OPEN PEER REVIEW

Open Peer Review (OPR)

As a new feature for progressing towards transparency, we decided to open a new window for all of our editors as well as authors titled "**Open Peer Review**". We hope by this new facility, our reviewers will be more motivated and authors will be more satisfied with the review process. We believe that publishing our peer review reports could make a transparent and clear environment for all our efforts within a journal, but not all reviewers tend to publish their comments.

What is "Open Peer Review" process?

An "Open Peer Review" process is making the details of all review process (including reviewers, associate editors, and EICs comments) as "Public" as it is agreed by EIC, Authors, and reviewers.

Advantages of "Open Peer Review" process

- More transparency, constructiveness and tactful comments of the peer review process: leads to an increase in the quality of reviews
- More motivations for all involved roles in the review process
- Authors' satisfactions from the review process: Increases honesty between authors and reviewers
- Education of both authors and new students
- Prevents reviewers from following their individual agendas and leading to the detection of reviewers' conflicts of interests

You can find out more at:
https://brieflands.com/briefland/knowledgebase/category/tree.html#opr.html

Iranian Journal of Radiology
Article DOI: 10.5812/iranjradiol.112562
Published in: Iranian Journal of Radiology: 19(1); e112562

Peer Review Report for "Early Identification of COVID-19 Progression to Its Severe Form Using Artificial Intelligence"

Author(s): Lei Yuan, Jia Chen, Hui Feng, Junwei Lv, Xuefang Lu, Mengyao Ji

Review Timeline:

Submit Date: 28 Dec 2020
Revised Date: 13 Dec 2021
Accepted Date: 18 Dec 2021

Revision (0)

Here, you can see the **Reviewers**, **Associate Editors** and **EICs'** comments from the beginning to the end of the revision process.

Hamed Dashti: Reviewer | **Revision** (0)

12 Feb 2021

Dear AE,

This is an interesting study and the authors have collected a unique dataset including clinical parameters, blood tests, physicians extracted, and AI extracted features of 786 patients. They compare the results of 4 prediction models to identify the COVID-19 progression of the patients. They also find new radiological signs that can help to understand prognosis prediction. However, in my opinion, the paper has some shortcomings in regards to some data analyses and text, and I feel this unique dataset has not been utilized to its full extent. Below I have provided numerous remarks on the text as it is often vague and long-winded.

1- In several instances, I also suggested citing more relevant and recent literature including

a- Mei, X., Lee, HC., Diao, Ky., et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 26, 1224–1228 (2020). https://doi.org/10.1038/s41591-020-0931-3

b- Hashmi, Hafiz Abdul Sattar, and Hafiz Muhammad Asif. "Early detection and assessment of covid-19." Frontiers in Medicine 7 (2020): 311.

c- Feng, Zhichao, et al. "Early prediction of disease progression in COVID-19 pneumonia

patients with chest CT and clinical characteristics." Nature communications 11.1 (2020): 1-9.

2- Boruta algorithm isn't a suitable choice for highly correlated features. The correlation between features should be explained to support the decision.

3- Table 2 should be reorganized to be more readable.

4- The implemented models should be discussed, especially deep models.

Given these shortcomings, the manuscript requires major revisions.

Thanks.

Faezeh khorasanizadeh: Reviewer | **Revision** (0) 14 Feb 2021

Dear AE,

The manuscript entitled '' Early identification of COVID-19 progression to severe illness using artificial intelligence '' is a well-conducted and well-organized study, combined different clinically, laboratory and radiological variables in individual models using AI to predicting COVID-19 progression with 786 sample volume and conclude that (30.3%) patients developed into severe or critical pneumonia during 14-day follow up and Urea, albumin and lesion size in basal segment of right lower lung lobe or the proportion of CT values of [-200~60) in left lung have significant predictive role in disease progression which introduce interesting and somehow new emerging concept about location of lesions at initial CT .

The manuscript has scientific merit to be accepted, with few minor revisions;

Comments:

1. Table S1, S2 and figure S1, S2 are not available for the reviewer .

2. Mild disease should be defined in definition section also it is better at part 3 of inclusion criteria in fig 1 to replace with non-severe illness at admission

3. Recent relevant published articles about using AI in covid-19 progression could be added in discussion and references, like:

Li, Y., Shang, K., Bian, W. et al. Prediction of disease progression in patients with COVID-19 by artificial intelligence assisted lesion quantification. Sci Rep 10, 22083 (2020). https://doi.org/10.1038/s41598-020-79097-1

4. Figure 2 included excess photos which added no more information, it seems to be fig a,b and g and are enough to show quantitative evaluation system with more detailed in zoom-in pictures

Thanks.

Madjid Shakiba: Reviewer | **Revision** (0) 30 Apr 2021

Dear AE,

Comments for the authors:

1- Please add objective part for abstract and paper according to journal format.

2- The time and protocol of CT scan taking is important. The study is retrospective and it is questionable if CT scans have been indicated based on the patient's clinical condition or based on a fix protocol, it has been indicated for all patients. Please explain about the CT scan indication criteria: Is it a mandatory exam in all patients? When is it indicated? Is it indicated in a fix time after beginning of disease? We know the corona CT manifestations are shown best after 6-7 days from beginning, however, some patients with rapid

progression show severe lung involvement earlier. In a patient in early phase, CT could show early findings and in a serious patient, it is severely involved and naturally, these findings affect the AI efficiency. In this regard, it seems the authors should implement a predefined and fix CT taking protocol, not based on the patients clinical condition. Please explain in this regards.

3- A considerable part of patients are the patients who develop pneumonia and severe disease while have no pneumonia at the time of positive PCR [patients with slowly progressing disease] and even they can show no pulmonary involvement in early CT. these patients have been excluded from the study [as the pneumonia in CT is inclusion criteria]. Please explain and discuss in this regard. They can mention a part that the study results are generalizable to patients with mild pneumonia in the first stages of disease or at the time of assessment.

4- This sentence is unclear. please clarify: “The high-risk group involved the COVID-19 patients progressed from no-severe to severe illness.” Does involved mean included?

5- How did the authors divide the whole patients to training and test subgroups? based on randomization? Please explain.

6- Why did the authors exclude patients with more than 14 days of treatment before enrollment?

7- What is the meaning of clinical course lower than 14 days in exclusion criteria? did these patients die or they did not have pneumonia?

8- This part: “ventilation, septic, septic shock” should be: “ventilation, sepsis, septic shock”

9- How did the authors rule out pulmonary thromboemboli among the patients? This could happen in a mild or moderate pneumonia and deteriorate the clinical condition and these patients are different from the patients that are critical due to respiratory distress syndrome due to severe lung involvement. It is questionable if we include both of these patients in one AI system.

10- What was the method for finalizing the CT evaluation regarding the fact that there were two radiologists assessing the images? how did they integrate the opinions of two radiologists? did they remark on images by a consensus or the third radiologist finalized the CT evaluation? Please explain. In addition, did the authors assess the reliability of radiologists in CT evaluations? [interrater and intrarater reliability]

11- What was the follow up period for patients? How and when were the authors sure that the patient is cured and recovered from covid? In fact, what was the cure criteria in a pneumonia patient with covid? [A patient can develop severe pneumonia in different days after beginning of the disease and we have many variations in this regard].

12- It is unclear if the authors include patients without dyspnea? Please note a patient could have mild pneumonia without dyspnea while he/she could develop severe lung involvement after a while and it seems the authors have not recruited these patients in the study [did the authors define any cases as mild pneumonia?] it seems the authors enrolled patients with compliant of dyspnea but this is not mentioned explicitly in inclusion criteria. [The authors mentioned: “Patients with pneumonia signs (fever, cough, dyspnoea, fast breathing) but no signs of severe ……. If the authors only enrolled the patients with compliant of dyspnea, they have ignored mild CT involvement cases without dyspnea at admission but these patients could progress to severe forms and so, the results of the study is not inclusive and comprehensive for all covid patients and could be applicable only for covid patients that have both CT involvement and dyspnea at admission. This could be a shortcoming for the

study. Please explain and define exactly in the paper which patients have been included in the study. If the study has been done only on patients in moderate condition [dyspnea and positive CT], then this point should be mention and clearly explain in the paper.

13- The authors mentioned: “Thereafter, radiological features (48 ones in total),….” What were these features? [it could be mentioned in a table]

14- The authors mentioned they used an artificial intelligence system for automatically quantification lesions. There are some considerations in this regard:

I- What is the name of this software or system? Is it precise and valid? Did you publish a paper on this software? please add the data of this software in the paper.

II- I think this is more an image processing software [rather than an AI software]. please explain.

III- The authors mentioned: “under the supervision of two board-certified radiologists with more than 5 years of experience”. What is the meaning of under supervision? Is the AI software semi-automated? what is the method of cooperation of radiologist and AI software or how do radiologists supervise the I software?

15- What is the left middle lung lobe? Do we have such anatomical part in left lung?

16- The authors mentioned Tables S [S1-S3] and figures S, but I could not find this table.

17- The authors mentioned in Model-AI-Mimic-Doc, they deliberately reclassified the numeric values into two categories as in Model-R-Doc. How did they implement this dichotomization? How did they select the cut off values for dichotomization? Please explain.

18- The authors mentioned: “The predictive performance of those models was evaluated by five-fold cross-validation first and further assessed in test set” this part is somehow unclear and it is better the sentence is written clearer. What is five-fold cross-validation?

19- The authors mentioned: “abnormally distributed variables”. It is better to write for variables without normal distribution.

20- Mean age is clinically and statistically different among train and test groups. Is it important? Please discuss.

21- What was the time of laboratory assessments? Did the authors evaluate the laboratory indicators at the time of CT and first evaluation?

22- It seems RLLL should be RLL.

23- What is difference between training and validation sets in figure 3?

24- How did the authors convert the output of models to ROC curve and analysis? In each model there are more than one variable and it is not evident how did the authors used variables in the models to a ROC analysis. [ROC analysis usually use a continuous variable for its analysis. Here we don’t know how did the authors convert more than a variable values in a single ROC curve analysis]. Did they use logistic regression output for ROC analysis?

25- What were the number of severe forms in training and test subsets?

26- As the urea and albumin are significant in all models [this point is unexpected], it reminds us if the severe cases are mostly SIRS cases that developed kidney and liver failure during sepsis and SIRS. Am I right? Authors defined severe pneumonia and critical disease and sepsis was the criteria of critical cases; not severe cases. Did all severe cases of the study develope SIRS? In this case, we can say most of severe cases developed eventually to critical pneumonia and thus, the urea and albumin [that could be increased in SIRS] were significant in all models differentiating moderate cases from severe

cases. Please explain in this regard. What were the number of severe and critical cases in the study?

27- Based on the point mentioned in question 26, I think the importance of urea and albumin in the models means only an association with severity of disease and they don't mean a causal effect regarding severity and they are not the reason for severity of disease [in fact they are results of SIRS due to severity of disease] and so, the conclusion of authors in this regard seems questionable: they concluded the protection of kidney and liver are so important [Therefore, early interventions to protect kidney function after hospital admission are highly recommended]. This conclusion may be misleading, I think the disease is reason for kidney failure [due to SIRS] and severity of disease could be unrelated to kidney malfunction. Please discuss in this regard. of course, when kidney and liver failure happens, the clinical course of disease could be worsen in a vicious cycle, but I think SIRS is cause of kidney failure at the first step. Of course, in SIRS step, conservative treatments for all organs [including kidneys and liver] is mandatory.

28- Routinely, we can omit variables with less importance in multivariate models by using regression models. The authors used Kendal tau method. This could be implemented in study by logistic regression analysis. Why did the authors used Kendal tau method?

Thanks.

REFEREE: Associate Editor | **Revision** (0)

14 May 2021

Dear EIC,

Thank you for submitting this study to IJR. This is an interesting study, however there are some concerns/comments that can improve the quality of manuscript.

The authors should consider the suggestions by the reviewers before proceeding to next step.

Thanks.

REFEREE: EIC | **Revision** (0)

15 May 2021

Dear authors,

Your article needs Major Revision.

Regards

OPEN PEER REVIEW

Revision (1)

Reply to Reviewers

Ideally, the reviewing process can significantly improve the submitted manuscripts by allowing the authors to take into account the advice of reviewers. Author(s) must reply to all reviewers' comments in a separate Word file, point by point. A "**Reply to Reviewers**" document is submitted along with revised manuscript during submission of revised files, summarizing the changes that the authors made in response to the reviewers' comments. The responses to reviewers' comments specifies how the authors addressed each comment the reviewers made.

You can read the authors' responses to the reviewers' comments in the next page.

Dear editor,

On behalf of my co-authors, we appreciate editor and reviewers very much for their positive and constructive comments and suggestions on our manuscript. We have resubmitted new version of manuscript in accordance with recommendations of the reviewers and editor. We have addressed the comments raised by the reviewers point-by-point, and the amendments are highlighted in red in the revised manuscript. We hope that the revision is acceptable and look forward to hearing from you soon.

Best wishes,

Below, please find the comments in black, followed by our responses in red. Exact changes in the manuscript are also highlighted in red.

Manuscript ID: 112562

Title: Early identification of COVID-19 progression to severe illness using artificial intelligence

Reviewer 1:

1. Please add objective part for abstract and paper according to journal format.

Response: We add the objective part in the abstract section, and also add the objective part in the paper, according to the journal format as follow.

The objective part was highlighted in red in the revised paper.

2. The time and protocol of CT scan taking is important. The study is retrospective and it is questionable if CT scans have been indicated based on the patient's clinical condition or based on a fix protocol, it has been indicated for all patients. Please explain about the CT scan indication criteria: Is it a mandatory exam in all patients? When is it indicated? Is it indicated in a fix time after beginning of disease? We know the corona CT manifestations are shown best after 6-7 days from beginning, however, some patients with rapid progression show severe lung involvement earlier. In a patient in early phase, CT could show early findings and in a serious patient, it is severely involved and naturally, these findings affect the AI efficiency. In this regard, it seems the authors should implement a predefined and fix CT taking protocol, not based on the patients clinical condition. Please explain in this regards.

Response: Actually, in this paper, patients were included in the study who were confirmed with COVID-19 by RT-PCR first, in non-severe pneumonia on admission and with their non-contrast chest CT within three days from when RT-PCR was done. Namely, patients were confirmed with COVID-19 by RT-PCR, then the mandatory CT exam will been indicated according to the clinical conditions within 3 days after admission. And the patients who were not indicated with CT exam were excluded in our study.

3. A considerable part of patients are the patients who develop pneumonia and severe disease while have no pneumonia at the time of positive PCR [patients with slowly progressing disease] and even they can show no pulmonary involvement in early CT. these patients have been excluded from the study [as the pneumonia in CT is inclusion criteria]. Please explain and discuss

in this regard. They can mention a part that the study results are generalizable to patients with mild pneumonia in the first stages of disease or at the time of assessment.

Response: Thank you for your constructive suggestion. We add this presentation at the end of the paper.

4. This sentence is unclear. Please clarify: “The high-risk group involved the COVID-19 patients progressed from no-severe to severe illness.” Does involved mean included?

Response: we have modified this unclear statement as ‘The high-risk group referred the COVID-19 patients progressed from no-severe to severe illness.’ This modification could be found at the participants section.

5. How did the authors divide the whole patients to training and test subgroups? based on randomization? Please explain.

Response: Patients were divided into training cohort (n=628) and test cohort (n=158), randomly.

6. Why did the authors exclude patients with more than 14 days of treatment before enrollment?

Response: Thank you for your constructive suggestion. Indeed, we actually excluded the patients with more than 14 days of treatment before enrollment. This is because the recent literature found that the time from illness onset to severe disease (sepsis, ARDS, ICU) was about 14 days [1].

7. What is the meaning of clinical course lower than 14 days in exclusion criteria? Did these patients die or they did not have pneumonia?

Response: As mentioned above, Zhou et. al found that the time from illness onset to severe disease (sepsis, ARDS, ICU) was about 14 days. For easy clinical promotion and application, we excluded the patients with clinical course lower than 14 days.

8. This part: “ventilation, septic, septic shock” should be: “ventilation, sepsis, septic shock”

Response: we have modified “ventilation, septic, septic shock” as “ventilation, sepsis, septic shock”.

9. How did the authors rule out pulmonary thromboemboli among the patients? This could happen in a mild or moderate pneumonia and deteriorate the clinical condition and these patients are different from the patients that are critical due to respiratory distress syndrome due to severe lung involvement. It is questionable if we include both of these patients in one AI system.

Response: Thank you for your constructive suggestion. Indeed, the pulmonary thromboemboli patients were excluded in this retrospective cohort. We have added this supplementary content in the inclusion/exclusion criteria.

10. What was the method for finalizing the CT evaluation regarding the fact that there were two radiologists assessing the images? How did they integrate the opinions of two radiologists? Did they remark on images by a consensus or the third radiologist finalized the CT evaluation? Please

explain. In addition, did the authors assess the reliability of radiologists in CT evaluations? [interrater and intrarater reliability]

Response: Thank you for your suggestions. In this paper, we first employed two in respiratory medicine (LXF and HLH, with over 5 years' experience) by double-blind method. Their results were combined by consensus. And a senior radiologist was invited for dealing with the discrepancies. (with over 15 years' experience). The Kappa index was utilized to measure the inter-observer and intra-observer variability between human readers. We also update our manuscript in Doctor-based and AI-based quantization of CT images section and lung lesion quantization section.

11. What was the follow up period for patients? How and when were the authors sure that the patient is cured and recovered from covid? In fact, what was the cure criteria in a pneumonia patient with covid? [A patient can develop severe pneumonia in different days after beginning of the disease and we have many variations in this regard].

Response: We are sorry to make you confused. In this paper, the follow-up period was defined as from hospital admission to 14-days' follow-up. The purpose of this paper was to identify which patients would develop into severe patients based on the clinical indicators based on the clinical indicators at hospital admission, and did not involve cure criteria or recurrence criteria.

12. It is unclear if the authors include patients without dyspnea? Please note a patient could have mild pneumonia without dyspnea while he/she could develop severe lung involvement after a while and it seems the authors have not recruited these patients in the study [did the authors define any cases as mild pneumonia?] it seems the authors enrolled patients with compliant of dyspnea but this is not mentioned explicitly in inclusion criteria. [The authors mentioned: "Patients with pneumonia signs (fever, cough, dyspnoea, fast breathing) but no signs of severe ……. If the authors only enrolled the patients with compliant of dyspnea, they have ignored mild CT involvement cases without dyspnea at admission but these patients could progress to severe forms and so, the results of the study is not inclusive and comprehensive for all covid patients and could be applicable only for covid patients that have both CT involvement and dyspnea at admission. This could be a shortcoming for the study. Please explain and define exactly in the paper which patients have been included in the study. If the study has been done only on patients in moderate condition [dyspnea and positive CT], then this point should be mention and clearly explain in the paper.

Response: Thank you for your suggestion. We have added this shortcoming at the end of the paper (limitation section).

13. The authors mentioned: "Thereafter, radiological features (48 ones in total),…." What were these features? [it could be mentioned in a table]

Response: The radiological features (48 ones in total) could be found in table S2 (Supplementary Material).

14. The authors mentioned they used an artificial intelligence system for automatically quantification lesions. There are some considerations in this regard:

I-What is the name of this software or system? Is it precise and valid? Did you publish a paper on

this software? Please add the data of this software in the paper.
II-I think this is more an image processing software [rather than an AI software]. please explain.
Response: The Quantitative Evaluation System of CT for COVID-19 (YT-CT-Lung, YITU Healthcare Technology Co., Ltd., China) was employed as the image automatic segmentation and analysis tool under supervision of two experienced radiologists (>5 years of experience in chest imaging diagnosis). The system combined a fully convolutional network with adaptive thresholding and morphological operations for segmentation of lungs and pneumonia lesions. After segmentation, the total volume and mean CT value of pneumonia lesions were quantitatively computed by AI. External validation with 383 axial CT images from 206 patients showed a Dice coefficient of 82.08% for COVID-19 pneumonia lesion segmentation, which is close to the lung nodules segmentation performance (82.15% and 80.02%) for two datasets reported in Wang et al [2].

III-the authors mentioned: “under the supervision of two board-certified radiologists with more than 5 years of experience”. What is the meaning of under supervision? Is the AI software semi-automated? What is the method of cooperation of radiologist and AI software or how do radiologists supervise the AI software?
Response: Thank you for your constructive suggestions. In fact, the AI software was not semi-automated. The two board-certified radiologists were employed to re-annotate the lesions, aiming to the performance of the AI software.

15. What is the left middle lung lobe? Do we have such anatomical part in left lung?
Response: Thank you for your suggestion. We omit this mistake and upload the supplementary material, including all the AI-driven radiological features.

16. The authors mentioned Tables S [S1-S3] and figures S, but I could not find this table.
Response: We have uploaded the Supplementary Material, including Table S1, S2 and figure S1, S2.

17. The authors mentioned in Model-AI-Mimic-Doc, they deliberately reclassified the numeric values into two categories as in Model-R-Doc. How did they implement this dichotomization? How did they select the cut off values for dichotomization? Please explain.
Response: Receiver operating characteristic (ROC) curves were constructed to evaluate the sensitivity and specificity in predicting disease progress and the cut off values were determined by the maximum Youden index and the two indicators were dichotomised accordingly.

18. The authors mentioned: “The predictive performance of those models was evaluated by five-fold cross-validation first and further assessed in test set” this part is somehow unclear and it is better the sentence is written clearer. What is five-fold cross-validation?
Response: in fact, in this paper, we first randomly divided the entire cohort into training cohort and test cohort first. The training cohort was employed to train the model using 5-fold cross validation. Namely, for the training cohort, 80% was used for training the model, and 20% was used for validating the models’ performances. This step was repeated for 5 times, with 5 different training cohorts and validation cohorts. And the test cohort was used to evaluate the

trained models. We have changed this unclear sentence in the paper and highlighted in red.

19. The authors mentioned: “abnormally distributed variables”. It is better to write for variables without normal distribution.

Response: We have modified the ‘abnormally distributed variables’ as ‘variables without normal distribution’ in statistical analysis section.

20. Mean age is clinically and statistically different among train and test groups. Is it important? Please discuss.

Response: In table 1, the mean age is clinically and statistically different between train and test groups, but we believe this difference only reflect the data distribution. Namely, data distribution is statistically different. And this difference will not affect the models’ performance.

21. What was the time of laboratory assessments? Did the authors evaluate the laboratory indicators at the time of CT and first evaluation?

Response: At the beginning of the epidemic, due to our limited understanding of COVID, in order to have a deeper understanding of the disease, patients' laboratory assessments were basically assessed every day. First, better understanding of COVID itself; second, providing more targeted and personalized treatment according to the indicators of the laboratory assessments.

22. It seems RLLL should be RLL.

Response: We have modified RLLL as RLL throughout the paper, according to nice your suggestion.

23. What is difference between training and validation sets in figure 3?

Response: In this paper, 5-fold cross validation was used to train the models. The entire cohort was divided into training cohort and test cohort, respectively. For the 5-fold cross validation, the 80% of the training set was used to train the models and the rest of the training set was used for validation set. In Figure 3, the 5-fold CV training set was 80% of the training set and the 5-fold validation set was 20% of the training set.

24. How did the authors convert the output of models to ROC curve and analysis? In each model there are more than one variable and it is not evident how did the authors used variables in the models to a ROC analysis. [ROC analysis usually use a continuous variable for its analysis. Here we don’t know how did the authors convert more than a variable values in a single ROC curve analysis]. Did they use logistic regression output for ROC analysis?

Response: I am sorry to make you in confusion. Indeed, in this paper the AUC was employed to evaluate the models’ performances. And the ROC curve was plot by the outputs of the models, using package ROCR, R version 3.6.2 software. Our source codes are freely available at https://github.com/yuanlei009/NCPRredict/.

25. What were the number of severe forms in training and test subsets?

Response: As illustrated in figure 1, the ratios of non-severe and severe patients are shown in training and test subsets. Also, we added this information in table 1.

26. As the urea and albumin are significant in all models [this point is unexpected], it reminds us if the severe cases are mostly SIRS cases that developed kidney and liver failure during sepsis and SIRS. Am I right? Authors defined severe pneumonia and critical disease and sepsis was the criteria of critical cases; not severe cases. Did all severe cases of the study develop SIRS? In this case, we can say most of severe cases developed eventually to critical pneumonia and thus, the urea and albumin [that could be increased in SIRS] were significant in all models differentiating moderate cases from severe cases. Please explain in this regard. What were the number of severe and critical cases in the study?

Response: Several recent literatures [3, 4] have found, indeed, the renal and hepatic involvement will be observed in the early stage of the pneumonia, and at this point the aberrant urea and albumin indicators will be witnessed. These laboratory indicators will be earlier predictive factors for the disease progression. We also admitted some cases will developed the critical cases, but will not affect the models' performances.

27. based on the point mentioned in question 26, I think the importance of urea and albumin in the models means only an association with severity of disease and they don't mean a causal effect regarding severity and they are not the reason for severity of disease [in fact they are results of SIRS due to severity of disease] and so, the conclusion of authors in this regard seems questionable: they concluded the protection of kidney and liver are so important [Therefore, early interventions to protect kidney function after hospital admission are highly recommended]. This conclusion may be misleading, I think the disease is reason for kidney failure [due to SIRS] and severity of disease could be unrelated to kidney malfunction. Please discuss in this regard. Of course, when kidney and liver failure happens, the clinical course of disease could be worsen in a vicious cycle, but I think SIRS is cause of kidney failure at the first step. Of course, in SIRS step, conservative treatments for all organs [including kidneys and liver] is mandatory.

Response: Thank you for constructive suggestions. Actually, the renal and hepatic involvement will be observed in the early stage of the pneumonia [3, 4], urea and albumin in the models will be predictable for the disease progression. Of course, we admitted the conclusion [Therefore, early interventions to protect kidney function after hospital admission are highly recommended] is borderline accepted. So, we modified this conclusion in our paper as [Therefore, early attentions to renal and hepatic involvement after hospital admission are highly recommended].

28. Routinely, we can omit variables with less importance in multivariate models by using regression models. The authors used Kendal tau method. This could be implemented in study by logistic regression analysis. Why did the authors used Kendal tau method?

Response: I am sorry to make you confused. In this paper, Boruta algorithm was used to select features, which is capable to select optimal features carrying usable information for model's prediction and is quite popular and highly recommended in exploration of high-dimensional data. Kendall's τ coefficient was used to measure correlations between clinical variables (such as age with albumin, comorbidity with urea), to explore their internal relations and thereby

explains why some of the common clinical indicators are not selected in the final models (seen in statistical analysis section).

Reviewer 2:

This is an interesting study and the authors have collected a unique dataset including clinical parameters, blood tests, physicians extracted, and AI extracted features of 786 patients. They compare the results of 4 prediction models to identify the COVID-19 progression of the patients. They also find new radiological signs that can help to understand prognosis prediction.

However, in my opinion, the paper has some shortcomings in regards to some data analyses and text, and I feel this unique dataset has not been utilized to its full extent. Below I have provided numerous remarks on the text as it is often vague and long-winded.

1. In several instances, I also suggested citing more relevant and recent literature including:

[a] Mei, X., Lee, HC., Diao, Ky., et al. Artificial intelligence-enabled rapid diagnosis of patients with COVID-19. Nat Med 26, 1224–1228 (2020). https://doi.org/10.1038/s41591-020-0931-3

[b] Hashmi, Hafiz Abdul Sattar, and Hafiz Muhammad Asif. "Early detection and assessment of covid-19." Frontiers in Medicine 7 (2020): 311.

[c] Feng, Zhichao, et al. "Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics." Nature communications 11.1 (2020): 1-9.

Response: We have included these recent literatures in our paper as your suggestion.

2. Boruta algorithm isn't a suitable choice for highly correlated features. The correlation between features should be explained to support the decision.

Response: Thank you for your constructive suggestion. We have added the correlation between features in the feature selection and models construction section.

3. Table 2 should be reorganized to be more readable.

Response: We have reorganized Table 2.

4. The implemented models should be discussed, especially deep models.

Response: Thanks a million for your suggestion. We add this discussion at the end of the paper.

Reviewer 3:

The manuscript entitled '' Early identification of COVID-19 progression to severe illness using artificial intelligence '' is a well-conducted and well-organized study, combined different clinically, laboratory and radiological variables in individual models using AI to predicting COVID-19 progression with 786 sample volume and conclude that(30.3%) patients developed into severe or critical pneumonia during 14-day follow up and Urea, albumin and lesion size in basal segment of right lower lung lobe or the proportion of CT values of [-200~60) in left lung have significant predictive role in disease progression which introduce interesting and somehow new emerging concept about location of lesions at initial CT .

The manuscript has scientific merit to be accepted, with few minor revisions;

Comments:

1. Table S1, S2 and figure S1, S2 are not available for the reviewer.

Response: We have uploaded the Supplementary Material, including Table S1, S2 and figure S1, S2.

2. Mild disease should be defined in definition section also it is better at part 3 of inclusion criteria in fig 1 to replace with non-severe illness at admission.
Response: we have added the mild disease definition in the definition section, and also modified the figure 1 as your suggestion.

3. Recent relevant published articles about using AI in covid-19 progression could be added in discussion and references, like: Li, Y., Shang, K., Bian, W. et al. Prediction of disease progression in patients with COVID-19 by artificial intelligence assisted lesion quantification. Sci Rep 10, 22083 (2020). https://doi.org/10.1038/s41598-020-79097-1
Response: We have included this article in the paper, according to your suggestion.

4. Fig 2 included excess photos which added no more information, it seems to be fig a, b and g and h are enough to show quantitative evaluation system with more detailed in zoom-in pictures.
Response: Thank you for your suggestion. However, in order to compare the non-progressed pneumonia group with the progressed pneumonia group. More figures were showed in figure 2. Actually, the figure is big, but it contains more comparative information, which is more readable for readers.

[1] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, and B. Cao, "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study," *Lancet*, vol. 395, no. 10229, pp. 1054-1062, Mar 28, 2020.

[2] S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert, and J. Tian, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Med Image Anal*, vol. 40, pp. 172-183, Aug, 2017.

[3] M. Ji, L. Yuan, W. Shen, J. Lv, Y. Li, M. Li, X. Lu, L. Hu, and W. Dong, "Characteristics of disease progress in patients with coronavirus disease 2019 in Wuhan, China," *Epidemiol Infect*, vol. 148, pp. e94, May 6, 2020.

[4] G. Pei, Z. Zhang, J. Peng, L. Liu, C. Zhang, C. Yu, Z. Ma, Y. Huang, W. Liu, Y. Yao, R. Zeng, and G. Xu, "Renal Involvement and Early Prognosis in Patients with COVID-19 Pneumonia," *J Am Soc Nephrol*, vol. 31, no. 6, pp. 1157-1165, Jun, 2020.

OPEN PEER REVIEW

Revision (1)

Here, you can see the **Reviewers**, **Associate Editors** and **EICs'** comments from the beginning to the end of the revision process.

Faezeh khorasanizadeh: Reviewer | **Revision (1)** 21 Jul 2021

Dear AE,
All requested changes have been made, and is accepted for publish in my opinion.
Thanks.

Hamed Dashti: Reviewer | **Revision (1)** 13 Aug 2021

Dear AE,
In response to the comments, the authors have fully improved their paper. I want to thank the authors for addressing my initial comments.
Some points remained.
1- Using the Knockoff feature selection method can improve the results and robustness.
2- Authors should explain how you extract AI-driven radiological features from CNN.

Madjid Shakiba: Reviewer | **Revision (1)** 12 Sep 2021

Dear AE,
The authors have corrected and responded many questions. some of responses contain points that are valuable if they could be mentioned in paper. I recommend the authors add the important and useful points of responses to questions 14 [part III], 17, 20, 23, 24, 26 to the paper [reviewer 1]. please consider that we try to prepare the best document for readers [not only for reviewer], so they should add any useful points that are mentioned in responses but are not mentioned in paper. this policy provide the best document for readers.
Thanks.

REFEREE: Associate Editor | **Revision (1)** 1 Oct 2021

Dear EIC,
Thanks to the authors for responding to the questions and proper revision of the manuscript. There are a few comments to improve this study. They should consider the comments before finalizing the process.
Thanks.

REFEREE: EIC | **Revision (1)** 2 Oct 2021

Dear authors,

Your article needs Minor Revision.

Regards

OPEN PEER REVIEW

Revision (2)

Reply to Reviewers

Ideally, the reviewing process can significantly improve the submitted manuscripts by allowing the authors to take into account the advice of reviewers. Author(s) must reply to all reviewers' comments in a separate Word file, point by point. A "**Reply to Reviewers**" document is submitted along with revised manuscript during submission of revised files, summarizing the changes that the authors made in response to the reviewers' comments. The responses to reviewers' comments specifies how the authors addressed each comment the reviewers made.

You can read the authors' responses to the reviewers' comments in the next page.

Dear editor,

On behalf of my co-authors, we appreciate editor and reviewers very much for their positive and constructive comments and suggestions on our manuscript. We have resubmitted new version of manuscript in accordance with recommendations of the reviewers and editor. We have addressed the comments raised by the reviewers point-by-point, and the amendments are highlighted in red in the revised manuscript. We hope that the revision is acceptable and look forward to hearing from you soon.

Best wishes,

Below, please find the comments in black, followed by our responses in red. Exact changes in the manuscript are also highlighted in red.

Manuscript ID: 112562-R1

Title: Early identification of COVID-19 progression to severe illness using artificial intelligence

Reviewer 1:

1. *All requested changes has been made, and is accepted for publish in my opinion.*

Response: Thank you for your recognition of our manuscript.

Reviewer 2:

In response to the comments, the authors have fully improved their paper. I would like to thank the authors for addressing my initial comments.

Some points remain.

1. Using the Knockoff feature selection method can improve the results and robustness.

Response: Thank you for your constructive suggestion. We admitted the knockoff feature selection method can improve the results and robustness based on some measure of variable importance, while guaranteeing finite-sample control of the false discovery rate. Indeed, we have used the knockoff feature selection method to select predictive features from potential features. However, we found that this feature selection method may not be suitable for this dataset. Actually, the Boruta algorithm also analyzed and selected features based on the importance of candidate variables.

2. Authors should explain how you extract AI-driven radiological features from CNN.

Response: Thank you for your constructive suggestion. Indeed, in this manuscript, the convolutional network (CNN) [1] with adaptive threshold and morphological operations was employed to segment the whole lung into different pulmonary regions firstly. Then a central focused convolutional neural networks (CF-CNN) [2] was utilized to locate and analyze the pneumonia lesions in each region. By thresholding on CT values in the pneumonia lesions, the CT value distribution in the lungs is calculated to obtain a histogram, which is subsequently used to calculate the pulmonary inflammation volume (pneumonia volume) with ranges of -700~-500 Hounsfield units (HU), -500~-200 HU, and -200~60 HU. We added this expression in the AI-based quantization of CT images section and heighted in red.

Reviewer 3:

The authors have corrected and responded many questions. Some of responses contain points that are valuable if they could be mentioned in paper. I recommend the authors add the important and useful points of responses to questions 14 [part III], 17, 20, 23, 24, 26 to the paper [reviewer 1]. Please consider that we try to prepare the best document for readers [not only for reviewer], so please add any useful points that are mentioned in responses but are not mentioned in paper. This policy provides the best document for readers.

Response: Thank you for your constructive suggestion. We have added these useful sections in our manuscript and highlighted in red.

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing And Computer-Assisted Intervention, Pt Iii*, vol. 9351, pp. 234-241, 2015.

[2] S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert, and J. Tian, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Med Image Anal*, vol. 40, pp. 172-183, Aug, 2017.

OPEN PEER REVIEW

Revision (2)

Here, you can see the **Reviewers**, **Associate Editors** and **EICs'** comments from the beginning to the end of the revision process.

REFEREE: Reviewer | **Revision** (2) — 15 Oct 2021

Dear AE,
The study is reviewed carefully. According to comments and revisions by the author, this manuscript is accepted.
Thanks

REFEREE: Associate Editor | **Revision** (2) — 5 Nov 2021

Dear EIC,
After appropriate revision, the manuscript could be accepted for the final process.
Thanks.

REFEREE: Reviewer | **Revision** (2) — 6 Nov 2021

Dear AE,
The study is reviewed carefully. According to comments and revisions by the author, this manuscript is accepted.
Thanks

REFEREE: Associate Editor | **Revision** (2) — 6 Dec 2021

Dear EIC
The reviewer(s) have recommended publication, but also suggest some minor revisions to the manuscript. Therefore, I invite the author to respond to the reviewer(s)' comments and revise his manuscript.
Kind regards,

REFEREE: EIC | **Revision** (2) — 6 Dec 2021

Dear authors,
The article needs Minor Revision.
regards

OPEN PEER REVIEW

Revision (3)

Reply to Reviewers

Ideally, the reviewing process can significantly improve the submitted manuscripts by allowing the authors to take into account the advice of reviewers. Author(s) must reply to all reviewers' comments in a separate Word file, point by point. A "**Reply to Reviewers**" document is submitted along with revised manuscript during submission of revised files, summarizing the changes that the authors made in response to the reviewers' comments. The responses to reviewers' comments specifies how the authors addressed each comment the reviewers made.

You can read the authors' responses to the reviewers' comments in the next page.

Dear editor,

On behalf of my co-authors, we appreciate editor and reviewers very much for their positive and constructive comments and suggestions on our manuscript. We have resubmitted new version of manuscript in accordance with recommendations of the reviewers and editor. We have addressed the comments raised by the reviewers and editors point-by-point, and the amendments are highlighted in red in the revised manuscript. We hope that the revision is acceptable and look forward to hearing from you soon.

Best wishes,

Below, please find the comments in black, followed by our responses in red. Exact changes in the manuscript are also highlighted in red.

Manuscript ID: 112562-R2

Title: Early identification of COVID-19 progression to severe illness using artificial intelligence

Reviewer 1:

1. *Accept*

Response: Thank you for your recognition of our manuscript.

Reviewer 2:

This paper is suitable for publication.

Response: Thank you for your recognition of our manuscript.

Associate Editor 1:

After appropriate revision, the manuscript could be accepted for final process.

Associate Editor 2:

1-please capitalize first characters of words in title and legends of tables.

Response: We have capitalized first characters of words in title and legends of tables throughout the paper.

2-please add full term of abbreviations for the first time usage in abstract and paper separately [for example, CI,....]

Response: We have added the full term of abbreviations for the first time usage.

3-please add sequential numbers for titles: 1-Background, 2-Objectives, 3-Patients and Merthods, 3-1-Participants, ...

Response: We have added the sequential numbers according to your suggestion.

4-there are 4 percent numbers in figure 1 including: 58.7%, 31.3%50% and 26.6%. it is unclear how did the authors calculated these percentages? Sum of them are not 100% and in addition, in training and test subsets, sum of progressed and non-progressed are not equal to 156 separately. Please clarify and correct.

Response: We have corrected and updated this according to your suggestions.

5-please introduce all abbreviations of tables in their footnotes. For example, add abbreviation of OR and CI in table 2 in footnote and AUC in table 3.

Response: We have added the abbreviations of tables in their footnotes, covering table 2 and table 3.

6-please follow journal format in references. For example, abbreviate journal names.

Response: we have changed the references style according to the journal format.

7-in footnote of table S1, please correct esions to lesions.

Response: We have corrected this mis-spelling.

8-in tables S1 and S2, the term factor seems to be better as feature or [characteristics]

Response: we used the 'feature' instead of 'factor' in table S1 and table S2.

9-please add space between words of variables column in table S2. For example; airBronchogram as: air Bronchogram, etc.

Response: we have added the space between words of variables in table S2 accordingly.

10-what is the word Thicken in table S2? Please check all parts for probable typographic errors

Response: We have changed this term to 'Vasodilatation sign' and highlighted by red in table S2.

11-please correct P=0.000 to P<0.001 [for example in figure S2]

Response: we corrected the P value accordingly in figure S2.

OPEN PEER REVIEW

Revision (3)

Here, you can see the **Reviewers**, **Associate Editors** and **EICs'** comments from the beginning to the end of the revision process.

REFEREE: Associate Editor | **Revision** (3) 16 Dec 2021

Dear EIC
Based on the authors' answers to the reviewers' comments, the reviewer(s) have recommended publication.
Kind regards.

REFEREE: EIC | **Revision** (3) 18 Dec 2021

Dear Authors,
This article is acceptable for publication in IJR.
regards