



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect

journal homepage: [www.jfma-online.com](http://www.jfma-online.com)

## Original Article

# The COVIDTW study: Clinical predictors of COVID-19 mortality and a novel AI prognostic model using chest X-ray

Chih-Wei Wu <sup>a,b,#</sup>, Bach-Tung Pham <sup>c,#</sup>, Jia-Ching Wang <sup>c</sup>,  
Yao-Kuang Wu <sup>a</sup>, Chan-Yen Kuo <sup>d</sup>, Yi-Chiung Hsu <sup>b,\*</sup>

<sup>a</sup> Division of Pulmonary Medicine, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan

<sup>b</sup> Department of Biomedical Sciences and Engineering, National Central University, Taoyuan, Taiwan

<sup>c</sup> Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan

<sup>d</sup> Department of Research, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan

Received 14 March 2022; received in revised form 19 August 2022; accepted 21 September 2022

**KEYWORDS**

COVID-19;  
Artificial intelligence;  
Chest X-rays;  
Prognosis;  
Mortality;  
Intensive care unit

**Background:** There is a lack of published research on the impact of the first wave of the COVID-19 pandemic in Taiwan. We investigated the mortality risk factors among critically ill patients with COVID-19 in Taiwan during the initial wave. Furthermore, we aim to develop a novel AI mortality prediction model using chest X-ray (CXR) alone.

**Method:** We retrospectively reviewed the medical records of patients with COVID-19 at Taipei Tzu Chi Hospital from May 15 to July 15 2021. We enrolled adult patients who received invasive mechanical ventilation. The CXR images of each enrolled patient were divided into 4 categories (1st, pre-ETT, ETT, and WORST). To establish a prediction model, we used the MobileNetV3-Small model with “Imagenet” pretrained weights, followed by high Dropout regularization layers. We trained the model with these data with Five-Fold Cross-Validation to evaluate model performance.

**Result:** A total of 64 patients were enrolled. The overall mortality rate was 45%. The median time from symptom onset to intubation was 8 days. Vasopressor use and a higher BRIXIA score on the WORST CXR were associated with an increased risk of mortality. The areas under the curve of the 1st, pre-ETT, ETT, and WORST CXRs by the AI model were 0.87, 0.92, 0.96, and 0.93 respectively.

\* Corresponding author. Department of Biomedical Sciences and Engineering, National Central University, No. 300, Zhongda Road, Zhongli District, Taoyuan City 32001, Taiwan. Fax: +886 3 425 3427.

E-mail address: [syic@ncu.edu.tw](mailto:syic@ncu.edu.tw) (Y.-C. Hsu).

# Chih-Wei Wu and Bach-Tung Pham contributed equally to the manuscript.

<https://doi.org/10.1016/j.jfma.2022.09.014>

0929-6646/Copyright © 2022, Formosan Medical Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Please cite this article as: C.-W. Wu, B.-T. Pham, J.-C. Wang et al., The COVIDTW study: Clinical predictors of COVID-19 mortality and a novel AI prognostic model using chest X-ray, Journal of the Formosan Medical Association, <https://doi.org/10.1016/j.jfma.2022.09.014>

**Conclusion:** The mortality rate of COVID-19 patients who receive invasive mechanical ventilation was high. Septic shock and high BRIXIA score were clinical predictors of mortality. The novel AI mortality prediction model using CXR alone exhibited a high performance. Copyright © 2022, Formosan Medical Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

The first coronavirus disease 2019 (COVID-19) outbreak occurred in Wuhan city in China in December 2019. The pathogen of COVID-19 was identified as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). COVID-19 is highly contagious and has rapidly led to a global pandemic. As of December 15, 2021, there were approximately 270 million confirmed cases and 5.3 million confirmed deaths worldwide.<sup>1</sup>

Due to effective public health policies, there were only small-scale outbreaks in Taiwan. The total population of Taiwan in 2021 was approximately 23.6 million people. As of December 15, 2021, there were only 16,759 confirmed COVID-19 cases and 849 COVID-19 deaths in Taiwan.<sup>2</sup> The first large-scale COVID-19 outbreak in Taiwan took place between May 15 and July 15, 2021. There were 14,052 new cases of COVID-19 during this period.<sup>2</sup> As of May 15, 2021, only 316,200 doses of the COVID-19 vaccine (AstraZeneca) were available in Taiwan.<sup>2</sup> Almost all patients with COVID-19 at the initial wave of infection were unvaccinated. Before the advent of the COVID-19 vaccines, a meta-analysis reported that the estimated mortality rate of patients with COVID-19 receiving invasive mechanical ventilation (IMV) was 45%.<sup>3</sup> Real-world data are limited in Taiwan. We aimed to investigate the outcomes of patients with COVID-19 receiving IMV in Taiwan.

Scoring systems of CXR are reproducible and reliable tools for predicting the risk of intensive care unit (ICU) admission or mortality among patients with COVID-19.<sup>4,5</sup> Such scoring systems include the BRIXIA<sup>6</sup> and percent opacification. We assessed whether these scoring could be applied to COVID-19 in Taiwan. Several mortality prediction scores exist, such as the DICE score<sup>7</sup> and the 4C Mortality Score.<sup>8</sup> These are based on clinical parameters, including sex, age, comorbidities, serum biomarkers, and blood oxygen levels. We aimed to identify such clinical predictors in the Taiwanese population. Artificial intelligence (AI) has recently demonstrated great applicability in medicine. AI models using CXR to prognosticate COVID-19 outcomes are limited. We aimed to build a novel AI prediction model to predict COVID-19 mortality based on CXR alone. Because this study presents data on the first large-scale COVID-19 outbreak in Taiwan, we named the study COVIDTW. The novel AI model was named the COVIDTW model.

## Materials and methods

We retrospectively reviewed the medical records of patients with COVID-19 at Taipei Tzu Chi Hospital from May 15 to July 15, 2021. All enrolled patients had reverse transcriptase-polymerase chain reaction (RT-PCR)-confirmed COVID-19. We excluded patients who did not

receive IMV or were not admitted to the ICU. Patients aged <18 y/o were also excluded. All patients underwent the first CXR (1st CXR) at the emergency department. We extracted baseline characteristics, including age, sex, body mass index (BMI), the first cycle threshold (CT) value of RT-PCR, the ability to perform activities of daily living (dependence or independence), smoking history, educational attainment (cut-point: bachelor's degree or higher), comorbidities, medications, COVID-19 complications, and serum biomarkers, including serum D-dimer level, C-reactive protein (CRP) level, and albumin level.

On the day of endotracheal intubation, we collected data on white blood cell count, CRP level, lactate, blood oxygen level (P/F ratio =  $\text{PaO}_2/\text{FiO}_2$ ), and ventilator setting {positive end respiratory pressure (PEEP) + pressure control level (PC) = peak inspiratory pressure (PIP)}. All patients received pressure-controlled ventilation. The maximum PIP was less than 40 cmH<sub>2</sub>O. We calculated the P/F ratio from the arterial blood gas (1st ABG) obtained about 2 h after endotracheal intubation. The physician and respiratory therapist adjusted the ventilator settings according to the 1st ABG. In the current study, we analyzed the adjusted ventilator settings to predict mortality.

The serial CXR images of each patient were labeled as the 1st CXR (the first CXR in the emergency room), the pre-ETT CXR (the CXR immediately before endotracheal intubation), the ETT CXR (the CXR immediately after endotracheal intubation), and the WORST CXR (the worst CXR during hospitalization). Dr. Chih-Wei Wu (a pulmonologist with 13 years of experience in thoracic radiology) reviewed all CXR images and calculated the BRIXIA and percent opacification scores of each. The BRIXIA score<sup>6</sup> is a semi-quantitative score and was also calculated by Dr. Yao-Kuang Wu (a pulmonologist with 33 years of experience in thoracic radiology). The mean BRIXIA score was used to predict the mortality. The interclass correlation coefficient (ICC) was used to evaluate the agreement between two experienced experts. We have demonstrated the representative CXR figures of the BRIXIA scores in [Supplementary Figs. 1–5](#). [Supplementary Fig. 4](#) shows the representative WORST CXR of a non-survivor with a high BRIXIA score. [Supplementary Fig. 5](#) shows the representative WORST CXR of a survivor with a relatively low BRIXIA score.

During the COVIDTW study, all patients were unvaccinated, and anti-SARS-CoV-2 monoclonal antibodies were unavailable. Systemic dexamethasone 6 mg QD up to 10 days was routinely administered to all patients receiving IMV according to the proven survival benefits.<sup>9</sup> Tocilizumab was routinely administered to patients with serum CRP levels  $\geq 7.5$  mg/dL according to the RECOVERY study.<sup>10</sup> Physicians used a combination of midazolam, fentanyl, or cisatracurium to achieve patient-ventilator synchrony. If  $\text{PaO}_2/\text{FiO}_2 < 100$ , the intensivist would consider prone

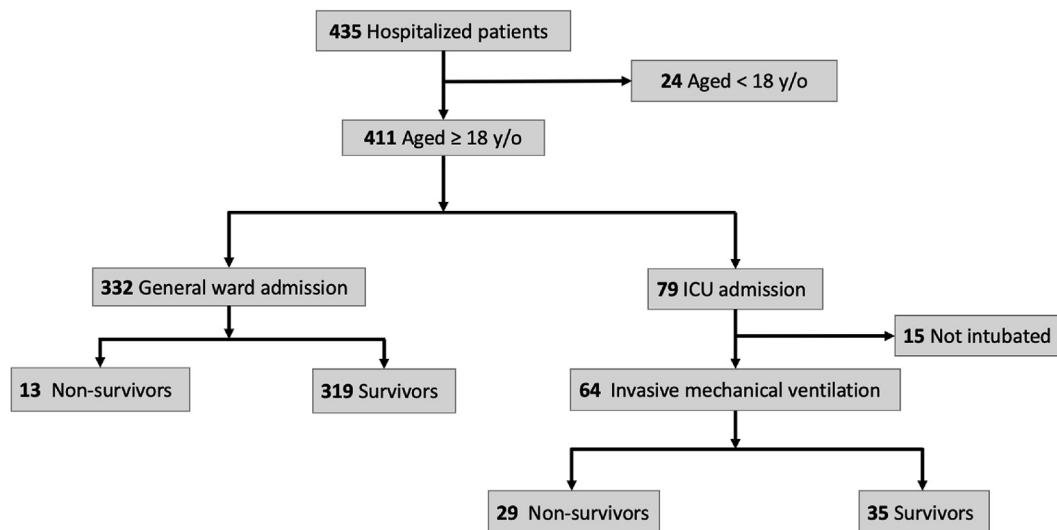


Fig. 1 Flowchart of patient enrollment.

positioning or implementation of extracorporeal membrane oxygenation (ECMO) in patients with severe acute respiratory distress syndrome (ARDS).

The natural history of COVID-19 critical illness was presented by a time-to-events table. Day 1 was defined as symptom onset. The events included tracheal intubation, WORST CXR, peak serum CRP level, peak serum D-dimer level, and nadir serum albumin level. We also present other clinical courses, including length of ICU stay, length of hospital stay, and duration of IMV, of deceased and survived patients.

We used Prism 9 statistical software to analyze the data. The Mann–Whitney U test was used to compare non-Gaussian continuous variables. Fisher’s exact test was used to compare the categorical variables. Logistic regression was used to analyze data with binary outcomes. We utilized the Kaplan–Meier method to plot the time-to-event figure and log-rank test to compare the differences. Differences between serial CXR images was tested by one-way analysis of variance (ANOVA). A  $p$  value  $< 0.05$  was considered statistically significant.

### Experimental setup for artificial intelligence

The small size of the dataset was a major challenge for this research. Deep learning models would easily overfit with the training data, and not perform well with the testing set.

To overcome the data limitation problem, the use of a lightweight model is important. In this work, we used the MobilenetV3-Small model with the “ImageNet” pretrained weights, followed by high Dropout regularization layers specifically to address the overfitting problem. We chose the sigmoid function (Supplementary functional equation 1) as the output activation to achieve binary classification.

Regarding MobileNetV3, in addition to the efficient last stage, the lightweight model introduces a combination of hardware-aware network architecture search (NAS) complemented by the NetAdapt algorithm. Moreover, the network design includes the use of a hard-swish activation and squeeze-and-excitation modules in the “MBCConv” blocks. The swish nonlinearity is listed in Supplementary functional

equation 2. This has been proven experimentally to improve accuracy. However, as the sigmoid function is computationally expensive, it was modified to produce the hard swish or h-swish function (Supplementary functional equation 3).

In addition, we deployed some preprocessing layers within the model, which only applies to the training procedure to conduct data augmentation. These preprocessing layers augment the image data by rotation, translation, flipping, and random contrast adjustment to increase the amount of relevant image data given our limited dataset. Early stopping was also applied to reduce overfitting.

We used the binary cross-entropy as the loss function (Supplementary functional equation 4). Adam was chosen for the optimizer; we first pretrained the model at a learning rate of 0.001 and then fine-tuned it with a slower learning rate of 0.0001.

### Data preparation for artificial intelligence

We collected the X-ray images of 64 COVID-19 patients. Each person’s data consisted of four X-ray images representing the four groups of diagnostic procedures: 1st CXR, pre-ETT, ETT, and WORST. Our goal was to predict the probability of mortality in each group. In the current research, we performed an experiment with the first three diagnostic groups.

For each state, we trained the model with five-fold cross-validation method to evaluate its performance. The performance was assessed as the average of the five-fold area under the receiver operating characteristic curve (AUROC), accuracy, positive predictive value, sensitivity, and F-1 score. For more details, please refer to the Supplementary section on data preparation for AI.

### Ethical statement

The study was approved by the Institutional Review Board of Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation (Protocol Number: 10-X-045), and the requirement

**Table 1** The baseline characteristics of patients with COVID-19 and clinical risk factors for mortality.

Characteristics	Deceased patients n = 29 (45%)	Survived patients n = 35 (55%)	P value	Total n = 64 (100%)	P value of univariate logistic regression	P value of multivariate logistic regression	Odds ratio with 95% CI of multivariate logistic regression
Sex, n (%)			0.690		0.690		
Female	11 (38%)	15 (43%)		26 (41%)			
Male	18 (62%)	20 (57%)		38 (59%)			
Age (years), median (IQR)	70 (64.5–79)	63 (51–70)	0.003 <sup>e</sup>	67 (59.5–74)	0.005 <sup>e</sup>	0.135	1.078 (0.985–1.208)
Body Mass Index, median (IQR)	25.3 (22.6–28.0)	25.8 (23.1–28.9)	0.5265	25.6 (22.8–28.5)	0.505		
1st CT value in RT-PCR, median (IQR)	21 (18.5–26)	22 (19–27)	0.589	21.5 (19–27)	0.527		
Functional status, n (%)			0.119		0.098		
independence	21 (72%)	31 (89%)		52 (81%)			
dependence	8 (28%)	4 (11%)		12 (19%)			
Ever-smoker, n (%)	8 (28%)	9 (26%)	0.866	17 (27%)	0.866		
Education level, n (%) (Bachelor's degree or higher)	3 (10%)	11 (31%)	0.07	14 (22%)	0.036 <sup>e</sup>	0.9611	1.086 (0.030–30.65)
Comorbidities, n (%)							
Any	22 (76%)	29 (83%)	0.489	51 (80%)	0.490		
Hypertension	17 (59%)	19 (54%)	0.728	36 (56%)	0.728		
Diabetes Mellitus	14 (48%)	15 (43%)	0.665	29 (45%)	0.665		
Dyslipidemia	5 (17%)	12 (34%)	0.160	17 (26%)	0.119		
Congestive heart failure	2 (7%)	2 (6%)	>0.999	4 (6%)	0.846		
Coronary artery disease	2 (7%)	1 (3%)	0.586	3 (5%)	0.446		
COPD	2 (7%)	2 (6%)	>0.999	4 (6%)	0.846		
ESRD	1 (3%)	1 (3%)	>0.999	2 (3%)	0.893		
Cancer	2 (7%)	0 (0%)	0.201	2 (3%)	NA <sup>b</sup>		
Stroke	3 (10%)	1 (3%)	0.321	4 (6%)	0.213		
Hypothyroidism	1 (3%)	3 (9%)	0.620	4 (6%)	0.387		
Prone positioning, n (%)	4 (14%)	2 (6%)	0.396	6 (9%)	0.269		
ECMO, n (%)	1 (3%)	0 (0%)	0.453	1 (2%)	NA <sup>b</sup>		
Medications, n (%)							
Enoxaparin use	26 (90%)	26 (74%)	0.198	52 (81%)	0.109		
Remdesivir use	20 (69%)	25 (71%)	0.830	45 (70%)	0.830		
Tocilizumab use	23 (79%)	22 (63%)	0.152	45 (70%)	0.147		
Triple anesthetics <sup>a</sup>	23 (79%)	26 (74%)	0.770	49 (75%)	0.636		
Total number of antibiotics	4 (2–5)	3 (1–4)	0.025 <sup>e</sup>	3 (2–5)	0.05		
Ventilator settings, median (IQR)						0.156	0.993 (0.981–1.002)
PaO <sub>2</sub> /FiO <sub>2</sub>	108.2 (77.2–221.9)	225 (137.4–316.7)	0.004 <sup>e</sup>	177.4 (93.5–283)	0.014 <sup>e</sup>		
Positive end expiratory pressure (cmH <sub>2</sub> O)	8 (8–10)	8 (8–10)	0.127	8 (8–10)	0.161		
Pressure control level (cmH <sub>2</sub> O)	18 (15–20)	18 (16–20)	0.345	18 (16–20)	0.614		
Peak inspiratory pressure (cmH <sub>2</sub> O)	28 (24–30)	26 (24–28)	0.159	26 (24–30)	0.240		

Complications, n (%)								
Vasopressor use	27 (93%)	10 (28%)	<0.001 <sup>e</sup>	37 (58%)	<0.001 <sup>e</sup>	0.040 <sup>e</sup>	26.81 (1.584–1003)	
Bacteremia	5 (17%)	1 (3%)	0.08	6 (9%)	0.04 <sup>e</sup>	0.213	78.26 (0.587- >10000)	
Acute kidney injury	26 (90%)	13 (37%)	<0.001 <sup>e</sup>	39 (61%)	<0.001 <sup>e</sup>	0.728	0.584 (0.023–11.67)	
Serum markers, median (IQR)								
White blood cell count (uL)	7190 (4915–11835)	8890 (5990–11240)	0.443	8080 (5225–11143)	0.497			
Lactate (mmol/L)	1.7 (0.9–2.3)	1.6 (1–2.1)	0.864	1.6 (1–2.1)	0.771			
Peak D-dimer value (ng/mL)	10000 <sup>d</sup> (6659–10000)	6672 <sup>d</sup> (3544–9696)	0.004 <sup>e</sup>	8983 <sup>d</sup> (4367–10000)	0.008 <sup>e</sup>	d	d	
1st D-dimer value (ng/mL)	1253 (662–3763)	1442 (675–7304)	0.8694	1314 (671–6141)	0.386			
Peak CRP value (mg/dL)	15.6 (10.1–20.4)	11.3 (6.7–16.9)	0.061	12.0 (9.1–18.5)	0.102			
ETT CRP value (mg/dL)	8.9 (3.8–15.0)	10.5 (2.9–16.9)	0.925	9.0 (3.7–16.5)	0.921			
1st CRP value (mg/dL)	8.5 (4.2–15.3)	8.8 (2.1–14.8)	0.730	8.7 (2.7–14.7)	0.444			
Nadir albumin value (g/dL)	2.5 (2.3–2.8)	2.9 (2.6–3.1)	<0.001 <sup>e</sup>	2.7 (2.5–3.1)	<0.001 <sup>e</sup>			
1st albumin value (g/dL)	3.3 (3.1–3.9)	3.4 (3.0–3.6)	0.574	3.4 (3.1–3.7)	0.418	0.311	0.223 (0.009–3.888)	
CXR categories, median (IQR)								
BRIXIA score of 1st CXR	7 (4–12)	6 (3–9)	0.151	7 (4–10)	0.125			
Percent opacification score of 1st CXR	50% (18%–80%)	35% (15%–50%)	0.104	40% (15%–64%)	0.081			
BRIXIA score of the pre-ETT CXR	11 (8–13)	10 (8–14)	0.997	10 (8–13)	0.859			
Percent opacification score of the pre-ETT CXR	70% (50%–83%)	65% (45%–80%)	0.379	70% (50%–80%)	0.323			
BRIXIA score of the ETT CXR	12 (10–16)	11 (9–14)	0.150	12 (9–15)	0.123			
Percent opacification score of the ETT CXR	80% (68%–90%)	80% (55%–90%)	0.257	80% (61%–90%)	0.134			
BRIXIA score of the WORST CXR	17 (15–18)	13 (10–16)	<0.001 <sup>e</sup>	15 (12–17)	<0.001 <sup>e</sup>	0.038 <sup>e</sup>	2.042 (1.107–4.394)	
Percent opacification score of the WORST CXR	95% (90%–100%)	80% (60%–90%)	<0.001 <sup>e</sup>	90% (75%–95%)	<0.001 <sup>e</sup>	0.7116	0.105 (0.001–17161)	

Data are presented as the number (percentage) or median  $\pm$  interquartile range.

Abbreviations: IQR = interquartile range, CT = cycle threshold, RT-PCR = reverse transcription-polymerase chain reaction, COPD = chronic obstructive pulmonary disease, ESRD = end stage renal disease, CRP = C-reactive protein, ETT-CRP = the CRP level on the day of endotracheal intubation, CI = confidence interval, CXR = chest X-ray, ECMO: extracorporeal membrane oxygenation, 1st CXR = the first CXR obtained at the emergency department, pre-ETT CXR = the CXR immediately before endotracheal intubation, ETT CXR = the CXR immediately after endotracheal intubation, WORST CXR = the worst CXR during hospitalization.

<sup>c</sup>: The odds ratio approaches 1.

<sup>a</sup> Triple anesthetics means that the patient received midazolam, fentanyl, and cisatracurium simultaneously.

<sup>b</sup> NA (not applicable): the logistic model was not fitted due to complete separation.

<sup>d</sup> If a D-dimer level is  $> 10000$  ng/mL, it is depicted as 10000 ng/mL (detection limit). Because 24 (38%) patients had a peak D-dimer  $> 10000$  ng/mL, peak D-dimer was not included in the multivariate logistic regression.

<sup>e</sup> Denotes statistical significance, which is also marked by gray shading.

**Table 2** Natural history of critically ill patients with COVID-19 (n = 64).

Time schedule (days), median + IQR	Deceased patients n = 29	Survived patients n = 35	All patients n = 64	Hazard ratio (95% CI)	P value
Day 1 = symptom onset					
Time to intubation	7 (4.5–11.5)	8 (13–31)	8 (5–12)	1.360 (0.820–2.254)	0.180
Time to the WORST CXR	19 (13–28.5)	11 (7–16)	15 (9–23.75)	0.4972 (0.300–0.825)	0.002 <sup>a</sup>
Time to peak serum CRP level	9 (4.5–20.5)	10 (5–12)	9 (5–13.75)	0.957 (0.586–1.565)	0.850
Time to peak serum D-dimer level	13 (8.5–16)	14 (10–21)	10 (13–17)	1.474 (0.885–2.455)	0.097
Time to nadir serum albumin level	15 (10.5–22)	15 (13–19)	15 (12–20)	1.024 (0.626–1.677)	0.919

Note: Data are presented as the median ± interquartile range or hazard ratio with 95% confidence interval.

Abbreviations: IQR = interquartile range, CI = confidence interval.

<sup>a</sup> Denotes statistical significance, which is also marked by gray shading.

**Table 3** Performances of the COVIDTW model.

Performance metric	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Average
AUROC						
1st CXR	0.898	0.905	0.845	0.875	0.814	0.868
Pre-ETT CXR	0.881	0.857	1.00	0.857	1.00	0.919
ETT CXR	1.00	0.952	0.976	0.929	0.943	0.960
WORST CXR	0.800	1.00	0.929	1.00	0.900	0.926
Accuracy						
1st CXR	92.9%	92.3%	84.6%	85.7%	83.3%	87.8%
Pre-ETT CXR	91.7%	92.3%	92.3%	92.3%	91.7%	92.1%
ETT CXR	92.3%	84.6%	92.3%	92.3%	100%	92.3%
WORST CXR	83.3%	100.0%	92.3%	100.0%	91.7%	93.5%

Abbreviation: AUROC = area under the receiver operating characteristic curve.

for informed consent was waived by the Institutional Review Board of Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation.

## Results

**Fig. 1** shows the flowchart of patient enrollment. A total of 435 patients were admitted for COVID-19. 79 patients were admitted to the ICU. Among the 64 patients receiving IMV in the ICU, 29 died, and 35 survived. The overall mortality of patients with COVID-19 receiving IMV was 45%. **Table 1** shows the baseline characteristics of patients with COVID-19 and the clinical risk factors for mortality. The first part of **Table 1** compares the differences between the deceased and survived patients. The second part of **Table 1** shows the P values of univariate logistic regression, P values of multivariate logistic regression, and 95% confidence interval of the odds ratio of possible risk factors for mortality. After univariate logistic regression, the statistically significant risk factors included older age, an education level below a bachelor's degree, a lower P/F ratio, the use of vasopressors, the presence of bacteremia, acute kidney injury, a higher peak serum D-dimer level, a lower nadir serum albumin level, and a higher BRIXIA or percent opacification score on the WORST CXR. The above nine risk factors were chosen for multivariate logistic regression and

D-dimer > 10,000 ng/mL was defined as 10,000 ng/mL (detection limit). Because 24 (38%) patients had a peak D-dimer > 10,000 ng/mL, peak D-dimer was not included in the multivariate logistic regression. After multivariate logistic regression, the statistically significant risk factors were vasopressor use and a higher BRIXIA score on the WORST CXR.

There were no statistically significant differences in the 1st, pre-ETT, and ETT CXR scores between the two groups. Only one 42-year-old woman received ECMO but unfortunately succumbed to progressive ARDS. The ICC of the 1st, pre-ETT, ETT, and WORST BRIXIA scores were 0.908, 0.820, 0.901, and 0.918, respectively.

**Table 2** presents the natural history of critically ill patients with COVID-19. The median time to intubation for all patients was 8 days. The deceased patients had a longer time to WORST CXR than the survived patients (19 days vs. 11 days, P = 0.002). **Supplementary Table 1** shows other clinical courses including the length of ICU stay, length of hospital stay, and duration of IMV.

**Table 3** presents the performance of the COVIDTW model. The average AUROCs were 0.87, 0.92, 0.96, and 0.93 for the 1st CXR, pre-ETT CXR, ETT, and WORST CXR, respectively. The average accuracies were 88%, 92%, 92%, and 94% for the 1st CXR, pre-ETT CXR, ETT, and WORST CXR, respectively. Other performance metrics for the COVIDTW model are shown in **Supplementary Table 2** (positive predictive value), **Supplementary Table 3** (sensitivity), and **Supplementary Table 4** (F-1 score).

Both the BRIXIA (**Fig. 2A**, P < 0.001) and percent opacification scores (**Fig. 2B**, P < 0.001) of the WORST CXR were significantly higher among deceased patients. There were no significant differences between the deceased and survived patients in other CXR stages (1st, pre-ETT, and ETT).

There were significant increases in the BRIXIA scores (**Fig. 3A**) and percent opacification scores (**Fig. 3B**) from the 1st CXR to pre-ETT, ETT, and WORST CXR.

## Discussion

The COVIDTW study presents the first large-scale COVID-19 outbreak in Taiwan. We investigated the outcomes of critically ill patients and clinical predictors of mortality. A novel AI model using CXR alone showed good performance in predicting mortality.

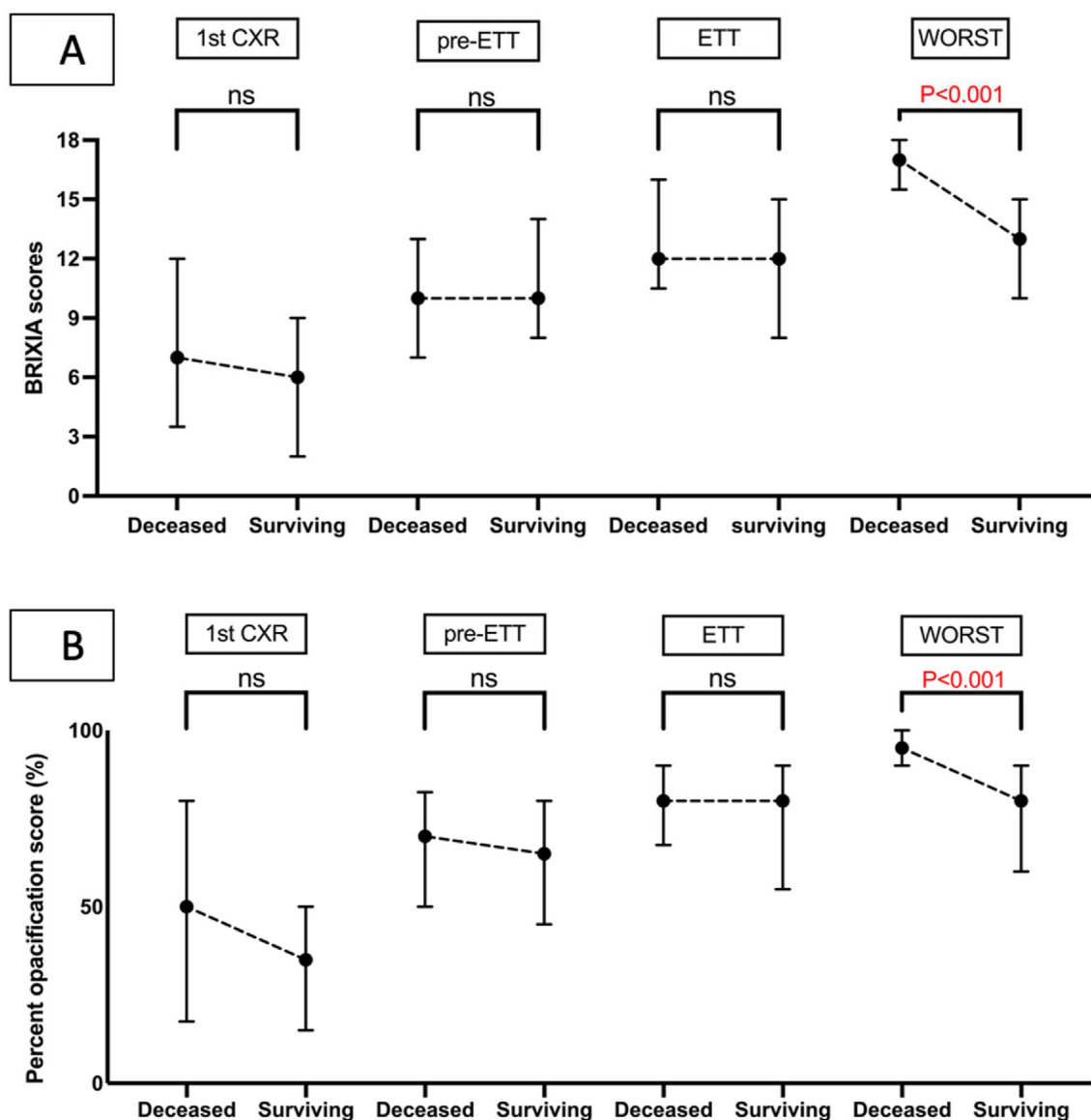


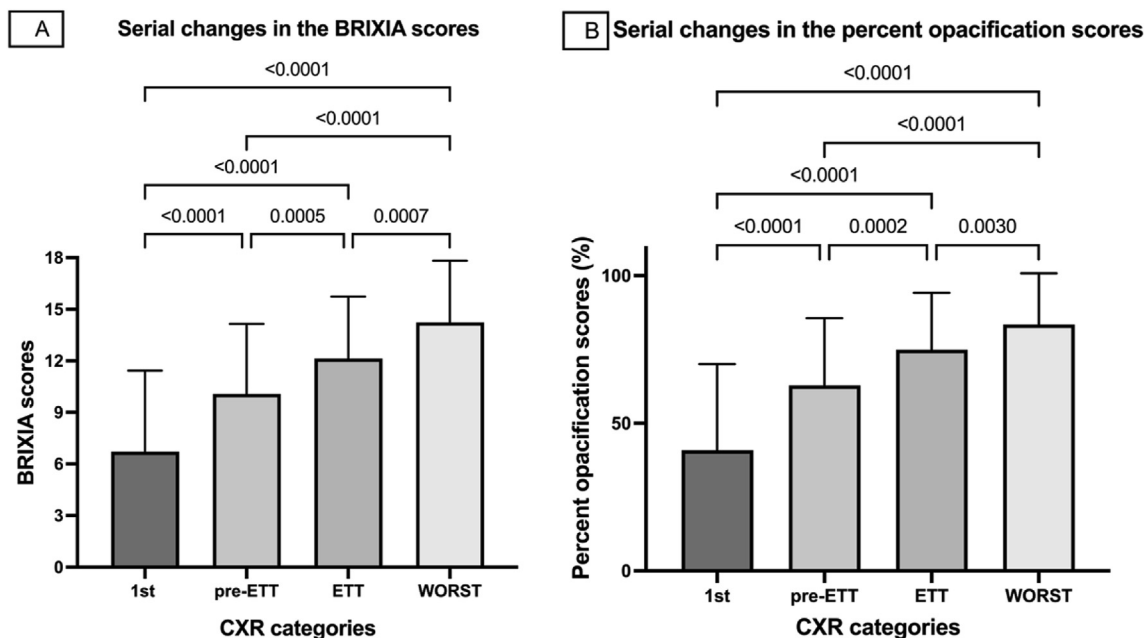
Fig. 2 (A) Comparison of BRIXIA scores between deceased and survived patients for different CXR images. (B) Comparison of percent opacification scores between deceased and survived patients for different CXR images. The plot shows the median with interquartile ranges. Abbreviation: ns = nonsignificant.

In this study, approximately 50 clinical parameters were extracted. After univariate logistic regression analysis, we identified nine significant risk factors. These nine risk factors were consistent with those reported in the literature. However, after multivariate logistic regression, only two risk factors were identified (use of vasopressors and a higher BRIXIA score on the WORST CXR.) as statistically significant. However, the factors like white blood cell counts, CRP, lactate, blood oxygen level, ventilator settings, and subsequent treatment are known to have a strong impact on mortality, but not in the COVIDTW study. The main reasons for the difference may result from difference in the study population, sample size, sampling time of serum biomarkers, time points of ventilator settings, etc. As for blood oxygen levels, in a retrospective study including 123 mechanically ventilated patients with COVID-

19 in China, the P/F ratio on the day of ICU admission was an independent risk factor for mortality.<sup>11</sup> In contrast to the abovementioned study, the time point of P/F ratio recording in the COVIDTW study was 2 h after endotracheal ventilation. Continuous changes in lung mechanics after COVID-19 infection are complicated. The most representative time point of the P/F ratio recording requires further investigation.

After endotracheal intubation in patients with COVID-19, few subsequent treatments have been shown to impact mortality. In the COVIDTW study, only patients with serum CRP level >7.5 mg/dL received Tocilizumab according to the mortality benefits from the RECOVERY study.<sup>10</sup> However, the optimal CRP cutoff value for mechanically ventilated patients with COVID-19 is unknown. A single-center study that included 154 mechanically ventilated patients





**Fig. 3** Serial changes in the BRIXIA scores (3A) and percent opacification scores (3B) for different CXR images. P values are shown in the figure. The plot shows the mean with standard deviation.

with COVID-19<sup>12</sup> did not delineate a specific CRP value for enrollment, and the results showed that Tocilizumab was associated with lower mortality. As for ventilator settings, limiting mechanical power<sup>13</sup> and use of low tidal volume ventilation<sup>14</sup> have mortality benefits for mechanically ventilated COVID-19 patients. However, in the COVIDTW study, the physicians did not calculate the mechanical power owing to limited resources, and there were no uniform protocols for low tidal volume ventilation. Prone positioning reduces mortality rates in moderate-to-severe ARDS due to COVID-19.<sup>15</sup> In the COVIDTW study, many real-world dilemmas, such as septic shock, acute gastrointestinal bleeding, and severe obesity, impeded routine use of prone positioning.

In the COVIDTW study, the severity of WORST CXR was associated with mortality. Previous studies focused on the first CXR image upon presentation and showed that the BRIXIA and percent opacification scores could predict COVID-19 mortality.<sup>4,5</sup> However, in the COVIDTW study, the 1st CXR scores were not significantly different between deceased and survived patients. The reason could be related to the different study populations. In the COVIDTW study, all patients received IMV and were admitted to the ICU, but the ICU admission rates were 8.3% (63/751)<sup>4</sup> and 17% (58/340)<sup>5</sup> in prior studies. In addition, in the COVIDTW study, the pre-ETT, and ETT CXR scores also had comparable severities between survived and deceased patients. This suggests that the physicians had similar judgments of respiratory failure and indications for tracheal intubation.

The COVIDTW study suggested that use of vasopressors was a risk factors for mortality. Our results are consistent with those reported in the literature. An observational study enrolling 217 critically-ill COVID-19 patients in the United States reported that vasopressor-requiring shock was significantly associated with mortality.<sup>16</sup> In total, 90%

of the deceased patients received vasopressors in contrast to the 54% of patients who survived. A retrospective study including 86 ICU patients with COVID-19 from Saudi Arabia revealed that septic shock was a significant predictor of death (odds ratio = 58,  $P < 0.001$ ).<sup>17</sup> Consequently, the health care system should focus on hemodynamically unstable patients and prevent complications in patients with COVID-19.

In the early era of the COVID-19 pandemic, the majority of AI studies focused on COVID-19 detection by CXR or chest computed tomography (chest CT). Later, AI-related studies attempted to generate a prognostic model based on chest CT images. However, CXR is more accessible and practical than chest CT in resource-limited health care systems. Few studies have developed AI models based on CXR images for predicting COVID-19 outcomes.<sup>18-20</sup> In the United States, Jiao et al. developed AI prediction models using the CXR images and clinical parameters of 1834 patients with COVID-19 to predict critical or noncritical outcome.<sup>18</sup> The AI performance (AUROC) using CXR alone was 0.753. A multicenter study in Italy developed 3 AI models to predict mild or severe outcomes based on the admission CXR images of 820 patients with COVID-19.<sup>19</sup> The AI performance (accuracy) using CXR alone ranged from 0.658 to 0.742. The EXAM model was built by using CXR images and clinical parameters of 16,148 patients.<sup>20</sup> The AUROC for predicting future oxygen requirement was  $>0.92$ . The above 3 studies utilized the earliest CXR (at ER or admission) to predict outcomes. However, the COVIDTW model employed 4 different stages of CXR (i.e., 1st CXR, pre-ETT, ETT, and WORST) to prognosticate the outcome. To our knowledge, the COVIDTW model is the first AI-based prediction model built using CXR images of intubated patients. A prognostic model based on the CXR obtained immediately after intubation (ETT CXR in the COVID model) is essential for critically ill patients.

Transportation of patients to the CT room has several disadvantages, such as catheter dislodgement, lack of oxygen support, and increased transmission risk. Further studies are required to develop reliable AI-based models using post-intubation CXR to prognosticate outcomes.

The COVIDTW study has inherent limitations, including its retrospective design and small sample size. Owing to small sample size, the COVIDTW model used five-fold cross-validation for internal validation. Internal validation usually leads to a higher performance of the prediction model than external validation. Optimally, a novel prognostic model should be externally validated using an independent dataset before incorporation into clinical practice. Further studies are required to enroll more patients with different baseline characteristics to improve the COVIDTW model.

In conclusion, the overall mortality rate of COVID-19 patients receiving IMV was 45%. The risk factors for COVID-19 mortality include the use of vasopressors and a higher BRIXIA score on the WORST CXR. The AI COVIDTW model uses CXR to predict COVID-19 mortality. The models built with the 1st, pre-intubation, post-intubation, and worst CXRs all achieved high performances.

## Access to data

The data are not publicly accessible. The corresponding author would consider sharing the data upon reasonable request.

## Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

## Acknowledgements

This study was supported by a grant from the Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation (TCRD-TPE-111-20).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jfma.2022.09.014>.

## References

1. WHO coronavirus (COVID-19) dashboard. 2021. at, <https://covid19.who.int/>.
2. Taiwan centers for disease control. at, <https://www.cdc.gov.tw/>.
3. Lim ZJ, Subramaniam A, Ponnappa Reddy M, Blecher G, Kadam U, Afroz A, et al. Case fatality rates for patients with COVID-19 requiring invasive mechanical ventilation. A meta-analysis. *Am J Respir Crit Care Med* 2021;203:54–66.
4. Au-Yong I, Higashi Y, Giannotti E, Fogarty A, Morling JR, Grainge M, et al. Chest radiograph scoring alone or combined with other risk scores for predicting outcomes in COVID-19. *Radiology* 2021;210986.
5. Balbi M, Caroli A, Corsi A, Milanese G, Surace A, Marco FD, et al. Chest X-ray for predicting mortality and the need for ventilatory support in COVID-19 patients presenting to the emergency department. *Eur Radiol* 2021;31:1999–2012.
6. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. *Radiol Med* 2020;125:509–13.
7. Nicholson CJ, Wooster L, Sigurssdottir HH, Li RH, Jiang W, Tian W, et al. Estimating risk of mechanical ventilation and in-hospital mortality among adult COVID-19 patients admitted to Mass General Brigham: the VICE and DICE scores. *EClinicalMedicine* 2021;33:100765.
8. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. 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.
9. Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, et al. Dexamethasone in hospitalized patients with covid-19. *N Engl J Med* 2021;384:693–704.
10. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet (London, England)* 2021;397:1637–45.
11. Gu Y, Wang D, Chen C, Lu W, Liu H, Lv T, et al. PaO<sub>2</sub>/FiO<sub>2</sub> and IL-6 are risk factors of mortality for intensive care COVID-19 patients. *Sci Rep* 2021;11:7334.
12. Somers EC, Eschenauer GA, Troost JP, Golob JL, Gandhi TN, Wang L, et al. Tocilizumab for treatment of mechanically ventilated patients with COVID-19. *Clin Infect Dis* 2020;73:e445–54.
13. Schuijt MTU, Schultz MJ, Paulus F, Neto AS, PROVENT–COVID Collaborative Group. Association of intensity of ventilation with 28-day mortality in COVID-19 patients with acute respiratory failure: insights from the PROVENT-COVID study. *Crit Care* 2021;25:283.
14. Nijbroek SGLH, Hol L, Ivanov D, Schultz MJ, Paulus F, Neto AS. Low tidal volume ventilation is associated with mortality in COVID-19 patients—insights from the PROVENT-COVID study. *J Crit Care* 2022;70:154047.
15. Shelhamer MC, Wesson PD, Solari IL, Jensen DL, Steele WA, Dimitrov VG, et al. Prone positioning in moderate to severe acute respiratory distress syndrome due to COVID-19: a cohort study and analysis of physiology. *J Intensive Care Med* 2021;36:241–52.
16. Auld SC, Caridi-Scheible M, Blum JM, Robichaux C, Kraft C, Jacob JT, et al. ICU and ventilator mortality among critically ill adults with coronavirus disease 2019. *Crit Care Med* 2020;48:e799–804.
17. Al Mutair A, Al Mutairi A, Zaidi ARZ, Salih S, Alhumaid S, Rabaan AA, et al. Clinical predictors of COVID-19 mortality among patients in intensive care units: a retrospective study. *Int J Gen Med* 2021;14:3719–28.
18. Jiao Z, Choi JW, Halsey K, Tran TML, Hsieh B, Wang D, et al. Prognostication of patients with COVID-19 using artificial intelligence based on chest x-rays and clinical data: a retrospective study. *Lancet Digit Health* 2021;3:e286–94.
19. Soda P, D'Amico NC, Tessadori J, Valbusa G, Guarraasi V, Bortolotto C, et al. AlforCOVID: predicting the clinical outcomes in patients with COVID-19 applying AI to chest-X-rays. An Italian multicentre study. *Med Image Anal* 2021;74:102216.
20. Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med* 2021;27:1735–43.