in non-COVID-19 CAP and COVID-19 patients not on these supports on the day of hospital admission. Logistic regression approaches without advanced variable selection strategies might be inferior to machine learning methods. Implemented appropriately, both could augment clinician judgment for triage and care of patients admitted to hospital for non-COVID-19 CAP.

1 *University of British Columbia, Vancouver, BC, Canada.*

2 *Centre for Advancing Health Outcomes, St. Paul's Hospital, University of British Columbia, Vancouver, BC, Canada.*

3 *Critical Care Medicine, Niagara Health Knowledge Institute, St Catharines, ON, Canada.*

4 *Critical Care Medicine, McMaster University, Hamilton, ON, Canada.*

5 *Critical Care Department, William Osler Health System, Brampton, ON, Canada.*

6 *Critical Care Medicine, Algarve Biomedical Centre, Faro, Portugal.*

7 *Critical Care Medicine, Centro Hospitalar Universitário do Algarve, Faro, Portugal.*

8 *Infectious Disease, Ottawa Research Institute, University of Ottawa, Ottawa, ON, Canada.*

9 *St. George Hospital, Levis, QC, Canada.*

10 *CHU de Québec-Université Laval Research Center, Population Health and Optimal Health Practices Unit, Trauma- Emergency- Critical Care Medicine, Québec City, QC, Canada.*

11 *Department of Anesthesiology and Critical Care Medicine, Division of Critical Care Medicine, Faculty of Medicine, Université Laval, Québec City, QC, Canada.*

12 *Critical Care Medicine, Humber River Hospital, Toronto, ON, Canada.*

13 *Critical Care Medicine, University of Sherbrooke, Sherbrooke, QC, Canada.*

14 *Mt. Sinai Hospital, University of Toronto, Toronto, ON, Canada.*

15 *Critical Care Medicine, Lion's Gate Hospital, North Vancouver, BC, Canada.*

16 *Infectious Disease, Memorial University of Newfoundland, St. John's, NL, Canada.*

17 *Critical Care Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada.*

18 *Department of Critical Care, Kingston General Hospital and Queen's University, Kingston, ON, Canada.*

19 *Departments of Critical Care Medicine, Medicine and Biochemistry and Molecular Biology, Foothills Medical Centre, University of Calgary, Calgary, AB, Canada.*

20 *Division of Infectious Disease, McGill University, Montreal, QC, Canada.*

21 *Division of General Internal Medicine, Vancouver General Hospital, Vancouver, BC, Canada.*

22 *Centre for Heart Lung Innovation, St. Paul's Hospital, University of British Columbia, Vancouver, BC, Canada.*

23 *Division of Critical Care Medicine, St. Paul's Hospital, University of British Columbia, Vancouver, BC, Canada.*

24 *Department of Surgery, St. Michael's Hospital, Toronto, ON, Canada.*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (https://journals.lww.com/ccejournal).

Mr. Chen was responsible for construction, validation, and statistical analysis of the machine learning model. He was also involved in writing the article. Dr. Lee was responsible for the logistic regression and other statistical analyses and maintaining the database. Dr. Russell is the supervising author, obtained funding, and writing the article. All other authors recruited patients at their sites and supervised research staff, and all contributed to the writing of the article.

This study was supported by generous grants to James A. Russell from the Canadian Institutes of Health Research and St. Paul's Foundation.

Dr. Russell reports patents owned by the University of British Columbia that are related to: 1) the use of proprotein convertase subtilisin/kexin type 9 inhibitor(s) in sepsis, 2) the use of vasopressin in septic shock, and 3) a patent owned by Ferring for use of selepressin in septic shock. Dr. Russell is an inventor on these patents. Dr. Russell is a co-founder, director, and shareholder of Resolve Nanotherapeutics. Dr. Russell was a co-founder, director, and shareholder in Cyon Therapeutics (now closed) and is a shareholder in Molecular You Corp. Dr. Russell is the Senior Research Advisor of the British Columbia, Canada Post-COVID-19 Interdisciplinary Clinical Care Network. Dr. Russell is no longer consulting for any industry. Dr. Russell reports receiving consulting fees in the last 3 years as a funded member of the Data and Safety Monitoring Board of an National Institutes of Health-sponsored trial of plasma in COVID-19 (Passive Immunity Trial for our Nation; PASS-IT-ON) (2020–2021). Since 2021, Dr. Russell has received grants for COVID-19

and for community-acquired pneumonia research: six from the Canadian Institutes of Health Research, three from the St. Paul's Foundation, one from the Rix Family Foundation, and one from the Canada's Immunoengineering and Biomanufacturing Hub (CIEBH) and the Canada Biomedical Research Fund (CBRF). Dr. Russell was a nonfunded Science Advisor and member, Government of Canada COVID-19 Therapeutics Task Force (June 2020–2021). The remaining authors have disclosed that they do not have any potential conflicts of interest.

For information regarding this article, E-mail: gschen@student.ubc.ca

REFERENCES

1. Dequin PF, Meziani F, Quenot JP, et al: Hydrocortisone in severe community-acquired pneumonia. *N Engl J Med* 2023; 388:1931–1941
2. Metlay JP, Waterer GW, Long AC, et al: Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med* 2019; 200:e45–e67
3. Barough SS, Safavi-Naini SAA, Siavoshi F, et al: Generalizable machine learning approach for COVID-19 mortality risk prediction using on-admission clinical and laboratory features. *Sci Rep* 2023; 13:2399
4. Chen R, Chen J, Yang S, et al: Prediction of prognosis in COVID-19 patients using machine learning: A systematic review and meta-analysis. *Int J Med Inform* 2023; 177:105151
5. Cilloniz C, Ward L, Mogensen ML, et al: Machine-learning model for mortality prediction in patients with community-acquired pneumonia: Development and validation study. *Chest* 2023; 163:77–88
6. Jeon ET, Lee HJ, Park TY, et al: Machine learning-based prediction of in-ICU mortality in pneumonia patients. *Sci Rep* 2023; 13:11527
7. Kamran F, Tang S, Otles E, et al: Early identification of patients admitted to hospital for COVID-19 at risk of clinical deterioration: Model development and multisite external validation study. *BMJ* 2022; 376:e068576
8. Park D, Jang R, Chung MJ, et al: Development and validation of a hybrid deep learning-machine learning approach for severity assessment of COVID-19 and other pneumonias. *Sci Rep* 2023; 13:13420
9. Patel MA, Knauer MJ, Nicholson M, et al: Organ and cell-specific biomarkers of long-COVID-19 identified with targeted proteomics and machine learning. *Mol Med* 2023; 29:26
10. Raman G, Ashraf B, Demir YK, et al: Machine learning prediction for COVID-19 disease severity at hospital admission. *BMC Med Inform Decis Mak* 2023; 23:46
11. Wang B, Li Y, Tian Y, et al: Novel pneumonia score based on a machine learning model for predicting mortality in pneumonia patients on admission to the intensive care unit. *Respir Med* 2023; 217:107363
12. Xu Y, Trivedi A, Becker N, et al: Machine learning-based derivation and external validation of a tool to predict death and development of organ failure in hospitalized patients with COVID-19. *Sci Rep* 2022; 12:16913
13. Best JR, Wang M, Lee T, et al: Early increases in anti-SARS-CoV-2 antibody isotypes associated with organ dysfunction and mortality in patients hospitalized with COVID-19. *Intensive Care Med* 2022; 48:616–618
14. Lee T, Cheng MP, Vinh DC, et al; ARBs CORONA I study investigators: Organ dysfunction and death in patients admitted to hospital with COVID-19 in pandemic waves 1 to 3 in British Columbia, Ontario and Quebec, Canada: A cohort study. *CMAJ Open* 2022; 10:E379–E389
15. Lee T, Cheng MP, Vinh DC, et al; for ARBs CORONA I: Outcomes and characteristics of patients hospitalized for COVID-19 in British Columbia, Ontario and Quebec during the Omicron wave. *CMAJ Open* 2023; 11:E672–E683
16. Lee T, Walley KR, Boyd JH, et al: Impact of the COVID-19 pandemic on non-COVID-19 community-acquired pneumonia: A retrospective cohort study. *BMJ Open Respir Res* 2023; 10:e001810
17. Mohammed Y, Goodlett DR, Cheng MP, et al; ARBs CORONA I: Longitudinal plasma proteomics analysis reveals novel candidate biomarkers in acute COVID-19. *J Proteome Res* 2022; 21:975–992
18. Mohammed Y, Tran K, Carlsten C, et al; ARBs CORONA I Consortium: Evolution of proteomics from acute to post-COVID-19 condition. *J Proteome Res* 2023; 23:52–70
19. Rocheleau GLY, Lee T, Mohammed Y, et al: Renin-angiotensin system pathway therapeutics associated with improved outcomes in males hospitalized with COVID-19. *Crit Care Med* 2022;50:1306–1317
20. Huang C, Wang Y, Li X, et al: Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; 395:497–506
21. Murthy S, Archambault PM, Atique A, et al; SPRINT-SARI Canada Investigators and the Canadian Critical Care Trials Group: Characteristics and outcomes of patients with COVID-19 admitted to hospital and intensive care in the first phase of the pandemic in Canada: A national cohort study. *CMAJ Open* 2021; 9:E181–E188
22. Piroth L, Cottenet J, Mariet AS, et al: Comparison of the characteristics, morbidity, and mortality of COVID-19 and seasonal influenza: A nationwide, population-based retrospective cohort study. *Lancet Respir Med* 2021; 9:251–259
23. 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
24. 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
25. Knight SR, Ho A, Pius R, et al; ISARIC4C investigators: Risk stratification of patients admitted to hospital with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol: Development and validation of the 4C Mortality Score. *BMJ* 2020; 370:m3339
26. Rocheleau GLY, Lee T, Mohammed Y, et al; for ARBs CORONA I Investigators: Renin-angiotensin system pathway therapeutics associated with improved outcomes in males hospitalized with COVID-19. *Crit Care Med* 2022; 50:1306–1317

27. Russell JA, Marshall JC, Slutsky A, et al: Study protocol for a multicentre, prospective cohort study of the association of angiotensin II type 1 receptor blockers on outcomes of coronavirus infection. *BMJ Open* 2020; 10:e040768
28. Hunter JD: Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 2008; 9:90–95.
29. Pedregosa F, Varoquaux G, Gramfort A, et al: Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011; 12:2825–2830.
30. Horby P, Lim WS, Emberson JR, et al; RECOVERY Collaborative Group: Dexamethasone in hospitalized patients with COVID-19. *N Engl J Med* 2021; 384:693–704
31. Tomazini BM, Maia IS, Cavalcanti AB, et al; COALITION COVID-19 Brazil III Investigators: Effect of dexamethasone on days alive and ventilator-free in patients with moderate or severe acute respiratory distress syndrome and COVID-19: The CoDEX randomized clinical trial. *JAMA* 2020; 324:1307–1316
32. Villar J, González-Martín JM, Hernández-González J, et al; Predicting Outcome and STratifiCation of severity in ARDS (POSTCARDS) Network: Predicting ICU mortality in acute respiratory distress syndrome patients using machine learning: The Predicting Outcome and STratifiCation of severity in ARDS (POSTCARDS) study. *Crit Care Med* 2023; 51:1638–1649