nature portfolio

Peer Review File

Development and Validation of a Real-time Prediction Model for Acute Kidney Injury in Hospitalized Patients

Corresponding Author: Dr Li Yang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
COMMENTS ON THE MANUSCRIPT
“Development and Validation of a Real-time Prediction Model for Acute Kidney Injury in Hospitalized Patients”
The authors have developed and validated a real-time prediction model for patients with acute kidney injury (AKI). Overall, the study is commendable for its large sample size and substantial number of events (AKI episodes) and controls in comparison to the number of predictors in the model (20 features). The model underwent development and validation using multicenter data to affirm its predictive validity. However, some concerns have been noted:
1. Questions concerning the missing data:
1-Given the extensive use of features into feature selection, it's conceivable that the level of missing data might be noteworthy. Did the authors establish any threshold for missing data proportion to exclude features with high levels of missingness?
2-It remains unclear whether the models were validated using datasets from hospitals 2-5 in a complete-case fashion (without imputation) or if they utilized already imputed datasets.
3-While the authors mentioned employing clinical-ready features, several of these features are not commonly included in routine investigations (such as procalcitonin, D-dimer, troponin). How should the model be applied in the absence of these predictors?
2. Questions concerning the selected features:
1-Among the 20 features, concerns may arise regarding the inclusion of "Use of diuretics," "Admission to ICU," and "Use of human albumin." These features are subject to variations between centers. However, they appear to have the most significant contribution to the prediction according to the SHAP plot. This could potentially limit the stability and generalizability of the model to other settings.
2-In Figure 3, it is unclear which dataset was used to plot the figure. If the dataset was from the predictive data timeframe (before AKI occurred), the authors should explain why some values of the "Scr" and "Lower Scr" plot are already high (ranging from 150 to 250 μmol/L).
3-Still regarding Figure 3, clarification is needed for the "latest change in Scr" plot. The unit of measurement for the x-axis scale is not clear. If μmol/L was used, it raises concerns that even small changes (0-2 μmol/L in Scr) might contribute significantly to SHAP values. Alternatively, if it is in mg/dL scale, such changes should typically suffice to diagnose AKI.
4. The authors should provide a list of definitions for the features used in the final model.
3. Other concerns
1-In the discussion section, the authors did not delve into the underlying relationship between the predictors employed in the model and AKI.
2-In Figure 3, the x-axis should include units of measurement. Additionally, all binary variables should be labeled as "Yes" or "No" instead of the current scale ranging from 0.0 to 1.0.
3-There are several types of cardiac troponin tests available. The specific type of troponin was not mentioned in the figures.
4-It is notable that almost all 95% confidence intervals (CIs) of the AUC values in the manuscript are either very narrow or similar to the mean estimate (Table 2, Table 3), which is surprising.
5-It is apparent that the models and AUC estimates were obtained through multiple and iterative testings, which may increase the likelihood of false positive findings.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This is a very interesting and important paper that seeks to bring machine learning techniques to predict inpatient AKI across 5 centers in China. while there are many merits to this work, including brining modern AKI risk prediction to the developing world, there are however several concerns about the paper that need to be addressed before acceptance of the manuscript

1) Why were patients w/ AKI in the first 24 hours excluded these seem like the patients you would want to be able to identify – consider changing this – especially since many of the patients who develop Stage 1 AKI witin 24 hours develop Stage 2 or 3 over the next 24-72 hours.
2) Why exclude patients with a baseline less than 0.6 mg/dl – these seems like ideal patients as it likely includes older patients, patients with heart and liver disease and cancer (with profound muscle wasting) all of whom are high risk for AKI.
3) There is no clear explaination for how duplicate patietns were handled ? if the same patient had AKI during the same admission or repeat admission? How was this handled ? did the baseline change ?
4) We are perplexed by the use of GFR rather than SCr – as GFR is great for those in steady state and when using outpatient values but for admitted pateints this seems like an issue as they are not in steady state and dividing the cohort by eGFR of 60 seems like it should be done by admission creatinine e.g 100, 200 mmol or 1mg/dl ,2 ,3)
5) Please provide the performance for the model to predict the receipt of RRT?
6) There needs to be extensive description of how “extreme” and non=physiologic values” were defined? – provide these cutoffs and the incidence of these – as some non-physiologic values can be accurate (low MAPs, BPs and very high and low electrolytes or BUNs etc…)
7) Similarly we did not see data round the extent of missing data in the cohorts – clearly this would vary across centers but it needs to be clearly described so we know how often the key components of the score are being tested – and then what was the rationale for imputing normal values when not tested. – if you are using tests like procalcitonin or troponin which are not commonly measured does it not make more sense to exclude these tests or leave them as missing – rather than imputing a normal value. We agree that a procalcitonin could be an important piece of info – and will be checked with more frequency than a sodium or BUN . and this should be accounted for in the model rather than assuming that 80% of the cohort has a normal procalcitonin or troponin. At the very least the numbers need to be present
8) In the descriptions of the cohort 5.4% having AKI acrose the total cohort seems low and not in line with prior publications and the fact that 40% of the AKI was in the ICU also seems low
9) Around line 286-293 when you discuss the performance in the subgroups were these changes in AUC significant?
10) The 5 features are intriguing – and it seems odd that ICU was so highly featured given the stress on ward based AKI
11) We find it interesting that rather than using 1 model across all sites that chose to refit the model at all cites – this is great - but make the expansion of the this model (there intent and something they discuss in the paper) problematic – as we love the idea of a base model or base components to the model (all models having the same 10 features) but then tailor the model to your patient population if you have more ICU perhaps you add in the procalcitionin or some other marker of critical illness versus a more cardiac based hospital …..
12) For table 3 – please provide the performance of the score based on gender (see comment 4 above as well)

Minor issues
1) Consider replacing critical AKI with the term “severe AKI” as this is more common in the literature
2) The abstract should make it is clear it is only SCr based AKI
3) We do not think the statement in the intro line 90 is correct – as there is actually more AKI on the wards than in the ICU – just because of the sheer volume of more ward patients than ICU patients

(Remarks on code availability)
i am not a coding expert -

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I feel very appreciated to the authors that can responding to all of my queries adequately. I do not have any more questions

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)

The authors have satisfactorily addressed the comments of Reviewer #2.

Clarity on prediction times:
It may be obvious to the authors but it is not clear from the paper from phrases such as "development of AKI within 48 hours" as to exactly when the predictions are made, and "within 48 hours" of what - admission or prior to AKI incident. The supplementary Figure S1 helped to understand it. It is suggested to include Figure S1 in the main paper with its description, or at least give a clear explanation of when the predictions are made and for AKI to occur within what interval. From Figure S1 and its description, it is still not clear how the prediction times (i.e. yellow dots) are decided. Are the predictions made every 24 hours for AKi to occur within the next 48 hours? Or are they made every 6 hours corresponding to blue dots (if not, then why not)? How was the first prediction point "A" in the figure decided? Or is that just a sample prediction point.

A few minor things:

- Line 95: "patient population of derivation" is unclear, either clarify it or improve the sentence.

- Line 107: "different hospitals convenient and economical", please improve the sentence.

- Typos: Line 101 and Line 407 "tired hospitals"

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

RESPONSE TO REVIEWERS

Reviewer #1 (Remarks to the Author):

COMMENTS ON THE MANUSCRIPT

"Development and Validation of a Real-time Prediction Model for Acute Kidney Injury in Hospitalized Patients"

The authors have developed and validated a real-time prediction model for patients with acute kidney injury (AKI). Overall, the study is commendable for its large sample size and substantial number of events (AKI episodes) and controls in comparison to the number of predictors in the model (20 features). The model underwent development and validation using multicenter data to affirm its predictive validity. However, some concerns have been noted:

1. Questions concerning the missing data:
1-Given the extensive use of features into feature selection, it's conceivable that the level of missing data might be noteworthy. Did the authors establish any threshold for missing data proportion to exclude features with high levels of missingness?

RESPONSE: We thank the Reviewer for raising this critical question. While conducting feature selection, we established a threshold of 50% for missing data proportion to

exclude features with high levels of missingness, except for several specific features that were usually selectively prescribed by physicians under disease conditions with high relevance to AKI. These specific features included procalcitonin, blood culture, serum myoglobin, fecal occult blood, ANCA, and Coombs test. In the cases with missing data of these features, we made imputations with normal values (negative for binary features, average or median within normal range for continuous features), assuming that patients with no test prescriptions had been assessed by the physicians and regarded as having low risks of developing these disease conditions. Among the 20 features that were finally included in the prediction model, procalcitonin remains the only feature with high missing data proportion.

We have included the threshold for missing data proportion and the way of imputation for specific features in the Supplementary Methods section (Page1, Item S1, paragraph 1) in the revised supplementary material. The missing data proportion of the 20 features in the prediction models for AKI and severe AKI in the derivation cohort are listed as follows.

Table 1-1-1 Missing Rate for 20 features included in the prediction model for AKI

Feature	**Missing Rate**
Latest changing rate of Scr (Originally named "Latest change in Scr")	*0*
Use of Diuretics	*0*
Percentage of lymphocyte	*0.028*
Admission to ICU	*0*
Use of human albumin	*0*
Lowest Scr	*0.026*
NE percentage/albumin ratio	*0.048*
Scr	*0.026*
Lowest BUN	*0.026*
Heart rate	*0.007*
Cardiac troponin I	*0.319*
Highest NE percentage	*0.028*
BUN	*0.026*
Procalcitonin	*0.780*
Baseline red blood cell count	*0.027*
Lowest D-dimer	*0.124*
Blood platelet count	*0.029*
Respiratory rate	*0.006*
Lactic dehydrogenase	*0.222*
Cardiovascular surgery	*0*

Table 1-1-2 Missing Rate for 20 features included in the prediction model for severe AKI

Feature	Missing Rate
Initial changing rate of Scr (Originally named "Initial change in Scr")	*0*
BUN	*0.026*
Percentage of lymphocyte	*0.027*
Use of Diuretics	*0*
Latest changing rate of Scr (Originally named "Latest change in Scr")	*0*
Highest D-dimer	*0.121*
Highest BUN	*0.026*
Procalcitonin	*0.769*
Heart rate	*0.006*
Lowest Scr	*0.026*
Scr	*0.026*
Admission to ICU	*0*
Baseline red blood cell count	*0.027*
White blood cell count	*0.027*
High-sensitive CRP/albumin ratio	*0.321*
Blood platelet count	*0.028*
Lactic dehydrogenase	*0.217*
Direct bilirubin	*0.053*
Initial changing rate of Scr (Originally named "Initial change in Scr")	*0*
BUN	*0.026*

2-It remains unclear whether the models were validated using datasets from hospitals 2-5 in a complete-case fashion (without imputation) or if they utilized already imputed datasets.

RESPONSE: We apologize for the unclarity of the Methods description in the previous manuscript. The models were validated using imputed datasets from hospitals 2-5 following the rules of imputation as described in Methods section. "Missing feature values were imputed by the mean (normally distributed data) or median (non-normally distributed data). For laboratory tests that were only prescribed when a certain clinical condition was highly suspected, such as cardiac troponin I and procalcitonin, missing values were imputed as normal values." For future application of the model, real-time data will be automatically imputed under rules set in advance and inputed into the model to generate prediction result timely.

3-While the authors mentioned employing clinical-ready features, several of these features are not commonly included in routine investigations (such as procalcitonin, D-dimer, troponin). How should the model be applied in the absence of these predictors?

RESPONSE: We appreciate the Reviewer for raising this very important issue. As

pointed out by the Reviewer, features including PCT, cTnI and D-dimer are not commonly included in routine investigations, yet they are frequently tested in patients with suspected infection, cardiac attack, or disease conditions with coagulation abnormalities. In patients with no prescription of PCT and cTnI, we made imputations with normal values, assuming that these patients had been assessed by the physicians and regarded as having low risks of developing these disease conditions. The missing rate of D-dimer range from 0.102-0.241 in five centers of our study, we imputed D-dimer with the median. If the model is being implemented in a hospital or a patient population without test of these features, it is indeed a very important issue of how the model should be applied. To answer the Reviewer's question, we conducted experiments to test the performance of the model assuming absence of these features and imputed the features of all the patients using normal values (median of normal values for procalcitonin, mean of normal values for cardiac troponin I) or median (for D-dimer). As shown below in Table 1-1-3, the AUCs for AKI prediction decreased slightly in the absence of one or two or three of these features in site 1-5. Similar results have been found while testing the performance of the model for severe AKI prediction in the absence of PCT and D-dimer (Table 1-1-4). Based on these experiments, we think the prediction models could be applied in the absence of these predictors.

Table 1-1-3 The performances of the AKI prediction model assuming absence of features including PCT, cTnI and D-dimer

	AUCs for Refitted Models									
	Site 1	***P***	**Site 2**	***P***	**Site 3**	***P***	**Site 4**	***P***	**Site 5**	***P***
No missing feature	*0.895(0.889-0.900)*	*-*	*0.863(0.860-0.866)*	*-*	*0.808(0.804-0.812)*	*–*	*0.885(0.881-0.890)*	*-*	*0.898(0.894-0.903)*	*-*
Absence of PCT	*0.887(0.881-0.892)*	*< 0.01*	*0.846(0.843-0.849)*	*< 0.01*	*0.807(0.803-0.811)*	*0.52*	*0.875(0.871-0.879)*	*< 0.01*	*0.890(0.886-0.895)*	*< 0.01*
Absence of cTnI	*0.885(0.879-0.890)*	*< 0.01*	*0.850(0.847-0.853)*	*< 0.01*	*0.801(0.797-0.805)*	*< 0.01*	*0.872(0.868-0.877)*	*< 0.01*	*0.893(0.888-0.897)*	*< 0.01*
Absence of D-dimer	*0.884(0.879-0.890)*	*< 0.01*	*0.856(0.853-0.859)*	*< 0.01*	*0.804(0.800-0.808)*	*< 0.01*	*0.880(0.876-0.885)*	*< 0.01*	*0.890(0.886-0.895)*	*< 0.01*
Absence of PCT and cTnI	*0.875(0.869-0.881)*	*< 0.01*	*0.835(0.832-0.838)*	*< 0.01*	*0.802(0.798-0.806)*	*< 0.01*	*0.860(0.856-0.865)*	*< 0.01*	*0.883(0.879-0.888)*	*< 0.01*
Absence of PCT and D-dimer	*0.875(0.869-0.881)*	*< 0.01*	*0.839(0.836-0.842)*	*< 0.01*	*0.803(0.799-0.807)*	*< 0.01*	*0.870(0.865-0.874)*	*< 0.01*	*0.882(0.878-0.887)*	*< 0.01*
Absence of cTnI and D-dimer	*0.872(0.866-0.878)*	*< 0.01*	*0.842(0.839-0.845)*	*< 0.01*	*0.798(0.794-0.802)*	*< 0.01*	*0.867(0.863-0.872)*	*< 0.01*	*0.885(0.881-0.889)*	*< 0.01*

Absence of PCT, cTnI, and D-dimer	*0.861(0.855-0.867)*	*＜0.01*	*0.827(0.824-0.831)*	*＜0.01*	*0.799(0.794-0.803)*	*＜0.01*	*0.855(0.850-0.859)*	*＜0.01*	*0.875(0.871-0.880)*	*＜0.01*

Table 1-1-4 The performances of the severe AKI prediction model assuming absence of features including PCT and D-dimer

	AUCs for Refitted Models									
	Site 1	***P***	**Site 2**	***P***	**Site 3**	***P***	**Site 4**	***P***	**Site 5**	***P***
No missing feature	*0.946(0.941-0.951)*	*-*	*0.921(0.917-0.924)*	*-*	*0.875(0.870-0.881)*	*－*	*0.932(0.926-0.937)*	*-*	*0.913(0.907-0.920)*	*-*
Absence of PCT	*0.942(0.936-0.948)*	*0.003*	*0.913(0.910-0.916)*	*＜0.01*	*0.876(0.871-0.882)*	*0.14*	*0.930(0.925-0.935)*	*0.05*	*0.908(0.902-0.915)*	*＜0.01*
Absence of D-dimer	*0.944(0.938-0.950)*	*0.08*	*0.918(0.915-0.922)*	*＜0.01*	*0.874(0.868-0.879)*	*0.09*	*0.928(0.922-0.933)*	*＜0.01*	*0.912(0.905-0.918)*	*0.07*
Absence of PCT and D-dimer	*0.939(0.933-0.945)*	*＜0.01*	*0.910(0.906-0.913)*	*＜0.01*	*0.875(0.870-0.880)*	*0.96*	*0.927(0.921-0.932)*	*＜0.01*	*0.906(0.899-0.912)*	*＜0.01*

2. Questions concerning the selected features:

1-Among the 20 features, concerns may arise regarding the inclusion of "Use of diuretics," "Admission to ICU," and "Use of human albumin." These features are subject to variations between centers. However, they appear to have the most significant contribution to the prediction according to the SHAP plot. This could potentially limit the stability and generalizability of the model to other settings.

RESPONSE: We thank the Reviewer for bringing out this critical comment and completely agree with the Reviewer that medical behaviors could vary greatly across different centers, particularly within different countries. It is intriguing to note that, in the 5 enrolled hospitals from distinct provinces and geographic districts and of different levels in China, "Use of diuretics," "Admission to ICU," and "Use of human albumin" were all among the significantly contributed features to the prediction. This may potentially imply a similarity of medical behaviors, in response to critically ill conditions, across different hospitals in China. Following Reviewer's question, we conducted experiments by retraining new models without one, two or three of the three features, and tested the performances of these models. The results showed that the predictive capacity of the model for AKI decreased markedly in all centers with a statistical significance by excluding all three features in 5 Sites, with AUC decreased from (0.808-0.898) to (0.777-0.895) for predicting AKI and from (0.875-0.946) to (0.861-0.939) for predicting severe AKI, suggesting a significance of medical practice in performing AKI prediction. (Table 1-2-1, Table 1-2-2) The comment raised by this

Reviewer indicates a very important issue, i.e. features of clinical practice which are subject to variations across centers should be carefully evaluated for their consistency while developing AKI prediction models in different countries, otherwise the stability and generalizability would be potentially limited. We have included this point in the Discussion section of the revised manuscript (Page 22).

"Use of diuretics and albumin, as well as admission to ICU, were all among the significantly contributed features to the prediction, which may potentially imply a similarity of medical behaviors in response to critically ill conditions across different hospitals in China; on the other hand, features of clinical practice which are subject to variations across centers should be carefully evaluated for their consistency while developing AKI prediction models in different countries, otherwise the stability and generalizability would be potentially limited."

Table 1-2-1 The performances of the retrained AKI prediction models without the three features (use of diuretics, admission to ICU, use of human albumin) respectively or in combinations

	AUC for Refitted model									
Models	**Site 1**	***P***	**Site 2**	***P***	**Site 3**	***P***	**Site 4**	***P***	**Site 5**	***P***
Original Model with 20 features	*0.895(0.889-0.900)*	*-*	*0.863(0.860-0.866)*	*-*	*0.808(0.804-0.812)*	*-*	*0.885(0.881-0.890)*	*-*	*0.898(0.894-0.903)*	*-*
Without feature "Use of diuretics"	*0.891(0.885-0.897)*	*< 0.01*	*0.848(0.845-0.851)*	*< 0.01*	*0.793(0.789-0.798)*	*< 0.01*	*0.878(0.874-0.883)*	*< 0.01*	*0.892(0.887-0.896)*	*< 0.01*
Without feature "Admission to ICU"	*0.888(0.882-0.893)*	*< 0.01*	*0.865(0.861-0.868)*	*0.003*	*0.797(0.792-0.801)*	*< 0.01*	*0.885(0.881-0.890)*	*0.63*	*0.893(0.889-0.898)*	*< 0.01*
Without feature "Use of human albumin"	*0.893(0.888-0.899)*	*0.11*	*0.862(0.859-0.865)*	*< 0.01*	*0.806(0.801-0.810)*	*0.04*	*0.865(0.861-0.869)*	*< 0.01*	*0.900(0.896-0.905)*	*0.02*
Without features "Use of diuretics" and "Use of human albumin"	*0.886(0.879-0.891)*	*< 0.01*	*0.849(0.846-0.852)*	*< 0.01*	*0.788(0.783-0.792)*	*< 0.01*	*0.844(0.840-0.849)*	*< 0.01*	*0.893(0.888-0.897)*	*< 0.01*
Without features "Use of human albumin" and "Admission to ICU"	*0.882(0.876-0.888)*	*< 0.01*	*0.863(0.860-0.866)*	*0.10*	*0.794(0.789-0.798)*	*< 0.01*	*0.860(0.857-0.864)*	*< 0.01*	*0.894(0.889-0.898)*	*< 0.01*
Without features "Use of diuretics" and "Admission to ICU"	*0.886(0.880-0.892)*	*< 0.01*	*0.853(0.850-0.856)*	*< 0.01*	*0.783(0.779-0.787)*	*< 0.01*	*0.879(0.875-0.883)*	*< 0.01*	*0.895(0.891-0.899)*	*< 0.01*
Without features	*0.879(0*	*<*	*0.852(0.*	*<*	*0.777(0.*	*<*	*0.848(0.*	*<*	*0.895(0.*	*<*

"Use of diuretics", "Use of human albumin", and "Admission to ICU"	*.873-0.885)*	*0.01*	*849-0.855)*	*0.01*	*773-0.781)*	*0.01*	*844-0.852)*	*0.01*	*891-0.899)*	*0.01*

Table 1-2-2 The performances of the retrained severe AKI prediction model without the two features (use of diuretics, admission to ICU) respectively or in combinations

Models	**AUCs for Refitted Models**									
	Site 1	***P***	**Site 2**	***P***	**Site 3**	***P***	**Site 4**	***P***	**Site 5**	***P***
Original Model with 20 features	*0.946(0.941-0.951)*	*-*	*0.921(0.917-0.924)*	*-*	*0.875(0.870-0.881)*	*-*	*0.932(0.926-0.937)*	*-*	*0.913(0.907-0.920)*	*-*
Without feature "Use of diuretics"	*0.944(0.938-0.950)*	*0.07*	*0.895(0.891-0.898)*	*< 0.01*	*0.867(0.862-0.873)*	*< 0.01*	*0.930(0.924-0.935)*	*0.004*	*0.912(0.905-0.918)*	*0.07*
Without feature "Admission to ICU"	*0.938(0.932-0.944)*	*< 0.01*	*0.919(0.916-0.922)*	*0.005*	*0.867(0.862-0.873)*	*< 0.01*	*0.931(0.926-0.936)*	*0.77*	*0.913(0.905-0.919)*	*0.27*
Without features "Use of diuretics" and "Admission to ICU"	*0.939(0.933-0.946)*	*< 0.01*	*0.893(0.890-0.896)*	*< 0.01*	*0.861(0.855-0.867)*	*< 0.01*	*0.926(0.921-0.931)*	*< 0.01*	*0.910(0.903-0.917)*	*0.003*

2-In Figure 3, it is unclear which dataset was used to plot the figure. If the dataset was from the predictive data timeframe (before AKI occurred), the authors should explain why some values of the "Scr" and "Lower Scr" plot are already high (ranging from 150 to 250 μmol/L).

RESPONSE: We apologize for the unclarity of the Methods description. The dataset was from timepoints in the predictive data timeframe, which was before the AKI identification for patients with AKI, and before hospital discharge for patients without AKI. "Scr" refers to the latest Scr value at the prediction timepoint (before the occurrence of AKI), "Lowest Scr" refers to the lowest Scr since admission to the prediction timepoint. As some of the patients developed AKI on a basis of chronic kidney disease with different baseline GFR, the corresponding "Scr" and "Lowest Scr" values of these patients were thus of different elevated levels. We have clarified the dataset inclusion in the revised manuscript in the annotations for Figure 3 (Page59 in manuscript) and Figure S8 (Page 40 in supplementary material), as follows:

"The dataset was from timepoints in the predictive data timeframe, which was before the AKI identification for patients with AKI, and before hospital discharge for patients without AKI."

3-Still regarding Figure 3, clarification is needed for the "latest change in Scr" plot. The unit of measurement for the x-axis scale is not clear. If μmol/L was used, it raises concerns that even small changes (0-2 μmol/L in Scr) might contribute significantly to SHAP values. Alternatively, if it is in mg/dL scale, such changes should typically suffice to diagnose AKI.

RESPONSE: We are very sorry for the confusing name and the missing unit of this feature. The term "Latest change in Scr" actually refers to the latest changing rate of Scr, i.e. the latest change in Scr within certain time interval (in hours), which is calculated using the difference between the latest two Scr values divided by the time interval (in hours) between two measurements, and expressed in units of "μmol/L/h". The "Latest Scr" value was documented at any time prior to the onset of AKI (excluding the Scr value when AKI was diagnosed). Consequently, the value of "Latest changing rate of Scr" does not have a direct correlation with the diagnosis of AKI. The clinical significance of 0-2μmol/L/h change in this feature is contingent upon the duration between the two Scr measurements. For instance, a latest changing rate of 2μmol/L/h over a 6-hour period equates to a 12μmol/L change in Scr, whereas the same rate over a 12-hour period corresponds to a 24μmol/L change in Scr. As we can see from figure 3, the majority of these values falls between the range of -1 to 1, and such variations can substantially influence the SHAP values. We have re-named this feature as "Latest changing rate of Scr" and described its definition (calculation method) in TableS4 (supplementary material, Page8-9). All the plots in Figure 3 and Figure S8 have been updated and the units for each feature have been included.

4. The authors should provide a list of definitions for the features used in the final model.

RESPONSE: We thank the Reviewer for the suggestion. We have listed the definitions and the units for the features used in the final model in TableS4 (supplementary material, Page 8-9), as follows:

Table 1-2-3 Features in the prediction model of AKI:

Feature	**Definition**	**Unit**
Latest changing rate of Scr (Originally named as "Latest change in Scr")	*The parameter is calculated using the difference between the latest two Scr values divided by the time interval (hour) between two measurements. ("Latest Scr" value was documented at any time prior to the onset of AKI, excluding the Scr value when AKI was diagnosed.)*	*μmol/L/h*
Use of Diuretics	*Use of Diuretics*	*-*
Percentage of lymphocyte	*Latest percentage of lymphocyte*	*%*
Admission to ICU	*Admission to ICU*	*-*
Use of human albumin	*Use of human albumin*	*-*
Lowest Scr	*Lowest Scr value since admission*	*μmol/L*

NE percentage/albumin ratio	*The ratio of latest neutrophils percentage to the latest albumin value*	*%/(g/L)*
Scr	*Latest value of Scr*	*μmol/L*
Lowest BUN	*Lowest BUN value since admission*	*mmol/L*
Heart rate	*Latest heart rate*	*/min*
Cardiac troponin I	*Latest value of cardiac troponin I*	*ng/ml*
Highest NE percentage	*Highest percentage of neutrophils since admission*	*%*
BUN	*Latest value of BUN*	*mmol/L*
Procalcitonin	*Latest value of procalcitonin*	*ng/ml*
Baseline red blood cell count	*Admission red blood cell count*	10^{12}*/L*
Lowest D-dimer	*Lowest value of D-dimer since admission*	*mg/L*
Blood platelet count	*Latest blood platelet count*	10^{9}*/L*
Respiratory rate	*Latest respiratory rate*	*/min*
Lactic dehydrogenase	*Latest value of lactic dehydrogenase*	*IU/L*
Cardiovascular surgery	*Receive cardiovascular surgery*	*-*

Table 1-2-3 Features in the prediction model of severe AKI:

Feature	**Definition**	**Unit**
Initial changing rate of Scr (Originally named as "Initial change in Scr")	*The parameter is calculated using the difference between the latest and the initial Scr values since admission divided by the time interval (hour) between two measurements ("Latest Scr" value was documented at any time prior to the onset of AKI, excluding the Scr value when AKI was diagnosed)*	*μmol/L/h*
BUN	*Latest value of BUN*	*mmol/L*
Percentage of lymphocyte	*Latest percentage of lymphocyte*	*%*
Use of Diuretics	*Use of Diuretics*	*-*
Latest changing rate of Scr (Originally named as "Latest change in Scr")	*The parameter is calculated using the difference between the latest two Scr values divided by the time interval (hour) between two measurements.*	*μmol/L/h*
Highest D-dimer	*Highest D-dimer value since admission*	*mg/L*
Highest BUN	*Highest BUN value since admission*	*mmol/L*
Procalcitonin	*Latest value of procalcitonin*	*ng/ml*
Heart rate	*Latest heart rate*	*/min*
Lowest Scr	*Lowest Scr value since admission*	*μmol/L*
Scr	*Latest value of Scr*	*μmol/L*
Admission to ICU	*Admission to ICU*	*-*
Baseline red blood cell	*Admission red blood cell count*	10^{12}*/L*

count		
White blood cell count	*Latest white blood cell count*	*10^9/L*
High-sensitive CRP/albumin ratio	*Latest value of high-sensitive CRP divided by latest albumin value*	*(mg/L)/(g/L)*
Blood platelet count	*Latest blood platelet count*	*10^9/L*
Lactic dehydrogenase	*Latest value of lactic dehydrogenase*	*IU/L*
Direct bilirubin	*Latest value of direct bilirubin*	*μmol/L*
High-sensitive CRP	*Latest value of high-sensitive CRP*	*mg/L*
Sodium	*Latest value of sodium*	*mmol/L*

Note: The data was from timepoints in the predictive data timeframe, which was before the AKI identification for patients with AKI, and before hospital discharge for patients without AKI.

3. Other concerns

1-In the discussion section, the authors did not delve into the underlying relationship between the predictors employed in the model and AKI.

RESPONSE: We thank the Reviewer for the suggestion and have added a paragraph discussing the underlying relationship between the predictors employed in the model and AKI in the Discussion section of the revised manuscript (Page21-22), as follows:

"The features that are employed in the model include vital signs (heart rate and respiratory rate), baseline laboratory parameters (red blood cell count), trending parameters (lowest D-dimer, highest neutrophil percentage, trending features of Scr and BUN), latest laboratory parameters (such as cardiac troponin I, procalcitonin, lymphocyte percentage, lactate dehydrogenase, BUN, Scr, and C reactive protein), and medical behaviors (use of diuretics and albumin, admission to ICU, cardiovascular surgery). The combination of these features may help discriminate disease conditions that are prone to AKI. For instance, an elevated heart rate and respiratory rate, in conjunction with a high level of cardiac troponin I, may indicate the presence of hemodynamic instability. While elevated procalcitonin, white blood cell count, percentage of neutrophils, and C reactive protein could be indicative of an infectious process. It is interesting to note that baseline red blood cell count is the only baseline laboratory parameter included in the model. Both excessively low and high baseline red blood cell counts are associated with an increased risk of AKI, which may suggest a significance of anemia or volume depletion in the development of AKI. Use of diuretics and albumin, as well as admission to ICU, were all among the significantly contributed features to the prediction, which may potentially imply a similarity of medical behaviors in response to critically ill conditions across different hospitals in China; on the other hand, features of clinical practice which are subject to variations across centers should be carefully evaluated for their consistency while developing AKI prediction models in different countries, otherwise the stability and generalizability

would be potentially limited."

2-In Figure 3, the x-axis should include units of measurement. Additionally, all binary variables should be labeled as "Yes" or "No" instead of the current scale ranging from 0.0 to 1.0.

RESPONSE: We have made the editing accordingly. (Page 59 of revised manuscript and Page 40 of supplementary material)

3-There are several types of cardiac troponin tests available. The specific type of troponin was not mentioned in the figures.

RESPONSE: We are sorry for the unclarity. The type of troponin in the prediction model referred to cardiac troponin I, we have changed the name of this feature accordingly both in the text and in the figures (Page 56-61 of revised manuscript and Page 37-43 of supplementary material).

4-It is notable that almost all 95% confidence intervals (CIs) of the AUC values in the manuscript are either very narrow or similar to the mean estimate (Table 2, Table 3), which is surprising.

RESPONSE: We thank the Reviewer for pointing out this issue. We calculated the 95% CIs for AUC based on 1000 AUC values generated by the bootstrap method. Previously, we presented mean AUC and the corresponding 95% CIs using the formula, "AUC_{Mean} ($AUC_{Mean} - Z_{\alpha/2}*SE(AUC)$, $AUC_{Mean} + Z_{\alpha/2}*SE(AUC)$), $\alpha=0.05$", under the assumption that AUC values are normally distributed (Bootstrap confidence intervals: A comparative simulation study, arXiv:2404.12967 2024*). We think that's the reason for the narrow CIs. We have now selected a more robust, generalizable, and accepted method to present the 95% CIs of AUC, namely, the 2.5thand 97.5th percentile of 1000 AUCs. The method also does not need the assumption that AUC values conform to normal distribution (Gentle, James E. 2002. Elements of Computational Statistics. Springer Science + Business Media, Inc.). We have now updated all 95% CIs of the AUC values and they turned out to become wider. We have updated the results for AUC in Results section in Table 2 on Page 38, Table 3 on Page 40, Table S3 on Page 7 of supplementary material, Figure 5 on Page 64, and Results section on Page 12-13.*

We have also updated the statistical analysis method for AUC and corresponding 95% confidence intervals in Methods section on Page 8, as follows:

"The discriminative effect of the predictive models was assessed using concordance statistics, namely, the area under the receiver operating characteristic curve (AUC), with 95% confidence intervals calculated using the bootstrap method to evaluate the uncertainty of model performance, which was done by sampling the validation dataset with replacement for 1000 times.

5-It is apparent that the models and AUC estimates were obtained through multiple and iterative testings, which may increase the likelihood of false positive findings.

RESPONSE: We thank the Reviewer for raising this critical concern. Given the low incidence of AKI, continuous predictions every six hours leads to issues of multiple testings, causing problems of raised false positive rates. The problem was also observed in previous studies. (X. Song, Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction, Nat Commun 2020 Vol. 11 Issue 1 Pages 5668; K. Kim, Real-Time Clinical Decision Support Based on Recurrent Neural Networks for In-Hospital Acute Kidney Injury: External Validation and Model Interpretation, J Med Internet Res 2021 Vol. 23 Issue 4 Pages e24120) In light of this, we endeavored in this revision to reduce false positive rates by optimizing model building, rather than employing strategies such as False Discovery Rate adjustment, which could not improve model performance essentially. In this version of revised manuscript, we have tried different ways of model refitting, including data fusion, model incremental learning, increasing the number of positive samples for refitting, and difficult sample mining (extracting more relatively ambiguous events), to decrease false positive rates (improve the specificity) of the model. As the updated results showing below in Table-1-3-1, the specificity for most of the hospitals (except for site 3) has increased to over 80% at the optimal cutoffs with sensitivity over 80% for AKI and 90% for severe AKI. In site 3, where only incremental learning was feasible for model refitting due to a local specific data protection protocol that precluded data fusion, the re-refitted models that incorporated an additional small proportion of data also demonstrated a significant improvement in specificity (Table-1-3-1). We have updated the relevant data in the revised manuscript (Results section Page 12-17, Table2-5 and Figure2-5) and supplementary material (Table S5-S7, Figure S3, S6-S10). However, although some extent of improvement has been achieved, the likelihood of false positive findings in the face of the extremely imbalanced number of positive and negative samples could not be avoided and remains an inevitable deficiency. We have included this limitation in the discussion section of the revised manuscript (Page 26), as follows.

"Finally, the low incidence of AKI episodes inevitably results in extremely imbalanced number of positive and negative samples during continuous predictions, which would lead to relatively high false positive rates. The same problem has also been observed in previous researches.[18,19] Future studies are warranted to develop practical strategies for minimizing false positive alerts, which may include setting different prediction thresholds and making specific alarm rules etc., while implementing the model into the EHR system."

Table 1-3-1 Comparison of results for the updated refitted models and previous refitted models

Cohort	*Previous Refitted Model*			*Updated Refitted Model*		
	Probability Cutoff	*Sensitivity (%)*	*Specificity (%)*	*Probability Cutoff*	*Sensitivity (%)*	*Specificity (%)*
AKI						
Site 1	*0.27*	*81.5*	*90.7*	*0.53*	*81.8*	*93.9*
Site 2	*0.31*	*81.4*	*84.1*	*0.55*	*81.4*	*80.4*
Site 3	*0.26*	*81.5*	*61.2*	*0.43*	*81.5*	*61.7*
Site 4	*0.24*	*81.3*	*81.7*	*0.51*	*80.5*	*83.1*
Site 5	*0.31*	*80.6*	*79.1*	*0.57*	*80.2*	*87.4*
Site 3- Re-refitted Model						
	0.28	*80.9*	*79.5*	*0.45*	*80.8*	*82.6*
Severe AKI						
Site 1	*0.11*	*90.9*	*87.9*	*0.40*	*90.9*	*93.9*
Site 2	*0.20*	*90.2*	*84.8*	*0.47*	*90.4*	*85.0*
Site 3	*0.12*	*90.8*	*60.0*	*0.24*	*90.7*	*62.2*
Site 4	*0.09*	*91.9*	*77.4*	*0.31*	*91.6*	*82.2*
Site 5	*0.12*	*91.4*	*66.2*	*0.40*	*91.0*	*81.3*
Site 3- Re-refitted Model						
	0.12	*90.3*	*71.1*	*0.27*	*90.1*	*74.2*

Reviewer #2 (Remarks to the Author):

This is a very interesting and important paper that seeks to bring machine learning techniques to predict inpatient AKI across 5 centers in China. while there are many merits to this work, including brining modern AKI risk prediction to the developing world, there are however several concerns about the paper that need to be addressed before acceptance of the manuscript

1) Why were patients w/ AKI in the first 24 hours excluded these seem like the patients you would want to be able to identify – consider changing this – especially since many of the patients who develop Stage 1 AKI within 24 hours develop Stage 2 or 3 over the next 24-72 hours.

RESPONSE: We appreciate the Reviewer for raising this very important issue. There were 110 patients diagnosed with AKI in the first 24 hours in the derivation cohort (85 patients with AKI stage 1), and 654 patients in the five validation cohorts (536 patients with AKI stage 1). Most of these patients were diagnosed as community-acquired AKI based on the serum creatinine values prior to admission. As we need features of dynamic changes to facilitate the prediction of AKI, we did not include these patients while developing the model due to the limited laboratory parameters in these patients. We consider the Reviewer's comment of great significance and made experiments to test the performance of the model for severe AKI (AKI stage 2-3) prediction on these patients. As shown below in Table 2-1-1, the AUCs of predicting a development of severe AKI (an upgrade of AKI) within 24/48/72 hours in these patients with admission

AKI were 0.77-0.88/0.76-0.85/0.74-0.83 in the five sites using the refitted models. In another experiment, we included these patients in the dataset and retrained and revalidated the model. As shown below in Table 2-1-2, the overall AUCs for predicting severe AKI within 48 hours were quite similar to the original model. We therefore did not redo all the analysis and remained the exclusion of these patients. We have listed this point as one of the limitations in Discussion section on Page 25-26, as follows, and we hope to improve the model performances for these patients by expanding the number of similar cases in future studies.

Discussion:

"Thirdly, due to lack of dynamic laboratory parameters for patients who developed AKI within the first 24 hours, coupled with a limited sample size of such cases, these patients were excluded from model development process. A refitted model, which incorporates an increased sample size of this patient population, may enhance the predictive performance of the model for the progression to severe AKI in these patients."

Table 2-1-1 The performances of the model predicting severe AKI (AKI stage 2-3) in patients excluded due to the development of AKI within 24 hours since admission

Transported Model	*Site 1 (2018-2020)*	*Site 2*	*Site 3*	*Site 4*	*Site 5*
24h	*0.785(0.745-0.826)*	*0.772(0.747-0.797)*	*0.737(0.700-0.776)*	*0.844(0.785-0.894)*	*0.717(0.575-0.861)*
48h	*0.728(0.691-0.763)*	*0.686(0.663-0.712)*	*0.691(0.657-0.723)*	*0.777(0.732-0.825)*	*0.736(0.632-0.847)*
72h	*0.696(0.664-0.726)*	*0.662(0.641-0.686)*	*0.681(0.649-0.710)*	*0.704(0.653-0.757)*	*0.603(0.500-0.706)*
Refitted Model	*Site 1*	*Site 2*	*Site 3*	*Site 4*	*Site 5*
24h	*0.781(0.741-0.820)*	*0.838(0.817-0.856)*	*0.805(0.776-0.834)*	*0.878(0.830-0.921)*	*0.765(0.629-0.896)*
48h	*0.758(0.726-0.790)*	*0.784(0.765-0.804)*	*0.757(0.726-0.786)*	*0.848(0.806-0.888)*	*0.788(0.688-0.889)*
72h	*0.741(0.714-0.767)*	*0.760(0.743-0.778)*	*0.785(0.761-0.809)*	*0.810(0.761-0.861)*	*0.827(0.748-0.898)*

For the combination of patients from fives sites, the AUC for prediction of severe AKI within 24/48/72 hours among those who developed AKI within 24 hours was 0.779(0.762-0.796)/0.714 (0.699-0.730)/0.684(0.670-0.699) for transported models, and 0.826(0.810-0.841)/0.788(0.775-0.800)/0.775(0.763-0.787) for refitted models.

Table 2-1-2 Comparison of model performances between including and not including patients who developed AKI within 24 hours

Transported Model for Severe AKI within 48h	*Site 1*	*Site 2*	*Site 3*	*Site 4*	*Site 5*
Original results without patients diagnosed with AKI within 24 hours	*0.902(0.895-0.908)*	*0.854(0.850-0.859)*	*0.830(0.825-0.836)*	*0.888(0.883-0.894)*	*0.862(0.855-0.867)*
With patients diagnosed with AKI within 24 hours included	*0.898(0.891-0.904)*	*0.855(0.851-0.860)*	*0.827(0.822-0.833)*	*0.888(0.883-0.893)*	*0.863(0.856-0.869)*

2) Why exclude patients with a baseline less than 0.6 mg/dl – these seems like ideal patients as it likely includes older patients, patients with heart and liver disease and cancer (with profound muscle wasting) all of whom are high risk for AKI.

RESPONSE: We apologize for the unclarity. While constructing the prediction model, we excluded patients with all Scr levels lower than 0.6mg/dL from 90 days prior to admission until discharge, rather than those with baseline Scr less than 0.6mg/dL. It has been acknowledged that fluctuations of Scr under the line of 0.6mg/dL is probably of no clinical significance as the variation of food or fluid intake may simply cause the change in these special patient population. (Xiaoxi Zeng, Sushrut S. Waikar, et al. Clin J Am Soc Nephrol 2014 Vol. 9 Issue 1 Pages 12-20)

3) There is no clear explanation for how duplicate patients were handled? if the same patient had AKI during the same admission or repeat admission? How was this handled? did the baseline change?

RESPONSE: We appreciate the Reviewer for raising this very important question. If AKI was diagnosed in the same patient during a subsequent hospital admission, it was treated as an isolated AKI event. The baseline Scr was defined, as illustrated in Methods section, as the lowest Scr value during the previous 90 days.
During the same hospital admission, due to the lack of accepted criteria, it is difficult for the current program in big data to automatically distinguish a new AKI episode following the recovery of a previous one from a deterioration of the initial AKI episode. (L. S. Chawla. Nat Rev Nephrol 2017 Vol. 13 Issue 4 Pages 241-257) Therefore, same as previous studies, we only included data pertaining to the initial AKI episode during the same hospital admission for training and validation of the model, and the baseline Scr was identified as described above. We have included the above method of handling duplicate patients in Methods section of the revised manuscript (Page 6). The limitation of the model in distinguishing a recurrent AKI in the same admission have been discussed in limitations of the Discussion section (Page 25), as follows:

Methods:

"If AKI was diagnosed in the same patient during a subsequent hospital admission, it was treated as an isolated AKI event; due to the lack of accepted criteria for

distinguishing a new AKI episode from a deterioration of the initial AKI episode, we only included data pertaining to the initial AKI episode during the same hospital admission.[27]"

Discussion:

"In line with previous studies,[16-18,21] we defined AKI using the Scr-based criteria of the KDIGO guidelines. Due to the lack of accepted criteria for distinguishing a new episode of AKI following the recovery of a previous one from a deterioration of the initial AKI, we have included only data pertaining to the initial AKI episode during the same hospital admission for model training and validation. Consequently, the predictive performance of the model for recurrent AKI within the same hospitalization remains unknown.[27]"

4) We are perplexed by the use of GFR rather than SCr – as GFR is great for those in steady state and when using outpatient values but for admitted patients this seems like an issue as they are not in steady state and dividing the cohort by eGFR of 60 seems like it should be done by admission creatinine e.g 100, 200 mmol or 1mg/dl ,2 ,3)

RESPONSE: We thank the Reviewer for the valuable suggestion. In the revised manuscript, we have updated the subgroup analysis using admission creatinine (μmol/L, unit of Scr used in China) instead of GFR. The descriptions in the Methods section (Page 8) and the Results section (Page 13-14, and Table 3 on Page 40-41) have been revised accordingly, as follows:

Methods:
Subgroup analyses were performed by admission Scr, gender, location, and prior operation, with AUCs compared using the Mann-Whitney U test.

Results:

"Table 3 shows the AUCs of the refitted models for the prediction of AKI and severe AKI within 48 hours, stratified by admission Scr, gender, location, and prior operation. Both models performed better in patients with lower admission Scr (< 88.4μmol/L), except for Site 3 and Site 1."

The results for subgroup analysis by admission Scr are listed as follows:

Table 2-4 The results for subgroup analysis by admission Scr

Subgroup	AUC (95%CI) for AKI									
	Site 1 (n=17074)	P	Site 2 (n=39605)	P	Site 3 (n=21239)	P	Site 4 (n=22603)	P	Site 5 (n=13605)	P
Admission Scr, μmol/L										
≥ 88.4	0.89(0.88-	0.61	0.85(0.84-		0.83(0.82-		0.85(0.84-		0.86(0.85-	

	0.89)		*0.85)*	*< 0.01*	*0.83)*	*< 0.01*	*0.86)*	*< 0.01*	*0.87)*	*< 0.01*
< 88.4	*0.89(0.88-0.90)*		*0.86(0.86-0.87)*		*0.77(0.77-0.78)*		*0.89(0.88-0.90)*		*0.91(0.90-0.91)*	
	AUC (95%CI) for Severe AKI									
Subgroup	*Site 1 (n=17074)*	*P*	*Site 2 (n=39605)*	*P*	*Site 3 (n=21239)*	*P*	*Site 4 (n=22603)*	*P*	*Site 5 (n=13605)*	*P*
Admission Scr, µmol/L										
≥ 88.4	*0.95(0.94-0.95)*	*0.05*	*0.92(0.91-0.92)*	*0.01*	*0.90(0.90-0.91)*	*< 0.01*	*0.90(0.89-0.91)*	*< 0.01*	*0.89(0.88-0.91)*	*< 0.01*
< 88.4	*0.94(0.93-0.95)*		*0.92(0.91-0.92)*		*0.83(0.82-0.84)*		*0.94(0.94-0.95)*		*0.92(0.91-0.93)*	

5) Please provide the performance for the model to predict the receipt of RRT?

RESPONSE: We thank the Reviewer for the suggestion.
We have retrained the model to predict the receipt of RRT using variables included in the model for AKI and severe AKI, respectively, and tested the performance of the models for the prediction of RRT. As shown below in Table 2-5, the model trained using 20 variables for predicting severe AKI showed better predictive performance for RRT in the future than the model trained using 20 variables for predicting AKI, with AUC ranging from 0.77-0.97 and 0.76-0.98 across five sites, respectively. However, given the limited number of patients receiving RRT within the entire cohort, particularly in the two local hospitals, we harbor reservations regarding the models' predictive performance for RRT, despite the reasonable AUCs. Besides, the models were initially developed for the purpose of predicting AKI, and all feature selection processes were based on the predictive goal of AKI rather than RRT. With these concerns, we did not include the model for RRT in the revised manuscript. We hope to develop a more robust model specifically targeting on the prediction of RRT by increasing sample size of patients receiving RRT in future studies.

Table 2-5 Performances of the models predicting RRT in the future

AUC	*Derivation*	*Site 1*	*Site 2*	*Site 3*	*Site 4*	*Site 5*
Model trained using 20 variables for Predicting AKI						
Transported Model	*0.944(0.936-0.953)*	*0.858(0.848-0.868)*	*0.723(0.716-0.730)*	*0.712(0.703-0.721)*	*0.953(0.945-0.961)*	*0.740(0.726-0.755)*
Refitted Model	*-*	*0.901(0.892-0.910)*	*0.763(0.756-0.769)*	*0.788(0.781-0.795)*	*0.979(0.976-0.982)*	*0.753(0.736-0.771)*
Model trained using 20 variables for predicting severe AKI						
Transported Model	*0.864(0.848-0.881)*	*0.845(0.837-0.852)*	*0.787(0.782-0.792)*	*0.830(0.826-0.834)*	*0.941(0.934-0.949)*	*0.744(0.730-0.758)*
Refitted Model		*0.877(0.869-0.884)*	*0.805(0.799-0.811)*	*0.856(0.852-0.861)*	*0.969(0.965-0.973)*	*0.773(0.759-0.788)*

6) There needs to be extensive description of how "extreme" and non-physiologic values" were

defined? – provide these cutoffs and the incidence of these – as some non-physiologic values can be accurate (low MAPs, BPs and very high and low electrolytes or BUNs etc…)

RESPONSE: We thank the Reviewer for raising this very important issue. The cutoffs for extreme values of the features were set in an extended wide range. We have included the cutoffs during data preprocessing in the derivation dataset in the supplementary methods section in Item S1 (Paragraph2) on Page 1 of supplementary material, as follows:

"The cutoffs for excluding extreme values include: height 100-230cm; weight 20-250kg; heart rate 0-200/min; respiratory rate 0-150/min; systolic blood pressure 0-270mmHg, diastolic blood pressure 0-270mmHg, systolic blood pressure > diastolic blood pressure; other vital signs≥0; hemoglobin > 25g/L, serum potassium≥0 and <19, other laboratory test values≥0."

The incidence of "extreme" and "non-physiologic values" for features included in AKI and severe AKI prediction models were all very low (below 0.26%), listed as follows:

Table 2-5-1 The incidence of "extreme" and "non-physiologic values" for features included in AKI prediction model

Feature	**Extreme/non-physiologic Rate**
Latest changing rate of Scr (Originally named "Latest change in Scr")	*0.059%*
Use of Diuretics	*0*
Percentage of lymphocyte	*0.161%*
Admission to ICU	*0*
Use of human albumin	*0*
Lowest Scr	*0.059%*
NE percentage/albumin ratio	*0.175%/0.055%*
Scr	*0.059%*
Lowest BUN	*0.054%*
Heart rate	*0.005%*
Cardiac troponin I	*0.079%*
Highest NE percentage	*0.175%*
BUN	*0.054%*
Procalcitonin	*0.044%*
Baseline red blood cell count	*0.116%*
Lowest D-dimer	*0.221%*
Blood platelet count	*0.256%*
Respiratory rate	*0.017%*
Lactic dehydrogenase	*0.118%*
Cardiovascular surgery	*0*

Table 2-5-2 The incidence of "extreme" and "non-physiologic values" for features

included in severe AKI prediction model

Feature	Extreme/non-physiologic Rate
Initial changing rate of Scr (Originally named "Initial change in Scr")	*0.059%*
BUN	*0.054%*
Percentage of lymphocyte	*0.161%*
Use of Diuretics	*0*
Latest changing rate of Scr (Originally named "Latest change in Scr")	*0.059%*
Highest D-dimer	*0.221%*
Highest BUN	*0.054%*
Procalcitonin	*0.045%*
Heart rate	*0.005%*
Lowest Scr	*0.059%*
Scr	*0.059%*
Admission to ICU	*0*
Baseline red blood cell count	*0.116%*
White blood cell count	*0.036%*
High-sensitive CRP/albumin ratio	*0.034%/0.055%*
Blood platelet count	*0.256%*
Lactic dehydrogenase	*0.118%*
Direct bilirubin	*0.095%*
High-sensitive CRP	*0.034%*
Sodium	*0.056%*

7) Similarly, we did not see data round the extent of missing data in the cohorts – clearly this would vary across centers but it needs to be clearly described so we know how often the key components of the score are being tested – and then what was the rationale for imputing normal values when not tested. – if you are using tests like procalcitonin or troponin which are not commonly measured, does it not make more sense to exclude these tests or leave them as missing – rather than imputing a normal value. We agree that a procalcitonin could be an important piece of info – and will be checked with more frequency than a sodium or BUN. and this should be accounted for in the model rather than assuming that 80% of the cohort has a normal procalcitonin or troponin. At the very least the numbers need to be present

RESPONSE: We appreciate the Reviewer for raising this very important issue. We have listed the missing rate of features included in the model for AKI and severe AKI as follows. (Table 2-7-1, Table 2-7-2) As pointed out by the Reviewer, features like PCT and cTnI are not commonly included in routine investigations, yet they are frequently tested in patients with suspected infection or cardiac attack. In patients with no prescription of these features, we made imputations with normal values, assuming that these patients had been assessed by the physicians and regarded as having low risks of developing these disease conditions. Following Reviewer's concerns, we conducted

experiments to test the performance of the model assuming that the model is being implemented in a hospital or a patient population without test of the two features, and imputed using normal values (median of normal values for PCT, mean of normal values for cardiac troponin I), as listed in Table 2-7-3 and Table 2-7-4. The AUCs for AKI decreased slightly in site 1-5 with both tests inaccessible, while the AUCs for severe AKI decreased slightly in 3 sites with PCT inaccessible, but the overall predictive performances were still good. Based on these experiments, we think the model could be applied in the absence of these predictors. Besides, as PCT and cTnI are clinically relevant with AKI, including these features in the model may let doctors pay more attention to these two indicators in application. Under these considerations, we did not exclude these two parameters in the model. We have mentioned this point in the discussion section on Page 23, as follows.

"Besides, for specific tests like procalcitonin or cardiac troponin I which are usually prescribed among patients at high risk of certain conditions, imputing missing data with normal values has been proved to be an efficient way for application without significant influence on model performances. Moreover, the model could be tailored to specific patient population in different hospitals, such as adding more markers of critical illness if the hospital has a large patient population in ICU, or adding more cardiological biomarkers in the model for a cardiac based hospital. This hospital "personalized" strategy could facilitate an efficient application of the model in real clinical practice."

Table 2-7-1 Missing rate for 20 features included in the prediction model for AKI

Feature	**Missing Rate**					
	Derivation cohort	**Site 1**	**Site 2**	**Site 3**	**Site 4**	**Site 5**
Latest changing rate of Scr (Originally named "Latest change in Scr")	*0*	*0*	*0*	*0*	*0*	*0*
Use of Diuretics	*0*	*0*	*0*	*0*	*0*	*0*
Percentage of lymphocyte	*0.028*	*0.027*	*0.093*	*0.115*	*0.079*	*0.055*
Admission to ICU	*0*	*0*	*0*	*0*	*0*	*0*
Use of human albumin	*0*	*0*	*0*	*0*	*0*	*0*
Lowest Scr	*0.026*	*0.027*	*0.055*	*0.058*	*0.073*	*0.060*
NE percentage/albumin ratio	*0.048*	*0.042*	*0.106*	*0.136*	*0.106*	*0.111*
Scr	*0.026*	*0.027*	*0.055*	*0.058*	*0.073*	*0.060*
Lowest BUN	*0.026*	*0.026*	*0.058*	*0.058*	*0.083*	*0.063*
Heart rate	*0.007*	*0.003*	*0.061*	*0.004*	*0.021*	*0.149*
Cardiac troponin I	*0.319*	*0.170*	*0.984*	*0.787*	*0.999*	*0.587*
Highest NE percentage	*0.028*	*0.027*	*0.094*	*0.118*	*0.079*	*0.055*
BUN	*0.026*	*0.026*	*0.058*	*0.058*	*0.083*	*0.063*
Procalcitonin	*0.780*	*0.706*	*0.625*	*0.884*	*0.752*	*0.735*

Baseline red blood cell count	*0.027*	*0.027*	*0.094*	*0.113*	*0.079*	*0.055*
Lowest D-dimer	*0.124*	*0.102*	*0.214*	*0.141*	*0.184*	*0.241*
Blood platelet count	*0.029*	*0.030*	*0.094*	*0.113*	*0.079*	*0.055*
Respiratory rate	*0.006*	*0.002*	*0.061*	*0.101*	*0.024*	*0.072*
Lactic dehydrogenase	*0.222*	*0.132*	*0.067*	*0.216*	*0.178*	*0.375*
Cardiovascular surgery	*0*	*0*	*0*	*0*	*0*	*0*

Table 2-7-2 Missing rate for 20 features included in the prediction model for severe AKI

Feature	**Missing Rate**					
	Derivation cohort	**Site 1**	**Site 2**	**Site 3**	**Site 4**	**Site 5**
Initial changing rate of Scr (Originally named "Initial change in Scr")	*0*	*0*	*0*	*0*	*0*	*0*
BUN	*0.026*	*0.026*	*0.056*	*0.056*	*0.082*	*0.061*
Percentage of lymphocyte	*0.027*	*0.026*	*0.090*	*0.111*	*0.078*	*0.053*
Use of Diuretics	*0*	*0*	*0*	*0*	*0*	*0*
Latest changing rate of Scr (Originally named "Latest change in Scr")	*0*	*0*	*0*	*0*	*0*	*0*
Highest D-dimer	*0.121*	*0.099*	*0.209*	*0.136*	*0.182*	*0.235*
Highest BUN	*0.026*	*0.026*	*0.056*	*0.056*	*0.082*	*0.061*
Procalcitonin	*0.769*	*0.694*	*0.608*	*0.881*	*0.745*	*0.725*
Heart rate	*0.006*	*0.002*	*0.063*	*0.004*	*0.021*	*0.152*
Lowest Scr	*0.026*	*0.026*	*0.053*	*0.055*	*0.072*	*0.058*
Scr	*0.026*	*0.026*	*0.053*	*0.055*	*0.072*	*0.058*
Admission to ICU	*0*	*0*	*0*	*0*	*0*	*0*
Baseline red blood cell count	*0.027*	*0.026*	*0.090*	*0.109*	*0.078*	*0.053*
White blood cell count	*0.027*	*0.026*	*0.091*	*0.109*	*0.078*	*0.053*
High-sensitive CRP/albumin ratio	*0.321*	*0.178*	*0.283*	*0.612*	*0.346*	*0.587*
Blood platelet count	*0.028*	*0.029*	*0.091*	*0.109*	*0.078*	*0.053*
Lactic dehydrogenase	*0.217*	*0.129*	*0.064*	*0.210*	*0.177*	*0.368*
Direct bilirubin	*0.053*	*0.055*	*0.094*	*0.063*	*0.107*	*0.157*
High-sensitive CRP	*0.317*	*0.173*	*0.276*	*0.608*	*0.330*	*0.578*
Sodium	*0.026*	*0.026*	*0.057*	*0.058*	*0.070*	*0.054*

Table 2-7-3 The performances of the AKI prediction model assuming one/two of the two features (PCT, cTnI) are missing

Missing Feature	AUCs for Refitted Models									
	Site 1	*P*	Site 2	*P*	Site 3	*P*	Site 4	*P*	Site 5	*P*
No missing feature	*0.895(0.889-0.900)*	*-*	*0.863(0.860-0.866)*	*-*	*0.808(0.804-0.812)*	*－*	*0.885(0.881-0.890)*	*-*	*0.898(0.894-0.903)*	*-*
Latest procalcitonin	*0.887(0.881-0.892)*	*＜0.01*	*0.846(0.843-0.849)*	*＜0.01*	*0.807(0.803-0.811)*	*0.51*	*0.875(0.871-0.879)*	*＜0.01*	*0.890(0.886-0.895)*	*＜0.01*
Cardiac troponin I	*0.885(0.879-0.890)*	*＜0.01*	*0.850(0.847-0.853)*	*＜0.01*	*0.801(0.797-0.805)*	*＜0.01*	*0.872(0.868-0.877)*	*＜0.01*	*0.893(0.888-0.897)*	*＜0.01*
Latest procalcitonin + Cardiac troponin I	*0.875(0.869-0.881)*	*＜0.01*	*0.835(0.832-0.838)*	*＜0.01*	*0.802(0.798-0.806)*	*＜0.01*	*0.860(0.856-0.865)*	*＜0.01*	*0.883(0.879-0.888)*	*＜0.01*

Table 2-7-3 The performances of the severe AKI prediction model assuming PCT is missing

Missing Feature	AUCs for Refitted Models									
	Site 1	*P*	Site 2	*P*	Site 3	*P*	Site 4	*P*	Site 5	*P*
No missing feature	*0.946(0.941-0.951)*	*-*	*0.921(0.917-0.924)*	*-*	*0.875(0.870-0.881)*	*－*	*0.932(0.926-0.937)*	*-*	*0.913(0.907-0.920)*	*-*
Latest procalcitonin	*0.942(0.936-0.948)*	*0.003*	*0.913(0.910-0.916)*	*＜0.01*	*0.876(0.871-0.882)*	*0.14*	*0.930(0.925-0.935)*	*0.05*	*0.908(0.902-0.915)*	*＜0.01*

8) In the descriptions of the cohort 5.4% having AKI across the total cohort seems low and not in line with prior publications and the fact that 40% of the AKI was in the ICU also seems low

RESPONSE: We thank the Reviewer for raising this critical comment. The incidence of AKI ranged from 3.6% to 9.1% in our study, which is close to the reported median incidence of AKI in China (2.03%) according to a nationwide, cross-sectional survey of adult patients admitted to hospitals in academic or local hospitals from 22 provinces in China (Lancet 2015 Vol. 386 Issue 10002 Pages 1465-71), the reported AKI detection rate was 2.03% (1.99-2.08) for all hospitals, 2.2% (2.14-2.26) in academic hospitals and 1.66% (1.59-1.74) for local hospitals. The reported incidence of AKI in China was much lower than those reported in the United States and United Kingdom, but is close to AKI incidence in many Asian countries (Kidney Dis (Basel) 2016 Vol. 2 Issue 3 Pages 95-102), possibly echoing conditions associated with differences in hierarchical medical systems between different countries.

As for the percentage of 40%, we apologize for the unclarity of description. As shown in the manuscript: "The total incidence of AKI in the whole cohort was 5.39% (8,721 patients), with severe AKI accounting for 41.4% of all AKI patients." 41.4% was the percentage of severe AKI, namely the percentage of AKI stage 2-3 among all AKI patients. The percentage of AKI in ICU was not reported in this study.

9) Around line 286-293 when you discuss the performance in the subgroups were these changes in AUC significant?

RESPONSE: We thank the Reviewer for raising this question. We have added P values for the comparison of AUCs between subgroups using the Mann-Whitney U test in Table 3 on Page 40-41, and the differences in AUCs between subgroups we discussed in this paragraph were all significant. We have also updated Methods section on Page 8 and Results section (Page 13-14, and Table 3 on Page 40-41), as follows.

Methods:
"Subgroup analyses were performed by admission Scr, gender, location, and prior operation, with AUCs compared using the Mann-Whitney U test."

Results：
"Table 3 shows the AUCs of the refitted models for the prediction of AKI and severe AKI within 48 hours stratified by admission Scr, gender, location, and prior operation. Both models performed better in patients with lower admission Scr (< 88.4μmol/L), except for Site 3 and Site 1. Gender based analyses showed a slight predictive advantage in female for the prediction of AKI, and a slight predictive advantage in male for severe AKI. Both models showed better predictive discrimination in the general ward than in the ICU in almost all five cohorts, except for Site 2, where the AUCs for the surgical ward was the lowest. Subgroup analysis by prior operation showed that the models performed better in patients who did not undergo surgical operation than in those who had just undergone surgical operation in almost all cohorts."

The updated results for subgroup analysis are listed as follows:

Table 2-9 The updated results for subgroup analysis by admission Scr, gender, location, and prior operation

Subgroup	*AUC (95%CI) for AKI*									
	Site 1 (n=17074)	*P*	*Site 2 (n=39605)*	*P*	*Site 3 (n=21239)*	*P*	*Site 4 (n=22603)*	*P*	*Site 5 (n=13605)*	*P*
Admission Scr, μmol/L										
≥ 88.4	*0.89(0.88-0.89)*	*0.61*	*0.85(0.84-0.85)*	*< 0.01*	*0.83(0.82-0.83)*	*< 0.01*	*0.85(0.84-0.86)*	*< 0.01*	*0.86(0.85-0.87)*	*< 0.01*
< 88.4	*0.89(0.88-0.90)*		*0.86(0.86-0.87)*		*0.77(0.77-0.78)*		*0.89(0.88-0.90)*		*0.91(0.90-0.91)*	
Gender										
Female	*0.89(0.88-*		*0.87(0.87-*		*0.82(0.81-*		*0.88(0.88-*	*0.02*	*0.91(0.90-*	

	0.90)	< 0.01	0.88)	< 0.01	0.82)	< 0.01	0.89)		0.91)	< 0.01
Male	0.90(0.89-0.90)		0.86(0.85-0.86)		0.80(0.80-0.81)		0.89(0.88-0.89)		0.89(0.88-0.90)	
Location										
Internal Medicine Ward	0.89(0.88-0.90)	< 0.01	0.88(0.88-0.89)	< 0.01	0.82(0.82-0.83)	< 0.01	0.87(0.87-0.88)	< 0.01	0.91(0.91-0.92)	< 0.01
Surgical Ward	0.89(0.88-0.90)		0.82(0.82-0.83)		0.80(0.80-0.81)		0.87(0.86-0.88)		0.89(0.89-0.90)	
ICU	0.82(0.81-0.84)		0.86(0.85-0.86)		0.80(0.79-0.81)		0.71(0.70-0.73)		0.71(0.69-0.73)	
Prior Operation										
Prior operation	0.87(0.86-0.88)	< 0.01	0.81(0.81-0.82)	< 0.01	0.80(0.79-0.81)	< 0.01	0.83(0.80-0.85)	< 0.01	0.88(0.87-0.89)	< 0.01
No prior operation	0.90(0.89-0.91)		0.89(0.89-0.89)		0.81(0.81-0.82)		0.89(0.89-0.89)		0.90(0.90-0.91)	
Subgroup	AUC (95%CI) for Severe AKI									
	Site 1 (n=17074)	P	Site 2(n=39605)	P	Site 3 (n=21239)	P	Site 4 (n=22603)	P	Site 5 (n=13605)	P
Admission Scr, μmol/L										
≥ 88.4	0.95(0.94-0.95)	0.05	0.92(0.91-0.92)	0.01	0.90(0.90-0.91)	< 0.01	0.90(0.89-0.91)	< 0.01	0.89(0.88-0.91)	< 0.01
< 88.4	0.94(0.93-0.95)		0.92(0.91-0.92)		0.83(0.82-0.84)		0.94(0.94-0.95)		0.92(0.91-0.93)	
Gender										
Female	0.95(0.94-0.95)	0.50	0.92(0.91-0.92)	< 0.01	0.86(0.86-0.87)	< 0.01	0.91(0.90-0.92)	< 0.01	0.91(0.90-0.92)	0.06
Male	0.95(0.94-0.95)		0.92(0.92-0.93)		0.88(0.88-0.89)		0.95(0.94-0.95)		0.91(0.90-0.92)	
Location										
Internal Medicine Ward	0.96(0.95-0.96)	< 0.01	0.94(0.93-0.94)	< 0.01	0.88(0.87-0.89)	< 0.01	0.92(0.91-0.93)	< 0.01	0.93(0.93-0.94)	< 0.01
Surgical Ward	0.95(0.94-0.96)		0.89(0.88-0.89)		0.87(0.86-0.88)		0.94(0.93-0.95)		0.90(0.89-0.91)	
ICU	0.85(0.82-0.87)		0.91(0.91-0.92)		0.88(0.87-0.89)		0.78(0.76-0.80)		0.77(0.74-0.80)	
Prior Operation										
Prior operation	0.93(0.91-0.95)	< 0.01	0.89(0.88-0.89)	< 0.01	0.87(0.86-0.88)	0.1	0.91(0.88-0.93)	< 0.01	0.90(0.88-0.92)	< 0.01
No prior operation	0.95(0.94-0.95)		0.94(0.93-0.94)		0.88(0.87-0.88)		0.93(0.93-0.94)		0.92(0.91-0.92)	

10) The 5 features are intriguing – and it seems odd that ICU was so highly featured given the stress on ward-based AKI

RESPONSE: We thank the Reviewer for raising this question. We agree with the Reviewer that the total number of AKI cases in general wards was higher than that in ICUs. Nevertheless, the incidence of AKI within ICUs markedly surpasses that in general wards. As reported in a previous nationwide survey in China (PLoS One 2017 Vol. 12 Issue 2 Pages e0171202), the detection rate of AKI was the highest in ICU (22.46%) among all clinical units, followed by medical (1.96%) and surgical departments (0.96%), despite the highest absolute number of cases in medical departments. The higher rate of AKI in ICUs indicates that these patients have higher risk of developing AKI than patients in general wards. In addition, we apologize for the possible unclarity for the results in Table 1, it shows the number of AKI in each clinical unit (internal medicine ward, surgical ward, or ICU) and their respective proportion within all AKI patients, rather than the incidence of AKI in each clinical unit. The incidence of AKI in general ward is much lower than ICU. Consequently, it is reasonable that admission to ICU is one of the top features contributing to the model.

11) We find it interesting that rather than using 1 model across all sites that chose to refit the model at all sites – this is great -but make the expansion of the this model (there intent and something they discuss in the paper) problematic – as we love the idea of a base model or base components to the model (all models having the same 10 features) but then tailor the model to your patient population if you have more ICU perhaps you add in the procalcitonin or some other marker of critical illness versus a more cardiac based hospital …..

RESPONSE: We thank the Reviewer for the supportive comment and the great suggestion. We have expanded the discussion on future application of the model in the revised manuscript (Page 23) following the Reviewer's comment:

"Moreover, the model could be tailored to specific patient population in different hospitals, such as adding more markers of critical illness if the hospital has a large patient population in ICU, or adding more cardiological biomarkers in the model for a cardiac based hospital. This hospital "personalized" strategy could facilitate an efficient application of the model in real clinical practice."

12) For table 3 – please provide the performance of the score based on gender (see comment 4 above as well)

RESPONSE: We have included the performance of the models in subgroup analysis based on gender in the revised manuscript. We have also updated Methods section on Page 8 and Results section (Page 13-14, and Table 3 on Page 40-41), as follows.

Methods:
"Subgroup analyses were performed by admission Scr, gender, location, and prior operation, with AUCs compared using the Mann-Whitney U test."

Results：

"Gender based analyses showed a slight predictive advantage in female for the prediction of AKI, and a slight predictive advantage in male for severe AKI."

The results for subgroup analysis by gender are listed as follows:

Table 2-12 The results for subgroup analysis by gender

Subgroup	*AUC (95%CI) for AKI*									
	Site 1 (n=17074)	*P*	*Site 2 (n=39605)*	*P*	*Site 3 (n=21239)*	*P*	*Site 4 (n=22603)*	*P*	*Site 5 (n=13605)*	*P*
Gender										
Female	*0.89(0.88-0.90)*	*< 0.01*	*0.87(0.87-0.88)*	*< 0.01*	*0.82(0.81-0.82)*	*< 0.01*	*0.88(0.88-0.89)*	*0.02*	*0.91(0.90-0.91)*	*< 0.01*
Male	*0.90(0.89-0.90)*		*0.86(0.85-0.86)*		*0.80(0.80-0.81)*		*0.89(0.88-0.89)*		*0.89(0.88-0.90)*	
Subgroup	*AUC (95%CI) for Severe AKI*									
	Site 1 (n=17074)	*P*	*Site 2(n=39605)*	*P*	*Site 3 (n=21239)*	*P*	*Site 4 (n=22603)*	*P*	*Site 5 (n=13605)*	*P*
Gender										
Female	*0.95(0.94-0.95)*	*0.50*	*0.92(0.91-0.92)*	*< 0.01*	*0.86(0.86-0.87)*	*< 0.01*	*0.91(0.90-0.92)*	*< 0.01*	*0.91(0.90-0.92)*	*0.06*
Male	*0.95(0.94-0.95)*		*0.92(0.92-0.93)*		*0.88(0.88-0.89)*		*0.95(0.94-0.95)*		*0.91(0.90-0.92)*	

Minor issues

1) Consider replacing critical AKI with the term "severe AKI" as this is more common in the literature

RESPONSE: We have replaced "critical" AKI with "severe AKI" in the revised manuscript as suggested.

2) The abstract should make it is clear it is only SCr based AKI

RESPONSE: We have clarified accordingly in the Abstract of the revised manuscript on Page 1, as follows:

"The model containing 20 readily available variables demonstrated consistent, high levels of predictive discrimination in the validation cohort, with AUCs for serum creatinine-based AKI and severe AKI within 48 hours ranging from 0.74-0.85 and 0.83-0.90 for the transported models and from 0.81-0.90 and 0.88-0.95 for the refitted models, respectively."

3) We do not think the statement in the intro line 90 is correct – as there is actually more AKI on the wards than in the ICU – just because of the sheer volume of more ward patients than ICU patients

RESPONSE: We completely agree with the Reviewer and have made revision of the statement and deleted the incorrect part, as follows:

"As revealed by a review, AKI in developed countries is usually associated with multiple organ failure, while in developing countries, AKI is frequently caused by a single clinical condition, such as infections, nephrotoxins, or hypoperfusion."

Reviewer #2 (Remarks on code availability):

i am not a coding expert –

RESPONSE: Thanks for your detailed comments on the manuscript.

RESPONSE TO REVIEWERS

Comments of the Editor:

Thank you for submitting your manuscript "Development and Validation of a Real-time Prediction Model for Acute Kidney Injury in Hospitalized Patients" to Nature Communications. I am delighted to say that we are happy, in principle, to publish it under an open access license. First, we ask you to revise your paper to address our editorial requests (in the attached Author Checklist) and any remaining comments from reviewers (included at the end of this email, if applicable).

RESPONSE: Thanks for giving us the opportunity. We have addressed all editorial requests and queries raised by the Reviewers and updated the manuscript as suggested. A point-by point response is presented as follows.

Reviewer #1 (Remarks to the Author):

I feel very appreciated to the authors that can responding to all of my queries adequately. I do not have any more questions.

RESPONSE: Thank you for your comments.

Reviewer #3 (Remarks to the Author):

The authors have satisfactorily addressed the comments of Reviewer #2.

RESPONSE: Thank you for your comments.

1. Clarity on prediction times:

It may be obvious to the authors but it is not clear from the paper from phrases such as "development of AKI within 48 hours" as to exactly when the predictions are made, and "within 48 hours" of what - admission or prior to AKI incident. The supplementary Figure S1 helped to understand it. It is suggested to include Figure S1 in the main paper with its description, or at least give a clear explanation of when the predictions are made and for AKI to occur within what interval. From Figure S1 and its description, it is still not clear how the prediction times (i.e. yellow dots) are decided. Are the predictions made every 24 hours for AKi to occur within the next 48 hours? Or are they made every 6 hours corresponding to blue dots (if not, then why not)? How was the first prediction point "A" in the figure decided? Or is that just a sample prediction point.

RESPONSE: We apologize for the unclarity of the description on prediction times in the previous manuscript. The predictions are generated every 6 hours since admission, corresponding to the blue circles that denote each prediction point in Figure S1(Figure

S1 is now named Figure S8 after formatting modification according to editorial request), to predict the development of AKI within the following 48 hours for each prediction point on a continuous basis. Highlighted by yellow circles, three sample prediction points (prediction point A, B, and C) exemplify dynamic predictions. The prediction point "A" is just a sample prediction point. To adhere to the journal's submission requirements, which specify a maximum limit on the number of figures and tables permissible in the main manuscript, we regret that we were unable to move Figure S1 (Figure S8) to the main manuscript. Consequently, we have added more detailed description on when the predictions are made and for AKI to occur within what interval, as well as more detailed illustration for Figure S1 (Figure S8). We have updated the Methods section on Page 25 of the manuscript and the description for Figure S1 (Figure S8) on Page 30 of supplementary material, as follows:

Methods:

"As illustrated in Figure S8, predictions are generated every 6 hours since admission, corresponding to the blue circles that denote each prediction point, to predict the development of AKI within the following 48 hours for each prediction point on a continuous basis."

Illustration for Figure S8:

"Predictions are generated every 6 hours since admission, corresponding to the blue circles that denote each prediction point. Data from EHR system was updated at each prediction point and inputted into the model to predict the development of AKI within the following 48 hours on a continuous basis. With time goes on and features updated every six hours, the predictive model continuously generate new AKI predictions at regular intervals, providing ongoing prediction over the entire hospitalization. Highlighted by yellow circles, three sample prediction points (prediction point A, B, and C) exemplify dynamic predictions. At each prediction point, data from admission to the prediction point were collected, including baseline features, updated latest features, and updated feature trends. Taken the above data as input, the model will return a prediction result for AKI within the following 48 hours for each prediction point. As illustrated in the form, the results remained negative (prediction point A and B) from admission until prediction point C. The prediction models for AKI within 24 and 72 hours shared similar patterns.

2. A few minor things:

- Line 95: "patient population of derivation" is unclear, either clarify it or improve the sentence.

RESPONSE: We apologize for the unclarity of the description, we have updated the sentence on Page 3 as follows:

"As the performance of AKI prediction models depends heavily on patient characteristics of the derivation cohort, the "Kidney Disease: Improving Global Outcomes" (KDIGO) guidelines call for research on the development and validation of AKI prediction models across diverse clinical settings."

- Line 107: "different hospitals convenient and economical", please improve the sentence.

RESPONSE: We apologize for the unclarity of the description, we have updated the sentence on Page 3-4 as follows:

"This model now enables convenient and cost-effective real-time prediction of AKI among general hospitalized patients across various healthcare institutions."

- Typos: Line 101 and Line 407 "tired hospitals"

RESPONSE: Thanks for pointing out this issue and we apologize for the typos. We have updated "tired" with "tiered" on Page 3 and Page 13 in the revised manuscript as suggested.