



Prediction of outpatient waiting time: using machine learning in a tertiary children's hospital

Xiaoqing Li^{1,2#}, Weiyu Liu^{3#}, Weiming Kong³, Wenqing Zhao⁴, Hansong Wang⁵, Dan Tian⁵, Jiali Jiao³, Zhangsheng Yu^{3,6,7,8}, Shijian Liu^{1,2}

¹Hainan Branch, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Sanya, China; ²School of Public Health, Shanghai Jiao Tong University, Shanghai, China; ³Center for Biomedical Data Science, Institute of Translational Medicine, Shanghai Jiao Tong University, Shanghai, China; ⁴Division of Information Department, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China; ⁵Division of Hospital Management, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China; ⁶Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China; ⁷SJTU-Yale Joint Center for Biostatistics and Data Science, Shanghai Jiao Tong University, Shanghai, China; ⁸Clinical Research Institute, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Contributions: (I) Conception and design: S Liu, Z Yu; (II) Administrative support: H Wang, D Tian; (III) Provision of study materials or patients: W Zhao; (IV) Collection and assembly of data: X Li, W Zhao; (V) Data analysis and interpretation: W Liu, W Kong, J Jiao, X Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Zhangsheng Yu, PhD. Center for Biomedical Data Science, Institute of Translational Medicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China; Department of Bioinformatics and Biostatistics, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China; SJTU-Yale Joint Center for Biostatistics and Data Science, Shanghai Jiao Tong University, Shanghai, China; Clinical Research Institute, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. Email: yuzhangsheng@sjtu.edu.cn; Shijian Liu, PhD. Hainan Branch, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, 399 Yingbin Road, Sanya 200127, China; School of Public Health, Shanghai Jiao Tong University, Shanghai, China. Email: liushijian@scmc.com.cn.

Background: Accurately predicting waiting time for patients is crucial for effective hospital management. The present study examined the prediction of outpatient waiting time in a Chinese pediatric hospital through the use of machine learning algorithms. If patients are informed about their waiting time in advance, they can make more informed decisions and better plan their visit on the day of admission.

Methods: First, a novel classification method for the outpatient clinic in the Chinese pediatric hospital was proposed, which was based on medical knowledge and statistical analysis. Subsequently, four machine learning algorithms [linear regression (LR), random forest (RF), gradient boosting decision tree (GBDT), and K-nearest neighbor (KNN)] were used to construct prediction models of the waiting time of patients in four department categories.

Results: The three machine learning algorithms outperformed LR in the four department categories. The optimal model for Internal Medicine Department I was the RF model, with a mean absolute error (MAE) of 5.03 minutes, which was 47.60% lower than that of the LR model. The optimal model for the other three categories was the GBDT model. The MAE of the GBDT model was decreased by 28.26%, 35.86%, and 33.10%, respectively compared to that of the LR model.

Conclusions: Machine learning can predict the outpatient waiting time of pediatric hospitals well and ease patient anxiety when waiting in line without medical appointments. This study offers key insights into enhancing healthcare services and reaffirms the dedication of Chinese pediatric hospitals to providing efficient and patient-centric care.

Keywords: Waiting time; artificial intelligence (AI); machine learning; pediatric

Submitted Feb 01, 2023. Accepted for publication Aug 18, 2023. Published online Nov 23, 2023.

doi: 10.21037/tp-23-58

View this article at: <https://dx.doi.org/10.21037/tp-23-58>

Introduction

The waiting time in the hospital is linked to the outpatient's satisfaction and impacts the quality of medical care provided (1). In China, many tertiary hospitals are overwhelmed, and patients have become accustomed to waiting in lines because of the deficient and uneven distribution of medical resources, chiefly in children's hospitals. Patients in European and American countries must book an appointment in advance unless they require emergency care. They must be strictly on time for their appointments, and any cancellations or changes in the schedule require advance notification. Therefore, in most European and American countries, the waiting time is usually expressed in days (2). Chinese hospitals do not require an appointment, and patients can choose to register and visit the hospital on the same day. The unpredictable nature of patient visits poses considerable challenges to the medical staff in China (3) and may be a waste of patients' valuable time. If the patient cannot momentarily stay in the waiting room because of an emergency, a turn-missing event may occur, as the waiting period is not clear, and the patient may lose their place in line. Because of this, some patients are fearful about missing their turn and do not leave the waiting area. The congested waiting area is detrimental to hospital infection prevention and management, especially during the

coronavirus disease 2019 (COVID-19) pandemic period (3).

Analyzing the determining factors that influence waiting time and proposing an effective forecast approach is critical from a practical standpoint. Various studies have been conducted to predict waiting times using a variety of machine learning prediction models in order to improve patient experience and care efficiency (4,5). In this manner, patients may be able to plan ahead and arrive at the hospital at the scheduled time, thus reducing their time in the hospital. However, several factors affect waiting time, and the time projected has generally relied on rolling average or median estimators, which may influence the accuracy (6). In other countries, due to differences in medical and health systems from those in China, waiting time is measured by the number of queuing days, which is not applicable to the situation in China (7).

The rapid development and implementation of artificial intelligence (AI) in the medical field has opened new possibilities for enhancing hospital management. Many machine learning algorithms (8), including deep learning (9) and random forest (RF) (10), have demonstrated excellent performance in time prediction. Studies have used AI to predict the onset time of illnesses and the time spent in the emergency department and operating room (11,12). However, compared with studies conducted elsewhere, those concerning the outpatient care situation in China entail greater difficulty. Online registration, machine registration, and window registration all coexist. Additionally, there is a large flow of patients and a diversity of diseases or illnesses, both of which complicate the implementation of AI technology in hospitals. Establishing an AI-based model to predict outpatient waiting time in pediatric hospitals may be novel solutions for better meeting the objectives of hospital development.

China currently has few models for predicting waiting times and even fewer models that are based on AI. Although two studies of prediction in medication for older adults (11) and chronic respiratory diseases (12) have been conducted in China, there is no specific research on patient waiting time prediction models for outpatients in Chinese children's hospitals. To better respond to hospital development and patient demands, this study developed a set of AI algorithm models capable of accurately predicting patient wait times

Highlight box

Key findings

- Artificial intelligence models were developed that were capable of predicting outpatient waiting time.

What is known and what is new?

- Unpredictable waiting times pose challenges for medical staff, wasting patients' time and potentially leading to missed appointments.
- Four machine learning algorithms were used to build prediction models, with the best-performing model identified for each category.

What is the implication, and what should change now?

- Through the use of prediction models, patients can be informed of likely wait times, allowing them to effectively arrange their schedules and make appropriate plans.

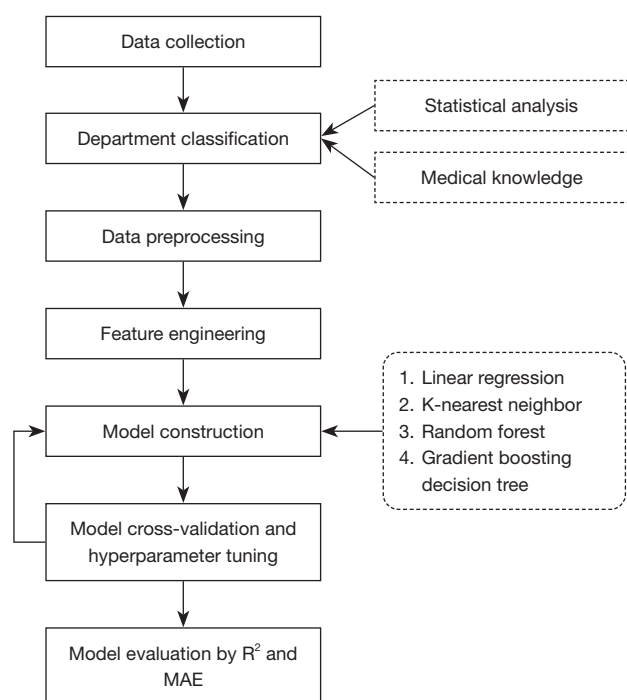


Figure 1 Flowchart of model construction and evaluation. MAE, mean absolute error.

using data from the hospital information system (HIS) of Shanghai Children's Medical Center (SCMC) and the characteristics of each department. In the future, we hope to be able to send real-time predictions to mobile devices in order to provide patients with reasonable and accurate time scheduling. This has considerable practical and social significance for improving patient satisfaction and reducing the burden on hospital management. We present this article in accordance with the TRIPOD reporting checklist (available at <https://tp.amegroups.com/article/view/10.21037/tp-23-58/rc>).

Methods

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the institute review board of SCMC (No. SCMCIRB-K2019020-2). This retrospective study did not require informed consent since it was not practicable.

Research flow

Figure 1 illustrates the modeling procedure. We planned to collect the relevant information in each department

from the HIS of SCMC over the past 5 years. Prior to modeling, relevant data were mined and examined, and several hospital departments were classified into a single category based on medical knowledge. Data within the same category were consistently modeled. To facilitate convergence of the model algorithms, preprocessing of the data was conducted to eliminate the effect of extreme values and null values. Subsequently, models for waiting time prediction were constructed using four algorithms: linear regression (LR), K-nearest neighbor (KNN), RF, and gradient boosting decision tree (GBDT). The test data set was used to evaluate the model's predictive ability, and the R^2 and mean absolute error (MAE) were used as indices for model evaluation.

Data collection and department classification

From 2015 to 2021, we gathered retrospective outpatient data at SCMC. However, the COVID-19 pandemic in China altered our hospital outpatient service, with certain emergency patients merged into the outpatient clinic. Additionally, outpatient visits had declined significantly since January 2020. Patient waiting time had also decreased significantly since the pandemic (Figure 2). As a result, this research selected data only from September 2020 to April 2021. SCMC had both specialty and general outpatient departments. Specialist departments had fewer patients than did the general departments. Furthermore, patients in specialty sections were treated separately and were not part of the general outpatient queue. The focus of this research was thus on the waiting time in the general outpatient departments.

SCMC had 24 outpatient departments; 17 of them had fewer people (<9,000) with short queues. There were seven departments with substantial patient flows and severe queue conditions whose waiting times were the focus of this study (13). These departments were internal or surgical. Internal departments included general internal medicine, endocrinology, and pneumology, while surgical departments included general surgery, orthopedics, otolaryngology, and cardiothoracic surgery, as seen in Table 1. The waiting time varied by department, as seen in Figure 2A. Meanwhile, each department's hours of operation and closure times varied. As a result, we classified these seven departments into four categories: Internal Medicine Departments I and II and Surgery Departments I and II. The general internal department was open 24 hours a day, 7 days a week, including on holidays. The endocrinology, pneumology,

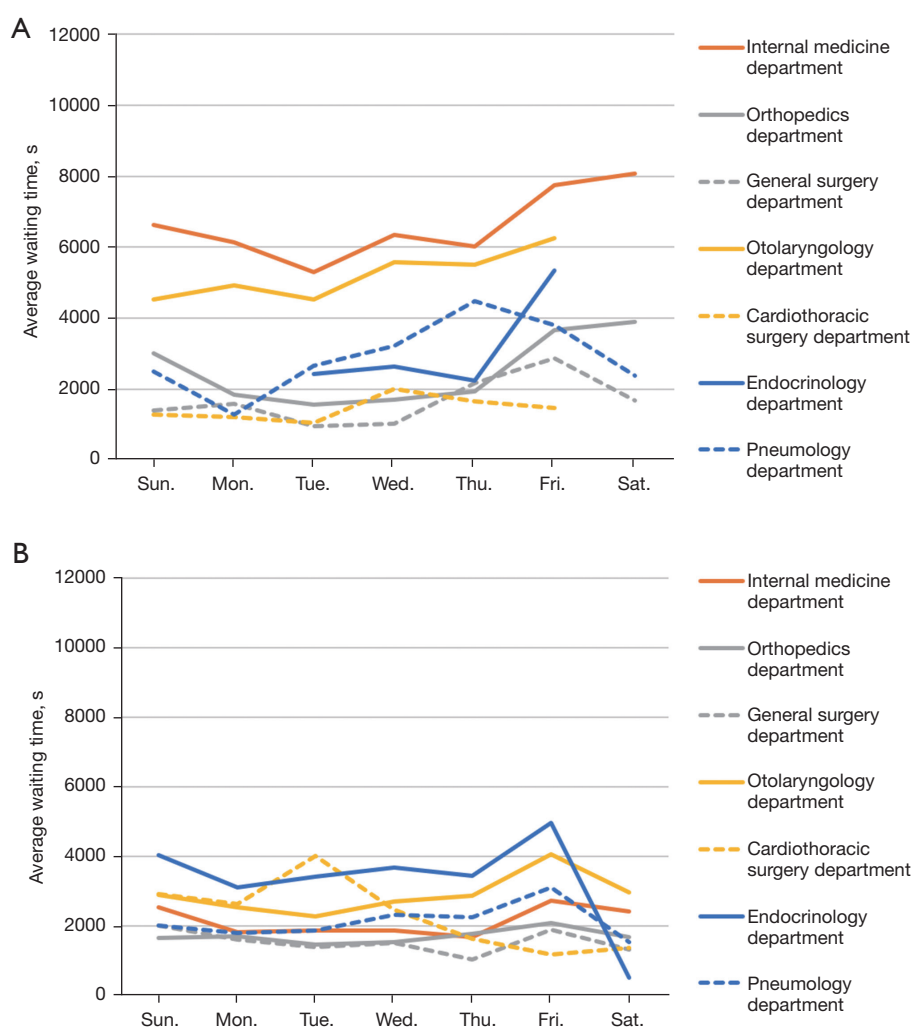


Figure 2 Average waiting time before (A) and during (B) the COVID-19 pandemic. COVID-19, coronavirus disease 2019.

otolaryngology, and cardiothoracic surgery departments were open from 7 am to 5 pm, while the orthopedics and general surgery departments were open from 7 am to 12 am. Thus, we classified general internal medicine as Internal Medicine Department I, endocrinology and pneumology as Internal Medicine Department II, orthopedics and general surgery as Surgery Department I, and otolaryngology and cardiothoracic surgery as Surgery Department II. Each category had its own model.

Data preprocessing

Due to the fact that certain patients' critical data (such as check-in time and starting time) were missing during exploratory analysis, we removed data containing null

values and erroneous data. For instance, the general surgery department was open from 7:00 am to 11:00 pm, yet a patient checked in at 6:00 am, which was not possible. After data cleaning and deletion of some outliers and missing data, 80% of the data were randomly chosen to compose the training set, while the remaining 20% were selected to compose the test set.

Feature engineering and value range

Foremost, the guardians of outpatients were required to register and wait for their turn with a doctor. As a result, we computed waiting time between the time required to register and the time beginning of the consultation (11,12). The dependent variable was the waiting time.

Table 1 Department classification

Category	Outpatient departments	Number of patients	Open time	Number of patients after preprocessing
Internal Medicine Department I	General internal medicine	97,908	00:00–23:59	97,908
	Total	97,908		97,908
Internal Medicine Department II	Endocrinology	14,724	07:00–16:59	14,644
	Pneumology	10,289		10,065
	Total	25,013		24,709
Surgery Department I	Orthopedics	33,520	07:00–23:59	33,272
	General surgery	9,460		9,383
	Total	42,980		42,655
Surgery Department II	Otolaryngology	18,548	07:00–16:59	18,497
	Cardiothoracic surgery	10,184		9,751
	Total	28,732		28,248

Table 2 Feature interpretation and value range

Feature	Value range
Registration week	Monday to Sunday
Registration day	1st to 31st
Registration time	0:00 to 23:00
The number of patients in line ahead	The number of patients who have signed in but have not yet seen the doctor when this patient registered
Patient gender	Girl/boy
Type of payment	Medical insurance/self-pay
Way of visit	Intraday/appointment
Turn missed	No/yes
Department	Internal Medicine Department I: general internal department
	Internal Medicine Department II: endocrinology department, pneumology department
	Surgery Department I: orthopedics department, general surgery department
	Surgery Department II: otolaryngology department, cardiothoracic surgery department

According to medical knowledge and the experience of physicians, independent variables were constructed. This study performed one-hot coding for categorical variables of gender, type of payment, method of registration, patient punctuality, and the type of department. We constructed the models using the features list in *Table 2* after completing a literature research and data interpretation. The day of the week on which guardians registered was the first feature considered since queueing took longer on Mondays

and the number of patients was comparatively lower on weekends. The second feature was the specific date of the registration. Different days might have had varying weather and temperature conditions, which might have impacted the patient's flow. The study included the specific time of the registration as the third major factor. The value was an integer ranging from 0 to 23. The peak time of visits in our hospital was around 8 am to 9 am and from 2 pm to 3 pm. As a result, registering during this time period would cause

longer waiting times. Another feature affecting the waiting time was the number of patients waiting ahead of a given patient at the time of registration and was the most direct influence on the time spent waiting. We also took gender, method of payment, appointment, and department into account.

Statistical analysis

Variance inflation factor and variable correlation were employed to examine the multicollinearity between variables. The variance inflation factors were all almost equal to 1, and the correlation coefficient between independent variables was about 0. Multiple correlations between the independent variables were not found. Following this, a significance test was completed for the variables in each category.

Model construction

We first attempted to establish the model in all different outpatient departments; if poor results were found, then dimension-reduction techniques would be used. LR, RF, GBDT, and KNN were used in constructing models for the four department categories to make the models more explanatory and diverse (14). Grid search was used as a parameter tuning method to list all the cases of hyperparameters in a one-by-one search; that is, to trial each possibility through a cycle among all candidate hyperparameter choices, with the parameter demonstrating the best performance being selected as the final result (15). The 5-fold cross-validation method was used to evaluate the effect of the model on the training set (16). The training set was divided into five subsets on average, with each subset in turn being used as a validation set while the other four self-subsets were used as training sets. Training and validation were repeated five times, and the result of the five average cross-validations was taken as the result of the training set. In this way, overfitting could be reduced to some extent, and effective information could be obtained as much as possible from limited data.

LR

A LR model was created as a reference for other algorithms with waiting time being used as the dependent variable.

RF

RF is a bagging algorithm that contains multiple weak

decision trees. The hyperparameters tuned by grid search included the number of subtrees as well as the maximum number of features and the minimum number of samples to split a node.

GBDT

GBDT (17,18) is a boosting algorithm that incorporates a number of weak decision trees. The learning rate, number of boosting iterations, maximum depth of each tree, maximum number of features, and minimum number of samples to split a node were tuned by grid search.

KNN

KNN (19) involves each sample being represented by its KNNs. The hyperparameters tuned were the number of neighbors and the type of weights.

Model evaluation

R^2 and the MAE were used to compare the model performances. R^2 measured the amount of variation in the dependent variable that could be explained by the independent variable. The nearer R^2 is to 1, the better the model performance. MAE is the average of the absolute values of the difference between the actual and expected waiting times for each patient. In practice, patients may encounter difficulties if the predicted waiting time is too lengthy or too short. Therefore, a lower MAE shows that the expected waiting time is closer to the actual time, which benefits patients. Additionally, predicted waiting time was compared against actual waiting time to demonstrate the disparities across models and departments. Throughout the investigation, data processing and analysis were carried out using Python (version 3.9.0, Python Software Foundation).

Results

Basic characteristics of data

Between September 1, 2020, and April 31, 2021, a total of 248,345 observations were gathered. After removal of null values and outliers, a total of 193,520 visits were recorded, comprising 97,908 visits to Internal Medicine Department I, 24,709 visits to Internal Medicine Department II, 42,655 visits to Surgery Department I, and 28,248 visits to Surgery Department II. According to *Table 3*, patients' guardians visited intraday or by appointment and were reimbursed by medical insurance or from their own pocket.

Table 3 Characteristics of the training set and testing set

Characteristics	Internal Medicine Department I		Internal Medicine Department II		Surgery Department I		Surgery Department II	
	Training set	Testing set	Training set	Testing set	Training set	Testing set	Training set	Testing set
Sample size	78,326	19,582	19,767	4,942	34,124	8,531	22,598	5,650
Gender								
Girl	35,286 (45.05)	8,753 (44.70)	10,453 (52.88)	2,643 (53.48)	13,856 (40.6049)	3,569 (41.84)	9,705 (42.9463)	2,487 (44.02)
Boy	43,040 (54.95)	10,829 (55.30)	9,314 (47.12)	2,299 (46.52)	20,268 (59.3951)	4,962 (58.16)	12,893 (57.0537)	3,163 (55.98)
Visit type								
Intraday	68,888 (87.95)	17,244 (88.06)	4,556 (23.05)	1,175 (23.78)	23,039 (67.5155)	5,729 (67.16)	11,352 (50.2345)	2,835 (50.18)
Appointment	9,438 (12.05)	2,338 (11.94)	15,211 (76.95)	3,767 (76.22)	11,085 (32.4845)	2,802 (32.84)	11,246 (49.7655)	2,815 (49.82)
Method of payment								
Medical insurance	50,755 (64.80)	12,797 (65.35)	14,586 (73.79)	3,642 (73.69)	20,943 (61.3732)	5,111 (59.91)	11,993 (53.0711)	2,943 (52.09)
Self-pay	27,571 (35.20)	6,785 (34.65)	5,181 (26.21)	1,300 (26.31)	13,181 (38.6268)	3,420 (40.09)	10,605 (46.9289)	2,707 (47.91)
Turned miss								
No	61,831 (78.94)	15,395 (78.62)	18,073 (91.43)	4,511 (91.28)	30,408 (89.1103)	7,615 (89.26)	19,297 (85.3925)	4,874 (86.27)
Yes	16,495 (21.06)	4,187 (21.38)	1,694 (8.57)	431 (8.72)	3,716 (10.8897)	916 (10.74)	3,301 (14.6075)	776 (13.73)

Data are presented as n (%). Internal Medicine Department I included the general internal department. Internal Medicine Department II included the endocrinology department and pneumology department. Surgery Department I included the orthopedics department and general surgery department. Surgery Department II included the otolaryngology department and cardiothoracic surgery department.

Gender, means of visit, method of payment, and turn missing were essentially the same across all areas. Internal Medicine Department I had 78,326 visits for the training set and 19,582 visits for the testing set, Internal Medicine Department II had 19,767 visits for the training set and 4,942 visits for testing set, Surgery Department I had 34,124 visits for the training set and 8,531 visits for the testing set, and Surgery Department II had 22,598 visits for the training set and 5,650 visits for the testing set. Data from May 2021 were collected as an external validation set, with 13,413 visits in Internal Medicine Department I, 3,717 visits in Internal Medicine Department II, 6,284 visits in Surgery Department I, and 4,245 visits in Surgery Department II.

Performance of models

Models for the departments of Internal Medicine Department I, Internal Medicine Department II, Surgery Department I, and Surgery Department II were constructed using four machine learning algorithms. *Table 4* summarizes the prediction performance of the training and testing set.

For Internal Medicine Department I, the R^2 of GBDT and RF on both the training and test set were 0.97, which was higher than the R^2 of LR (training set: $R^2=0.91$; testing set: $R^2=0.91$) and KNN (training set: $R^2=0.94$; testing set: $R^2=0.95$). The R^2 of GBDT was the largest on the test sets for Internal Medicine Department II, Surgery Department I, and Surgery Department II, with values of 0.82, 0.89, and 0.85, respectively, while the MAE values were 14.62, 8.73, and 14.11 minutes, respectively.

With the R^2 and MAE of the LR algorithm being used as a baseline, other algorithms were compared (*Table 5*). The MAE of RF on the testing set for Internal Medicine Department I was 5.03 minutes, accounting for just 13.80% of the overall average wait time. When RF was compared to LR, the R^2 increased by 6.59% and the MAE decreased by 47.60%. Accordingly, it was found that the RF model was most effective in predicting outcomes in Internal Medicine Department I. The best predictive effect was achieved with the GBDT algorithm Internal Medicine Department II, Surgery Department I, and Surgery Department II, and in comparison to the LR model, the R^2 was increased by 22.39%, 23.61%, and 19.72%, respectively, while the

Table 4 Performance evaluation of four prediction models in different departments

Department category	Model	Training set		Test set	
		R ²	MAE (min)	R ²	MAE (min)
Internal Medicine Department I	LR	0.91	9.59	0.91	9.60
	KNN	0.94	6.49	0.95	6.47
	GBDT	0.97	5.27	0.97	5.28
	RF	0.97	5.06	0.97	5.03
Internal Medicine Department II	LR	0.68	19.8	0.67	20.38
	KNN	0.74	16.39	0.74	16.74
	GBDT	0.83	14.15	0.82	14.62
	RF	0.79	15.07	0.74	15.43
Surgery Department I	LR	0.70	13.55	0.72	13.61
	KNN	0.78	10.82	0.79	10.86
	GBDT	0.88	8.76	0.89	8.73
	RF	0.87	8.95	0.86	8.85
Surgery Department II	LR	0.70	20.65	0.71	21.09
	KNN	0.77	16.67	0.77	17.12
	GBDT	0.85	13.62	0.85	14.11
	RF	0.84	13.89	0.81	14.07

Internal Medicine Department I included the general internal department. Internal Medicine Department II included the endocrinology department and pneumology department. Surgery Department I included the orthopedics department and general surgery department. Surgery Department II included the otolaryngology department and cardiothoracic surgery department. MAE, mean absolute error; LR, linear regression; KNN, K-nearest neighbor; GBDT, gradient boosting decision tree; RF, random forest.

MAE was decreased by 28.26%, 35.86%, and 33.10%, respectively.

The optimal prediction model for each category was used to predict the data from the external validation set, and the results are shown in *Table 6*. The MAE of four models was within 14 minutes, and MAE of the RF model for Internal Medicine Department I was only 2.46 minutes. The prediction results on the external validation set indicated that the optimal model for each category had good generalization performance. The variable coefficients and significance tables of the four category models are shown in *Tables S1-S4*.

Table 5 Performance evaluation of the three algorithms compared with LR

Department category	Model	Training set		Test set	
		R ² (%)	MAE (%)	R ² (%)	MAE (%)
Internal Medicine Department I	KNN	3.30	−32.33	4.40	−32.60
	GBDT	6.59	−45.05	6.59	−45.00
	RF	6.59	−47.24	6.59	−47.60
Internal Medicine Department II	KNN	8.82	−17.22	10.45	−17.86
	GBDT	22.06	−28.54	22.39	−28.26
	RF	16.18	−23.89	10.45	−24.29
Surgery Department I	KNN	11.43	−20.15	9.72	−20.21
	GBDT	25.71	−35.35	23.61	−35.86
	RF	24.29	−33.95	19.44	−34.97
Surgery Department II	KNN	10.00	−19.27	8.45	−18.82
	GBDT	21.43	−34.04	19.72	−33.10
	RF	20.00	−32.74	14.08	−33.29

As a control, the linear regression algorithm was used. Internal Medicine Department I included the general internal department. Internal Medicine Department II included the endocrinology department and pneumology department. Surgery Department I included the orthopedics department and general surgery department. Surgery Department II included the otolaryngology department and cardiothoracic surgery department. LR, linear regression; MAE, mean absolute error; KNN, K-nearest neighbor; GBDT, gradient boosting decision tree; RF, random forest.

Visualization of predicted time versus real time

Figure 3 presents the relationship between actual waiting time and predicted waiting time. In the test set of Internal Medicine Department I, the waiting time predicted with the RF algorithm and the actual waiting time were basically distributed near the axis of symmetry. Meanwhile, a similar tendency was observed for the RF algorithm in the training set and test set, which proved that there was no overfitting problem. Therefore, for Internal Medicine Department I, we chose the RF model as the final prediction model.

For Internal Medicine Department II, Surgery Department I, and Surgery Department II, the predicted and actual values of the LR method were distributed at opposite ends of the symmetry axis; meanwhile, for the GBDT algorithm, the data were more evenly spread along

Table 6 Model prediction performance on the external validation set

Department	Number of patients	Average waiting time (min)	MAE (min)
Internal Medicine Department I	13,413	25.6	2.46
Internal Medicine Department II	3,717	55.6	13.08
Surgery Department I	6,284	28.14	8.29
Surgery Department II	4,245	56.67	13.18

Internal Medicine Department I included the general internal department. Internal Medicine Department II included the endocrinology department and pneumology department. Surgery Department I included the orthopedics department and general surgery department. Surgery Department II included the otolaryngology department and cardiothoracic surgery department. MAE, mean absolute error of the optimal model of each category in the external validation set.

the symmetry axis. As a result, GBDT was chosen as the best prediction model, Internal Medicine Department II, Surgery Department I, and Surgery Department II.

Discussion

Waiting time is one of the indicators used to assess the overall quality of hospital health services. It is influenced by patients, hospitals, and society, and it has an element of unpredictability. In this work, a waiting time prediction model was established using the LR, RF, GBDT, and KNN algorithms, and the model was compared and analyzed using R^2 and MAE. GBDT was found to perform the best among these AI algorithms, and Internal Medicine Department I showed the greatest R^2 of these departments, at 0.97.

The impact of data sets on results

In this study, the outpatient and emergency flowchart of our hospital were changed before and during the COVID-19 pandemic. To differentiate fever patients from nonfever patients, the hospital included emergency nonfever patients in the normal outpatient clinic. The emergency nonfebrile patient queue sequence overlapped with the outpatient queue sequence. Following check-in, emergency patients would be positioned right in front of all out-patients, putting a strain on outpatient medical services (20). In order to account for changes in outpatient and emergency visiting practices, this study did not include outpatient data from before COVID-19, but rather only data from after September 2020 (21). At the start of the pandemic, patient flow was fluctuating but steadily stabilized until the second half of 2020 (22). Despite the abandoning of a substantial amount of early data, the situation in the second half of 2020 data was more accurate (23) and may be integrated

more easily into the hospital's outpatient system in the future, providing ease for patients and aiding in hospital decision-making. The waiting time was defined as the period from registration and to be beginning of consultation with a physician. After swiping the hospital card, the patient would be added to the queue, which eliminated the disruption caused by scheduling an appointment.

Although this study did not cover all hospital departments, it focused on departments with a greater number of people (>9,000). The findings of the exploratory analysis showed that the waiting time in departments with lower volumes (patients <9,000) was brief, and there was virtually no waiting. As a result, their data were not taken into consideration. Inadequate sample size may have an influence on the model's training effect. The general internal medicine department had a sizable outpatient population, and hence the training's sample size was the greatest. The model of Internal Medicine Department I outperformed that of other departments. As a significant component determining the waiting time, turn missing was also included as an independent variable in the model.

Not only did patients who missed their turn have to wait longer for a doctor's consultation, but they also shortened the waiting time for the patient right behind of them, who could see a doctor immediately. There were a significant number of patients who missed a in the data set, and in actuality, the scenario of patients missing turns happened often in real life. As a consequence, since the direct deletion of registered patients might impact the extrapolation of the findings, we retained these data.

The effect of the algorithm on the results

Among the models in the four categories, the optimal model for Internal Medicine Department I was the RF

model, and the optimal model for the other three categories was the GBDT model. The training algorithm for RF applied the technique of bagging in which a random sample was repeatedly selected, with replacement of the training set and the fitting of trees to these samples. Although the predictions of a single tree were highly sensitive to noise in its training set, the average of many trees was not as long as the trees were not correlated. Therefore, the bagging technique reduced the influences of outliers and noise on the model, thereby reducing the variance. Simply training many trees on a single training set would give strongly correlated trees, but bagging has a way of decorrelating the trees by showing them different training sets. Internal Medicine Department I involved only one department, and there was the certain linear relationship between the independent variables and the dependent variable (*Figure 3*). Therefore, the RF algorithm had the best performance in this category because it decreased the variance without increasing the bias.

In the other three categories, there were two small departments, and there was no obvious linear relationship between the independent variables and the dependent variable. Therefore, in this study, we needed to train the model to reduce the bias between the training value and true value. Gradient boosting combined weak “learners” into a single strong learner in an iterative fashion, with the aim of reducing the residual between the predicted value and the observed value in each iteration. As a result, the GBDT algorithm yielded a smaller bias than did the RF algorithm. The optimal models for Internal Medicine Department II, Surgery Department I, and Surgery Department II were the GBDT models.

The prediction performance for Internal Medicine Department I was the best among all kinds according to our findings after training and verification. The R^2 of these four AI methods were all more than 0.9. It was possible to deduce the causes for this discrepancy from the following observations. First, the data volume of the internal medicine clinic was comparatively broad, and then there were more training samples, resulting in high accuracy. Second, Internal Medicine Department I only included a single department, general internal medicine, while the other categories consisted of multiple departments. The heterogeneity between departments might also be one of the reasons for the low accuracy (24). In our earlier understanding, merging and modeling different departments might adequately expand the sample size and reduce the working time. However, it would have a negative impact on

the accuracy of the results (25). Therefore, each department should be represented independently in the subsequent phase. However, departments must first determine if they need AI to estimate wait times before proceeding. If the waiting time in the department is unusually short or the patient’s willingness for use AI is low, then there is no urgent need to predict the waiting time (26).

In the future, we plan to embed the models into a mobile social media app, the WeChat mini-program, allowing patients to access the predicted and potential wait time on mobile phone. This will enable outpatients to plan their own schedule and participate in other activities during the waiting period. Moreover, this would allow hospitals to allocate doctors according to the different waiting times of each department, improving the hospital management process. We also plan to develop a patient feedback program to assess patients’ satisfaction with the prediction system, thereby improving the AI system for predicting patient waiting time (27).

Comparison with the average method

At present, Chinese hospitals only provide patients with the number of patients waiting in line ahead of them. 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 ahead in line. 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 ahead in line. 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 was increased over 35% compared to the MAE of the average method.

Strengths and limitations

The advantage of this study is that departments with a large number of outpatient visits were selected based on real-world data, the possibility of using four algorithms to predict postepidemic waiting time was explored, and the performance of different models was compared (28,29). The use of AI in predicting patient wait times holds significant clinical implications, providing valuable insights

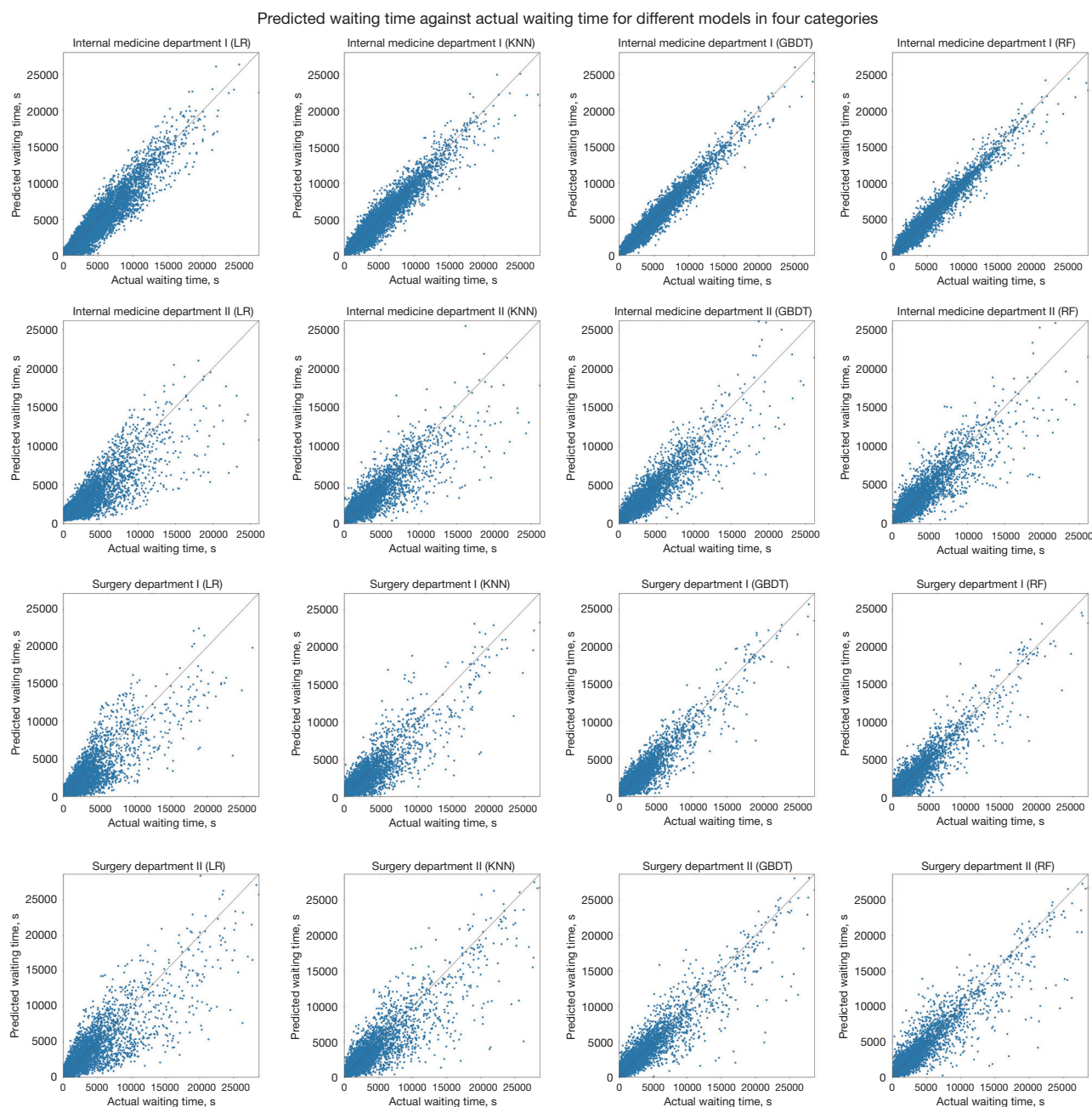


Figure 3 Actual waiting time and predicting waiting time of four algorithms for four types of departments. Internal Medicine Department I included the general internal medicine department. Internal Medicine Department II included the endocrinology and pneumology departments. Surgery Department I included the orthopedics and general surgery departments. Surgery Department II included the otolaryngology and cardiothoracic surgery departments. LR, linear regression; KNN, K-nearest neighbor; GBDT, gradient boosting decision tree; RF, random forest.

for healthcare providers. By accurately estimating waiting durations, hospitals managing staff can optimize workflows 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

Table 7 The comparison of the optimal prediction model and the average method for each category

Department	MAE (optimal model)	MAE (average method)	Improvement ratio
Internal Medicine Department I	5.06	10.03	50%
Internal Medicine Department II	14.15	22.58	37%
Surgery Department I	8.76	13.48	35%
Surgery Department II	13.62	22.09	38%

Internal Medicine Department I included the general internal department. Internal Medicine Department II included the endocrinology department and pneumology department. Surgery Department I included the orthopedics department and general surgery department. Surgery Department II included the otolaryngology department and cardiothoracic surgery department. Improvement ratio: (MAE of the optimal model – MAE of the average method)/MAE of the average method. MAE, mean absolute error.

facilities can effectively ensure available personnel and facilities, minimizing overcrowding and reducing delays. Furthermore, patients can benefit from the transparency and predictability offered by this model. Access to estimated waiting times through a user-friendly interface, such as a mobile application, empowers individuals in planning their schedules accordingly. This informed decision-making not only reduces frustration and anxiety related to uncertain wait times but also improves patient satisfaction.

The main limitation to this study was that due to the single-center design, validation using data from other external sources is needed. Second, in order to expand the sample size, some departments were combined, which might have led to insufficient accuracy of prediction. Finally, the outpatient procedure of the hospital changed before and after the epidemic, and a massive amount of pre-epidemic data were not used.

Conclusions

Machine learning can predict the outpatient waiting time of pediatric hospitals and ease patient anxiety when queuing without medical appointments.

Acknowledgments

We are grateful to all the children and their guardians who provided the data to this study.

Funding: This study was supported by the National Natural Science Foundation of China (Nos. 82173534, 81872637, 12171318); Program of the Shanghai Science and Technology Committee (No. 19441904400); Program for Artificial Intelligence Innovation and Development of the Shanghai Municipal Commission of Economy and Informatization (No. 2020-RGZN-02048); Foundation of

National Facility for Translational Medicine, Shanghai (No. TMSK-2020-124); Key Discipline Construction Project of the Three-year Action Plan of Shanghai Public Health System (No. GWV-10.1-XK07); Construction of alliance based on Artificial Intelligence for Pediatric Common Diseases (No. SHDC12020605); Shanghai Science and Technology Development Fund (No. 21ZR1436300); Project of “Unveiling the Top” for Sanya Women and Children Hospital (No. SYFY-JBGS-202201); Major Science and Technology Projects of Fujian Province (No. 2021YZ034011); and Shanghai Jiao Tong University, Star Grant (No. 20190102).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://tp.amegroups.com/article/view/10.21037/tp-23-58/rc>

Data Sharing Statement: Available at <https://tp.amegroups.com/article/view/10.21037/tp-23-58/dss>

Peer Review File: Available at <https://tp.amegroups.com/article/view/10.21037/tp-23-58/prf>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://tp.amegroups.com/article/view/10.21037/tp-23-58/coif>). S.L. reports grants paid for attending meetings. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was

conducted in accordance with the Declaration of Helsinki (as revised in 2013). This study was approved by the institute review board of Shanghai Children's Medical Center (No. SCMCIRB-K2019020-2). This retrospective study did not require informed consent since it was not practicable.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Tang L. The Chinese community patient's life satisfaction, assessment of community medical service, and trust in community health delivery system. *Health Qual Life Outcomes* 2013;11:18.
2. Chaou CH, Chen HH, Tang P, et al. Traffic Intensity of Patients and Physicians in the Emergency Department: A Queueing Approach for Physician Utilization. *J Emerg Med* 2018;55:718-25.
3. Bundgaard H, Bundgaard JS, Raaschou-Pedersen DET, et al. Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers : A Randomized Controlled Trial. *Ann Intern Med* 2021;174:335-43.
4. Cheng N, Kuo A. Using Long Short-Term Memory (LSTM) Neural Networks to Predict Emergency Department Wait Time. *Stud Health Technol Inform* 2020;272:199-202.
5. Jancauskas V, Piontek T, Kopta P, et al. Predicting queue wait time probabilities for multi-scale computing. *Philos Trans A Math Phys Eng Sci* 2019;377:20180151.
6. Pak A, Gannon B, Staib A. Predicting waiting time to treatment for emergency department patients. *Int J Med Inform* 2021;145:104303.
7. Kuo YH, Chan NB, Leung JMY, et al. An Integrated Approach of Machine Learning and Systems Thinking for Waiting Time Prediction in an Emergency Department. *Int J Med Inform* 2020;139:104143.
8. Sapiertein Silva JF, Ferreira GF, Perosa M, et al. A machine learning prediction model for waiting time to kidney transplant. *PLoS One* 2021;16:e0252069.
9. Ayoobi N, Sharifrazi D, Alizadehsani R, et al. Time series forecasting of new cases and new deaths rate for COVID-19 using deep learning methods. *Results Phys* 2021;27:104495.
10. Cha GW, Moon HJ, Kim YM, et al. Development of a Prediction Model for Demolition Waste Generation Using a Random Forest Algorithm Based on Small DataSets. *Int J Environ Res Public Health* 2020;17:6997.
11. Hu Q, Tian F, Jin Z, et al. Developing a Warning Model of Potentially Inappropriate Medications in Older Chinese Outpatients in Tertiary Hospitals: A Machine-Learning Study. *J Clin Med* 2023;12:2619.
12. Peng J, Chen C, Zhou M, et al. Peak Outpatient and Emergency Department Visit Forecasting for Patients With Chronic Respiratory Diseases Using Machine Learning Methods: Retrospective Cohort Study. *JMIR Med Inform* 2020;8:e13075.
13. Raita Y, Goto T, Faridi MK, et al. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019;23:64.
14. Joseph JW. Queueing Theory and Modeling Emergency Department Resource Utilization. *Emerg Med Clin North Am* 2020;38:563-72.
15. Sukhpal K, Himanshu A, Rani R. Hyper-parameter optimization of deep learning model for prediction of Parkinson's disease. *Mach Vis Appl* 2020;31:32.
16. Wong TT, Yeh PY. Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Trans Knowl Data Eng* 2020;32:1586-94.
17. Liang W, Luo S, Zhao G, et al. Predicting Hard Rock Pillar Stability Using GBDT, XGBoost, and LightGBM Algorithms. *Mathematics* 2020;8:765.
18. Zhu Z, Li J, Huang J, et al. An intelligent prediagnosis system for disease prediction and examination recommendation based on electronic medical record and a medical-semantic-aware convolution neural network (MSCNN) for pediatric chronic cough. *Transl Pediatr* 2022;11:1216-33.
19. Taunk K, De S, Verma S, et al. A Brief Review of Nearest Neighbor Algorithm for Learning and Classification. 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India. 2019. doi: 10.1109/ICCS45141.2019.9065747.
20. Macchi ZA, Ayele R, Dini M, et al. Lessons from the COVID-19 pandemic for improving outpatient neuropsychiatric care: A qualitative study of patient and caregiver perspectives. *Palliat Med* 2021;35:1258-66.
21. Ayele R, Macchi ZA, Dini M, et al. Experience of

- Community Neurologists Providing Care for Patients With Neurodegenerative Illness During the COVID-19 Pandemic. *Neurology* 2021;97:e988-95.
22. McCarthy ML, Ding R, Pines JM, et al. Comparison of methods for measuring crowding and its effects on length of stay in the emergency department. *Acad Emerg Med* 2011;18:1269-77.
 23. Begaz T, Elashoff D, Grogan TR, et al. Initiating Diagnostic Studies on Patients With Abdominal Pain in the Waiting Room Decreases Time Spent in an Emergency Department Bed: A Randomized Controlled Trial. *Ann Emerg Med* 2017;69:298-307.
 24. Grant RW, Lyles C, Uratsu CS, et al. Visit Planning Using a Waiting Room Health IT Tool: The Aligning Patients and Providers Randomized Controlled Trial. *Ann Fam Med* 2019;17:141-9.
 25. Harding KE, Snowdon DA, Lewis AK, et al. Staff perspectives of a model of access and triage for reducing waiting time in ambulatory services: a qualitative study. *BMC Health Serv Res* 2019;19:283.
 26. Ebert JF, Huibers L, Christensen B, et al. Does an emergency access button increase the patients' satisfaction and feeling of safety with the out-of-hours health services? A randomised controlled trial in Denmark. *BMJ Open* 2020;10:e030267.
 27. Mackert M, Mandell D, Donovan E, et al. Mobile Apps as Audience-Centered Health Communication Platforms. *JMIR Mhealth Uhealth* 2021;9:e25425.
 28. Almalki M, Giannicchi A. Health Apps for Combating COVID-19: Descriptive Review and Taxonomy. *JMIR Mhealth Uhealth* 2021;9:e24322.
 29. Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, et al. A Mobile Phone-Based Intervention to Reduce Mental Health Problems in Health Care Workers During the COVID-19 Pandemic (PsyCovidApp): Randomized Controlled Trial. *JMIR Mhealth Uhealth* 2021;9:e27039.

Cite this article as: Li X, Liu W, Kong W, Zhao W, Wang H, Tian D, Jiao J, Yu Z, Liu S. Prediction of outpatient waiting time: using machine learning in a tertiary children's hospital. *Transl Pediatr* 2023;12(11):2030-2043. doi: 10.21037/tp-23-58

Table S1 Coefficient and significance of predictors in the model for Internal Medicine Department I

Feature	Coefficient	<i>t</i>	P> <i>t</i>	Lower limit	Upper limit
Constant	-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, male	-113.3659	-28.922	0	-121.048	-105.683
Patient gender, female	-117.2743	-30.58	0	-124.791	-109.758
Payment method, medical insurance	-123.6747	-31.171	0	-131.451	-115.898
Payment method, self-pay	-106.9655	-26.336	0	-114.926	-99.005
Visit type, intraday	-25.8462	-4.83	0	-36.335	-15.358
Visit type, appointment	-204.794	-35.608	0	-216.067	-193.521
Turned miss, no	-225.2856	-47.689	0	-234.545	-216.026
Turned miss, yes	-5.3546	-1.24	0.215	-13.82	3.111

Table S2 Coefficient and significance of predictors in the model for Internal Medicine Department II

Feature	Coefficient	<i>t</i>	P> <i>t</i>	Lower limit	Upper limit
Constant	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, male	237.1917	14.442	0	205	269.383
Patient gender, female	205.0198	12.949	0	173.985	236.055
Payment method, medical insurance	98.753	5.805	0	65.41	132.096
Payment method, self-pay	343.4585	19.685	0	309.26	377.657
Visit type, intraday	584.8047	29.729	0	546.247	623.362
Visit type, appointment	-142.5932	-8.742	0	-174.564	-110.622
Turned miss, no	160.3576	7.454	0	118.191	202.524
Turned miss, 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

Table S3 Regression coefficient and significance of predictors in the model for Surgery Department I

Feature	Coefficient	<i>t</i>	P> <i>t</i>	Lower limit	Upper limit
Constant	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, male	117.7354	13.022	0	100.015	135.456
Patient gender, female	150.1669	16.968	0	132.821	167.513
Payment method, medical insurance	84.295	9.179	0	66.295	102.295
Payment method, self-pay	183.6074	20.523	0	166.072	201.143
Visit type, intraday	227.3864	21.995	0	207.123	247.65
Visit type, appointment	40.5159	4.504	0	22.886	58.146
Turned miss, no	126.6662	10.647	0	103.347	149.985
Turned miss, 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

Table S4 Regression coefficient and significance of predictors in the model for Surgery Department II

Feature	Coefficient	<i>t</i>	P> <i>t</i>	Lower limit	Upper limit
Constant	-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, male	-96.808	-5.746	0	-129.831	-63.785
Patient gender, female	-53.902	-3.296	0.001	-85.952	-21.852
Payment method, medical insurance	-182.7064	-9.561	0	-220.164	-145.249
Payment method, self-pay	31.9964	1.845	0.065	-1.995	65.988
Visit type, intraday	26.6591	1.475	0.14	-8.771	62.089
Visit type, appointment	-177.3692	-11.421	0	-207.808	-146.93
Turned miss, no	-122.6211	-5.28	0	-168.145	-77.097
Turned miss, 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