Peer Review File

Article information: https://dx.doi.org/10.21037/tp-23-58

Reviewer A

The study "Prediction of Outpatient Waiting Time: Using Machine Learning in a Tertiary Children Hospital" aimed to identify factors responsible for outpatient waiting time using machine learning models to improve hospital management. It is a critical topic to be studied; however, the manuscript has several limitations that must be addressed.

Response to Reviewer: We sincerely thank the reviewer for your positive evaluation of our manuscript. We especially appreciate the reviewer for the following very careful and valuable comments, which help us to improve the present study. We have revised the manuscript according to the editor's and reviewer's comments. The detailed point-by-point responses are listed below. In order to help you easily find the correction places, the revised manuscript is marked using the "Track Changes" function and detailed page/line number was included in the resples to the comments.

Comment 1: The manuscript is poorly written and poorly formatted. Professional English services should be utilized to improve writing.

Reply 1: Thank you for pointing it out. We have improved the English writing as suggested. In addition, we would like to accept the official AME Editing Service and pay 15% additional charge.

Changes in the text: We have revised our manuscript according to the reviewer's and editor's comments.

Comment 2: The gap in knowledge is not very clear. For example, the authors state that prediction models are developed in relevant areas but not using machine learning. What difference does it make? If the model developed without "machine learning" works well, why do we need to develop something new?

Reply 2: Thank you very much for raising this important question and your insightful comments. In recent years, we do mean many machine learning algorithms have been used extesively in various fields. First, machine learning including deep learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. Second, machine learning algorithms such as random forest, show excellent performance on handling data that are multi-dimensional and multi-variety, and do this in dynamic or uncertain environments. Meanwhile, as machine learning algorithms gain experience, they keep improving in accuracy and efficiency. This study developed the prediction model of patient waiting time for multiple departments in a tertiary children hospital. The outpatient care situation in China is more challenging compared to western countries. Online registration, machine registration and counter manul registration coexist. Additionally, the patient flow is large while the diseases are diverse, both of which complicate the prediction of patient waiting time in children hospitals. In addition, we planned to input newly generated waiting time data into the model every day to promote continuous updates to adapt to new situations and make more accurate predictions. Compared with traditional models, machine learning methods have the advantages of being suitable for larger and more complex data, and being able to automatically and continuously update. Therefore, we used machine learning methods to develop the prediction models.

Comment 3: Gaps in the existing prediction models developed in China must be clarified.

Reply 3: Thank you very much for your kind suggestion. In China, tertiary children hospitals are often overwhelmed due to insufficient medical resources and uneven distribution, and patients have become accustomed to waiting in lines.

The waiting time of outpatient patients is linked to patients' satisfaction and affects the quality of medical services. Analyzing the determining factors that influence outpatient waiting time and proposing the effective methods for predicting patient waiting time is critical from a practical standpoint in Chinese hospitals. We found two studies of prediction in adult (11) or chronic respiritary diseases (12) in China. However, there is currently no specific research on patient waiting time prediction models for outpatients in Chinese Children's hospitals. Referring to relevant research on predicting the waiting time of emergency patients in European and American hospitals, we have attempted to use machine learning algorithms to develop the prediction models for outpatient waiting time in China.

Reference

11. Hu, Q.; Tian, F.; Jin,Z.; Lin, G.; Teng, F.; Xu, T. Developing a Warning Model of Potentially Inappropriate Medications in Older Chinese Outpatients in Tertiary Hospitals: A Machine-Learning Study. J Clin Med. 2023;12(7):2619.

12. Peng, J.; Chen, C.;Zhou, M.;Xie, X.; Zhou, Y.; Luo, C. Peak Outpatient and Emergency Department Visit Forecasting for Patients With Chronic Respiratory Diseases Using Machine Learning Methods: Retrospective Cohort Study. JMIR Med Inform. 2020 Mar 30;8(3):e13075.

Changes in the text: We have added "There are two studies of prediction in adult (11) or chronic respiratory diseases (12) in China. However, there is currently no specific research on patient waiting time prediction models for outpatients in Chinese Children's hospitals." On Page 3, Lines 89-91.

Comment 4: Information between lines 132 to 156 could be better presented in Tables.

Reply 4: Thank you very much for the constructive suggestion. We summarize these contents in Table 1, including Number of patients, Open time and Number of patients after preprocessing.

Changes in the text: We have added the Departments Classification Table as suggested onPage 4, line 126; Page 20, Table 1.

Table 1 Departments Classification Table

Categories	Outpatient departments	Number of patients	Open time	Number of patients after preprocessing
Internal Medicine	General Internal Medicine	97908	00:00-11:59	97908
Departments I	Total	97908		97908
Internal	Endocrinology	14724	07:00-16:59	14644
Medicine	Pneumology	10289		10065
Departments II	Total	25013		24709
Surgery	Orthopedics	33520	07:00-11:59	33272
Departments I	General Surgery	9460		9383
	Total	42980		42655
Surgery	Otolaryngology	18548	07:00-16:59	18497
Departments II	Cardiothoracic Surgery	10184		9751
	Total	28732		28248

Comment 5: The models should be trained separately in all different outpatient departments first. If poor results are found, then dimension-reduction techniques can be utilized.

Reply 5: Thank you very much for your reasonable suggestion. As you expected, we first tried to build the model in all different outpatient departments, but we found the results were unexpected. For example, the coefficient of determination is only 0.58 for pneumology, the coefficient of determination is only 0.51 for the cardiothoracic surgery, and it is only 0.64 for general surgery. So, we analyzed the potential reason, and we found it was due to the diverse range of outpatient departments in our pediatric hospital, some departments catered to a narrower audience of children, with patient volumes for an entire year being significantly lower than those of popular departments in just a month. This posed a challenge when training the models separately for each department, as limited data availability in certain departments can potentially impact the overall model performance. Therefore, we developed dimension-reduction techniques to classify the departments into Internal Medicine Departments I/II, and surgical I/II., etc.

The model only contains 8 independent variables, and there was no collinearity between them. So, there was no gradient explosion problem. If we obtain more

variable data in the future research. Thank you for your good suggestion.
Changes in the text: We have added "We first tried to build the model in all different outpatient departments, if poor results are found, then dimension-reduction techniques will be utilized." On page 6, Lines 172-173.

Comment 6: Why validation dataset was not used to test the models' performance?
Reply 6: Thank you very much for your constructive suggestion. The model was trained using five-fold cross-validation method (lines 178 to 179). The training set was divided into five subsets on average, and each subset took turns to do a validation set, and the other four self-subsets were used as training sets.
The external validation dataset was very important to prove the performance of the model. As you suggested, we have added the external validation set to prove the generalization of the model. The prediction performance was shown in **Table 6.**

Table 6 Model prediction performance on external validation sets

Departments categories	Number of patients	Average Waiting Time(min)	MAE (min)
Internal Medicine Departments I	13413	25.60	2.46
Internal Medicine Departments II	3717	55.60	13.08
Surgery Departments I	6284	28.14	8.29
Surgery Departments II	4245	56.67	13.18

MAE: mean absolute error of the optimal model of each category in the external validation set.

Changes in the text: We have added the Departments Classification Table 6 as suggested (See Page 8, lines 248-252; Page 24, Table 6).

Comment 7: What are the significant predictors, and how clinically these predictors are relevant?
Reply 7: Thank you very much for raising this important question. Firstly, we regret that we do not include the variable coefficients and significance tables in the main text due to space limitations. The variable coefficients and significance tables of four category models were shown in **Supplementary Table** 1-4.
(1). There were eight independent variables in internal medicine I model. The regression coefficients and p-values for each independent variable were shown in **Supplementary Table** 1. All the other variables were significant except Which missed the turn (Yes).

Supplementary Table 1 Coefficient and significance of predictors in Internal Medicine Departments I model

Features	Coefficient	t	P>|t|	[0.025	0.975]
const	-230.6402	-53.775	0	-239.047	-222.234
Registration week	27.0594	17.013	0	23.942	30.177
Registration time	61.6357	80.24	0	60.13	63.141
Registration day	-2.3413	-6.253	0	-3.075	-1.607
The number of patients in line ahead	40.7901	860.144	0	40.697	40.883
Patient gender(girl)	-113.3659	-28.922	0	-121.048	-105.683
Patient gender(boy)	-117.2743	-30.58	0	-124.791	-109.758
Type of payment(Medical insurance)	-123.6747	-31.171	0	-131.451	-115.898
Type of payment(Self pay)	-106.9655	-26.336	0	-114.926	-99.005
Way of visit(Intraday)	-25.8462	-4.83	0	-36.335	-15.358
Way of visit(Appointment)	-204.794	-35.608	0	-216.067	-193.521
Whether miss the turn(No)	-225.2856	-47.689	0	-234.545	-216.026
Whether miss the turn(Yes)	-5.3546	-1.24	0.215	-13.82	3.111

(2). There were nine independent variables in internal medicine II model. The regression coefficients and p-values for each independent variable were shown in **Supplementary Table** 2. All the other variables were significant except Registration time and Registration Dayday.

Supplementary Table 2 Coefficient and significance of predictors in Internal Medicine Departments II model

Features	Coefficient	t	P>|t|	[0.025	0.975]
const	442.2115	23.398	0	405.167	479.256
Registration week	-97.2636	-13.267	0	-111.633	-82.894
Registration time	6.0641	1.264	0.206	-3.337	15.466
Registration day	-1.7085	-1.176	0.239	-4.555	1.138
The number of patients in line ahead	243.0529	184.727	0	240.474	245.632
Patient gender(girl)	237.1917	14.442	0	205	269.383

Patient gender(boy)	205.0198	12.949	0	173.985	236.055
Type of payment(Medical insurance)	98.753	5.805	0	65.41	132.096
Type of payment(Self pay)	343.4585	19.685	0	309.26	377.657
Way of visit(Intraday)	584.8047	29.729	0	546.247	623.362
Way of visit(Appointment)	-142.5932	-8.742	0	-174.564	-110.622
Whether miss the turn(No)	160.3576	7.454	0	118.191	202.524
Whether miss the turn(Yes)	281.8539	10.441	0	228.94	334.768
Department(endocrinology)	229.2454	12.916	0	194.455	264.035
Department(pneumology)	212.9662	12.938	0	180.703	245.229

(3). There were nine independent variables in Surgery Departments I model. The regression coefficients and p-values for each independent variable were shown in **Supplementary Table** 3. All the other variables were significant.

Supplementary Table 3 Regression coefficient and significance of predictors in Surgery Departments I model

Features	**Coefficient**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	267.9023	27.828	0	249.033	286.772
Registration week	31.5194	8.332	0	24.105	38.934
Registration time	-20.7405	-9.556	0	-24.995	-16.486
Registration day	-3.8288	-4.436	0	-5.52	-2.137
The number of patients in line ahead	171.0756	273.777	0	169.851	172.3
Patient gender(girl)	117.7354	13.022	0	100.015	135.456
Patient gender(boy)	150.1669	16.968	0	132.821	167.513
Type of payment(Medical insurance)	84.295	9.179	0	66.295	102.295
Type of payment(Self pay)	183.6074	20.523	0	166.072	201.143
Way of visit(Intraday)	227.3864	21.995	0	207.123	247.65
Way of visit(Appointment)	40.5159	4.504	0	22.886	58.146
Whether miss the turn(No)	126.6662	10.647	0	103.347	149.985
Whether miss the turn(Yes)	141.2361	10.222	0	114.154	168.318
Department(general surgery)	599.0419	56.592	0	578.294	619.789
Department(orthopedics)	-331.1396	-32.578	0	-351.063	-311.217

(4). There were 9 independent variables in Surgery Departments II model. The regression coefficients and p-values for each independent variable were shown in

Supplementary Table 4. All the other variables were significant except Registration day, Way of visit (Intraday) and Which missed the turn (Yes).

Supplementary Table 4 Regression coefficient and significance of predictors in Surgery Departments II model

Features	Coefficient	t	P>|t|	[0.025	0.975]
const	-150.71	-7.725	0	-188.952	-112.468
Registration week	110.4145	14.751	0	95.743	125.086
Registration time	94.4458	16.888	0	83.484	105.407
Registration day	-1.8643	-1.199	0.23	-4.912	1.183
The number of patients in line ahead	205.5732	224.004	0	203.774	207.372
Patient gender(girl)	-96.808	-5.746	0	-129.831	-63.785
Patient gender(boy)	-53.902	-3.296	0.001	-85.952	-21.852
Type of payment(Medical insurance)	-182.7064	-9.561	0	-220.164	-145.249
Type of payment(Self pay)	31.9964	1.845	0.065	-1.995	65.988
Way of visit(Intraday)	26.6591	1.475	0.14	-8.771	62.089
Way of visit(Appointment)	-177.3692	-11.421	0	-207.808	-146.93
Whether miss the turn(No)	-122.6211	-5.28	0	-168.145	-77.097
Whether miss the turn(Yes)	-28.089	-1.455	0.146	-65.935	9.757
Department(cardiothoracic surgery)	142.5413	7.542	0	105.499	179.584
Department(otolaryngology)	-293.2513	-14.689	0	-332.381	-254.122

Changes in the text: We have added "The variable coefficients and significance tables of four category models were shown in Supplementary Table 1-4" on page 8 lines 252-253. (Supplement material, Supplement Tables 1-4).

Comment 8: How could these predictions be used to inform hospitals to improve the workflow and reduce wait time?

Reply 8: The practical application of our model in clinical settings was indeed the primary objective of our research. we plan to embed the model into a mobile instant social media app, named WeChat mini-program, allowing patients to access predicted and potential wait times by mobile phone. This enables outpatients to plan their own schedule and engage in other activities during waiting period. At the same time, hospitals can also allocate doctors according to the different waiting time of each department, improving the hospital management process. We also plan to develop a patient feedback program to collect patients' satisfaction with the prediction system, thereby improving the artificial intelligence system for predicting patient waiting time.

Changes in the text: We have added "In the future, we plan to embed the models into the mobile instant social media app, named WeChat mini-program, allowing patients to access predicted and potential wait time by mobile phone. This will enable outpatients to plan their own schedule and participate in other activities during waiting period. At the same time, hospitals can also allocate doctors according to the different waiting time of each department, improving the hospital management process. We also plan to develop a patient feedback program to collect patients' satisfaction with the prediction system, thereby improving the artificial intelligence system for predicting patient waiting time (27)" on Page 11 lines 342-349.

Comment 9: Clinical significance is poorly described in both results and discussions.

Reply 9: The utilization of artificial intelligence in predicting patient waiting time holds significant clinical implications, providing valuable insights for healthcare providers. By accurately estimating waiting time, hospitals manage staff can optimize workflow and enhance the overall patient experience.

The primary clinical significance of this predictive model lies in its ability to proactively manage resources and allocate staff effectively. By leveraging precise waiting time predictions, healthcare facilities can ensure adequate personnel and facilities are available, minimizing overcrowding and reducing delays.

Furthermore, patients benefit from the transparency and predictability offered by this model. Access to estimated waiting time through a user-friendly interface, such as a mobile application, empowers individuals to arrange their schedules accordingly. This informed decision-making not only reduces frustration and anxiety related to uncertain waiting time, but also will improve patient satisfaction.

Additionally, the integration of this model into the clinical workflow enables hospitals manage staff to optimize patient flow and streamline scheduling. Leveraging the AI-generated predictions, healthcare providers can strategically prioritize cases, allocate resources efficiently, and enhance overall operational efficiency. This, in turn, leads to reduced waiting time, increased patient throughput, and improved healthcare delivery. In summary, the clinical significance of this predictive model lies in its potential to enhance healthcare delivery by providing accurate waiting time predictions. By empowering both patients and healthcare providers, it improves resource allocation, streamlines operations, and ultimately enhances the overall quality of care.

Changes in the text: We have added in the discussion section as "The utilization of artificial intelligence in predicting patient wait times holds significant clinical implications, providing valuable insights for healthcare providers. By accurately estimating waiting durations, hospitals managing staff can optimize workflow and enhance the overall patient experience. The primary clinical significance of this predictive model lies in its ability to proactively manage resources and allocate staff effectively. By leveraging precise waiting time predictions, healthcare facilities can ensure available personnel and facilities effectively, minimizing overcrowding and reducing delays. Furthermore, patients benefit from the transparency and predictability offered by this model. Access to estimated waiting time through a user-friendly interface, such as a mobile application, empowers individuals to plan their schedules accordingly. This informed decision-making not only reduces frustration and anxiety related to uncertain waiting time but also improves patient satisfaction"on Page 12 lines 365-376.

Comment 10: Reasons for the good/moderate performance of the models need some elaboration.

Reply 10: Thank you very much for your kind suggestion. The coefficient of determinations of the optimal models for predicting the outpatient waiting time in four categories all reached above 0.82 on both the training and testing sets. The reasons for the good performance of the random forest method in Internal Medicine I, because Internal Medicine I only included one department, and there was the certain linear relationship between the independent variables and the dependent variable. So, the random forest which used bagging technique to reduce the variance was the best method. But in other three categories, there were no obvious linear relationship between the independent variables and the dependent variable. So, the Gradient Boosting Decision Tree was the best method to reduce the bias. For specific explanations, please refer to lines 315 to 321 of the manscript.

Comment 11: In discussion, it is important to compare this study's results with other tools and how it outperformed.

Reply 11: Thank you very much for the constructive suggestion. We have added the comparison between this study's results and the prediction result using the Average Method. The average method predicts the waiting time of patients by calculating the average waiting time of patients in each department in the dataset, and then multiplying it by the number of patients waiting in line ahead. The average of the absolute values of the difference between the predicted value and the true value are the MAE of the Average Method. The comparison of the optimal prediction model and the Average Method prediction capability for each category was shown in Table 7. The MAE of the optimal model in each category had an improvement of over 35% compared to the MAE of the Average Method.

Table 7 The comparison of the optimal prediction model and the Average Method for each category

Departments	MAE (Optimal Model)	MAE (Average Method)	Improvement Ratio
Internal Medicine Departments I	5.06	10.03	50%
Internal Medicine Departments II	14.15	22.58	37%
Surgery Departments I	8.76	13.48	35%
Surgery Departments II	13.62	22.09	38%

MAE: Mean Absolute Error

Improvement Ratio: (MAE of the optimal model – MAE of Average Method)/MAE of Average Method

Changes in the text: We have added as the following:

Compare with the Average Method

At present, Chinese hospitals only provide patients with the number of patients waiting in line ahead. Therefore, the most intuitive way to estimate waiting time is to multiply the average waiting time of each patient by the number of patients waiting in line ahead. The average method predicts the waiting time of patients by calculating the average waiting time of patients in each department in the dataset, and then multiplying it by the number of patients waiting in line ahead. The average of the absolute values of the difference between the predicted value and the true value are the MAE of the Average Method. The comparison of the optimal prediction model and the Average Method prediction capability for each category is shown in *Table 7*. The MAE of the optimal model in each category has an improvement of over 35% compared to the MAE of the Average Method on Page 11, lines 350-360; Page 25, Table 7.

Reviewer B

1. If available, please update your reference list by including related literatures published within a year. Some of the references are outdated.

Reply: We have updated our reference lists as we can as possible.

2. Each figure should have its own X- and Y-axis title and unit. Please revise Figure 3.

Reply: We have revised Figure 3 as required on page 19.

3. Please increase the length of your abstract.

Reply: We have revised Abstract from 186 words to 239 words.

4. Please include Statistical Analysis in "Methods" Section.

Reply: We have added the subheading "Statistical Analysis" in the "Methods" Section and revised the content on pages 6-7 lines 176-215.

5. The citation of reference 14 is missing in the article.

Reply: We have added reference 14 on line 252.

6. Please add the auxiliary lines in Figure 2 and rotate the title of the Y-axis by 180 degrees.

Reply: We have add the auxiliary lines and rotate the title of the Y-axis by 180 degrees in Figure 2.

7. The open time does not match with that in Table 1.

categories: Internal Medicine Departments I and II, Surgery Departments I and II. The general internal department was open 24 hours a day, seven days a week, including on holidays. From 7 a.m. to 4 p.m., the endocrinology, pneumology, otolaryngology, and cardiothoracic surgery departments were open, while the orthopedics and general surgery departments were open from 7 a.m. to 11p.m.

Reply: We have revised "From 7 a.m. to 5 p.m., the endocrinology, pneumology, otolaryngology, and cardiothoracic surgery departments were open, while the orthopedics and general surgery departments were open from 7 a.m. to 12 p.m" on lines 204-206 according to the Table 1. We have revised 11:59 to be 23:59 in second row of Table 1. We have reploaced "p.m." with "a.m." on line 211.

8. The numbers do not match with that in Table 3.

34,124 visits for the training set and 8,531 visits for the testing set. At last, Surgery Department II, including 19,767 visits for the training set, and 4,942 visits for the testing set. Data from May 2021 was collected as an external validation set, with

Reply: We have revised "At last, Surgery Department II, including 34,124 visits for the training set, and 8,531 visits for the testing set " on lines 304-306 according to the Table 3.

9. The numbers do not add up in Table 3.

Reply: We have added up and updated the total number as 34,124 visits for the training set, and 8,531 visits for the testing set in Table 3.

10. Please check if these should be MAE.

established using the LR, RF, GBDT, and KNN algorithms, and the model was compared and analyzed using R^2 and MSE. GBDT was found to perform the best of these AI algorithms. And Internal Medicine Department I showed the greatest R^2

Category of department	Model	Training Set		Testing Set	
		R^2 (%)	MSE (%)	R^2 (%)	MSE (%)

Reply: Thanks for your careful check, we have revised MSE to MAE in Table 5 and on line 360, and we have checked the whole manuscript.

11. Please explain what the numbers indicate.

Monday to Sunday (0~6)

Reply: The numbers 0~6 indicate Monday (0), Tuesday (1), Wednesday (2), Thursday (3), Friday (4), Saturday (5) and Sunday (6). We have delete the number in the bracket to avoid misunderstanding.