



## OPEN ACCESS

## EDITED BY

Yi-Ju Tseng,  
National Yang Ming Chiao Tung  
University, Taiwan

## REVIEWED BY

Vikas Chaurasia,  
Veer Bahadur Singh Purvanchal  
University, India  
Wallace Casaca,  
São Paulo State University, Brazil  
Vijay Anant Athavale,  
Walchand Institute of  
Technology, India  
Hanana Salem Marie,  
Delta University for Science and  
Technology, Egypt

## \*CORRESPONDENCE

Tongquan Liao  
liaotq@126.com  
Hao Chen  
chenh99@foxmail.com  
Caiping Song  
scp1974@163.com  
Chunji Huang  
hcjwyhws@163.com  
Yazhou Wu  
asiawu@tmmu.edu.cn  
Zhi Xu  
xu\_zhi999@163.com;  
xuzhihxk@tmmu.edu.cn

†These authors have contributed  
equally to this work

## SPECIALTY SECTION

This article was submitted to  
Infectious Diseases - Surveillance,  
Prevention and Treatment,  
a section of the journal  
Frontiers in Medicine

RECEIVED 24 July 2022

ACCEPTED 30 September 2022

PUBLISHED 02 November 2022

## CITATION

Xu Y, Ye W, Song Q, Shen L, Liu Y,  
Guo Y, Liu G, Wu H, Wang X, Sun X,  
Bai L, Luo C, Liao T, Chen H, Song C,  
Huang C, Wu Y and Xu Z (2022) Using  
machine learning models to predict  
the duration of the recovery of  
COVID-19 patients hospitalized in  
Fangcang shelter hospital during the  
Omicron BA. 2.2 pandemic.  
*Front. Med.* 9:1001801.  
doi: 10.3389/fmed.2022.1001801

# Using machine learning models to predict the duration of the recovery of COVID-19 patients hospitalized in Fangcang shelter hospital during the Omicron BA. 2.2 pandemic

Yu Xu<sup>1†</sup>, Wei Ye<sup>2†</sup>, Qiuyue Song<sup>2†</sup>, Linlin Shen<sup>1†</sup>, Yu Liu<sup>1,3</sup>,  
Yuhang Guo<sup>1,3</sup>, Gang Liu<sup>1,3</sup>, Hongmei Wu<sup>1,3</sup>, Xia Wang<sup>1</sup>,  
Xiaorong Sun<sup>1</sup>, Li Bai<sup>1</sup>, Chunmei Luo<sup>3,4</sup>, Tongquan Liao<sup>3,5\*</sup>,  
Hao Chen<sup>3,6\*</sup>, Caiping Song<sup>3,7\*</sup>, Chunji Huang<sup>3,8\*</sup>, Yazhou Wu<sup>2\*</sup>  
and Zhi Xu<sup>1,3\*</sup>

<sup>1</sup>Respiratory and Critical Care Medical Center, Xinqiao Hospital, Army Medical University, Chongqing, China, <sup>2</sup>Department of Health Statistics, Army Medical University, Chongqing, China, <sup>3</sup>National Exhibition and Convention Center Fangcang Shelter Hospital, Shanghai, China, <sup>4</sup>Department of Orthopedics, Xinqiao Hospital, Army Medical University, Chongqing, China, <sup>5</sup>Department of Medical Administration, Xinqiao Hospital, Army Medical University, Chongqing, China, <sup>6</sup>Academic Affairs Office, Army Medical University, Chongqing, China, <sup>7</sup>Xinqiao Hospital, Army Medical University, Chongqing, China, <sup>8</sup>Army Medical University, Chongqing, China

**Background:** Factors that may influence the recovery of patients with confirmed SARS-CoV-2 infection hospitalized in the Fangcang shelter were explored, and machine learning models were constructed to predict the duration of recovery during the Omicron BA. 2.2 pandemic.

**Methods:** A retrospective study was conducted at Hongqiao National Exhibition and Convention Center Fangcang shelter (Shanghai, China) from April 9, 2022 to April 25, 2022. The demographics, clinical data, inoculation history, and recovery information of the 13,162 enrolled participants were collected. A multivariable logistic regression model was used to identify independent factors associated with 7-day recovery and 14-day recovery. Machine learning algorithms (DT, SVM, RF, DT/AdaBoost, AdaBoost, SMOTEENN/DT, SMOTEENN/SVM, SMOTEENN/RF, SMOTEENN+DT/AdaBoost, and SMOTEENN/AdaBoost) were used to build models for predicting 7-day and 14-day recovery.

**Results:** Of the 13,162 patients in the study, the median duration of recovery was 8 days (interquartile range IQR, 6–10 d), 41.31% recovered within 7 days, and 94.83% recovered within 14 days. Univariate analysis showed that the administrative region, age, cough medicine, comorbidities, diabetes, coronary artery disease (CAD), hypertension, number of comorbidities, CT value of the ORF gene, CT value of the N gene, ratio of ORF/IC, and ratio of N/IC were associated with a duration of recovery within 7 days. Age, gender, vaccination dose, cough medicine, comorbidities, diabetes, CAD, hypertension, number

of comorbidities, CT value of the ORF gene, CT value of the N gene, ratio of ORF/IC, and ratio of N/IC were related to a duration of recovery within 14 days. In the multivariable analysis, the receipt of two doses of the vaccination vs. unvaccinated (OR = 1.118, 95% CI = 1.003–1.248;  $p = 0.045$ ), receipt of three doses of the vaccination vs. unvaccinated (OR = 1.114, 95% CI = 1.004–1.236;  $p = 0.043$ ), diabetes (OR = 0.383, 95% CI = 0.194–0.749;  $p = 0.005$ ), CAD (OR = 0.107, 95% CI = 0.016–0.421;  $p = 0.005$ ), hypertension (OR = 0.371, 95% CI = 0.202–0.674;  $p = 0.001$ ), and ratio of N/IC (OR = 3.686, 95% CI = 2.939–4.629;  $p < 0.001$ ) were significantly and independently associated with a duration of recovery within 7 days. Gender (OR = 0.736, 95% CI = 0.63–0.861;  $p < 0.001$ ), age (30–70) (OR = 0.738, 95% CI = 0.594–0.911;  $p < 0.001$ ), age (>70) (OR = 0.38, 95% CI = 0.292–0.494;  $p < 0.001$ ), receipt of three doses of the vaccination vs. unvaccinated (OR = 1.391, 95% CI = 1.12–1.719;  $p = 0.0033$ ), cough medicine (OR = 1.509, 95% CI = 1.075–2.19;  $p = 0.023$ ), and symptoms (OR = 1.619, 95% CI = 1.306–2.028;  $p < 0.001$ ) were significantly and independently associated with a duration of recovery within 14 days. The SMOTEEN/RF algorithm performed best, with an accuracy of 90.32%, sensitivity of 92.22%, specificity of 88.31%, F1 score of 90.71%, and AUC of 89.75% for the 7-day recovery prediction; and an accuracy of 93.81%, sensitivity of 93.40%, specificity of 93.81%, F1 score of 93.42%, and AUC of 93.53% for the 14-day recovery prediction.

**Conclusion:** Age and vaccination dose were factors robustly associated with accelerated recovery both on day 7 and day 14 from the onset of disease during the Omicron BA. 2.2 wave. The results suggest that the SMOTEEN/RF-based model could be used to predict the probability of 7-day and 14-day recovery from the Omicron variant of SARS-CoV-2 infection for COVID-19 prevention and control policy in other regions or countries. This may also help to generate external validation for the model.

#### KEYWORDS

COVID-19, omicron, Fangcang shelter, machine learning model, vaccination

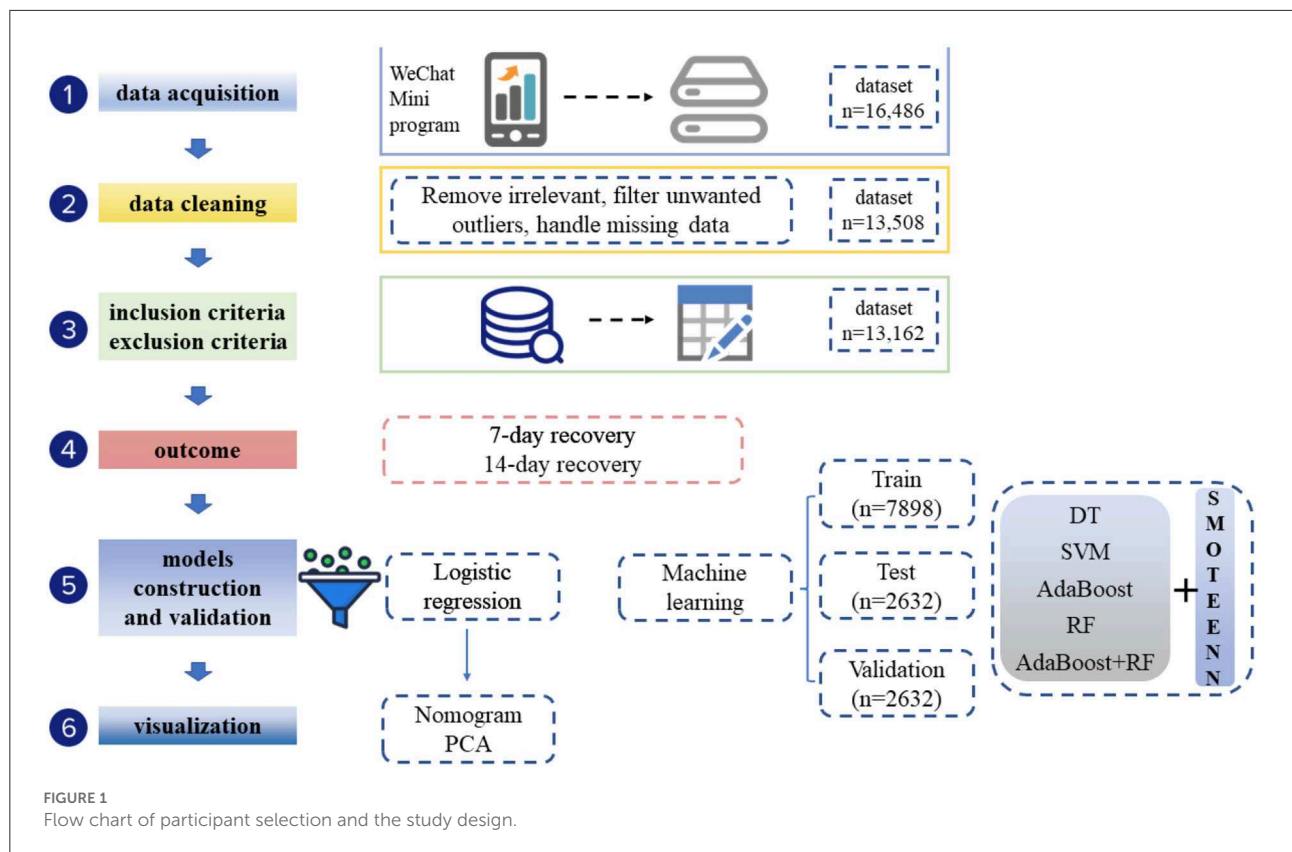
## Introduction

Since the first case of the Omicron variant of the SARS-CoV-2 infection was detected in Shanghai, China on March 1, 2022, the epidemic has spread rapidly, with the largest number of daily new confirmed cases reaching 5,487 on April 28 and asymptomatic infections reaching 25,173 on April 10. To maximize the protection of people's health and embody the concept of people first and life first, Fangcang shelter hospitals were quickly built and put into use in Shanghai. The National Exhibition Center Fangcang shelter hospital was set up with the capacity to accommodate up to 50,000 asymptomatic carriers and patients with mild pneumonia. The hospital has admitted and discharged more than 170,000 patients with asymptomatic and mild Omicron infections since April 9, 2022.

The dominant strain of the current epidemic wave was SARS-CoV-2 Omicron BA. 2.2, which represented a small sub-lineage of BA.2 worldwide seen in Hong Kong, the UK, and

Australia previously. Omicron BA. 2.2 is more transmissible, but tends to be less virulent, with the majority of diseases observed in Shanghai being asymptomatic infection or mild illness. Despite this, the crude case fatality rate in people over 60 was 2.7% in Hong Kong (1). The dynamic model of SARS-CoV-2 transmission from the School of Public Health of Fudan University estimated that the Omicron epidemic spread in Shanghai would require 15.6 times the available capacity of the intensive care unit, and would cause –1.55 million deaths if the pandemic control strategies were lifted (2).

The pressure for the control of the omicron pandemic wave in China is still stressed until there is a comprehensive understanding of the clinical characteristics of the BA.2.2 strain infections in the Chinese population. The duration of viral shedding is decisive in considering quarantine strategies and reducing transmission. For the original SARS-CoV-2, the medium duration of viral RNA shedding was 12 days after



the onset of illness in hospitalized patients (3). In a recent study, researchers reported that the duration of viral RNA shedding for the Omicron variant was 7 days (IQR = 5–8 days) among fully vaccinated national football league players and staff members in the United States (3). However, there are few serial virus testing data available to analyze the viral clearance time for Omicron variants in the general population. Older age was reported to be a determinant of RNA persistence in nasopharyngeal samples (4). Treatment plans, including corticosteroid usage and Lopinavir/ritonavir administration, also led to prolonged viral RNA shedding (5). The Chinese government has enacted a COVID-19 vaccination program over the past 2 years, with booster shots for fully vaccinated people. However, whether vaccination is associated with the duration of virus clearance, specifically for the omicron variant, has not been reported.

In this study, we aim to describe factors that may influence the recovery of patients with confirmed SARS-CoV-2 Omicron BA. 2.2 infections hospitalized in the National Exhibition and Convention Center Fangcang hospital and to build machine learning models to predict the probability of recovery on day 7 and day 14 from the onset of disease during the Omicron BA. 2.2 wave. The findings provide a theoretical basis and reference for prevention and control strategies for the COVID-19 epidemic.

## Materials and methods

### Patients

Participants were retrospectively screened at the National Exhibition and Convention Center Fangcang shelter (Shanghai, China) from April 9, 2022 to April 25, 2022 during the Omicron BA. 2.2 strain pandemic period. This shelter is responsible for admitting confirmed COVID-19 patients classified as having an asymptomatic infection or mild illness based on the WHO COVID-19 treatment guidelines (6). The eligibility criteria required individuals to have a positive nucleic acid amplification test (NAAT) for COVID-19. The exclusion criteria included the following: (1) lack of key clinical data or inoculation information; and (2) moderate, severe, or critical illness. This study was approved by the Ethics Committee of Xinqiao Hospital (No. 2022-197-01). Informed consent was waived according to the retrospective nature of the study.

### Data collection and outcome

Medical documents were recorded using a WeChat Mini Program that enables patients' self-reported information to be assembled under the guidance of medical staff. Data including

TABLE 1 Clinical characteristics of 13,162 cases of Omicron infection in Shanghai.

Variables		Variables	
<b>Sex</b>		N	12,956 (98.43)
Male	7,538 (57.27)	Y	206 (1.57)
Female	5,624 (42.73)	<b>Cold medicine</b>	
<b>Age</b>		N	12,887 (97.91)
<30	3,068 (23.31%)	Y	275 (2.09)
30-70	8,710 (66.18%)	<b>Comorbidities</b>	
>70	1,384 (10.51%)	N	11,964 (90.90)
<b>Marriage</b>		Y	1,198 (9.10)
Married	12,746 (96.84)	<b>Diabetes</b>	
Single	389 (2.96)	N	12,805 (97.29)
Divorced	27 (0.21)	Y	357 (2.71)
<b>Diagnosis</b>		<b>Coronary artery disease</b>	13,142 (99.85)
Asymptomatic infection	12,575 (95.54)	N	20 (0.15)
Mild illness	587 (4.46)	Y	
<b>Vaccination dose</b>		<b>Hypertension</b>	
0 None	1,995 (15.16)	N	12,210 (92.77)
1 dose	491 (3.73)	Y	952 (7.23)
2 doses	4,352 (33.06)	<b>Cancer</b>	
3 doses	6,324 (48.05)	N	13,155 (99.95)
<b>Chinese medicine prescription</b>		Y	7(0.05)
N	10,100 (76.74)	<b>Other comorbidities</b>	
Y	3,062 (23.26)	N	13,149 (99.90)
<b>Lianhua Qingwen</b>		Y	13(0.10)
N	10,548 (80.14)	<b>Number of comorbidities</b>	
Y	2,614 (19.86)	0	11,964 (90.90)
<b>Kangbingdu Granules</b>		1	1,017 (7.73)
N	13,095 (99.49)	2	181 (1.38)
Y	67 (0.51)	<b>Fever</b>	
<b>Other Chinese medicine prescription</b>		N	12,925 (98.20)
N	12,738 (96.78)	Y	237 (1.80)
Y	424 (3.22)	<b>CT value of ORF gene</b>	39.60 [34.65- 40.00]
<b>Cough medicine</b>		<b>CT value of N gene</b>	38.59 [33.19, 40.00]
N	12,180 (92.54)	<b>ORF/IC</b>	1.32 [1.21, 1.40]
Y	982 (7.46)	<b>N/IC</b>	1.30 [1.16, 1.38]
<b>Antipyretics</b>			

age, gender, living districts, occupation, oral medication, comorbidities, SARS-CoV-2-related symptoms (fever, headache, cough, sputum, malaise, stuffy/running nose, scratchy throat, muscle pain, nausea, vomiting, diarrhea, and loss of taste and smell), vaccination history, rounds of NAAT tests, and Ct values of the ORF lab gene, N gene, and internal control (IC) in every round were collected. The primary outcome was recovery on day 7 and day 14 from disease onset. Recovery was defined as being allowed to leave quarantine or hospital with two consecutive negative tests (at least 24 h apart) or Ct values  $\geq 35$  according to

the 9th edition of COVID-19 diagnosis and treatment guidance issued by the Nation Health Commission of China (7).

### Statistics and model construction

Statistical analysis was performed using SPSS Statistics (IBM, version 26.0) and R (Version 4.1.1). Continuous variables were expressed as the median with IQR or mean  $\pm$  SD, and categorical variables were expressed as the number and

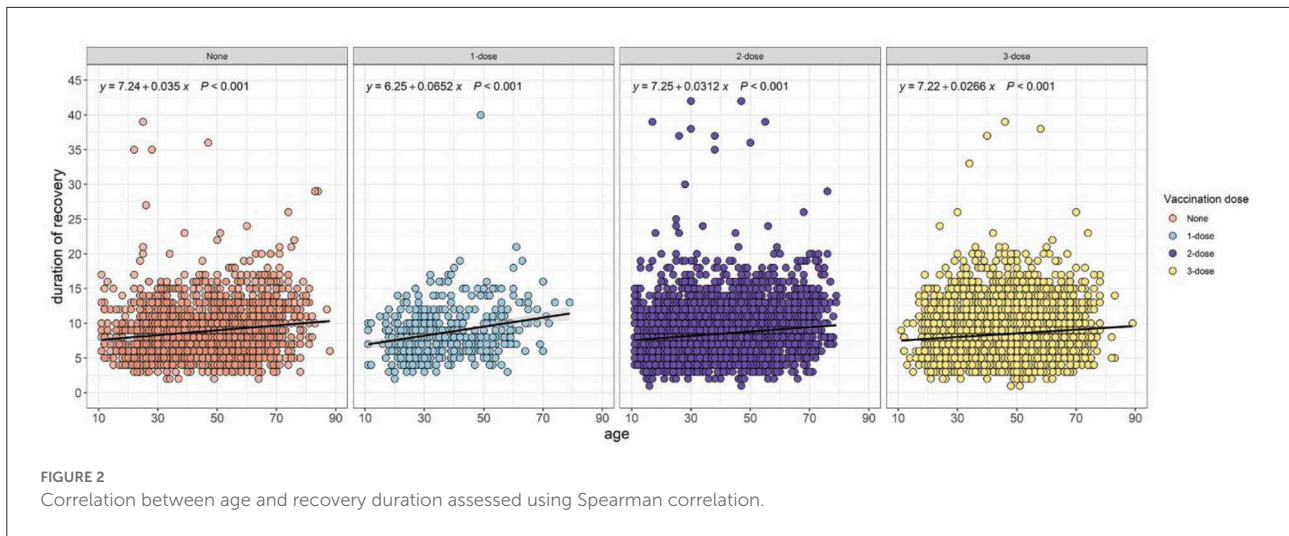


TABLE 2 Multivariable logistic regression models for Omicron infection recovery.

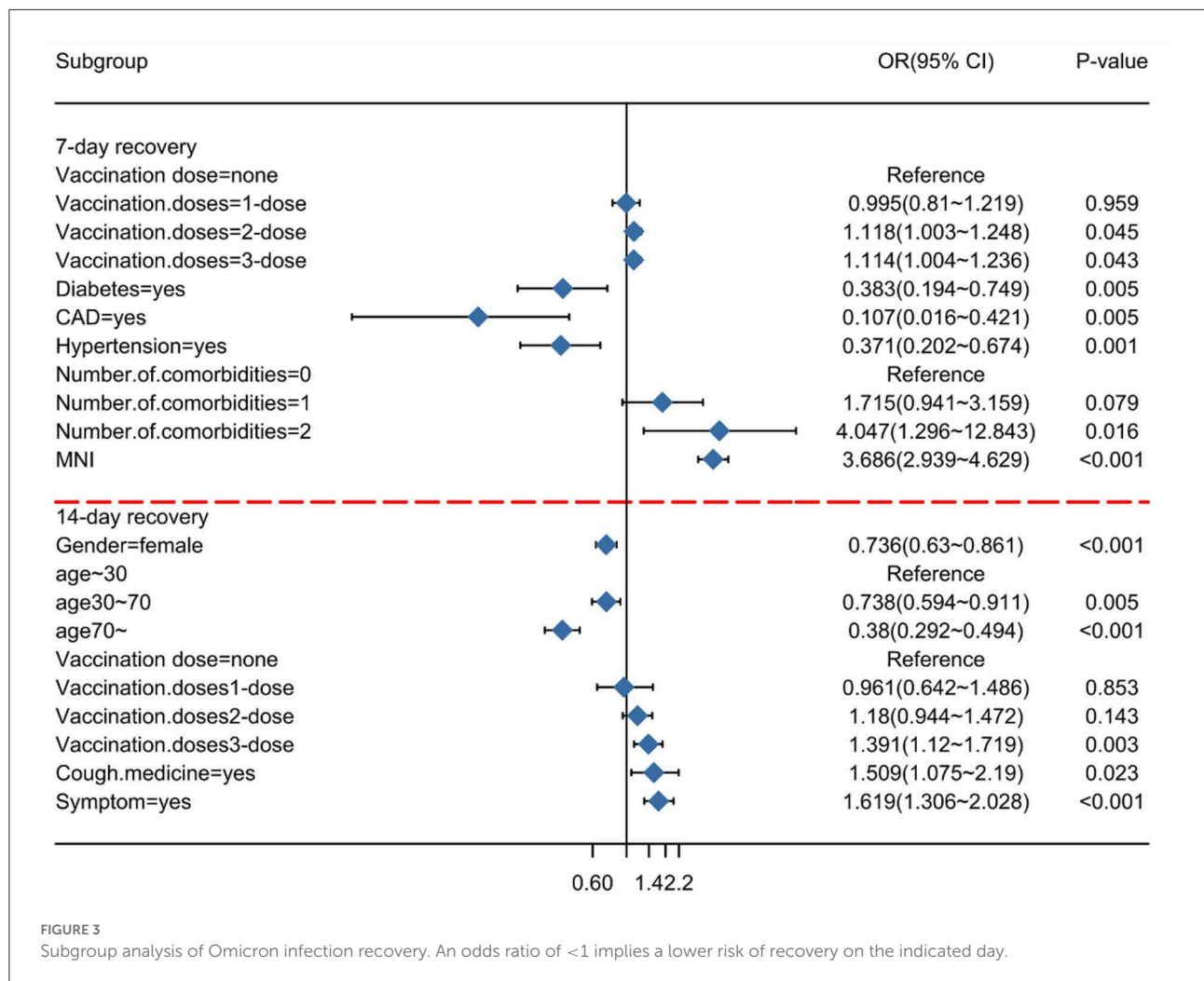
Variable	7-day recovery	P-Value	14-day recovery	P-Value
	OR (95%CI)		OR (95%CI)	
Vaccination dose = none	Reference		Reference	
Vaccination doses = 1-dose	0.995 (0.81–1.219)	0.959	0.961 (0.642–1.486)	0.853
Vaccination doses = 2-dose	1.118 (1.003–1.248)	0.045	1.18 (0.944–1.472)	0.143
Vaccination doses = 3-dose	1.114 (1.004–1.236)	0.043	1.391 (1.12–1.719)	0.003
Diabetes = yes	0.383 (0.194–0.749)	0.005	–	–
CAD = yes	0.107 (0.016–0.421)	0.005	–	–
Hypertension = yes	0.371 (0.202–0.674)	0.001	–	–
Number of comorbidities = 0	Reference		–	–
Number of comorbidities = 1	1.715 (0.941–3.159)	0.079	–	–
Number of comorbidities = 2	4.047 (1.296–12.843)	0.016	–	–
MNI	3.686 (2.939–4.629)	< 0.001	–	–
Gender = female			0.736 (0.63–0.861)	< 0.001
age–30			Reference	
Age 30–70			0.738 (0.594–0.911)	0.005
Age 70–			0.38 (0.292–0.494)	< 0.001
Cough medicine = yes			1.509 (1.075–2.19)	0.023
Symptom = yes			1.619 (1.306–2.028)	< 0.001

percent (%). Variables were compared using Fisher’s exact tests,  $\chi^2$  statistic, or Student’s *t*-test. Variables with a *p*-value of <math>< 0.05</math> in the univariate test were included in the multivariable logistic regression model to screen out variables that had an independent impact on the recovery of Omicron infection on day 7 and day 14. Based on the variables screened using logistic regression, two consecutive negative rt-PCR results on 7-day or 14-day were used as outcome variables, and the dataset was randomly divided into a training set and test set in the ratio of 7:3. Predictive models were constructed to predict the 7-day and 14-day recovery rates of COVID-19 patients

hospitalized in Fangcang shelter during the Omicron BA. 2.2 pandemic. Principal component analysis (PCA) was used to identify separate outcome groups with variables selected using logistic regression.

Python 3.7.6 was used to build machine learning predictive models. Five machine learning algorithms, that is, the decision tree (DT), support vector machine (SVM), random forest (RF), AdaBoost algorithm, and AdaBoost algorithm with DT as a classifier, were used. The outcomes of rehabilitation on day 7 and day 14 were used as dependent variables, and all the collected variables were included as independent variables. The





dataset was divided into a training set, test set, and validation set in the ratio of 6:2:2. The training set was used to train different machine learning models and the test set was used to optimize the parameters of different machine learning models using a grid search method. Finally, the validation set was used to validate the efficacy of different machine learning models with optimal parameters. The prediction results were evaluated using the accepted accuracy, sensitivity, specificity, and F1 score, area under the curve (AUC), and Matthews correlation coefficient (MCC) as evaluation indicators. Finally, the SMOTEENN algorithm was used to balance the two datasets. The SMOTEENN algorithm is a combination of the SMOTE algorithm and the ENN algorithm. SMOTEENN is a balanced algorithm that combines oversampling and undersampling the dataset through the SMOTE algorithm and then uses the ENN algorithm to clean it. Batista et al. proposed it in 2004 (8). After the data were balanced, the prediction was made by the five aforementioned methods, and the prediction results before and after balancing were compared.

## Results

### Patients' characteristics

A total of 13,162 COVID-19 nucleic acid-positive cases from Hongqiao International Exhibition Center Fangcang shelter (Shanghai) were included (Figure 1). These patients came from 14 administrative regions of Shanghai. The geographic distribution of these Omicron-infected patients is visualized in Figure 1. Table 1 shows the details of the baseline characteristics of the enrolled patients. The median age of these patients was 43.0 (IQR, 31.0–53.0), 7,358 patients (57.27%) were male, and 12,575 patients (95.54%) and 587 (4.466%) patients were categorized as having asymptomatic and mild COVID-19, respectively. The common comorbidities included diabetes (357, 2.71%), CAD (20, 0.15%), hypertension (952, 7.23%), and cancer (7, 0.05%). The median CT value was 39.60 (IQR, 34.65–40.00) for the ORF gene and 38.59 (IQR, 33.19–40.00) for the N gene. The median ORF/IC ratio was 1.32 (IQR, 1.21–1.40) and median

TABLE 3 Evaluation indices of the classification models for full variables.

Model	7-day						14-day					
	Sen	Sp	Acc	F-measure	MCC	AUC	Sen	Sp	Acc	F-measure	MCC	AUC
DT	0.2973	0.8291	0.6092	0.3849	0.1468	0.5622	1.0000	0.0000	0.9502	0.9745	0.0000	0.5000
RF	0.4109	0.7374	0.6031	0.4599	0.1560	0.5742	0.9956	0.0229	0.9472	0.9729	0.0553	0.5093
SVM	0.2188	0.9026	0.6213	0.3222	0.1685	0.5607	1.0000	0.0000	0.9502	0.9745	0.0000	0.5000
AdaBoost	0.3564	0.8110	0.6240	0.4381	0.1883	0.5837	1.0000	0.0000	0.9502	0.9745	0.0000	0.5000
DT+AdaBoost	0.3694	0.7793	0.6103	0.4375	0.1600	0.5734	0.9980	0.0076	0.94873	0.9737	0.0257	0.5028
DT+SMOTEENN	0.8689	0.8481	0.8588	0.8631	0.7173	0.8585	0.7248	0.8895	0.8095	0.7871	0.6246	0.8072
RF+SMOTEENN	0.9222	0.8831	0.9032	0.9071	0.8066	0.8975	0.9340	0.9381	0.9361	0.9342	0.8721	0.9353
SVM+SMOTEENN	0.8311	0.6098	0.7232	0.7548	0.4532	0.7205	0.5461	0.8479	0.7013	0.6399	0.4147	0.6970
AdaBoost+SMOTEENN	0.7644	0.7103	0.7380	0.7495	0.4756	0.7374	0.6038	0.6181	0.6112	0.6014	0.2218	0.6110
DT+AdaBoost+SMOTEENN	0.8489	0.8294	0.8394	0.8442	0.6972	0.8392	0.8166	0.9589	0.8897	0.8780	0.7881	0.8877

N/IC ratio was 1.30 (1.16–1.38). Among the patients, 11,167 cases (84.84%) had a history of SARS-CoV-2 vaccination, 6,324 (48.05%) participants received three doses, 4,352 (33.06%) cases received two doses, and 491 (3.73%) received one dose.

### Identifying prognostic factors associated with 7-day recovery and 14-day recovery

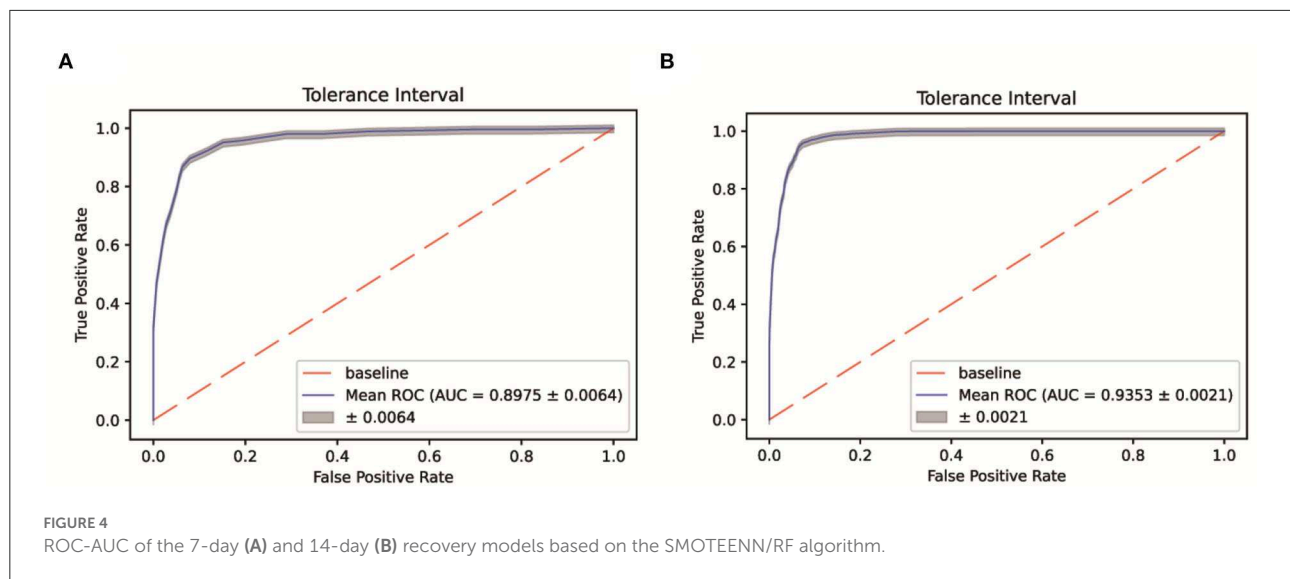
The univariate analysis results revealed the potential prognostic factors associated with 7-day and 14-day recovery, and are presented in Table 1. The analysis demonstrated that the administrative region, age, cough medicine, comorbidities, diabetes, CAD, hypertension, number of comorbidities, CT value of the ORF gene, CT value of the N gene, ratio of ORF/IC, and ratio of N/IC were potential prognostic factors that affected the 7-day recovery of COVID-19 patients. Data regarding age, gender, vaccination dose, cough medicine, comorbidities, diabetes, CAD, hypertension, number of comorbidities, CT value of the ORF gene, CT value of the N gene, ratio of ORF/IC, and ratio of N/IC were potential prognostic factors that affected 14-day recovery. The relationship between the duration of recovery and age was performed using Spearman correlation stratified by the vaccination dose. Regardless of the vaccination status, a positive correlation was observed in unvaccinated ( $p < 0.001$ ), one-dose ( $p < 0.001$ ), two-dose ( $p < 0.001$ ) and three-dose ( $p < 0.001$ ) participants (Figure 2).

To minimize the effect of confounders, these prognostic factors were subsequently included in the multivariate logistic regression analysis (Table 2). In the multivariable analysis, the receipt of two doses of the vaccination vs. unvaccinated (OR = 1.118, 95% CI = 1.003–1.248;  $p = 0.045$ ), receipt of three doses of the vaccination vs. unvaccinated (OR = 1.114, 95% CI = 1.004–1.236;  $p = 0.043$ ), diabetes (OR = 0.383, 95% CI =

0.194–0.749;  $p = 0.005$ ), CAD (OR = 0.107, 95% CI = 0.016–0.421;  $p = 0.005$ ), hypertension (OR = 0.371, 95% CI = 0.202–0.674;  $p = 0.001$ ), and the ratio of N/IC (OR = 3.686, 95% CI = 2.939–4.629;  $p < 0.001$ ) were significantly and independently associated with a duration of recovery within 7 days. Gender (OR = 0.736, 95% CI = 0.63–0.861;  $p < 0.001$ ), age (30–70) (OR = 0.738, 95% CI = 0.594–0.911;  $p < 0.001$ ), age (>70) (OR = 0.38, 95% CI = 0.292–0.494;  $p < 0.001$ ), receipt of three doses of the vaccination vs. unvaccinated (OR = 1.391, 95% CI = 1.12–1.719;  $p = 0.0033$ ), cough medicine (OR = 1.509, 95% CI = 1.075–2.19;  $p = 0.023$ ), and symptoms (OR = 1.619, 95% CI = 1.306–2.028;  $p < 0.001$ ) were significantly and independently associated with a duration of recovery within 14 days. A forest plot was used to visualize the results of the multivariate logistic regression analysis (Figure 3).

### Logistic regression model and PCA

According to the variables screened above, first, a logistic regression model was constructed to predict recovery on day 7, with variables including vaccination doses, diabetes, CAD, hypertension, number of comorbidities, and ratio of N/IC. However, the predictive value of the model was not acceptable, and had an AUC of 56.8% for the test set. Similarly, when variables including vaccination dose, gender, age, cough medicine, and symptoms were used to construct the 14-day predictive model, this yielded an AUC of 63.01% for the test set. Then PCA showed there was a large degree of overlap between the two groups for the 7-day and 14-day recovery outcomes based on the selected variables in the logistic regression. Some sample points were scattered in other areas, which indicates that there were discrete recovered patients and isolated patients or outliers (Figure 2).



## Machine learning models

To build a more robust predictive model, five machine learning-based prediction models were established (DT, RF, SVM, AdaBoost, and DT+AdaBoost) for 7-day recovery and 14-day recovery from Omicron infection. Table 3 shows the accuracy, sensitivity, specificity, F-measure, AUC, and MCC of each model evaluated on the validation set. The five models performed poorly, with AUC <60% for both 7-day and 14-day recovery models based on the original dataset.

Patients were sampled from a single Fangcang shelter. The selection bias and data imbalance contributed to the low accuracy of the 7-day recovery models and falsely high accuracy of the 14-day recovery models for the above algorithms. Therefore, the SMOTEENN algorithm was used to perform the balance process for the original dataset. The balanced sample size of the recovered vs. unrecovered cohort was 835 vs. 1,996 for the 7-day recovery dataset, and 10,599 vs. 435 for the 14-day recovery dataset. Then five machine learning algorithms were applied to develop prediction models based on the balanced dataset. The SMOTEENN+RF algorithm achieved the best prediction results for both the 7-day and 14-day recovery models, that is, AUCs of 89.75% and 93.53% for the validation set, respectively (Figure 4).

## Discussion

During the Omicron BA. 2.2 pandemic wave in Shanghai, over 100 Fangcang shelter hospitals were developed and served to isolate patients with mild to moderate COVID-19. However, no detailed published reports have investigated the demographic and clinical information, viral load shedding time, or factors that

influenced the recovery of patients with confirmed SARS-CoV-2 infection hospitalized in the Fangcang shelter.

In the present study, 13,162 COVID-19 patients composed of asymptomatic carriers and mild cases were enrolled from the National Exhibition Center Fangcang shelter in Shanghai. The collected data, including age, gender, comorbidities, vital signs, therapeutic drugs, results of nucleic acid detection, and viral load shedding time, were analyzed to identify risk factors that may influence SARS-CoV-2 RNA shedding. The results showed that the median duration of disease recovery was 8 days (IQR 6–10 d). At 7 days, the rate of recovery from omicron infection was 41.31%. Vaccination, diabetes, CAD, hypertension, ratio of N/IC were independent factors associated with the 7-day recovery of patients with confirmed SARS-CoV-2 infection hospitalized at the Fangcang shelter. At 14 days, the rate of recovery was 94.83%. Gender, age, three doses of the vaccination, cough medicine, and symptoms were observed to be independently associated with 14-day recovery. Machine learning models were constructed to predict the recovery of COVID-19 on day 7 and day 14.

The Omicron variant was first found in South Africa at the end of November 2021. As of May 2022, the Omicron variant of concern became the dominant variant circulating globally, and evolved into many distinct sub-lineages: BA.1, BA.2, BA.3, BA.4, and BA.5 (9). According to WHO epidemiological update reports, BA.2 and its descendent lineages (i.e., BA.2.X) were the dominant variants from the 1st week of May 2022, comprising 97% of all sequences submitted to GISAID. The prevalence of BA.1 and its descendent lineages (i.e., BA.1.X), BA.3, and the B.1.617.2 (Delta) variant significantly decreased, falling below a global prevalence of <1% (9).

In several studies, researchers reported that Omicron variants had 70-fold higher replication competence in the



human bronchus, but low viral replication in the lungs compared with the original COVID-19 strain or Delta strain 24 h post-infection (10). The BA.2 variant, with eight unique spike alterations compared with the BA.1 variant (11), has a higher effective reproduction number, higher fusogenicity, and higher pathogenic potential than the BA.1 variant (12). This evidence reasoned the rapid spread of Omicron BA. 2.2 in Shanghai with most infected victims being asymptomatic or having mild illness. In our study, 95.59% of the included patients had asymptomatic infections (12,911/13,508). Of the remaining 597 cases with mild illness, only 240 cases presented with fever.

As reported by Maslo et al., a significantly decreased demand in oxygen therapy was observed during the Omicron wave compared with the oxygen need during the Delta wave in South Africa (17.6 vs. 74%;  $P < 0.001$ ); this was also the case for the percentage of mechanical ventilation requirement (1.6 vs. 12.4%;  $P < 0.001$ ) (13). Maslo et al. suggested that the decreased severity of Omicron infection may be associated with vaccination. Now, this new sub-lineage of BA.2 is causing mayhem in many countries, such as the UK, Australia, and China (14). Previously, the outbreak of Omicron BA. 2.2 in Hong Kong claimed 6,356 lives, and 90% of these deceased people had not received the COVID-19 vaccination (15).

In many studies, researchers have shown that Omicron variants have substantial immunologic escape ability from neutralizing antibodies induced by vaccination (16–19). However, a booster dose of mRNA vaccines, either Moderna or Pfizer, can still induct consistent neutralizing antibody titers against either BA.1 or BA.2 (19, 20). In a recent study, researchers also suggested that promoting the booster shot is an effective means to prevent the transmission of SARS-CoV-2, particularly to withstand the transmission of the Omicron strain. Moreover, a retrospective study in which 23,391 COVID-19 cases and 46,64 controls from a nationwide pharmacy-based testing program in the US were analyzed and showed that the prior receipt of three mRNA vaccine doses prevented the development of symptomatic SARS-CoV-2 for both Omicron and Delta. In the present study, it was first confirmed that two doses of SARS-CoV-2 vaccine (BBIBP-CorV or CoronaVac) increased the probability of SARS-CoV-2 nucleic acid negative conversion within 14 days after infection. Three doses of CoronaVac or a booster dose of BBIBP-CorV shortened the duration of virus clearance within 7 days and 14 days after infection. Hence, it makes sense to devise different vaccination rollouts optimally (21).

In the present study, the median duration of RNA shedding was 8 days, which is shorter than the 2 weeks of RNA shedding of the original SARS-CoV-2 strain reported previously (4). Nationwide vaccination may contribute to the accelerated viral clearance of the Omicron BA. 2.2 wave. It was noticed that age is associated with prolonged viral RNA shedding, which is consistent with a previous report (4). This can also be

attributed to the age-dependent impairment of innate and adaptive immunity, which could make it more difficult for older patients to eradicate pathogens. In contrast to earlier findings, however, in this study, the results showed that female patients tended to have prolonged RNA shedding compared with male patients at day 14, but tended to become negative at day 7 compared with male patients. This contradictory finding may involve an underlying mechanism for sex-related differences in SARS-CoV-2 clearance related to the presence of sex hormones, which influence different components of the immune system.

Understanding the recovery of COVID-19 during Omicron BA. 2.2 is of great importance in terms of enabling the government to set up a quarantine and lockdown policy to inform public health guidance. Fangcang shelter hospitals provide isolation for asymptomatic infection or mild illness, and are characterized by rapid construction, massive scale, and low cost (2). In the present study, it was noticed that cough medicine increased the probability of recovery on day 7, which reinforces the importance of oral prescription during quarantine in Fangcang shelter Hospitals. At the summit of the Shanghai pandemic, the average number of new cases was >20,000 per day, and the timely turnover of Fangcang shelter Hospitals was important for preventing panic caused by medical shortages. Using the 7-day and 14-day recovery machine learning-based predictive model (SMOTEENN/RF), the recovery of patients was predicted and it was possible to prepare for available beds within the subsequent 1 or 2 weeks. Machine learning-based models have been used in the screening, diagnosis, severity assessment, prognosis, epidemiology, and even dietary guidance of COVID-19 (21–26) (Table 2). In the present study, a machine learning-based model was used to predict the recovery of Omicron BA. 2.2 infected COVID-19 patients, and the results can be used as a reference to guide quarantine or isolation policy.

The study has several limitations. First, although it has been reported that the virus strains detected in Shanghai were all Omicron BA. 2.2, personalized genotyping of the virus was not available. Second, information about the vaccination type was not accessible because clinical data were collected using a self-report WeChat Mini Program. Third, external validation is required to test the accuracy of the proposed models.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

## Ethics statement

The studies involving human participants were reviewed and approved by Medical Ethics Committee of the Second

Affiliated Hospital of the PLA Army Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

## Author contributions

ZX is the guarantor of the manuscript. YX wrote the manuscript and generated the figures and tables. WY and QS researched the data. LS reviewed/edited the manuscript before submission. YW managed the result analysis. TL, HC, CS, and CH contributed to a discussion on the content and writing of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Key Support Object Training Project of Army Medical University (Third Military Medical University) (No. 2019R025) and the National Natural Science Foundation of China (No. 82173621).

## References

- Zhang X, Zhang W, Chen S. Shanghai's life-saving efforts against the current omicron wave of the COVID-19 pandemic. *Lancet*. (2022) 399:2011–2. doi: 10.1016/S0140-6736(22)00838-8
- Cai J, Deng X, Yang J, Sun K, Liu H, Chen Z, et al. Modeling transmission of SARS-CoV-2 Omicron in China. *Nat Med*. (2022) 28:1468–1475. doi: 10.1038/s41591-022-01855-7
- Mack CD, Wasserman EB, Killerby ME, Soelaeman RH, Hall AJ, MacNeil A, et al. Results from a test-to-release from isolation strategy among fully vaccinated national football league players and staff members with COVID-19 - United States, December 14–19:2021. *MMWR Morb Mortal Wkly Rep*. (2022) 71:299–305. doi: 10.15585/mmwr.mm7108a4
- Cogliati Dezza F, Oliva A, Cancelli F, Savelloni G, Valeri S, Mauro V, et al. Determinants of prolonged viral RNA shedding in hospitalized patients with SARS-CoV-2 infection. *Diagn Microbiol Infect Dis*. (2021) 100:115347. doi: 10.1016/j.diagmicrobio.2021.115347
- Chen X, Zhu B, Hong W, Zeng J, He X, Chen J, et al. Associations of clinical characteristics and treatment regimens with the duration of viral RNA shedding in patients with COVID-19. *Int J Infect Dis*. (2020) 98:252–260. doi: 10.1016/j.ijid.2020.06.091
- WHO. *COVID-19 Treatment Guidelines* (2022). Available at: <https://www.covid19treatmentguidelines.nih.gov/> (accessed May 26, 2022).
- National Health Commission & National Administration of Traditional Chinese Medicine. *Diagnosis and treatment guidelines of COVID-19 (Version 9.0)*. (2022). Available at: [http://www.gov.cn/zhengce/zhengceku/2022-03/15/content\\_5679257.htm](http://www.gov.cn/zhengce/zhengceku/2022-03/15/content_5679257.htm) (accessed May 26, 2022).
- Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *Sigkdd Explor*. (2022) 6:20–9. doi: 10.1145/1007730.1007735
- Weekly. *WHO epidemiological update on COVID-19* (2022). Available at: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19> (accessed on |May 18, 2022).
- Hui KPY, Ho JCW, Cheung MC, Ng KC, Ching RHH, Lai KL, et al. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature*. (2022) 603:715–720. doi: 10.1038/s41586-022-04479-6
- Iketani S, Liu L, Guo Y, Liu L, Chan JF, Huang Y, et al. Antibody evasion properties of SARS-CoV-2 Omicron sublineages. *Nature*. (2022) 604:553–556. doi: 10.1038/s41586-022-04594-4
- Yamasoba D, Kimura I, Nasser H, Morioka Y, Nao N, Ito J, et al. Virological characteristics of the SARS-CoV-2 Omicron BA.2 spike. *Cell*. (2022) 185:2103–2115.e19. doi: 10.1016/j.cell.2022.04.035
- Maslo C, Friedland R, Toubkin M, Laubscher A, Akaloo T, Kama B. Characteristics and outcomes of hospitalized patients in South Africa during the COVID-19 omicron wave compared with previous waves. *JAMA*. (2022) 327:583–584. doi: 10.1001/jama.2021.24868
- CoVariants 2022. *Overview of Variants/Mutations* (2022). Available at: <https://covariants.org/per-variant> (accessed May 19, 2022).
- Cheung PH, Chan CP, Jin DY. Lessons learned from the fifth wave of COVID-19 in Hong Kong in early 2022. *Emerg Microbes Infect*. (2022) 11:1072–8. doi: 10.1080/22221751.2022.2060137
- Schmidt F, Muecksch F, Weisblum Y, Da Silva J, Bednarski E, Cho A, et al. Plasma Neutralization of the SARS-CoV-2 Omicron Variant. *N Engl J Med*. (2022) 386:599–601. doi: 10.1056/NEJMc2119641
- Liu L, Iketani S, Guo Y, Chan JF, Wang M, Liu L, et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature*. (2022) 602:676–681. doi: 10.1038/s41586-021-04388-0
- Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, Tegally H, et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature*. (2022) 602:654–656. doi: 10.1038/s41586-021-04387-1
- Yu J, Collier AY, Rowe M, Mardas F, Ventura JD, Wan H, et al. Neutralization of the SARS-CoV-2 Omicron BA.1 and BA.2 Variants. *N Engl J Med*. (2022) 386:1579–1580. doi: 10.1056/NEJMc2201849

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.1001801/full#supplementary-material>

20. Accorsi EK, Britton A, Fleming-Dutra KE, Smith ZR, Shang N, Derado G, et al. Association between 3 doses of mRNA COVID-19 vaccine and symptomatic infection caused by the SARS-CoV-2 omicron and delta variants. *JAMA*. (2022) 327:639–651. doi: 10.1001/jama.2022.0470
21. Parino F, Zino L, Calafiore GC, Rizzo A. A model predictive control approach to optimally devise a two-dose vaccination rollout: A case study on COVID-19 in Italy. *Int J Robust Nonlinear Control*. (2021) 2021. doi: 10.1002/rnc.5728
22. Verde L, De Pietro G, Ghoneim A, Alrashoud M, Al-Mutib KN, Sannino G. Exploring the use of artificial intelligence techniques to detect the presence of coronavirus Covid-19 through speech and voice analysis. *IEEE Access*. (2021) 9:65750–65757. doi: 10.1109/ACCESS.2021.3075571
23. Park MS, Jo H, Lee H, Jung SY, Hwang HJ. Machine learning-based COVID-19 patients triage algorithm using patient-generated health data from nationwide multicenter database. *Infect Dis Ther*. (2022) 11:787–805. doi: 10.1007/s40121-022-00600-4
24. ElAraby ME, Elzeki OM, Shams MY, Mahmoud A, Salem H. A novel gray-scale spatial exploitation learning net for COVID-19 by crawling internet resources. *Biomed Signal Process Control*. (2022) 73:103441. doi: 10.1016/j.bspc.2021.103441
25. Amaral F, Casaca W, Oishi CM, Cuminato JA. Towards providing effective data-driven responses to predict the Covid-19 in São Paulo and Brazil. *Sensors (Basel)*. (2021) 21:540. doi: 10.3390/s21020540
26. Shams MY, Elzeki OM, Abouelmagd LM, Hassanien AE, Elfattah MA, Salem H. HANA: |covid-19 pandemic. *Comput Biol Med*. (2021) 135:104606. doi: 10.1016/j.combiomed.2021.104606

## COPYRIGHT

© 2022 Xu, Ye, Song, Shen, Liu, Guo, Liu, Wu, Wang, Sun, Bai, Luo, Liao, Chen, Song, Huang, Wu and Xu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.