Can artificial intelligence predict the need for oxygen therapy in early stage COVID-19 pneumonia?

Hirofumi Obinata (h-obinata@nms.ac.jp)
Nippon Medical School https://orcid.org/0000-0002-2219-3507
Peiying Ruan
Nvidia Japan
Hitoshi Mori
Self-Defense Forces Central hospital
Wentao Zhu
Nvidia
Hisashi Sasaki
Self-Defense Forces Central hospital
Kodama Tatsuya
Self-Defense Forces Central hospital
Murakami Wakana
Japan Self-Defense Forces Hospital Yokosuka
Masumi Tanaka
Department of Internal Medicine, Japan Self-Defense Forces Hospital Yokosuka
Pin-Lun Hsu
Nvidia Taiwan
Dong Yang
Nvidia
Ziyue Xu
Nvidia
Daguang Xu
Nvidia
Kaku Tamura
Self-Defense Forces Central hospital
Shoji Yokobori
Nippon Medical School

Research Article

Keywords: COVID-19, artificial intelligence, medical triage, disease progression

Posted Date: June 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-33150/v1

Abstract

This study investigated the utility of artificial intelligence in predicting disease progression. We analysed 194 patients with COVID-19 confirmed by reverse transcription polymerase chain reaction. Among them, 31 patients had oxygen therapy administered after admission. To assess the utility of artificial intelligence in the prediction of disease progression, we used three machine learning models employing clinical features (patient's background, laboratory data, and symptoms), one deep learning model employing computed tomography (CT) images, and one multimodal deep learning model employing a combination of clinical features and CT images. We also evaluated the predictive values of these models and analysed the important features required to predict worsening in cases of COVID-19. The multimodal deep learning model had the highest accuracy. The CT image was an important feature of multimodal deep learning model. The area under the curve of all machine learning models employing clinical features and the deep learning model employing CT images exceeded 90%, and sensitivity of these models exceeded 95%. C-reactive protein and lactate dehydrogenase were important features of machine learning models. Our machine learning model, while slightly less accurate than the multimodal model, still provides a valuable medical triage tool for patients in the early stages of COVID-19.

Introduction

A novel coronavirus disease (COVID-19), first reported in Wuhan, China in December 2019, has led to a global health crisis [1]. The World health organization (WHO) declared the disease a pandemic in March 2020 [2]. COVID-19 has affected 210 countries and territories around the world. The number of confirmed cases has exceeded six million and the number of confirmed deaths exceeded 360 000. In some countries, a medical crisis has occurred due to the exponential increase in incidence of COVID-19 [3], resulting in a shortage of medical resources and staff to treat patients with COVID-19. Moreover, some medical personnel have become infected and died. Hence, the number of preventable deaths has increased due to the inability to deliver lifesaving treatments to patients who could be saved if adequate resources were available.

There are no definitive therapies, vaccines or specific antiviral medications, to prevent or treat COVID-19 [4]. However, some drugs have shown promise in the treatment and supportive care of patients with the disease. These include the oxygen therapy and antibiotic treatments, which have become common treatment regimens for moderate and severe cases [5]. In contrast, careful observation remains the standard management strategy for asymptomatic and mild cases. Nevertheless, some asymptomatic or mild cases have been shown to silently progress toward pneumonia [6] and may rapidly develop into severe respiratory conditions [7, 8].

Under pandemic medical conditions, medical triage for these worsening asymptomatic or mild cases as well as for moderate and severe cases is crucial for successful patient management in the early phase of the disease. In order to achieve the best outcomes, the prediction of disease progression, particularly among asymptomatic or mild cases, at an early stage is indispensable. Several factors related to the

severity of COVID-19 have been reported [8, 9]. Furthermore, a simple "Call Score" model for the prediction of disease progression based on medical history and laboratory findings has been advocated in China [10] and the results of this model have been tested using traditional statistical analysis. Recently, artificial intelligence, machine learning, or deep learning, have been shown to offer superior results in prediction tasks compared to traditional statistical methods [11]. The aims of this study were to identify the value of artificial intelligence techniques for predicting disease progression of COVID-19 and to test the sensitivity of these models.

Materials And Methods

Study Participants and Design

This study was approved by the Institutional Review Board of Self-Defense Forces Central Hospital (02-014) and the requirement to obtain informed consent from the individual patients was waived because of its retrospective design. All procedures were performed according to the principals of the Declaration of Helsinki.

The study was performed at Self-Defense Forces Central Hospital between February 1, 2020 and April 15, 2020. We retrospectively examined all patients with COVID-19 confirmed by reverse transcription polymerase chain reaction (RT-PCR) who were admitted to our hospital. We have already reported some of this cohort in another paper [6]. In Japan, except for emergency cases, the hospitalisation of patients with COVID-19 is regulated by local public health centres. Exclusion criteria for this study were as follows: patients <18 years old, pregnancy, past medical history of severe respiratory disease, or home oxygen therapy before administration. Emergency cases who needed oxygen therapy before admission were also excluded. For purposes of this study, the need for initiation of oxygen therapy after admission was defined as an indicator of disease progression.

Data Collection

We reviewed electronic medical records and extracted the following data: background characteristics, clinical symptoms, laboratory findings, and chest computed tomography (CT) images. Data other than CT images were reviewed by two physicians. CT images were reviewed by two radiologists (W.M. and Y.S.) with 5 and 11 years of experience, blinded to the clinical information. The final decision was reached by consensus. The CT images of all patients were evaluated semi-quantitatively using a scoring system for all the outcomes affected by COVID-19. The axial images were visually scored on the basis of the previous studies [12]. Each lung was divided into six zones without regard to anatomical lobes. For the cranial/caudal dimension, three zones were defined as the upper (above the carina), middle (below the carina and above the inferior pulmonary vein), and lower (below the inferior pulmonary vein) zones. For medial/lateral dimension, the one-third of the outer area of axial slice was defined as the peripheral zone and the inner area as the central zone. Thus, a total of 12 lung zones were defined in each patient. Each zone was graded according to distribution of the involvement as 0 (0%), 1 (1%–5%), 2 (6%–25%), 3 (25%–50%), 4 (51%–75%), and 5 (75%–100%). The final score was the sum of the scores from all 12

zones and ranged from 0 (no involvement) to 60 (maximum involvement). Discrepancies were resolved by consensus.

Patient Management

On admission, laboratory tests and chest CT scans were performed for all cases. Newly initiated oxygen therapy, which indicated disease progression, was started when patients complained severe dyspnoea, tachypnoea (respiratory rate > 30) or hypoxia (pulse oximetry arterial saturation < 93%). Antiviral therapy or corticosteroids were administered after oxygen therapy by the attending physician.

Machine Learning Analysis of Clinical Metadata

Logistic Regression, Random Forest, and XGBoost machine learning models were used to analyse the clinical metadata in order to predict disease progression. We performed 10-fold cross validation in which we randomly and evenly divided the dataset into 10 different folds preserving the ratio of disease progression and patients. The average result was recorded as the final performance value.

The clinical metadata analysed included background characteristics including gender, age, admission from onset of disease, body height, body weight, body mass index (BMI), and smoking; past medical history including cardiovascular disease, respiratory disorders, and diabetes mellitus; laboratory findings including blood urea nitrogen, creatinine (Crea), aspartate transaminase (AST), alanine aminotransferase (ALT), total bilirubin (T-Bil), gamma-glutamyl transpeptidase (□GTP), amylase, lactate dehydrogenase (LDH), albumin (Alb), c-reactive protein (CRP), red blood cell count, haemoglobin (Hb), white blood cell count, platelets (Plt), percentage of neutrophils (Neutrophil [%]), percentage of lymphocytes (Lymphocyte [%]), percentage of monocytes (Monocyte [%]), percentage of eosinocyte (Eosinocyte [%]), absolute count of lymphocytes (Lymphocyte [absolute count]), neutrophil to lymphocyte ratio, activated partial thrombin time, and international normalised ratio of prothrombin time (PT -INR); and clinical symptoms including fever, cough, arthralgia, and abdominal symptoms.

Deep Learning Model Based on CT Images

A deep learning model from NVIDIA Clara Train [13] was used to segment the lung regions of CT images. EfficientNet-B7 [14], the best pre-trained model for image classification, was used to complement limited training data. Thereafter, we employed a pointwise convolution layer with 50 different filters and a swish activation function, followed by another pointwise convolution layer with 2 filters and a softmax activation function. A two-dimensional spatial max pooling along one channel was conducted to obtain the final oxygen therapy probability for the CT image (Figure 1). A stochastic gradient descent optimiser was used with a learning rate of 1×10^{-3}, momentum of 0.9, and weight decay of 1×10^{-4}. The total number epochs was set as 200, and the training was stopped if the area under the curve (AUC) did not improve in 10 epochs.

Multimodal Deep Learning Model

Since clinical metadata data and CT images are likely to be complementary, we designed a novel deep learning model which employed both clinical metadata and CT images (Figure 1). We used the probability of oxygen therapy from the deep learning model based on CT images and added an extra feature. We concatenated the oxygen therapy probability with the clinical metadata of 39 dimensions and used a fully connected layer with a softmax activation function and two outputs for non-oxygen therapy and oxygen therapy probabilities. The loss function, optimiser and hyper-parameter settings were the same as those for the deep learning model based only on CT images.

Evaluation Metrics of Machine Learning

We performed 10-fold cross validation in which we randomly and evenly divided the dataset into 10 different folds preserving the ratio of disease progression and patients. One of the folds was taken as validation data, and the remaining nine folds were taken as training data. We iteratively took one of the ten folds as validation data and the average validation performance was calculated as the final result.

To measure the prediction value of each model, we created receiver operating curves (ROCs) and computed the AUC, sensitivity, specificity, and accuracy for correctly distinguishing disease progression. Moreover, we included Youden's J statistic for comparison to identify the optimal cut-off point on the ROC, indicating the power of the model, balancing sensitivity and specificity [15]. To explain and analyse the prediction value of our deep learning model, we used visualisation to predict the need for oxygen therapy using a heatmap of each zone of two CT slices from a randomly selected case [16].

Results

Patient Enrollment and Characteristics

Two-hundred and fourteen patients with COVID-19 were admitted to the hospital between February 1, 2020 and April 20, 2020, and their diagnoses were confirmed by reverse transcription polymerase chain reaction (RT-PCR). A total of 20 patients were excluded for the following reasons: age (n = 8), history of severe respiratory disease (n = 3) and received oxygen therapy before admission (n = 2), CT was not performed on admission (n = 7). Hence, the study enrolled 194 patients. Thirty-one of these patients had oxygen therapy initiated after admission.

Baseline characteristics for the patients are shown in Table 1.

Performance of Machine Learning Models and Deep Learning Models

Figure 2 shows the AUC curve of each model. The AUC of all models exceeded 90%. Table 2 shows a comparison of the AUC, sensitivity, specificity, and accuracy of machine learning models (logistic regression, Random Forest, and XGBoost) based on clinical metadata, deep learning model based on CT images and multimodal deep learning models based on CT images and clinical metadata . The sensitivity of all five models were equal to or exceeded 95%. In three machine learning models, the specificity was highest in the XGBoost model (90%), followed by the logistic regression (88%), and the

Random forest model (86%). The XGBoost model was most accurate in three machine learning models. (AUC 93%, sensitivity 95%, specificity 90%, accuracy 88%) The multimodal deep learning model based on CT images and clinical metadata achieved better results than the three conventional machine learning models based on clinical metadata. The multimodal deep learning model achieved 100% sensitivity with 94% specificity and highest accuracy of all five models (95%). The sensitivity of the deep learning model based on CT images alone (97%) was higher than that of the machine learning models, but the specificity (84%) was lower than that of the machine learning models.

Importance of Features in Machine Learning Models

Figures 3 indicate the features of importance in each machine learning model, logistic regression, XGBoost, and Random Forest. CRP and LHD were very important features in all models. Other important features included time from onset to admission, BMI, Eosinocyte (%), age, and Lymphocyte (absolute count) in the logistic regression model; BMI and Lymphocyte (absolute count) in the XGBoost model; Lymphocyte (absolute count), Lymphocyte (%), and ALT in the Random forest model.

Visualisation of CT Images

We randomly selected a test case and used visualisation to predict the need for the oxygen therapy based on a heatmap of two CT slices (Figure 4). The deep learning model located the ground-glass opacities, and the high response of those regions led to a positive prediction for the case.

Discussion

This is the first study to predict the disease progression at an early stage of COVID-19 pneumonia by artificial intelligence employing chest CT scans and clinical data. By combining clinical data and CT, the accuracy of our machine learning model was significantly improved, achieving an accuracy of 95%. Few studies have investigated the value of artificial intelligence in the diagnosis or progress prediction of COVID-19, and these reports examined either CT or clinical metadata alone but not a combination of both [17]. We previously reported on the utility of early chest CT scans in the diagnosis of early stage COVID-19 [18] and chest CT scans have also reported to be useful in the diagnosis of severe cases [12].

Some centres use chest CT as a medical triage tool for the management of COVID-19 patients [19]. However, CT scans do not offer the necessary ease of accessibility in the light of growing concern over infection control. Blood sampling and medical interviews are more readily accessible and can be easily be performed either in the clinic or during home medical consultations.

The predictive performance of the three machine learning models tested in this study, while not as accurate as the multimodal deep learning model, would nonetheless provide sufficiently accurate predictions of COVID-19 progression in the clinical setting, and would certainly be more accurate than the current “Call Score” method [10].

Machine learning models are also useful in determining the importance of certain features in predicting patient prognoses. The three machine learning models in this study all revealed CRP, Lym, and LDH as major predictive factors. CRP is increased by the systemic inflammation; Lym reflects the severity and duration of illness [20]; and lung damage increases the serum level of LDH. The findings regarding these parameters support previously reports which categorised them as risk factors of disease severity [8, 9].

Under the current pandemic conditions, hospitalisation may not be possible for all COVID-19 patients due to the shortage of medical resources. Therefore, the WHO guidelines recommend home care for patients presenting with mild COVID-19 symptoms [21]. However, several patients have died under the home care due to rapid deterioration in their respiratory condition. Our machine learning model provides a useful medical triage tool for both home care and hospital settings. This model is able to accurately predict the severity cases with 95% sensitivity and 90% specificity. Patients who have a low risk according to our model could be safely managed with home care performed under careful observation.

Without the need for a single specific biomarker, our machine learning models are able to accurately predict the necessity of oxygen therapy based solely on admission data. This could enable clinicians to start medical intervention at an early stage, before patients progress to the severe stage. Although several clinical trials are underway [5], there is, as yet, no definitive antiviral therapy for this emerging disease. Remdesivir, which is expected to be most effective treatment for COVID-19 [22], and another antiviral drug, favipiravir, act by suppressing the virus's ability to replicate. Hence, these drugs are effective in the early stages of the disease and are more effective if administered at an early stage rather than at a later stage. However, these antiviral drugs can cause serious adverse effects such as renal impairment, liver impairment, or teratogenicity [22, 23]. Hence, the administration of these drugs to all COVID-19 patients is not advisable. Prediction of disease progression at an early stage would provide an insight into appropriate use of these "double-edged sword" drugs and would contribute to a reduction in mortality rates.

Our study has some limitations. Firstly, the sample size is relatively small (n = 194). Small sample size causes overfitting issues in AI modelling. Although we used 10-fold cross validation to validate the model and we achieved high performance, larger studies should be conducted to validate the findings regarding these machine learning models. Secondly, our model was designed to predict disease progression only to the stage where oxygen therapy is required. This study did not attempt to predict progression to the more severe stages of disease that require intubation or the use of extracorporeal membrane oxygenation therapy. Therefore, careful observation remains indispensable during and after hospitalisation or home care.

Conclusion

Artificial intelligence models are useful in predicting the need for oxygen therapy in the early stages of COVID-19. The artificial intelligence model employing a combination of clinical metadata with CT images predicted the necessity for oxygen therapy with the greatest accuracy (95%). Although the accuracy of the

machine learning model was lower than that of the multimodal deep learning model, it remained sufficiently accurate for medical triage under the pandemic conditions.

Declarations

Acknowledgements

The authors would like to thank the all task members of Japan Self-Defense Forces who are involved in the management and treatment of COVID-19 patients, especially members of Self-Defense Forces Central Hospital and Japan Self-Defense Forces Hospital Yokosuka. We are also grateful to Yohsuke Suyama (National Defense Medical College) for radiological interpretation, and Yutaka Igarashi (Nippon Medical School) for scientific advice.

Author Contributions

HO and PR wrote the manuscript; PR, WZ, PH, DY, ZX, and DX constructed the AI models; HM, HS, TK and MT collected the clinical data; WM interpreted the CT images; KT and SY supervised the study; All authors approved the final manuscript.

Additional Information

Competing interests

The author(s) declare no competing interests.

Source of Funding: None

References

1. World Health Organization. Novel coronavirus, 2019–. *nCoV*) situation reports, Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/

2. World Health Organization. Press briefings, Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/press-briefings

3. Remuzzi, A. & Remuzzi, G. COVID-19 and Italy: what next? *Lancet* **395**, 1225–1228 (2020). 1016/S0140-6736(20)30627-9, Pubmed:32178769

4. Sanders, J. M., Monogue, M. L., Jodlowski, T. Z. & Cutrell, B. Pharmacologic treatments for coronavirus Disease 2019 (COVID-19): a review. *JAMA* **323**, 1824–1836 (2020). 10.1001/jama.2020.6019, Pubmed:32282022

5. Centers for Disease Control and Prevention. Interim clinical guidance for management of patients with confirmed coronavirus disease (COVID-19), Available from: https://www.cdc.gov/coronavirus/2019-

ncov/hcp/clinical-guidance-management-patients.html

6. Tabata, S. *et al. The Clinical Characteristics of COVID-19: a Retrospective Analysis of 104 Patients from the Outbreak on Board the Diamond Princess Cruise Ship in Japan. medRxiv*. Preprint 2020. 1101/2020.03.18.20038125

7. Wang, D., *et al.* Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020). 1001/jama.2020.1585, Pubmed:32031570

8. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054–1062 (2020). 1016/S0140-6736(20)30566-3, Pubmed:32171076

9. Huang, C. *et al.* Li Xet a. *Lancet* **395**, 497–506 (2020). 1016/S0140-6736(20)30183-5, Pubmed:31986264

10. Ji, *et al.* Prediction for progression risk in patients with COVID-19 pneumonia: the CALL score. *Clin. Infect. Dis.* (2020). 10.1093/cid/ciaa414, Pubmed:32271369

11. Liu, Y., Chen, P. C., Krause, J. & Peng, How to read articles that use machine learning: users' guides to the medical literature. *JAMA* **322**, 1806–1816 (2019). 10.1001/jama.2019.16489, Pubmed:31714992

12. Pan, F. *et al.* Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia. *Radiology* **295**, 715–721 (2020)

13. Clara imaging, Available from: https://developer.nvidia.com/clara-medical-imaging

14. Tan, M. & Le, Q. V. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv 2019*

15. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950). 1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3, Pubmed:15405679

16. Ronneberger, O. & Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv 2015*

17. Wynants, *et al.* Prediction Models for Diagnosis and Prognosis of covid-19 Infection: systematic Review and Critical Appraisal. *BMJ* **369**, m1328 (2020). 10.1136/bmj.m1328, Pubmed:32265220

18. Amalou, *et al.* Targeted early chest CT in COVID-19 outbreaks as diagnostic tool for containment of the pandemic- A multinational opinion. *Diagn. Interv. Radiol.* (2020). 10.5152/dir.2020.20231, Pubmed:32352918

19. Orsi, M. A., Oliva, A. G. & Cellina, M. Radiology Department preparedness for COVID-19: facing an unexpected outbreak of the disease. *Radiology* **295**, E8 (2020). 1148/radiol.2020201214, Pubmed:32228364

20. Tan, L. *et al.* Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. *Signal Transduct. Ther.* **5**, 33 (2020). 10.1038/s41392-020-0148-4

21. World Health Organization. Home care for patients with COVID-19 presenting with mild symptoms and management of their contacts, Available from: https://www.who.int/publications-detail/home-care-for-patients-with-suspected-novel-coronavirus-(ncov)-infection-presenting-with-mild-symptoms-and-management-of-contacts

22. Grein, *et al.* Compassionate use of remdesivir for patients with severe Covid-19. *N. Engl. J. Med.* (2020). 10.1056/NEJMoa2007016, Pubmed:32275812

23. Furuta, Y., Komeno, T. & Nakamura, T. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.* 93, 449–463 (2017). 2183/pjab.93.027, Pubmed:28769016

Tables

Table 1. Performance of machine learning models: a comparison of the AUC, sensitivity, specificity, and accuracy of logistic regression, Random Forest, and XGBoost Models.

Data are shown as number (percentage) or median [interquartile range].

Clinical characteristics (n=194)	
Age, years	55 [40–71]
Male, n (%)	112 (57.7)
Body height, cm	163.0 [158.0–170.0]
Body weight, kg	62 [53–72]
BMI	22.9 [20.1–25.9]
Medical history, n (%)	
cardiovascular	39 (20.1)
respiratory	19 (9.7)
DM	11 (5.7)
others	61 (31.4)
Body temperature, ℃	36.8 [36.5–37.2]
Respiratory rate, bpm	18 [16-20]
Systolic blood pressure, mmHg	127 [114–145]
Diastolic blood pressure, mmHg	81 [74–90]
Heart rate, bpm	83 [75–92]
SpO2	97 [96–98]
Fever (≥37.5℃) , n(%)	61 (31.4)
Cough, n (%)	73 (37.6)
Arthralgia, n (%)	21 (10.8)
Abdominal symptom, n (%)	26 (13.4)
Admission from onset, day	7 [5–10]
CT findings	
Score	5.0 [1.0–13.0]

Laboratory findings	
Blood urea nitrogen, mg/dL	13.5 [11.0–17.0]
Creatinine, mg/dL	0.78 [0.64–0.91]
Aspartate transaminase, IU/L	26.5 [21.0–35.0]
Alanine aminotransferase, IU/L	25.5 [16.0–40.0]
Total bilirubin, mg/dL	0.50 [0.40–0.70]
γ-glutamyl transpeptidase, IU/L	31.0 [19.0–62.0]
Amylase, IU/L	70.0 [57.0–88.0]
lactate dehydrogenase, IU/L	209.0 [176.0–260.0]
Albumin, mg/dL	4.1 [3.8–4.3]
C-reactive protein, mg/dL	0.7 [0.1–2.8]
Red blood cell count, $\times 10^4/\mu l$	487 [445–520]
Hemoglobin, g/dL	14.6 [13.4-15.6]
White blood cell count, /μl	4950 [4046–6410]
Platelet count, $10^4/\mu l$	20.7 [17.4–26.4]
Neutrophil, (%)	65.4 [57.5–72.0]
Lymphocyte, (%)	25.4 [17.2–33.4]
Monocyte, (%)	7.6 [5.5–9.6]
Eosinocyte, (%)	0.7 [0.2–1.7]
Lymphocyte count, /μl	1234 [860–1655]
Neutrophil-to-lymphocyte ratio	2.54 [1.78–4.09]
Activated partial thromboplastin time, s	31.9 [29.3–34.8]
Prothorombin time international normalized ratio	1.0 [1.0–1.1]

Table 2. Performance of the Deep Learning Models and Machine Learning Models: a comparison of the AUC, sensitivity, specificity and accuracy of the Deep Learning Model (based on CT images) with the Multimodal Deep Learning Model (combining clinical features and CT images).

Model	Logistic Regression	XGBoost	Random Forest	Deep Learning	Multimodal Deep Learning
AUC	0.93	0.93	0.91	0.91	0.97
Sensitivity	0.95	0.95	0.95	0.97	1.00
Specificity	0.88	0.9	0.86	0.84	0.94
Accuracy	0.86	0.88	0.82	0.87	0.95

Figures

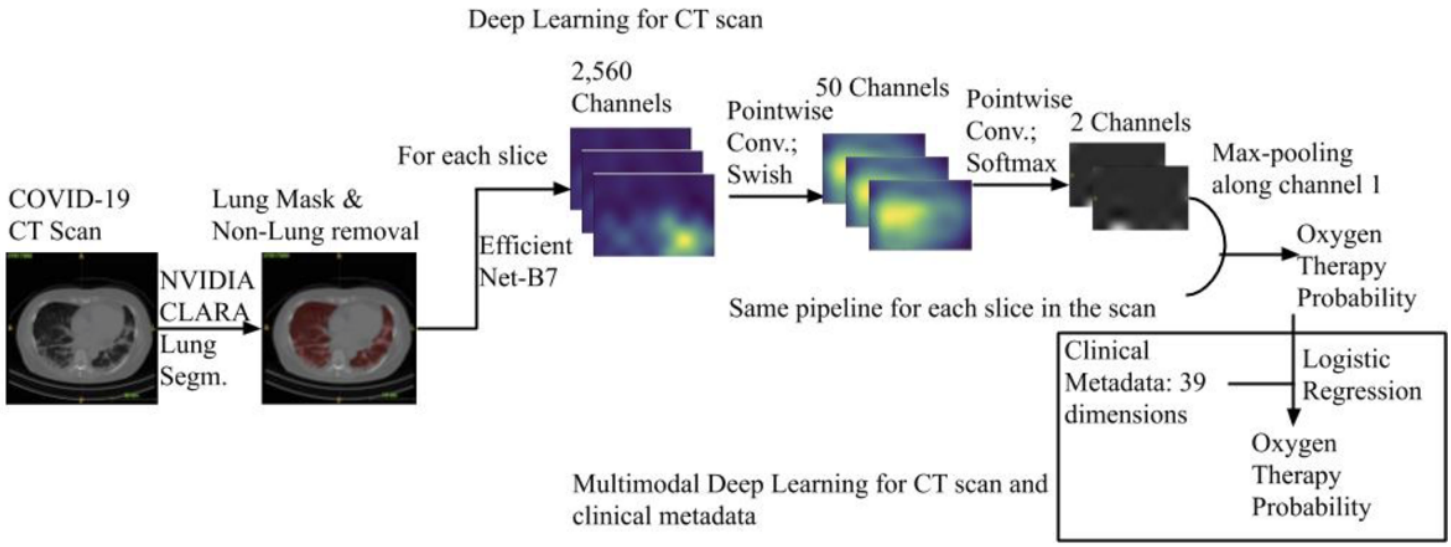

Figure 1

A Flowchart of the Deep Learning Model based on CT images (upper portion without the bottom right rectangular box) and Multimodal Deep Learning Model based on CT images and Clinical metadata (whole pipeline)

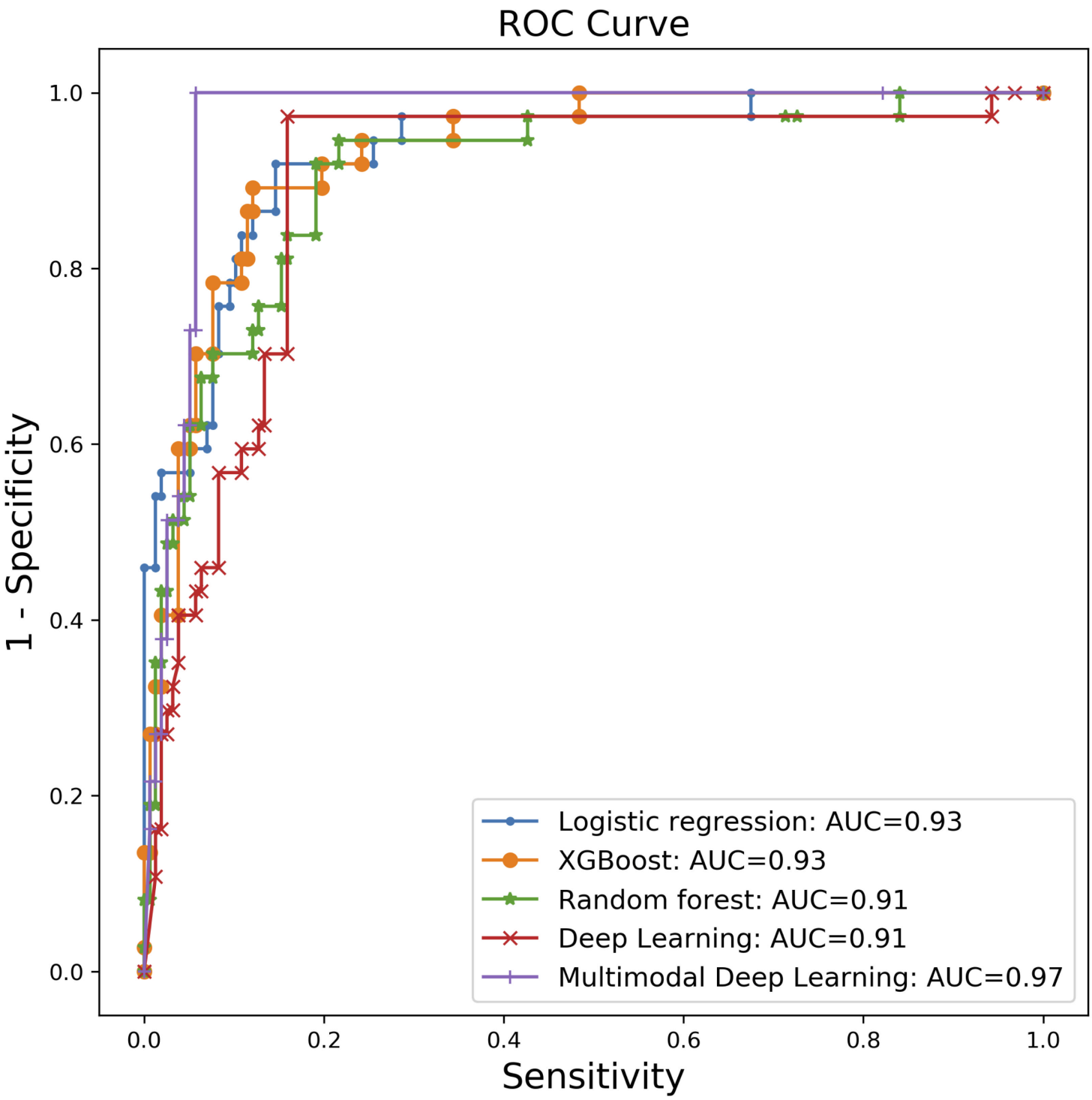

Figure 2

The ROC Curve of Each Machine Learning Method, including Logistic Regression Model, Xgboost Model, and Random Forest Model.

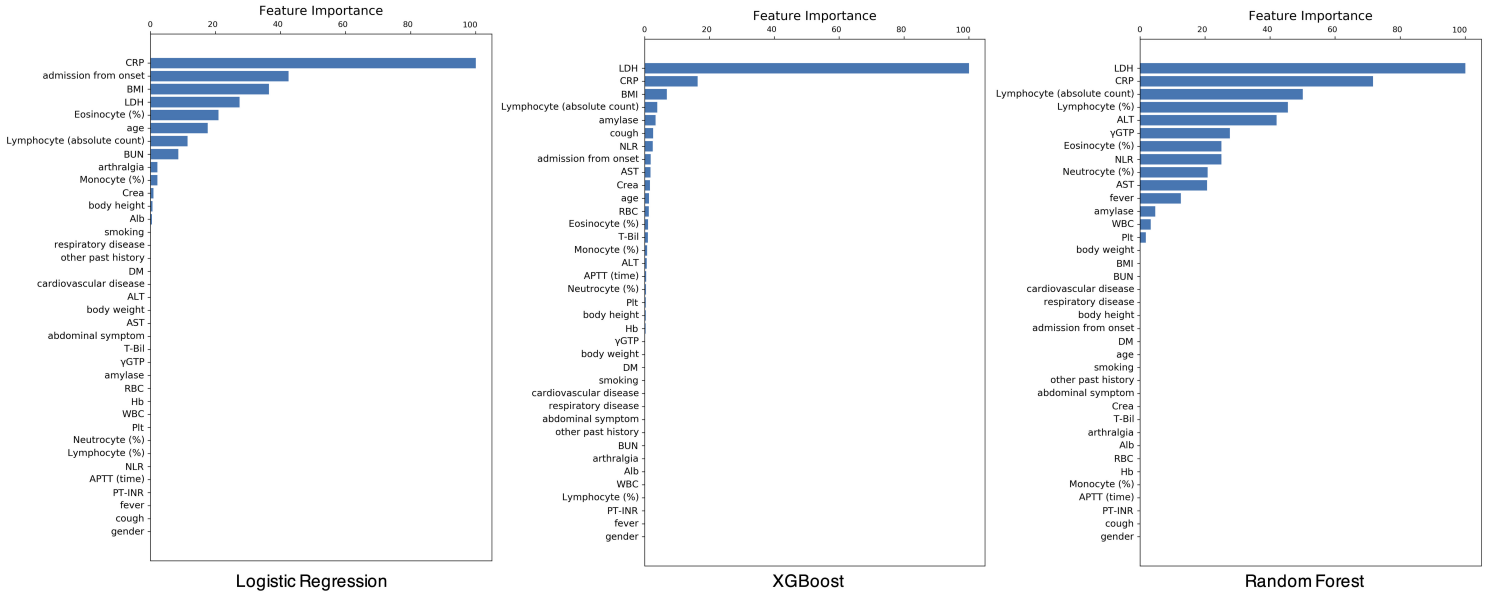

Figure 3

The Features of Importance in Logistic Regression Model

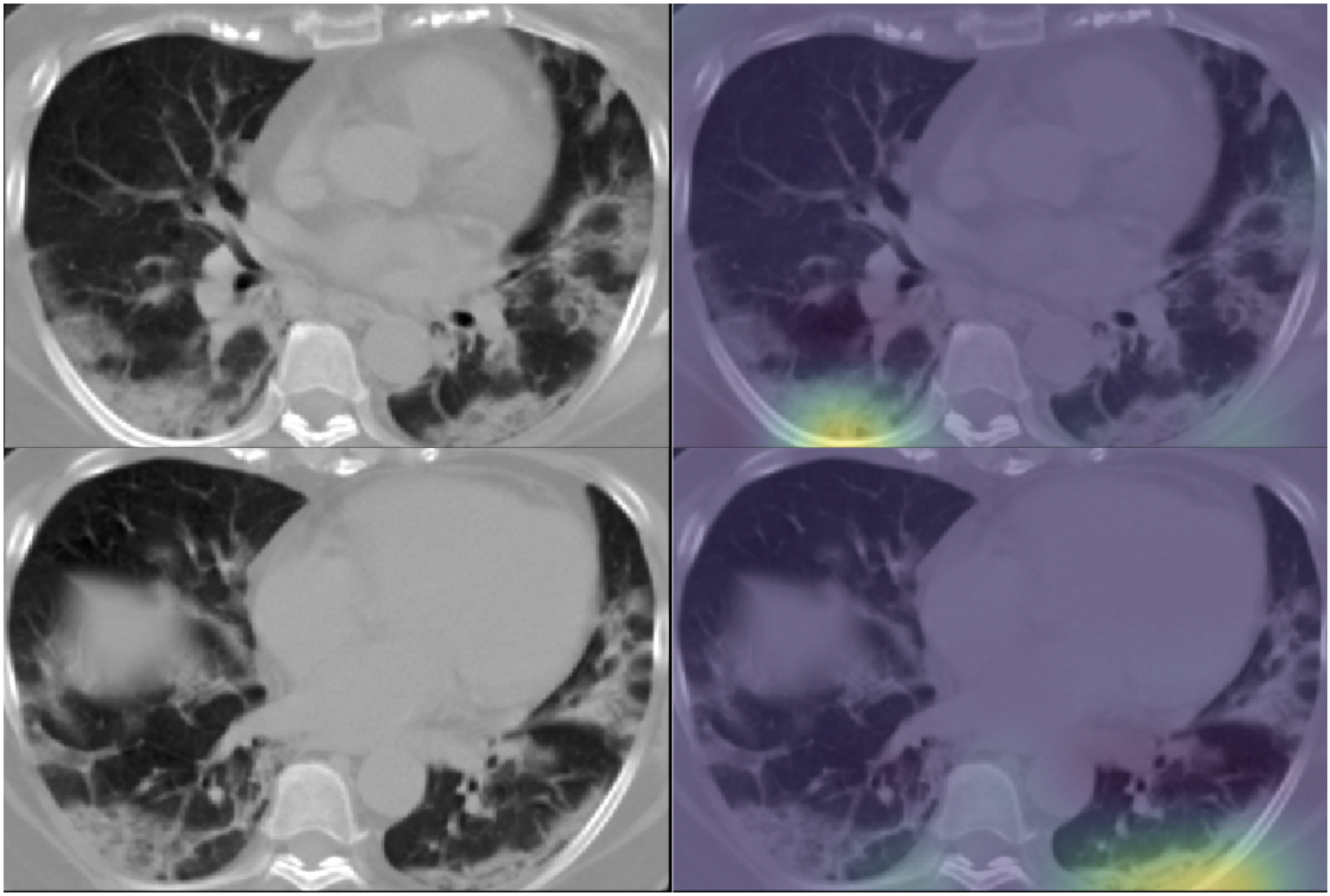

Figure 4

Heatmap of Deep Learning Visualisation of CT images