

Nowcast Deep Learning Models For Constraining Zero-Day Pathogen Attacks – Application on Chest Radiographs to Covid-19

Wan Hang Keith Chiu

Medical Artificial Intelligence Laboratory Program (MAIL), Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong

Varut Vardhanabhuti

Medical Artificial Intelligence Laboratory Program (MAIL), Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong

Dymtro Poplavskiy

Ensemble-group.com, Scottsdale, AZ

Philip LH Yu

Department of Statistics and Actuarial Sciences, The University of Hong Kong

Richard Du

Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong

Alistair Yun Hee Yap

Medical Artificial Intelligence Laboratory Program (MAIL), Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong

Sailong Zhang

Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong

Ambrose Ho-Tung Fong

Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong

Thomas Wing-Yan Chin

Department of Radiology and Imaging, Queen Elizabeth Hospital, Hong Kong SAR

Jonan Chun Yin Lee

Department of Radiology and Imaging, Queen Elizabeth Hospital, Hong Kong SAR

Siu Ting Leung

Department of Radiology, Pamela Youde Nethersole Eastern Hospital, Hong Kong SAR

Christine Shing Yen Lo

Department of Radiology, Queen Mary Hospital, Hong Kong SAR

Macy Mei-Sze Lui

8Department of Medicine, Queen Mary Hospital, Hong Kong SAR

Benjamin Xin Hao Fang

Department of Radiology, Queen Mary Hospital, Hong Kong SAR

Ming-Yen Ng

Medical Artificial Intelligence Laboratory Program (MAIL), Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong

Michael D. Kuo (✉ mdkuo@hku.hk)

Medical Artificial Intelligence Laboratory Program (MAIL), Department of Diagnostic Radiology, LKS Faculty of Medicine, The University of Hong Kong and Ensemble-group.com, Scottsdale, AZ, USA

Research Article

Keywords: Covid-19, Nowcast DL models, general pneumonia (PNA) CXR Deep Learning (DL) model (MAIL1.0), a second-generation DL model (MAIL2.0), detection of Covid-19 on chest radiographs

Posted Date: April 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-22078/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Outbreaks due to emergent pathogens like Covid-19 are difficult to contain as the time to gather sufficient information to develop a detection system is outpaced by the speed of transmission. Here we develop a general pneumonia (PNA) CXR Deep Learning (DL) model (MAIL1.0) followed by a second-generation DL model (MAIL2.0) for detection of Covid-19 on chest radiographs (CXR). We validate the models on two prospective cohorts of high-risk patients screened for Covid-19 reverse transcriptase-polymerase chain reaction (RT-PCR). MAIL1.0 has an Area Under the Receiver Operating Characteristics (AUC) of 0.93, sensitivity and specificity of 90.5% and 76.7% in detection of visible pneumonia and MAIL2.0 has an AUC of 0.81, sensitivity and specificity of 84.7% and 71.6%, significantly outperforming radiologists, especially amongst asymptomatic and patients presenting with early symptoms. Nowcast DL models may be an effective tool in helping to constrain the outbreak, particularly in resource-stretched healthcare systems.

Introduction

The first case of what is now known as Covid-19 was described on December 31, 2019 and was later revealed to be due to a heretofore never described betacoronavirus resembling SARS-CoV (Sars-CoV2)¹. Evidence of person-to-person transmission raised concerns of rapid spread and by January 30, 2020 the World Health Organisation (WHO) labelled this a global health emergency². By March 11, the WHO had declared the rapidly spreading Covid-19 outbreak a pandemic³. As of March 31st, 2020, there were over 784,000 cases with 37,638 deaths in over 178 countries/regions, sparking concerns of a global shutdown as the world struggled to contain the threat⁴.

Rapid deployment of detection systems in the early phases of outbreaks have been shown to have a profound impact in containing pathogen spread before they can become global threats⁵. However, our abilities to rapidly put in place critical diagnostic systems involving novel causative pathogens are severely limited due to sparse knowledge of the disease at the onset.

Much akin to cyber-attacks, the longer the window of vulnerability remains open—the period between first human infection (“zero-day”) and development of a diagnostic test to identify and then isolate infected hosts—the more likely an outbreak will expand to an epidemic or pandemic (Fig 1a).

In the case of Covid-19, reverse-transcriptase-polymerase-chain-reaction (RT-PCR) testing, the diagnostic “gold standard”, has been severely hampered by kit availability, test reproducibility, test turn-around time and low sensitivity⁶. As a result, hospitals in China resorted to using computed tomography (CT) imaging to aid diagnosis; while high sensitivity has been initially reported, its inherent high costs and low throughput render it impractical as a global front-line screening tool⁷. Fast, cheap, and globally scalable means to rapidly detect Covid-19 cases are critical to contain further spread of the virus.

Chest radiographs (CXR) have long been employed as a screening tool for infectious diseases due to its low costs, wide availability and physician familiarity⁸. In this study, we developed two “nowcasting” Deep

Learning (DL) models to detect Covid-19 on CXR using available real time knowledge and data. We reasoned that due to a lack of ground truth data in the initial phases of “zero day” attacks, descriptive data from the preliminary case definition of COvid-19, in this case, radiographic pneumonia, could be used to rapidly train a highly scalable initial threat detection DL model. Our first DL model MAIL1.0 achieved high sensitivity and specificity in detection of radiologically visible pneumonia. As the outbreak unfolds, and as confirmed cases accumulated, these were then used to further update and specify the DL model (Fig 1b). Our second model MAIL2.0 is superior to radiologists in detection of Covid-19 patients, especially amongst asymptomatic patients and those presenting with early symptoms.

Results

Study Population Characteristics

The demographics of the CheXpert, PadChest and HKU1 datasets are described in the supplement. The study population characteristics of VC1 and VC2 are detailed in Fig. 2 and Table 1.

Validation Cohort 1 (VC1)

During the recruitment period, 9,941 patients attended QMH A&E department. Of those 566 fulfilled the reporting criteria and had undergone RT-PCR testing for suspected Covid-19 pneumonia. In total, 460 patients with a corresponding CXR were identified consisting of 455 RT-PCR negative patients and 5 Covid-19 confirmed cases. Additionally, 54 Covid-19 RT-PCR confirmed cases were collected from collaborating hospitals. This cohort represented (59/101) 58% of all confirmed cases in the Hong Kong SAR as of March 2, 2020. The prevalence of respiratory diseases as a cause of admission based on International Classification of Diseases 9th version (ICD-9) was 191/514 (37.2%), of which 77 (15.0%) were pneumonia, 71 (13.8%) were upper respiratory tract infection (URTI) and 15 (2.9%) confirmed influenza by molecular testing. Amongst the positive cases, 40/59 (67.8%) presented with fever or respiratory systems, while 5/59 (8.5%) were asymptomatic. The median time from onset of symptoms to CXR in symptomatic patients was 7.5 days (range 0- 20 days).

Validation Cohort 2 (VC2)

Of the 1456 patients attending QMH A&E department during the subsequent data collection period, 218 had undergone Covid-19 RT-PCR testing (3 of these were positive and included in VC1). Another 6 were excluded as they were paediatric cases and 33 did not have CXR performed on admission, resulting in total of 176 RT-PCR tested negative cases. Another 72 positive cases were recruited as per VC1. The prevalence of disease of the respiratory system accounted for 69/248 (27.8%), of which 43 (17.3%) were pneumonia, 12 (4.8%) were upper respiratory tract infection (URTI); there were no influenza cases (Table 1). Similarly, 50/72 (69.4%) of Covid-19 positive cases presented with fever or respiratory symptoms

while 9/72 (12.5%) were asymptomatic. There was a decrease in the time from onset of symptoms to CXR being performed (median 2.0 days range 0–15 days, $p < 0.001$) in VC2 relative to VC1.

Radiologists Performance

In VC1, the incidence of PNA identified by consensus radiologists reads was 261/460 (56.7%) with 113/191 (59.2%) of patients in the respiratory disease group. Interobserver reliability was moderate (κ 0.506 95%CI 0.504, 0.509). The sensitivity and specificity for detection of Covid–19 were 61.0% and 58.2% respectively with moderate interobserver reliability (κ 0.480 95%CI 0.477, 0.482). The proportion of PNA in the Covid–19 cohort was

36/59 (61.0%). In VC2, the incidence of PNA identified by radiologists was 128/248 (51.6%) with 47/69 (68.1%) in the respiratory disease group. Interobserver reliability was again moderate (κ 0.466 95%CI 0.462, 0.470). The sensitivity and specificity for detection of Covid–19 was 31.9% and 62.5% respectively and the reliability was only fair (κ 0.264 95%CI 0.260, 268) (Table 2). The proportion of Covid–19 patients with PNA was significantly lower 23/72 (32.0%, $p = 0.001$) compared with VC1.

Model Performance

MAIL1.0

The original Kaggle Model 2 achieved an AUC of 0.87 (95%CI 0.82, 0.92) and 0.81 (95%CI 0.77, 0.86) for CheXpert1 and HKU1. The corresponding AUC of MAIL1.0 was 0.87 (95%CI 0.82, 0.92) and 0.81 (95%CI 0.77, 0.85) on CheXpert and HKU1. Additionally, we tested MAIL1.0 on a selected independent public database (PadChest) which achieved an AUC of 0.83 (95%CI 0.80, 0.86). The throughput of MAIL1.0 had a 35.0% increase from 0.5 scans/s to 0.6 scans/s compared with Kaggle Model 2 and the power consumption also reduced at inference by 45.0% (Supplement TableS1–4, FigS1–3).

Validation of MAIL1.0 for CXR PNA detection on VC1 demonstrated an AUC 0.93 (95%CI 0.91, 0.95) with an accuracy of 84.6% against consensus radiologists reads (Fig 3a).

Evaluation of MAIL1.0 detection of Covid–19 confirmed cases in VC1 however revealed the model only had a sensitivity of 54.2% and specificity of 37.1% and AUC of 0.45 (95%CI 0.37, 0.53), compared to radiologists (61.0% and 58.2% for sensitivity and specificity respectively). MAIL1.0 missed 27/59 (45.8%) of Covid–19 positive patients versus 23/59 (40.0%) for radiologists. Feature dimensionality reduction visualization confirmed that while the features learned by MAIL1.0 were suitable for radiological PNA prediction and importantly, could detect >50% of initial Covid–19 patients, they were still insufficiently informative to capture a sizable fraction of Covid–19 cases (Fig 3b).

MAIL2.0

The MAIL2.0 model achieved an AUC of 0.81 (95%CI 0.75, 0.87) on VC2 and significantly outperformed expert radiologists for Covid-19 detection (AUC 0.47, 95%CI 0.39, 0.55, $p < 0.001$) (Fig 4a). The accuracy, sensitivity and specificity of MAIL2.0 were 75.4%, 84.7% and 71.6% compared with 53.6%, 31.9% and 62.5% for radiologists, respectively. The model correctly identified 40/49 (81.6%) Covid-19 positive cases missed by radiologists; conversely, radiologists identified 2/11 (18.2%) missed by the model (Fig. 4b-f). Eleven of seventy-two (15.3%) of Covid-19 positive cases also underwent high resolution computed tomography (HRCT) of the thorax during the same admission with a median time between CXR and HRCT of 2 days (range 0-6). Of the 10 HRCT that showed radiological appearances consistent with Covid-19, MAIL2.0 correctly identified 9/10 (90.0%) compared with 3/10 (30.0%) for radiologists.

Amongst Covid-19 patients with fever or respiratory symptoms (50/72, 69.4%), MAIL2.0 correctly identified 40/50 (80.0%) compared with 16/50 (32.0%) for radiologists ($p < 0.001$). MAIL2.0 was also significantly more sensitive in identifying asymptomatic patients (8/9, 88.9% vs 2/9, 22.2% $p = 0.02$) and patients with early symptoms compared with radiologists (i.e. with less than 2 days of symptoms on presentation for CXR) (27/33, 81.8% vs 10/33, 30.3% $p < 0.001$) (Fig. 5a,b). Additionally, there were 3 patients in VC2 that had initial RT-PCR negative tests which subsequently converted to positive: radiologists correctly identified 1/3 while MAIL2.0 identified 3/3.

Radiologists and model performances are summarized in Table 3.

Discussion

In this study, we developed 2 nowcast DL models for detection of Covid-19 on CXR, coinciding with available knowledge and data during the ongoing Covid-19 outbreak. On a prospectively collected independent validation set representing a real-world mix of patients screened for Covid-19 (VC1), our initial model, MAIL1.0, designed to detect PNA and reflecting up to date radiological case description knowledge of Covid-19 at that time, achieved high sensitivity and specificity in detecting radiologically visible PNA, but low sensitivity and specificity for Covid-19. With gradual accumulation of local Covid-19 confirmed cases, a second updated nowcast DL model, MAIL2.0, was developed and when validated on a second prospectively collected independent cohort (VC2) significantly outperformed radiologists, most notably amongst asymptomatic and patients early in their clinical presentation.

At the onset of any outbreak such as the current Covid-19 crisis, there is lack of ground truth or experience amongst doctors to reliably diagnose the disease and diagnosis therefore must rely on the published literature. However, this period is crucial for any containment strategy and to mitigate local spread. CXR has long been considered an excellent screening tool for chest pathology, including for infectious diseases such as tuberculosis. It is fast, easy, readily available even in resource limited areas, and utilizes only a small dose of ionizing radiation relative to CT⁹. Given the available knowledge as of early February 2020 (the initial stage of the outbreak) describing ground-glass opacities or consolidations

(PNA) on imaging, which were highly conserved and present in up to 88% of cases, we focused on developing a rapidly deployable DL solution to aid in Covid-19 triage based on detecting PNA-like features on CXR (MAIL1.0) ¹⁰.

As speed to deployment of defensive countermeasure is of critical importance in zero-day pathogen attacks, we developed a first generation nowcast DL model that was efficient and highly accurate in detecting radiologically visible pneumonia. This model shows the feasibility of generating an initial DL rapid-response countermeasure utilizing initial emerging information of the pathogen when there is an initial lack of ground truth cases.

However, as new information and local cases of Covid-19 became available as the outbreak progressed, it was clear that a general PNA model was no longer sufficient. First, peer-reviewed literature now reported several key radiographic features that were absent in Covid-19 PNA and that up to 14% of Covid-19 cases had “normal” CXR imaging ^{11, 12, 13}. Second, while accurate at its designed function (PNA detection), MAIL1.0 had only moderate sensitivity and specificity in independent validation (VC1) missing 27/59 (45.8%) of Covid-19 cases. Feature dimensionality reduction visualization of these cases further confirmed that the PNA features learned by MAIL1.0 were informationally insufficient to capture the broader feature space describing Covid-19, and therefore, an update to our model was necessary.

Using VC1, we trained a second model which include features other than PNA. Due to the continual and rapid escalation of the outbreak now reaching pandemic level, we were also able to prospectively collect another independent cohort (VC2) and the performance of MAIL2.0 was significantly superior to radiologists. Interestingly, radiologist performance dropped sharply in VC2 despite having the experience of reviewing Covid-19 cases in VC1. One of the reasons we speculate could be the temporal shift in the Covid-19 screening distribution, with much earlier presentations seen in VC2, as the timing of the recruitment coincided with aggressive contact tracing and screening of patients in Hong Kong. As radiological signs in early stages can be subtle or limited, radiologists clearly had difficulties in detecting these cases ¹⁴. Indeed, when the interobserver variability of PNA and H Covid were analyzed, there was a significant decrease in agreement in H Covid in VC2 only. This reflects the unfamiliarity amongst radiologists in distinguishing this disease although our results are comparable to other published literature ¹⁵. On the other hand, MAIL2.0 performed significantly better than radiologists in VC2, despite also having been trained on VC1, suggesting that MAIL2.0 is more generalizable.

Nevertheless, there are several limitations. We did not measure the performance of our model in distinguishing Covid-19 from other forms of pneumonia. As reported in the literature, radiological features of Covid-19 can overlap with other types of pneumonia, most notably other viral PNA ¹⁶. However, our results clearly show that the features of Covid-19 are very different than general PNA, and that MAIL2.0 does well at distinguishing non-Covid-19 PNA from Covid-19 PNA. This is further supported through the validation of MAIL2.0 in a prospective cohort of RT-PCR screened patients, where it is able to identify Covid-19 with superior accuracy to radiologists who had significant difficulty at distinguishing PNA from Covid-19 on CXR. Finally, while our model performed well in our region, and

appeared to generalize well to the shifting dynamics of symptoms and time to presentation, we do not know if it will generalize to all age groups or geographic regions as the disease continues to evolve. Further studies evaluating this are currently underway.

Due to the expected poor sensitivity of radiologists in early stages of Covid-19 PNA, the ability to effectively detect cases both earlier and unappreciable to expert radiologists is critical to constrain the Covid-19 outbreak as early isolation and careful infection control in this group can significantly reduce human-to-human transmission¹⁷. Unlike many of the triage strategies developed in China where haematological, biochemical and radiological (usually CT-based) results are essential in the diagnostic pathway, they are not readily available, nor scalable in many developing countries, thus severely limiting their use^{18, 19}. Even in the West, issues with RT-PCR kit reliability, sub-optimal sensitivity, and kit availability have rendered aggressive Covid-19 case detection by RT-PCR difficult²⁰. While prospective validation will be needed, it is promising that our models were able to perform well in actual prospectively collected Covid-19 screening cohorts. Thus, our algorithm, together with the widespread availability of CXR, could lend itself as a potentially fast, sensitive and cost effect triaging tool, reducing the burden of already severely stretched healthcare systems.

In summary, we present a novel nowcast DL strategy where we were able to rapidly build 2 generations of Covid-19 DL detection models with emerging and evolving Covid-19 knowledge and data at hand. We evaluated these models in the context of prospective collected real-world data from an at-risk geographic territory attempting to avert an uncontrolled epidemic with performance surpassing radiologists. Our results have applicability to real-world containment strategies especially in the context of the limited and stretched healthcare resources we are currently experiencing. Nowcast DL models represent a potentially powerful means to rapidly implement successive defensive barriers to efficiently constrain zero-day pathogen attacks.

Materials And Methods

Patients

The study protocol was approved by the Institutional Review Boards (IRB) of the respective contributing hospitals (IRB ref no: UW20-134, UW20-115, HKECREC-2020-012 and KC/KE-20-0052/ER-3). Informed patient consent was not required as the data used in this study were anonymized. Study datasets consisted of one retrospectively collected and two separate prospectively collected cohorts: HKU1, Validation cohorts 1 and 2 (VC1 and VC2).

HKU1 was acquired retrospectively from January 1, 2014 to December 31, 2018 from the Queen Mary Hospital (QMH), Hong Kong SAR, consisting of a mixture of acute and chronic patients with frontal CXRs. VC1 consisted of prospectively collected patients who presented consecutively to the A&E Department of QMH of sufficiently high clinical concern warranting Covid-19 reverse transcription polymerase chain reaction (RT-PCR) testing between December 31, 2019 to February 15, 2020 and Covid-19 RT-PCR

confirmed cases across 4 local hospitals: Queen Mary, Ruttonjee, Queen Elizabeth and Pamela Youde Nethersole Eastern Hospitals through March 2, 2020. VC2 was also a prospectively collected dataset in similar fashion as VC1 consisting of patients presented to the QMH A&E Department between February 16 to 22, 2020 that had undergone RT-PCR with additional Covid-19 RT-PCR confirmed cases from the same 4 hospitals between March 3, 2020 and March 22, 2020 (supplement).

After the appropriate cohort was identified, clinical records of the patients, including A&E admissions notes, discharge letters and past medical history were collected. Epidemiological and clinical characteristics of these patients were analysed²¹. The frontal CXR of all patients on admission (i.e. the first CXR performed during the admission in which the patient underwent RT-PCR) in DICOM format were retrieved and used for evaluation and model training (Fig 2).

General PNA CXR detection DL Model (MAIL1.0)

A general PNA CXR detection model (MAIL1.0) was developed by leveraging and modifying a top performing PNA detection model from a recent RSNA Kaggle competition (Kaggle model 2). MAIL1.0 was then evaluated on three independent datasets: the CheXpert Validation set (n = 200), the PadChest dataset (n = 1000), and HKU1 (n = 921)^{22, 23, 24}. To assess the potential effect of this model in detecting Covid-19 pneumonia in a real-world scenario, the model was also tested on VC1.

Covid-19 specific DL model (MAIL2.0)

MAIL2.0 used an Imagenet pre-trained network based on SE-ResNeXt-50-32x4d. The first model used this pre-trained network and was then trained on VC1 to predict Covid-19 status. A second model used the Imagenet pre-trained network and was then trained on the full NIH dataset utilizing the 14 weakly labelled classes provided²⁵ and then trained on VC1 to predict Covid-19 status. Specifically, we replaced the last fully connected layer to predict a single binary Covid-19 RT-PCR label. Training was performed in two stages: (1) keeping all weights frozen except for the new classification layer, for 3 epochs; (2) training the last ResNet layer plus the new classification layer for 3 more epochs. To avoid overfitting, we applied moderate augmentations (horizontal flip, small rotation, shear, resize, gamma change, blur and noise) and added 75% dropout. We trained on 4 folds for each model (random splits with the same patient saved to the same fold). Each resulting model was trained on a different fold of the VC1 dataset, with predictions of the 8 models averaged. Threshold was selected based on maximizing sensitivity and specificity based on the out of fold predictions from VC1.

All models were trained on NVIDIA 2080ti and 1080ti graphics cards with 11Gb memory; trained 4 folds and features averaged between models trained on different folds.

CXR PNA and Covid-19 PNA radiologist evaluation

CXRs of all cases in the VC1 and VC2 datasets were evaluated by two board-certified radiologists (WHKC and VV each with >10 years' experience) independently with disagreements resolved by a third radiologist independently (MN). CXRs were evaluated for both the presence and absence of: (1) radiologically visible pneumonia (PNA), and (2) PNA with radiological appearances consistent with Covid-19. Radiologists evaluation of VC1 occurred when the radiologists were aware of the literature describing the radiological features of Covid-19 but had not yet seen any local Covid-19 cases whilst VC2 occurred after all radiologists critically evaluated their CXR Covid-19 detection performance on VC1 against the ground truth RT-PCR results at least 2 weeks prior to evaluation of VC2.

Statistical Analysis:

All analyses were performed using python packages for sklearn and matplotlib and commercially available software (SPSS version 26, SPSS Inc., Chicago, IL, USA). To evaluate the stand-alone performance of the models, the AUC-ROC was estimated using the normalized Wilcoxon (Mann-Whitney) U statistics²⁶. Non-parametric confidence intervals on the AUC were computed with DeLong's method by Python²⁷. We compared the sensitivity and specificity of the expert radiologists' consensus reads with that of the threshold scores from the models. In the reader study, each reader reviewed the CXR using a forced protocol (i.e. either positive or negative) for assessment for the radiological characteristics of Covid-19 pneumonia based on published literature^{13,28}. We used nonparametric Wilcoxon test to evaluate the difference between the median times in VC1 and VC2. The primary comparisons of this study were: sensitivity, specificity, positive and negative predictive values for consensus radiologists reads against MAIL1.0 and 2.0 against VC1 and VC2; the AUC-ROC in the reader studies. Kappa analyses were performed to assess interrater variability²⁹. A statistical significance threshold of 0.05 was used.

Declarations

Acknowledgements. We would like to acknowledge Chu Hong Yin, Chan Wing Ying, Benedicte Coiffier, the Departments of Radiology, Infectious Disease, and Pulmonary and Critical Care at Queen Elizabeth Hospital, Ruttonjee Hospital, Queen Mary Hospital, and Pamela Youde Nethersole Eastern Hospital along with all the patients and their families.

Author contributions: M.D.K. developed the concept with assistance from W.H.K.C., W.H.K.C., and M.D.K designed the study and related experiments, and analyzed and interpreted the data with assistance from V.V., D.P., R.D., A.Y.H.Y., and S.Z.. M.D.K and D.P. designed the models with assistance from W.H.K.C.. W.H.K.C. and M.D.K. wrote the manuscript with assistance from V.V., D.P., R.D., A.Y.H.Y., and S.Z.. W.H.K.C., V.V., and M.Y.N. performed radiological reads, and S.Z., A.H.T.F, T.W.Y.C., J.C.Y.L, S.T.L., C.S.Y.L, M.M.S.L., B.X.H.Y. and M.Y.N. contributed to acquisition of the data.

Disclosures of Conflicts of Interest: M.D.K. is a founder and shareholder in Ensemble- group. D.Y. is an employee of Ensemble-group.

Data Availability. The datasets from HKU (HKU1, VC1, and VC2) contain personal information and are not publicly available. The NIH, CheXpert, and PadChest datasets are public datasets, each with their own terms and conditions for use; their use as specified in this study are described in Supplementary Methods.

Code availability. The software for training the models is available upon request at (radiology.hku.hk/MAIL/covid) and the most recent deployable models are available at (www.ensemblehealth.ai/covid).

References

1. Huang C, *et* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, (2020).
2. Nishiura H, Linton NM, Akhmetzhanov AR. Initial Cluster of Novel Coronavirus (2019-nCoV) Infections in Wuhan, China Is Consistent with Substantial Human-to- Human Transmission. *Journal of clinical medicine* **9**, (2020).
3. World Health Coronavirus disease 2019 (COVID-19) Situation Report https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19pdf?sfvrsn=1ba62e57_10, (2020).
4. World Health Coronavirus disease 2019 (COVID-19) Situation Report https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200322-sitrep-62-covid-19pdf?sfvrsn=755c76cd_2, (2020).
5. Kelly-Cirino CD, *et* Importance of diagnostics in epidemic and pandemic preparedness. *BMJ global health* **4**, e001179 (2019).
6. Corman VM, *et* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **25**, (2020).
7. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing. *Radiology*, 200343 (2020).
8. Al Abri S, *et* Tools to implement the World Health Organization End TB Strategy: Addressing common challenges in high and low endemic countries. *International Journal of Infectious Diseases* **92**, S60-S68 (2020).
9. van't Hoog AH, Onozaki I, Lonnroth Choosing algorithms for TB screening: a modelling study to compare yield, predictive value and diagnostic burden. *BMC Infectious Diseases* **14**, (2014).

10. Zu ZY, *et* Coronavirus Disease 2019 (COVID-19): A Perspective from China. *Radiology* **0**, 200490 (2020).
11. Pan F, *et* Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia. *Radiology*, 200370 (2020).
12. Ng M-Y, *et* Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. *Radiology: Cardiothoracic Imaging* **2**, e200034 (2020).
13. Chung M, *et* CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV). *Radiology*, (2020).
14. Pan F, *et* Time Course of Lung Changes On Chest CT During Recovery From 2019 Novel Coronavirus (COVID-19) Pneumonia. *Radiology*, (2020).
15. Hopstaken RM, Witbraad T, van Engelshoven JMA, Dinant GJ. Inter-observer variation in the interpretation of chest radiographs for pneumonia in community- acquired lower respiratory tract *Clinical Radiology* **59**, 743-752 (2004).
16. Li L, *et* Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. *Radiology*, (2020).
17. Kanne JP, Little BP, Chung JH, Elicker BM, Ketani Essentials for Radiologists on COVID-19: An Update—Radiology Scientific Expert Panel. *Radiology*, (2020).
18. Zhang J, Zhou L, Yang Y, Peng W, Wang W, Chen X. Therapeutic and triage strategies for 2019 novel coronavirus disease in fever *The Lancet Respiratory Medicine* **8**, e11-e12 (2020).
19. Ayebare RR, Flick R, Okware S, Bodo B, Lamorde Adoption of COVID-19 triage strategies for low-income settings. *The Lancet Respiratory Medicine*, (2020).
20. Fisher D, Wilder-Smith The global community needs to swiftly ramp up the response to contain COVID-19. *The Lancet*, (2020).
21. Chen N, *et* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, (2020).
22. RSNA Pneumonia Detection Challenge.) (2019).
23. Irvin J, *et* CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. *Proceedings of the AAAI Conference on Artificial Intelligence* **33**, 590-597 (2019).
24. Bustos A, Pertusa A, Salinas J-M, de la Iglesia-Vayá PadChest: A large chest x- ray image dataset with multi-label annotated reports. Preprint at <https://ui.adsabs.harvard.edu/abs/2019arXiv190107441B> (2019).

25. Hu J, Shen L, Albanie S, Sun G, Wu Squeeze-and-Excitation Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1 (2019).
26. Hanley JA, McNeil The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29-36 (1982).
27. DeLong ER, DeLong DM, Clarke-Pearson Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837-845 (1988).
28. Guan W-j, et Clinical Characteristics of Coronavirus Disease 2019 in China. *New England Journal of Medicine*, (2020).
29. McHugh Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* **22**, 276-282 (2012).

Tables

		VC1 (%)	VC2 (%)
Patients (n)		514	248
<u>Characteristics</u>			
Age, years	Median (IQR)	65 (44-82)	61 (39-79)
Sex			
	Male	269 (52.3)	126 (50.8)
	Female	245 (47.7)	122 (49.2)
Travel History to China		154 (30.0)	43 (17.3)*
Contact History		41 (8.0)	26 (10.5)
Chronic medical illnesses			
	Cardiovascular diseases	164 (31.9)	102 (41.3)*
	Respiratory system diseases	94 (18.3)	38 (15.3)
	Digestive system diseases	100 (19.5)	35 (14.1)
	Malignant tumours	91 (17.7)	42 (16.9)
	Cerebrovascular diseases	93 (18.1)	29 (11.7)*
	Endocrine system diseases	120 (23.6)	72 (29.0)
<u>Signs and symptoms on admission</u>			
	Fever	246 (47.9)	108 (43.6)
	Cough	242 (47.1)	115 (46.4)
	Shortness of Breath	145 (28.2)	60 (24.2)
	Chest Pain	21 (4.1)	13 (5.2)
	Fever or any respiratory symptoms	407 (79.2)	182 (73.4)
<u>Clinical Diagnoses</u>			
	Diseases of the respiratory system	191 (37.2)	69 (27.8)
	URTI	71 (13.8)	12 (4.8)*
	Influenza	15 (2.9)	0 (0.0)*
	Pneumonia	77 (15.0)	43 (17.3)
	COVID-19 pneumonia	59 (11.5)	72 (29.0)

Table 1. Patient characteristics of validation cohorts 1 & 2 (VC1 and VC2). *denotes statistical significance ($p < 0.05$) between the two groups.

Cohort		n	Main Analysis		
			Agreement / %	Kappa	95% CI
VC1	PNA	514	78.2	0.506	0.504, 0.509
	H Covid	514	74.3	0.480	0.477, 0.483
VC1 (GT Covid -ve)	PNA	455	78.2	0.507	0.415-0.599
	H Covid	455	74.1	0.477	0.386-0.569
VC2	PNA	242	73.4	0.466	0.462, 0.470
	H Covid	242	67.3	0.264	0.260, 0.268
VC2 (GT Covid -ve)	PNA	176	76.7	0.515	0.367-0.663
	H Covid	176	67.7	0.276	0.129-0.424

Table 2. Interobserver agreement for pneumonia (PNA) and radiological appearances consistent with Covid-19 (H Covid) in VC1 and VC2 between radiologists. Agreement was moderate for PNA and H Covid detection in VC1 as well as PNA in VC2. Agreement was only fair for H Covid in VC2.

Cohort	Model	Prediction	Sensitivity	Specificity	AUC (95% CI)	P-value
VC1 (n=514)	MAIL1.0	PNA	90.5%	76.7%	0.93 (0.91, 0.95)	N/A
		GT Covid	54.2%	37.1%	0.45 (0.37, 0.53)	ns
	Radiologists	GT Covid	61.0%	58.2%	0.60 (0.52, 0.67)	
VC2 (n=248)	MAIL2.0	GT Covid	84.7%	71.6%	0.81 (0.75, 0.87)	< 0.001
	Radiologists		31.9%	62.5%	0.47 (0.39, 0.55)	

Table 3. The performance of DL models and radiologists on two independent validation cohorts (VC1 and VC2).

Figures

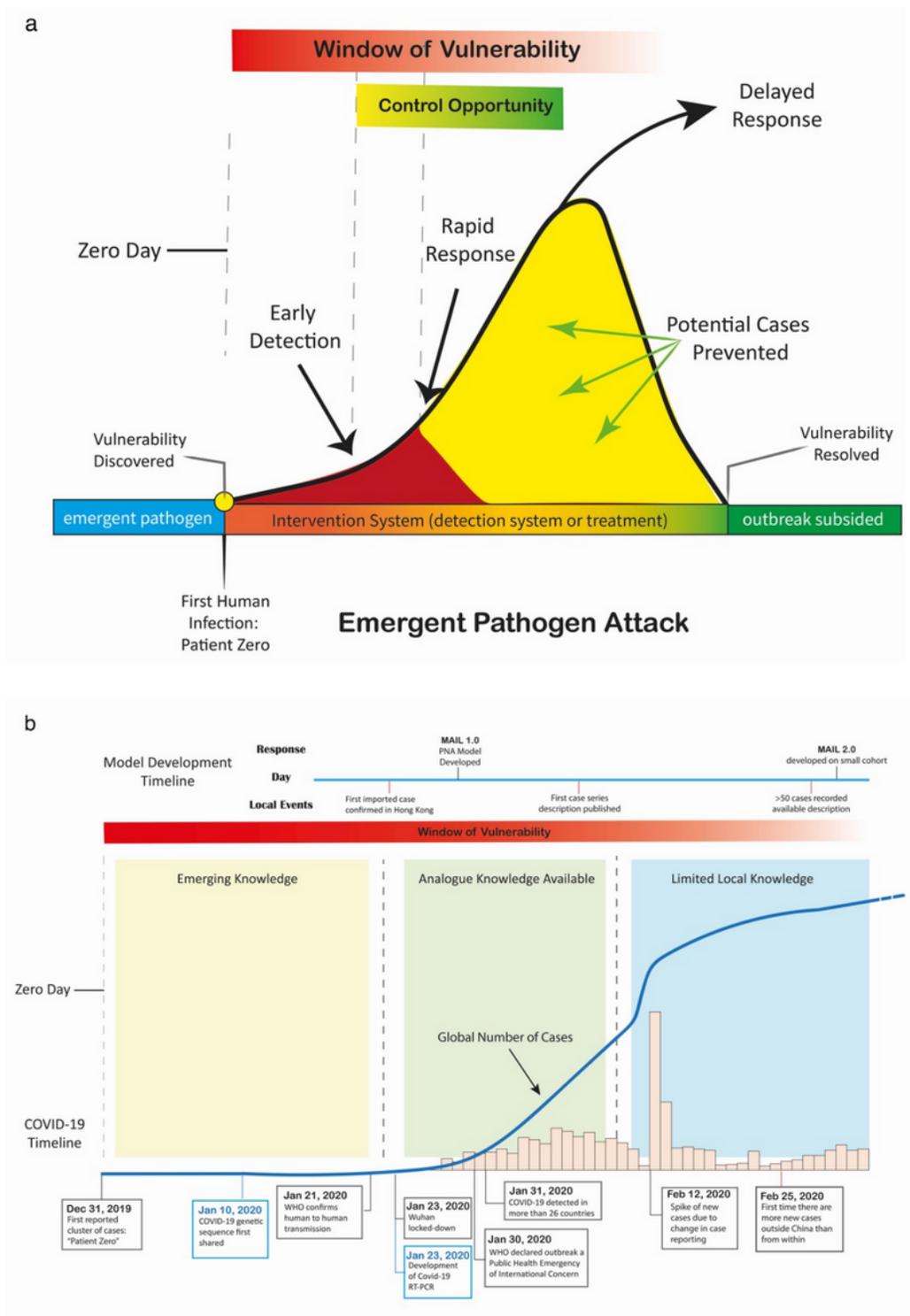


Figure 1

Constraining Emergent Pathogen Attacks. a, Rapid deployment of detection countermeasures are crucial in mitigating the effects of emergent pathogen attacks by shortening the “Window of Vulnerability”. b, Timeline of Covid-19 events plotted against the “Window of Vulnerability” and the progressive accumulation of Covid-19 knowledge and case data, and ‘nowcast’ DL model development.

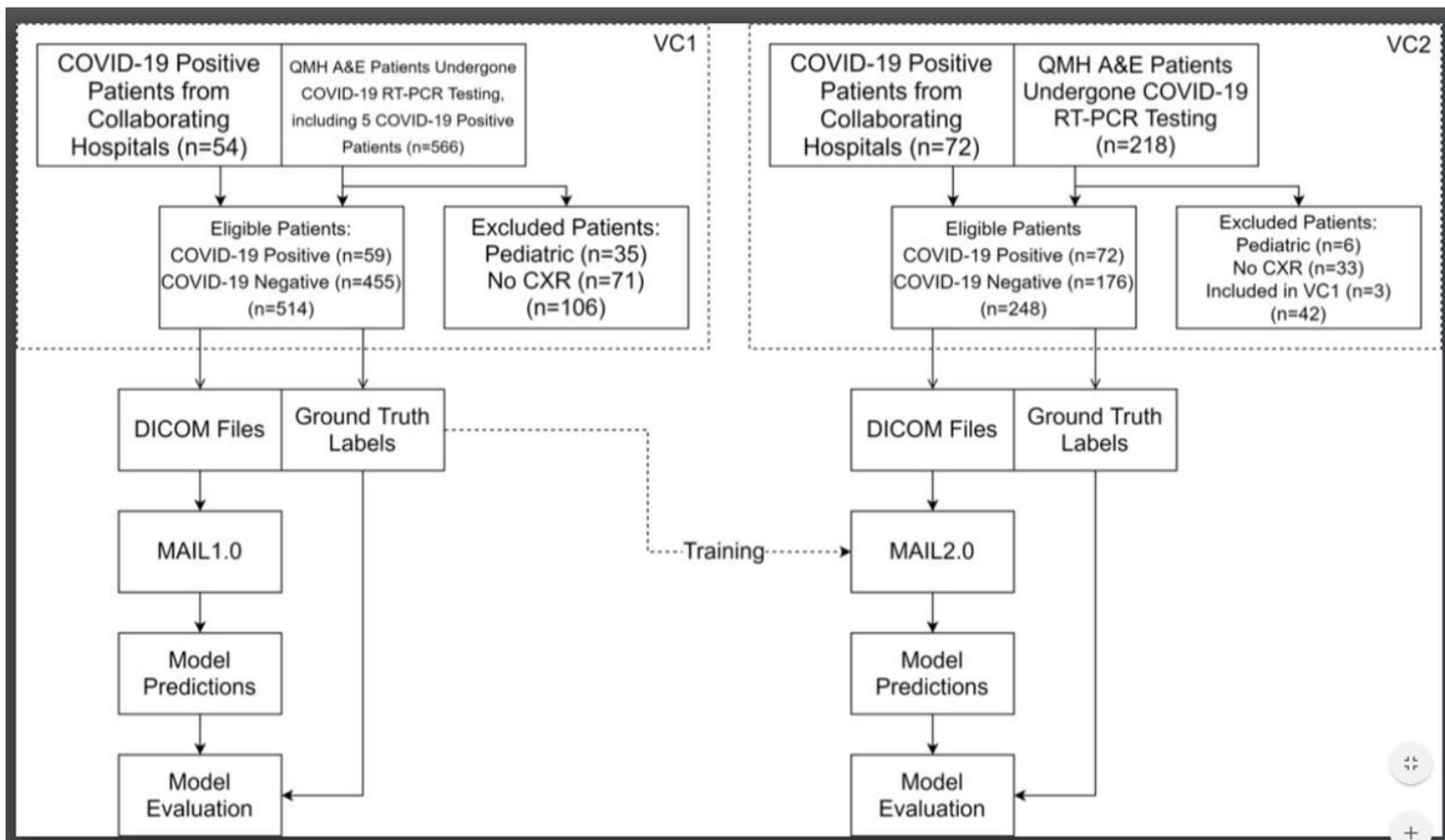


Figure 2

Flow diagram. VC1 consisted of 514 patients (455 RT-PCR negative and 59 RT-PCR positive cases). VC2 consisted of 248 patients (176 RT-PCR negative and 72 RT-PCR positive cases).

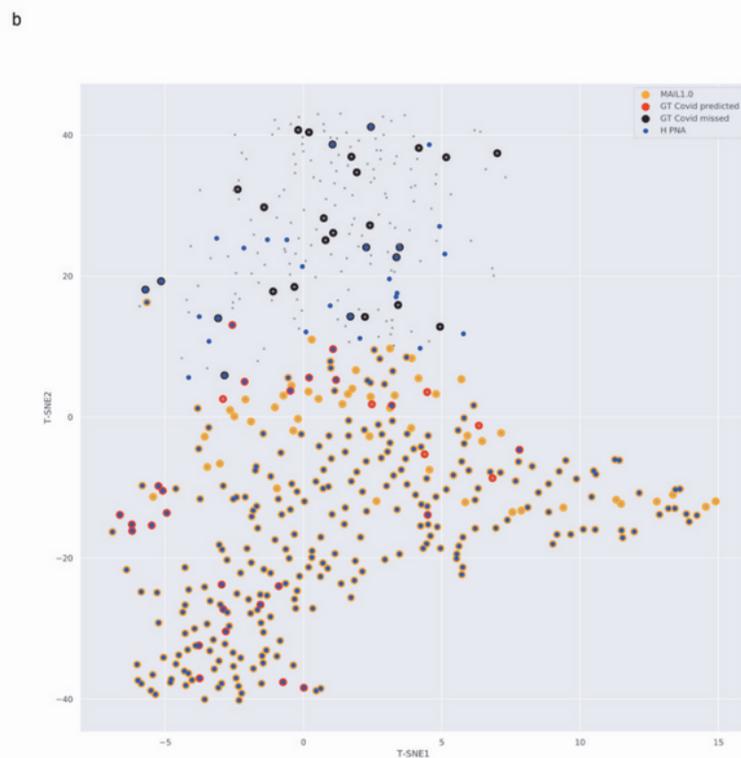
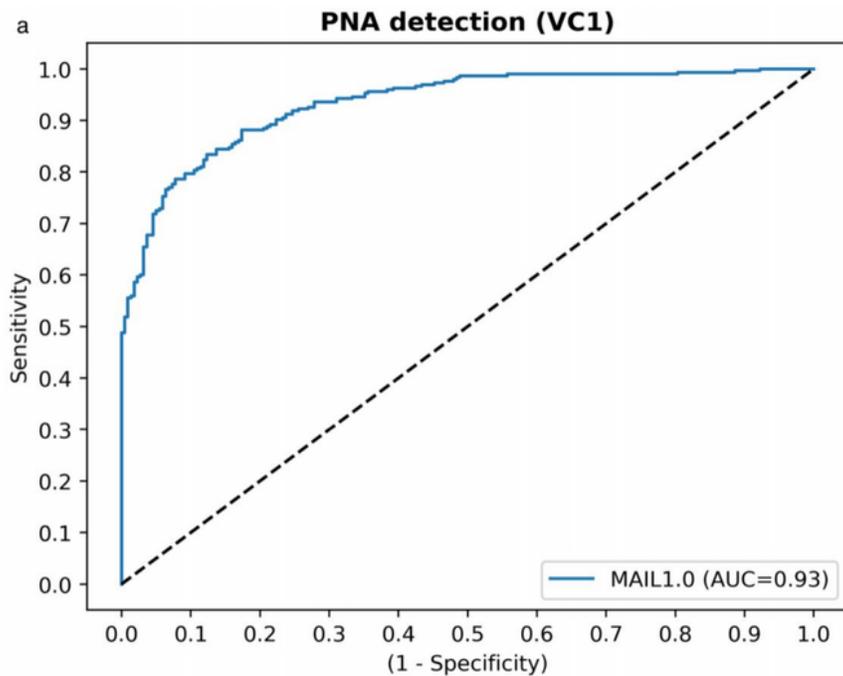


Figure 3

Performance of MAIL1.0 nowcast DL model. a, receiver operating characteristic (ROC) curve analysis of MAIL1.0 for detecting PNA on CXR as identified by radiologists in VC1 ($n=514$). MAIL1.0 achieved an area under the curve (AUC) of 0.93 (95%CI 0.91, 0.95). b, t-SNE projection illustrates substantial overlap between MAIL1.0 prediction and radiologist interpretation of pneumonia on CXR (H PNA). However, 45.8% of Covid-19 cases were missed by MAIL1.0 (GT Covid missed) and form a distinct cluster from Covid-19

cases detected by MAIL1.0 (GT Covid predicted) suggesting Covid-19 features are different from conventional PNA CXR features

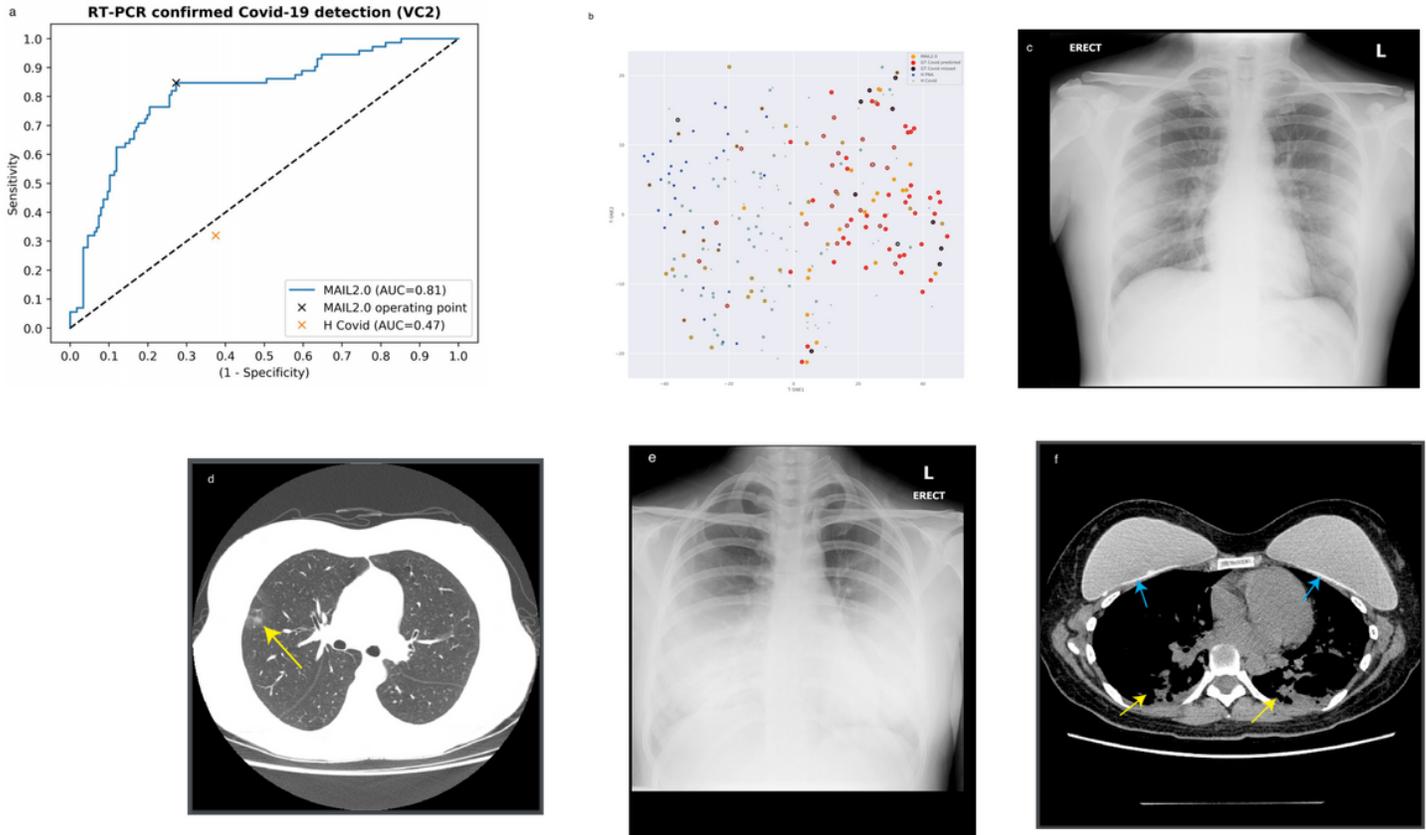


Figure 4

Performance of MAIL2.0 nowcast DL model. a, The ROC curve of MAIL2.0 on the VC2 independent validation set compared to expert radiologist (H Covid). The AUC is 0.81 (95%CI 0.75, 0.87) for MAIL2.0 versus 0.47 (95%CI 0.39, 0.55) for radiologist. b, t- SNE visualization shows relationships between all cases in VC2 that MAIL2.0 predicted as Covid-19 positive (MAIL2.0), Covid-19 confirmed cases correctly predicted (GT Covid predicted) and missed (GT Covid missed) and radiologist Covid-19 (H Covid) and PNA (H PNA) predictions. c, CXR of a Covid-19 RT-PCR confirmed patient considered normal by expert radiologists but correctly predicted positive by MAIL2.0. d, Corresponding selected axial CT image showing a small area of ground glass changes peripherally in the right upper lobe (yellow arrow). e, CXR showing subtle bilateral lower zone consolidative changes masked by breast implants in a Covid-19 RT-PCR confirmed patient that was correctly predicted Covid-19 positive by expert radiologists but incorrectly predicted negative by the model. f, Corresponding CT image showing bi-basal peripheral consolidation (yellow arrows) masked by breast implants (blue arrows).

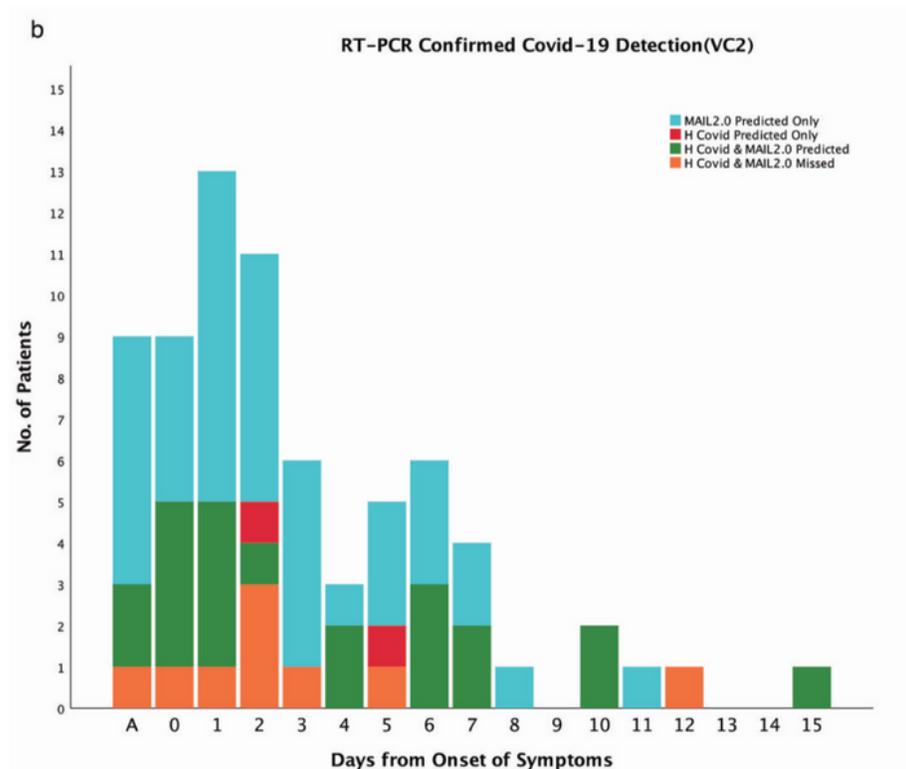
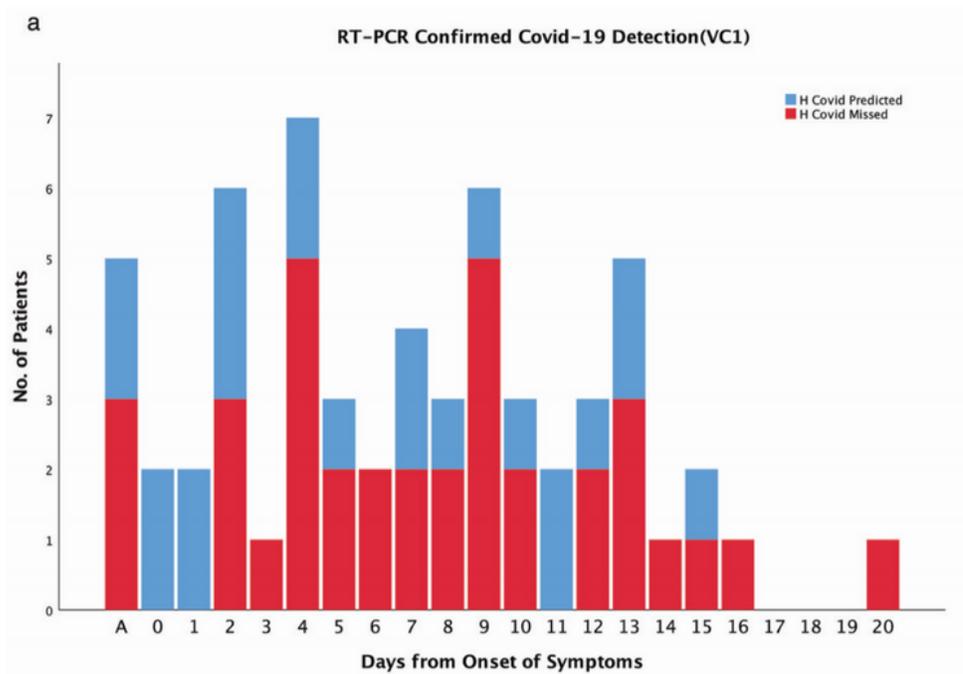


Figure 5

Frequency distribution plot. Time from onset of symptoms to CXR performed in RT-PCR confirmed cases versus that of expert radiologists correctly predicted and missed (H Covid Predicted Only and H Covid Missed Only, respectively), or that were correctly predicted or missed by both radiologists and MAIL2.0 (H Covid & MAIL2.0 Predicted Only and H Covid & MAIL2.0 Missed Only, respectively) in a. VC1 and b. VC2. A denotes asymptomatic patients.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryinformation.pdf](#)