Soaring cases of coronavirus disease (COVID-19) are pummeling the global health system. Overwhelmed health facilities have endeavored to mitigate the pandemic, but mortality of COVID-19 continues to increase. Here, we present a mortality risk prediction model for COVID-19 (MRPMC) that uses patients’ clinical data on admission to stratify patients by mortality risk, which enables prediction of physiological deterioration and death up to 20 days in advance. This ensemble model is built using four machine learning methods including Logistic Regression, Support Vector Machine, Gradient Boosted Decision Tree, and Neural Network. We validate MRPMC in an internal validation cohort and two external validation cohorts, where it achieves an AUC of 0.9621 (95% CI: 0.9464–0.9778), 0.9760 (0.9613–0.9906), and 0.9246 (0.8763–0.9729), respectively. This model enables expeditious and accurate mortality risk stratification of patients with COVID-19, and potentially facilitates more responsive health systems that are conducive to high risk COVID-19 patients.
Management of the surging infections of the coronavirus disease (COVID-19) is a huge clinical challenge. Currently, the pandemic is pummeling the global health system, with 18,902,735 people infected as of August 7, 2020. The overwhelmed health facilities are unable to curb the increasing mortality of COVID-19. Moreover, without proven effective treatments to date, patients who rapidly deteriorate into a refractory state harbor significantly higher risks of death. Third, advanced COVID-19 is characterized by heterogeneous clinical features and multiorgan damage, which requires an effective triage and intensive monitoring. Therefore, an early warning system that enables stratification of COVID-19 patients by risk of death on admission holds enormous promise to assist in the management of COVID-19.

Electronic health records (EHRs) abound with valuable information generated from routine clinical practices, which can be useful for mortality risk prediction of COVID-19. However, data in EHRs are complex, multidimensional, nonlinear, and heterogeneous. Using models more effective than traditional statistical methods (univariate or multivariate Cox regressions and logistic regression (LR)) for analysis can help to fully utilize the clinical data in EHRs. Machine learning (ML), a subfield of artificial intelligence, encapsulates statistical and mathematical algorithms that enable facts interrogation and complex decision-making. Therefore, combinatory uses of ML algorithms and EHRs for prognosis prediction in the context of COVID-19 pandemic are worth exploring.

ML algorithms have been explored in myriad fields of COVID-19 including, but not limited to, detecting outbreaks, identification and classification of COVID-19 medical images, rapid diagnosis, severity risk prediction, and prognosis prediction. For COVID-19 patients and clinicians, the greatest concern is whether the patients can survive. Available ML models that focus on this exhibit promising prognostic implications, but are still impeded by the paucity of external validations and limited follow-ups, and lack the capability of predicting prognosis as early as the time of admission.

In this study, we aim to develop a mortality risk prediction model for COVID-19 (MRPMC) that utilizes clinical data in EHRs to stratify patients by mortality risk on admission. The validated capability of enabling expeditious and accurate mortality risk stratification of COVID-19 may facilitate more responsive health systems that are conducive to high-risk COVID-19 patients via early identification, and ensuing instant intervention as well as intensive care and monitoring, thus, hopefully assisting to save lives during the pandemic.

Results

Study design and baseline characteristics. To train and validate the MRMPC for prognosis prediction of COVID-19, we included 2520 consecutive COVID-19 patients with known outcomes (discharge or death) from two affiliated hospitals of Tongji Medical College, Huazhong University of Science and Technology, including Sino-French New City Campus of Tongji Hospital (SF) and Optical Valley Campus of Tongji Hospital (OV), and The Central Hospital of Wuhan (CHWH) between January 27, 2020 and March 21, 2020. As a total of 360 patients were excluded with definite reasons, 2160 COVID-19 patients met eligibilities. For detailed exclusions, see Fig. 1 and “Methods,” participants. We randomly partitioned 50 and 50% of participants from SF into the training cohort (SFT cohort) and internal validation cohort (SFV cohort), respectively. Participants from OV and CHWH were used as two external validation cohorts (OV cohort and CHWH cohort). Compositions of the four cohorts are displayed in Fig. 1 and “Methods,” cohorts. The study design has been schematically presented in Fig. 1 and Supplementary Fig. 1.

Table 1 shows the baseline characteristics of the four cohorts. The median age of the participants was 62 years (interquartile range [IQR]: 51–71) in the SFT cohort, 63 years (IQR: 51–70) in the SFV cohort, 63 years (IQR: 50–70) in the OV cohort, and 62.5 years (IQR: 55–72) in the CHWH cohort. The male patients accounted for 50.7, 50.0, 46.7, and 54.3% of all participants in the SFT, SFV, OV, and CHWH cohorts, respectively. Hypertension (37.1–40.3%) was the most prevalent comorbidity and fever (61.2–86.0%) remained the most common symptom. The median time from admission to death or discharge ranged from 17 to 23 days among all four cohorts.

Features selected by least absolute shrinkage and selection operator (LASSO). Among 53 raw features extracted from EHRs (Supplementary Table 1), those with a proportion of missing values greater than or equal to 5% in each cohort were filtered (Supplementary Fig. 2), resulting in 34 features, including 18

---

**Fig. 1 Study design.** MRMPC mortality risk prediction model for COVID-19, SFT training cohort of Sino-French New City Campus of Tongji Hospital, SFV internal validation cohort of Sino-French New City Campus of Tongji Hospital, OV Optical Valley Campus of Tongji Hospital, CHWH The Central Hospital of Wuhan.
Table 1 Baseline characteristics of individuals by cohort.

| Characteristics | SFT cohort \((n = 621)\) | SFV cohort \((n = 622)\) | OV cohort \((n = 801)\) | CHWH cohort \((n = 116)\) |
|----------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Age            | 62 (51–71)               | 63 (51–70)               | 63 (50–70)               | 62.5 (55–72)             |
| Sex            |                          |                          |                          |                          |
| Female         | 306 (49.3%)              | 311 (50.0%)              | 427 (53.3%)              | 53 (45.7%)              |
| Male           | 315 (50.7%)              | 311 (50.0%)              | 374 (46.7%)              | 63 (54.3%)              |
| Comorbidity number | 1 (0–2)          | 1 (0–2)                  | 1 (0–2)                  | 2 (1–3)                  |
| Comorbidity    |                          |                          |                          |                          |
| Hypertension   | 245 (39.5%)              | 244 (39.2%)              | 321 (40.3%)              | 43 (37.1%)              |
| Diabetes       | 110 (17.7%)              | 110 (17.7%)              | 121 (15.2%)              | 16 (13.8%)              |
| CHD            | 72 (11.6%)               | 59 (9.5%)                | 68 (8.5%)                | 16 (13.8%)              |
| CLD            | 26 (4.2%)                | 19 (3.1%)                | 33 (4.1%)                | 7 (6.0%)                |
| Tumor          | 22 (3.5%)                | 21 (3.4%)                | 20 (2.5%)                | 5 (4.4%)                |
| HBV            | 16 (2.6%)                | 13 (2.1%)                | 24 (3.0%)                | 1 (1.0%)                |
| CKD            | 13 (2.1%)                | 8 (1.3%)                 | 11 (1.4%)                | 1 (0.9%)                |
| COPD           | 4 (0.6%)                 | 7 (1.1%)                 | 7 (0.9%)                 | 1 (0.9%)                |
| Fever          | 533 (86.0%)              | 527 (84.9%)              | 584 (73.0%)              | 71 (61.2%)              |
| Temp \((\text{max}) ≥ 39°C\) | 169 (27.4%)               | 194 (31.5%)               | 158 (26.0%)               | 18 (8.6%)               |
| Cough          | 450 (72.6%)              | 436 (70.2%)              | 601 (75.1%)              | 63 (54.3%)              |
| Dyspnea        | 313 (50.5%)              | 283 (45.6%)              | 274 (34.2%)              | 37 (31.9%)              |
| Sputum         | 233 (37.6%)              | 228 (36.7%)              | 344 (43.0%)              | 32 (27.6%)              |
| Fatigue        | 253 (40.8%)              | 233 (37.5%)              | 250 (31.2%)              | 43 (37.1%)              |
| Diarrhea       | 186 (30.0%)              | 167 (26.9%)              | 135 (19.6%)              | 9 (7.8%)                |
| Myalgia        | 133 (21.5%)              | 144 (23.2%)              | 129 (16.1%)              | 20 (17.2%)              |
| Vomiting       | 30 (4.8%)                | 31 (5.0%)                | 32 (4.0%)                | 3 (2.6%)                |
| Conscious at admission | 595 (95.8%)          | 600 (96.5%)              | 786 (98.1%)              | 79 (68.1%)              |
| Respiratory rate, per min | 20 (20–22)            | 21 (20–24)               | 21 (20–24)               | 21 (20–24)              |
| MAP, mmHg      | 96.7 (88.7–104.7)        | 97.2 (89.7–105.6)        | 96.3 (87.7–106.7)        | 93.3 (86.9–101.5)       |
| SpO2, %        | 95 (91–97)               | 95 (91–97)               | 96 (94–97)               | 95.5 (93–97.3)          |
| Vital status   |                          |                          |                          |                          |
| Death          | 86 (13.8%)               | 89 (14.3%)               | 60 (7.5%)                | 16 (16.4%)              |
| Discharge      | 535 (86.2%)              | 533 (85.7%)              | 741 (92.5%)              | 97 (83.6%)              |
| Follow-up, days | 23 (15–30)             | 21 (15–29)               | 19 (14–26)               | 17 (12–24)              |

Continuous variables are presented as median (interquartile ranges [IQR]), while categorical variables as counts and percentages (%).

SFT cohort training cohort of Six-French New City Campus of Tongji Hospital, SFV cohort internal validation cohort of Six-French New City Campus of Tongji Hospital, OV cohort external validation cohort of Optical Valley Campus of Tongji Hospital, CHWH cohort external validation cohort of The Central Hospital of Wuhan. Follow-up time from admission to death or discharge, CHD coronary heart disease, CLD chronic liver disease, HBV hepatitis B virus, CKD chronic kidney disease, COPD chronic obstructive pulmonary disease, MAP mean arterial pressure.

categorical features and 16 continuous ones (Supplementary Fig. 3 and 4) that underwent feature selection by the LASSO (Fig. 2a). Only 14 of the 34 features were eventually chosen for modeling (Fig. 2b), among which 8 features had a positive association with mortality (high risk: consciousness, male sex, sputum, blood urea nitrogen [BUN], respiratory rate [RR], D—dimer, number of comorbidities, and age) and 6 features were negatively correlated with mortality (low risk: platelet count [PLT], fever, albumin [ALB], SpO2, lymphocyte, and chronic kidney disease [CKD]). Multivariable Cox analysis using raw data of the 34 features proved that the features selected by LASSO exhibited similar prognostic implications (Supplementary Fig. 5 and Supplementary Table 2). High-risk features identified by LASSO were also significant unfavorable prognostic indicators recognized via multivariable Cox analysis (hazard ratio [HR] > 1 and \(p < 0.05\)). Similarly, low-risk features accorded with favorable prognostic indicators (HR < 1 and \(p < 0.05\)).

Model performance. In general, six ML models including LR, support vector machine (SVM), gradient boosted decision tree (GBDT), neural network (NN), K-nearest neighbor (KNN), and random forest (RF) all displayed varying but promising performances to predict mortality risk in the three validation cohorts in terms of discrimination and calibration. To build a predictive model with augmented prognostic implications, we integrated the top four best predictive models (LR, SVM, GBDT, and NN) to create an ensemble model called MRPMC. MRPMC outputted a normalized probability of mortality risk ranging from 0 to 1. We selected the threshold of 0.6 to assign the predicted mortality risk label by optimizing F1 score on the training cohort (Supplementary Fig. 6). Probabilities of less than 0.6 were assigned to low risk and otherwise to high risk for all ML methods across all cohorts. The procedures of establishing the MRPMC are elaborated in Methods, Model development. As expected, MRPMC exhibited greater capability of predicting mortality risk of COVID-19 than the four contributive models alone in the SFV and CHWH cohorts, though the differences between SVM and MRPMC were nuanced in the SFV cohort (Fig. 3a–c). MRPMC achieved an area under the receiver operating characteristics (ROC) curve (AUC) of 0.9621 (95% confidence interval [CI]: 0.9464–0.9778) in identification of nonsurvivors with an accuracy of 92.4% (95% CI: 90.1–94.4%) in SFV cohort. For OV cohort, MRPMC demonstrated an AUC of 0.9760 (95% CI: 0.9613–0.9906) and an accuracy of 95.5% (95% CI: 93.8–96.8%) to predict prognosis of COVID-19. An AUC of 0.9906 (95% CI: 0.9872–0.9940) and an accuracy of 96.8% (95% CI: 95.2–98.4%) for prognosis prediction were observed for CHWH cohort (Table 2). The calibration curve of MRPMC in the three validation cohorts are depicted in Supplementary Fig. 7, showing that MRPMC displayed a Brier score of 0.051 for SFV.
Performances of four contributing algorithms are listed in Table 2, and that of the other two ML models (KNN and RF) in Supplementary Fig. 8 and Supplementary Table 3. Moreover, with the time from admission to death or discharge as the endpoint, Kaplan–Meier analysis further confirmed that MRPMC could robustly stratify patients by mortality risk. High-risk COVID-19 patients labeled by MRPMC were significantly less likely to survive than low-risk patients in the SFV, OV, and CHWH validation cohorts (Fig. 3d; p < 0.0001) with an HR of 41.42, 32.83 (95% CI: 19.70–54.70), and 12.81 (95% CI: 5.09–32.24), respectively, highlighting the capability of MRPMC to accurately predict prognosis of COVID-19.

Analyzing features included in models. Eight continuous features included in MRPMC exhibited correlation to varying degrees (Fig. 4a). Relative importance rank of all 14 variables for mortality prediction in MRPMC and the four contributive models are illustrated in Fig. 4b and Supplementary Table 4. The top weighted features (elevated D-dimer, decreased SpO2, increased RR, and lymphocytopenia) coincided with previously reported risk factors that were highly correlated with poor outcome in COVID-19. Standard box plots presented all differential continuous variables between survivors and nonsurvivors (Fig. 4c). Nonsurvivors had significantly (p < 0.001) advanced age, higher levels of BUN and D-dimer, and lower levels of SpO2, lymphocyte, ALB, and PLT (Fig. 4c and Supplementary Table 5). These findings were also parallel to risk factors of mortality of COVID-19 delineated previously, indicating that the selected features were highly relevant to prognosis.

Discussion
In this multicenter retrospective study, we built the MRPMC, an ensemble model derived from four ML algorithms (LR, SVM, GBDT, and NN), that enabled accurate prediction of physiological deterioration and death for COVID-19 patients up to 20 days in advance using clinical information in EHRs on admission, and validated it both internally and externally. Importantly, the MRPMC displayed an AUC ranging from 0.9186 to 0.9762 in the three validation cohorts. The prognostic implications of MRPMC might facilitate more responsive health systems that are conducive to high-risk COVID-19 patients via early identification, and ensuing instant intervention as well as intensive care and monitoring, thus, hopefully assisting to save lives during the pandemic.

Generalizability was the first advantage of MRPMC. Initially, the SFV and OV cohorts comprised patients from two designated campuses for COVID-19, where 40 top-level medical teams across China collaborated to eradicate the crisis. Patients in the CHWH cohort were treated in a general hospital. Therefore, medical records on admission were more comprehensive in SFV and OV cohorts than in CHWH, and the treatments that patients received throughout hospitalization were more parallel between SFV and OV cohorts. Second, 44% of participants in CHWH cohort were COVID-19 patients with malignancy who were more vulnerable to COVID-19 and less likely to survive than non-cancerous COVID-19 patients. Validation of MRPMC in an external validation cohort with heterogenous baseline characteristics. Importantly, although the
settings of the validation cohorts varied, MRPMC exhibited an AUC of 0.9186 (95% CI: 0.8686–0.9687) to identify high-risk patients in the CHWH cohort, indicating that the prognostic implications of MRPMC were not confined to cohorts similar to SFT, but could also be successfully validated in an inhomogeneous cohort.

Strengths of MRPMC also include its stability and practicability upon COVID-19 patients with several missing features. To begin with, the 14 features for prognosis prediction were readily accessible and frequently monitored in routine clinical practice. Age and sex were basic information. Fever, sputum, and consciousness were easily observed symptoms, while RR and SpO2 were physical signs available at hand. Presence of CKD and number of comorbidities could be ascertained by referring to previous EHRs and patients or their family doctors. PLT, BUN, D-dimer, ALB, and lymphocytes were low-cost laboratory tests and conveniently determined. Unlike self-reported symptoms, these features were relatively more objective and solid, and less susceptible to memory bias. Though the 14 features were readily accessible, we appreciated the differences in medical procedures and uneven distribution of medical resources among different regions, countries, and continents. The missing features may thwart those who imminently need MRPMC. Importantly, with the imputation method we adopted (see Methods), MRPMC could still perform well in patients with several missing features.

In addition, MRPMC had certain interpretability. Features contributing to mortality risk prediction in this study were tangible and many of them had been proven intimately correlated with mortality in COVID-19 patients. Advanced age, male sex, and presence of multiple comorbidities were identified as risk factors associated with death in COVID-19 patients. Sputum, supraphysiologic RR, and decreased SpO2 were directly related to pulmonary abnormalities in COVID-19. Elevated BUN, increased D-dimer, and lymphocytopenia might indicate extrapulmonary disorders and were potentially correlated with multiorgan damage caused by COVID-19.
|                              | AUC (95% CI) | Accuracy (95% CI) | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | F1       | Kappa    | Brier   |
|------------------------------|--------------|-------------------|----------------------|----------------------|--------------|--------------|----------|----------|---------|
| **Internal validation cohort (SFV)** |              |                    |                      |                       |              |              |          |          |         |
| MRPMC                        | 0.9621       | (0.9464-0.9778)   | (90.1-94.4%)         | (46.4-67.7%)          | 98.3%        | (96.8-99.2%) | 85.0%    | 93.2%    | 0.685   |
| SVM                          | 0.9594       | (0.9424-0.9764)   | (90.1-94.4%)         | (49.8-70.9%)          | 97.8%        | (96.1-98.8%) | 81.8%    | 93.7%    | 0.697   |
| GBDT                         | 0.9454       | (0.9246-0.9662)   | (89.0-93.6%)         | (49.8-70.9%)          | 96.6%        | (94.7-98.0%) | 75.0%    | 93.6%    | 0.696   |
| LR                           | 0.9614       | (0.9456-0.9772)   | (89.7-94.1%)         | (45.3-66.7%)          | 98.1%        | (96.6-99.1%) | 83.3%    | 93.1%    | 0.671   |
| NN                           | 0.9615       | (0.9456-0.9774)   | (89.7-94.1%)         | (40.8-62.4%)          | 98.9%        | (97.6-99.6%) | 88.5%    | 92.5%    | 0.653   |
| **External validation cohort (OV)** |              |                    |                      |                       |              |              |          |          |         |
| MRPMC                        | 0.9760       | (0.9613-0.9906)   | (93.8-96.8%)         | (32.1-58.4%)          | 99.6%        | (98.8-99.9%) | 90.0%    | 95.7%    | 0.600   |
| SVM                          | 0.9774       | (0.9640-0.9908)   | (94.1-97.0%)         | (36.8-63.2%)          | 99.5%        | (98.6-99.9%) | 88.2%    | 96.1%    | 0.638   |
| GBDT                         | 0.9536       | (0.9279-0.9793)   | (93.0-96.2%)         | (35.2-61.6%)          | 98.5%        | (97.4-99.3%) | 72.5%    | 95.9%    | 0.580   |
| LR                           | 0.9721       | (0.9568-0.9875)   | (93.7-96.7%)         | (32.1-58.4%)          | 99.5%        | (98.6-99.9%) | 87.1%    | 95.7%    | 0.593   |
| NN                           | 0.9754       | (0.9602-0.9906)   | (94.0-96.9%)         | (33.7-60.0%)          | 99.6%        | (98.8-99.9%) | 90.3%    | 95.8%    | 0.615   |
| **External validation cohort (CHWH)** |              |                    |                      |                       |              |              |          |          |         |
| MRPMC                        | 0.9246       | (0.8763-0.9729)   | (80.6-93.2%)         | (20.3-66.5%)          | 96.9%        | (91.2-99.4%) | 72.7%    | 89.5%    | 0.533   |
| SVM                          | 0.9067       | (0.8482-0.9652)   | (81.6-93.9%)         | (33.5-79.8%)          | 94.6%        | (88.4-98.3%) | 68.8%    | 92.0%    | 0.629   |
| GBDT                         | 0.9021       | (0.8347-0.9694)   | (80.6-93.2%)         | (12.6-56.6%)          | 99.0%        | (94.4-100.0%) | 85.7%    | 88.1%    | 0.462   |
| LR                           | 0.9231       | (0.8710-0.9717)   | (79.6-92.6%)         | (16.3-61.6%)          | 96.9%        | (91.2-99.4%) | 70.0%    | 88.7%    | 0.483   |
| NN                           | 0.9202       | (0.8700-0.9705)   | (81.6-93.9%)         | (24.5-71.1%)          | 96.9%        | (91.2-99.4%) | 75.0%    | 90.4%    | 0.581   |

SFV, internal validation cohort of Sino-French New City Campus of Tongji Hospital, OV Optical Valley Campus of Tongji Hospital, CHWH The Central Hospital of Wuhan; MRPMC mortality risk prediction model for COVID-19; SVM support vector machine; GBDT gradient boosted decision tree; LR logistic regression; NN neural network; AUC area under the receiver operating characteristics curve; PPV positive predictive value; NPV negative predictive value; 95% CI 95% confidence interval.
Available ML-based studies on prognosis prediction of COVID-19 patients are impeded by limited sample size, category of variables for prediction, short-term follow-ups for outcomes, and paucity of independent external validation. To overcome these obstacles, we included 2520 consecutive inpatients with definite outcomes and detailed baseline characteristics within a specific time period for training and multiple validations of MRPMC to avoid overfitting and ensure general applicability, reproducibility, and credibility. Meanwhile, the features contributing to prognosis prediction were collected and proposed by a multidisciplinary team including experienced clinicians, epidemiologists, and informaticians, which guaranteed the representativeness of features. Importantly, time from admission to death or discharge was 21 (IQR: 15–29) days, 19 (IQR: 14–26) days, and 17 (IQR: 12–24) days in the SFV, OV, and CHWH validation cohorts, respectively. As MRPMC displayed impressive AUCs to predict mortality risk in the validation cohorts, it could predict death ~20 days in advance. Last, since the characteristics of datasets could affect the validity of the classification strategies of ML algorithms, we proposed an ensemble model derived from four ML algorithms for more accurate prediction of mortality risk in COVID-19 patients.

Although most cases of COVID-19 are not life-threatening, those that underwent physiological deterioration harbored significantly higher mortality (49.0% for critically ill patients versus 2.3% for overall patients). As the pandemic causes more infections, our understandings of the risk factors for mortality and the role that supportive, targeted, and immunological therapies play in treating COVID-19 continue to improve. The aim of developing MRPMC is to mitigate the huge burden derived from COVID-19 on global health system and help to optimize clinical decision makings. MRPMC could automatically identify patients having high mortality risk as early as the time of admission when related symptoms are mild and nonspecific. This
group of patients needs intensive monitoring and instant treat-
ment when unfavorable prognostic indicators are observed, thus,
hopefully improving patient outcomes. However, multiple eva-
uations of MRPMC in larger cohorts, prospective settings, and
clinical trials are needed before elucidating its contribution to
improving outcome of COVID-1913,14.

This study had some limitations. Patients included were pri-
marily local residents from Wuhan, China. The predictive per-
formance of the ML models merits investigation in other regions
and ethnicities. Besides, the prognostic implications of MRPMC
have not been evaluated in prospective cohorts due to the re-
trospective nature of this study.

In conclusion, combinatorial applications of MRPMC and
EHRs with readily available features can enable timely and
accurate risk stratification of COVID-19 patients on admission.
MRPMC can potentially assist clinicians to promptly target
the high-risk patients on admission, and accurately predict physio-
logical deterioration and death up to 20 days in advance.

Methods
Participants. We included 2520 consecutive COVID-19 patients with known
dates of death and discharge (or death) from two affiliated hospitals of Tongji Medical
College, Huazhong University of Science and Technology (Sino–New City
Campus of Tongji Hospital, SF and Optical Valley Campus of Tongji Hospital,
OV) and The Central Hospital of Wuhan (CHWH) between January 27, 2020 and
March 21, 2020. A total of 360 patients were excluded for various reasons,
including 72 patients who failed to accord with the defined diagnosis of COVID-19
in the 7th edition of the Diagnosis and Treatment Protocol of COVID-19 released
by the National Health Commission of China10, 217 patients who were transferred
from Fangcang shelter hospitals for isolation, 33 patients who died within 24 h of
admission, and 38 patients who were under 18 years of age, were pregnant, or were
re-hospitalized or discharged for special reasons such as dialysis (Fig. 4). Eventu-
ally, 2160 patients were included for model training and validations.

Cohorts. We randomly partitioned 50 and 50% of participants from SF into
training cohort (SFT cohort) and internal validation cohort (SFV cohort),
respectively. Participants from OV and CHWH were used as two external vali-
dation cohorts (OV cohort and CHWH cohort). Specifically, as Fig. 1 indicates,
SFT cohort comprised 621 patients (535 survivors and 86 nonsurvivors); SFV, 622
patients (533 survivors and 89 nonsurvivors); OV, 801 patients (741 survivors and
60 nonsurvivors); and CHWH, 116 patients (97 survivors and 19 nonsurvivors). Patients
with malignancy were reportedly more susceptible and vulnerable to
COVID-19 owing to their immunocompromised states caused by the cancer itself,
cachexia, and cancer treatments31. They were also less likely to survive than non-
cancerous COVID-19 patients16-19, making COVID-19 patients with cancer an
intriguing group of population for prognosis prediction. To investigate the cap-
ability of ML models to predict prognosis in this population, we consecutively
included 54 malignant COVID-19 patients from the Cancer Center of CHWH and
62 cancerous COVID-19 patients from the Department of Respiratory of
CHWH to constitute another external validation cohort. The detailed baseline
characteristics of the cohorts are shown in Table 1.

Ethics. This study was approved by the Research Ethics Commission of Tongji
Medical College, Huazhong University of Science and Technology (TJ-
IRB202000406) with waived informed consent by the Ethics Commission mentioned
above. This study was part of the observational clinical trial titled “A retrospective
study for evolution and clinical outcomes study of novel coronavirus pneumonia
(COVID-19) patients,” which was registered in the Chinese Clinical Trial Registry
(ChiCTR2000032161). The clinical trial partly aimed to investigate the independent
risk factors for adverse outcomes of COVID-19. The detailed information can be
accessed in http://www.chictr.org.cn/showproj.aspx?proj=52561.

Data collection. Under the guidance of a multidisciplinary team including
experienced clinicians, epidemiologists, and informaticians, we extracted 53 fea-
tures including epidemiological, demographic, clinical, laboratory, radiological, and
outcome data from EHRs using identical data collection forms on the
first day of admission (Supplementary Table 1). Trained researchers entered and double-
checked the data independently. To ensure the alarming function and subjective
outcome data from EHRs using identical data collection forms on the
viability of the MRPMC model, the performance was validated by an open-source
model development. We trained the models to predict mortality risk with the 14
variables and outcomes of COVID-19 patients. During model training, we fitted six
baseline ML models, including LR, SVM, KNN, RF, GBDT, and NN, into the SFT
cohort with tenfold cross validation to fine-tune the model parameters. Increasing
the weight of minority categories in the model can increase the punishment for
wrong classification of minority categories during training, and improve the
model’s ability to recognize minority categories17. Therefore, we adopted weighted
classification loss, which is a class of cost functions that multiplies the loss
classifiers (LR, RF, GBDT, and NN). Subsequently, an ensemble model derived from
four baseline models of best predictive performance (LR, SVM, GBDT, and NN),
named MRPMC, was proposed by weighted voting. Specifically, the mortality risk
probability of each individual estimator (LR, SVM, GBDT, and NN) was integrated
by manually assigning weights with 0.25, 0.3, 0.1, and 0.35, respectively. After all
ML models were well fitted, they were internally and externally evaluated in SFV,
OV, and CHWH cohorts. Herein, we modeled the mortality prediction task as a
binary classification problem. All included ML models output a normalized
probability of mortality risk range from 0 to 1. We selected the threshold of 0.6 to
assign the predicted mortality risk label by optimizing F1 score on the training
cohort (Supplementary Fig. 6). Probabilities of less than 0.6 were assigned to low
risk and otherwise to high risk for all ML methods across all cohorts. R library caret
was utilized for model training and prediction. The LR, SVM, KNN, RF, GBDT,
and NN models were called with method bayesglm, svmLinear, knn, rf, gbm, and
avNNet with default settings, respectively. We standardized the features data with
BoxCox, center, and scale function before training and prediction. Especially, we
first adopted BoxCox transformation to make the data distribution more Gaussian-
like34, and then standardized features by subtracting the mean and scaling to unit
variance. Variable z was calculated as: $z = (x - \mu) / \sigma$, where $\mu$ is the mean and $\sigma$ is
the standard deviation of the variable.

Model evaluation. The predictive performance of the models was evaluated
by ROC curve, Kaplan–Meier curve, calibration curve, and evaluation metrics
including area under the ROC curve (AUC), accuracy, sensitivity, specificity,
positive predictive value (PPV), negative predictive value (NPV), F1 score, Cohen’s
Kappa coefficient (Kappa), and Brier score. The relative feature importance of each
model was calculated using variable importance methods for RF, KNN, and CART
classifier had no built-in importance score, the AUC for each feature was utilized as
the importance score.

Statistical analysis. Statistical analysis was performed in R (version 3.6.2). For
descriptional analysis, median (IQR) and frequencies (%) were assessed for con-
tinuous and categorical variables, respectively. The ROC curve and AUC analysis
were conducted with R pROC package. Accuracy, sensitivity, specificity, PPV,
NPV, Kappa, and F1 score were calculated with R caret and e1071 packages. The
calibration curve and Brier score were obtained with R package rms. Relative
feature importance was calculated using R package caret. Survival curves were
developed by Kaplan–Meier method with log-rank test, and plotted with R package
survminer. Comparison of continuous variables was achieved by the
Mann–Whitney U test using R package table1. Odds ratio and corresponding 95% CI
from LR were calculated with R-package stats. The significance level was set at a
two-sided p value below 0.05. Univariate and multivariate Cox regression was
utilized to calculate the HR with R-package survival. All dry lab experiments
were conducted in three different computing servers with consistent result.

Reporting summary. Further information on research design is available in the Nature
Research Reporting Summary linked to this article.

Data availability. Data pertaining to the patients’ features used for modeling are available to researchers
upon reasonable request via contacting the corresponding author. Patient current vital
status and follow-up information are not publically available due to privacy concerns.
The remaining data are available in the article and supplementary files. Source data are provided with this paper.

### Code availability
The code used to develop and evaluate the model is available on GitHub with R (version 3.6.2)39, https://doi.org/10.5281/zenodo.3991113.

Received: 22 May 2020; Accepted: 27 August 2020; Published online: 06 October 2020

### References
1. Katz, J. N. et al. Disruptive modifications to cardiac critical care delivery during the Covid-19 pandemic: an international perspective. J Am Coll Cardiol. https://doi.org/10.1016/j.jacc.2020.04.029 (2020).
2. World Health Organization. Coronavirus 2019 (COVID-19) (World Health Organization, 2020). https://covid19.who.int/.
3. Phelan, A. L., Katz, R. & Gostin, L. O. The novel coronavirus originating in Wuhan, China: challenges for global health governance. J. Am. Med. Assoc. 323, 709–710 (2020).
4. Guan, W. J. et al. Clinical characteristics of coronavirus disease 2019 in China. N. Engl. J. Med. 382, 1708–1720 (2020).
5. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506 (2020).
6. Wang, D. et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA https://doi.org/10.1001/jama.2020.1585 (2020).
7. Valerik, K. et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. Lancet Digit. Health 1, e53–e77 (2019).
8. Denaxas, S. et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. J. Am. Med. Inform. Assoc. 26, 1545–1559 (2019).
9. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. N. Engl. J. Med. 380, 1347–1358 (2019).
10. Vollmer, S. et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ 368, I6927 (2020).
11. Cho, A. AI systems aim to snuff out coronavirus outbreaks. Science 368, 810–811 (2020).
12. Bai, H. X. et al. Artificial augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT. Radiology, 201491 https://doi.org/10.1148/radiol.2020201491 (2020).
13. Mei, X. et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat. Med. https://doi.org/10.1038/s41591-020-0931-3 (2020).
14. Wu, G. et al. Development of a clinical decision support system for severity risk prediction and triage of COVID-19 patients at hospital admission: an international multicenter study. Eur. Respir. J. https://doi.org/10.1183/13993003.01044-2020 (2020).
15. Wynants, L. et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 369, m1328 (2020).
16. Zhou, F. et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 395, 1054–1062 (2020).
17. Tian, J. et al. Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. Lancet Oncol. 21, 893–903 (2020).
18. Yang, K. et al. Clinical characteristics, outcomes, and risk factors for mortality in patients with cancer and COVID-19 in Hubei, China: a multicentre, retrospective, cohort study. Lancet Oncol. 21, 904–913 (2020).
19. Yuan, M., Yin, W., Tao, Z., Tan, W. & Hu, Y. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. PLoS ONE 15, e0230548 (2020).
20. Shi, Y. et al. Host susceptibility to severe COVID-19 and establishment of a host risk score: findings of 487 cases outside Wuhan. Crit. Care 24, 108 (2020).
21. Wynants, L. et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 369, m1328 (2020).
22. Bai, X. et al. Predicting COVID-19 malignant progression with AI techniques. Preprint at https://www.medrxiv.org/content/10.1101/2020.03.20.20037325v2 (2020).
23. Gong, J. et al. A tool for early prediction of severe coronavirus disease 2019 (COVID-19): a multicenter study using the risk nomogram in Wuhan and Guangdong, China. Clin. Infect. Dis. 71, 833–840 (2020).
24. Lu, J. et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. Preprint at https://www.medrxiv.org/content/10.1101/2020.02.20.20025510v1 (2020).
25. Pourhomayoun, M. & Shakibi, M. Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making. Preprint at https://www.medrxiv.org/content/10.1101/2020.03.30.20047308v1 (2020).
26. Yue, H. et al. Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study. Ann Transl Med 8, 859 (2020).
27. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in china: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 323, 1239–1242 (2020).
28. Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J. & Prescott, H. C. Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. JAMA https://doi.org/10.1001/jama.2020.12839 (2020).
29. Wu, C. et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Internal Med. https://doi.org/10.1001/jamainternalmed.2020.0994 (2020).
30. National Health Commission of the People's Republic of China. Interim Diagnosis and Treatment of 2019 Novel Coronavirus Pneumonia 7th edn (National Health Commission of the People’s Republic of China, 2020). http://www.nhc.gov.cn/yzygj/s7653p/202003/46c929a47dfe4e88c755912e1b899.shtml.
31. Yu, J., Ouyang, W., Chua, M. L. K. & Xie, C. SARS-CoV-2 transmission in patients with cancer at a tertiary care hospital in Wuhan, China. JAMA Oncol. https://doi.org/10.1001/jamaoncol.2020.0980 (2020).
32. Liang, W. et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. JAMA Internal Med. https://doi.org/10.1001/jama.2020.2033 (2020).
33. Stekhoven, D. J. & Buhlmann, P. MissForest—non-parametric missing value imputation for mixed-type data. Bioinformatics 28, 112–118 (2012).
34. Waljee, A. K. et al. Comparison of imputation methods for missing laboratory data in medicine. BMJ Open 3 https://doi.org/10.1136/bmjopen-2013-002847 (2013).
35. Van Buuren, S. Flexible Imputation of Missing Data (CRC Press, 2018).
36. Fu, H. et al. Identification and validation of stromal immunotype predict survival and benefit from adjuvant chemotherapy in patients with muscle-invasive bladder cancer. Clin. Cancer Res. 24, 3069–3078 (2018).
37. He, H. & Garcia, E. A. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21, 1263–1284 (2009).
38. Osborne, J. Improving your data transformations: applying Box-Cox transformations as a best practice. Pr. Assess. Res. Eval. 15, 1–9 (2010).
39. Chen, L. X. Machine Learning Based Early Warning System Enables Accurate Mortality Risk Prediction for COVID-19 (Github, 2020). https://doi.org/10.5281/zenodo.3991113.

### Acknowledgements
We thank all health workers and people behind them that fight against COVID-19. This study was funded by the National Science and Technology Major Sub-Project (2018ZX10010-002), the Technical Innovation Special Project of Hubei Province (2018ACA138), the National Natural Science Foundation of China (81777287, 81873452, and 81974405), the Fundamental Research Funds for the Central Universities (2019kyXMBZ024), and Nature Science Foundation of Hubei Province (2019CFB453).

### Author contributions
Y.G designed the study. G.-Y.C., W.F., and L.C did the analysis. Y.G., G.-Y.C., W.F., H.-Y.L., and L.C interpreted the data and wrote the paper. Y.Yu, S.-Y.W., D.L., P.-F.C., Q.-L.G., F.Y., D.M., and C.-R.L. advised on the conception and design of the study. All authors vouch for the respective data and analysis, approved the final version, and agreed to publish the manuscript.

### Competing interests
The authors declare no competing interests.
