Indian J Hematol Blood Transfus
https://doi.org/10.1007/s12288-020-01333-5

ORIGINAL ARTICLE

RBC Inventory-Management System Based on XGBoost Model

Xiaolin Sun[1] · Zhenhua Xu[2] · Yannan Feng[1] · Qingqing Yang[2] · Yan Xie[2] ·
Deqing Wang[1] · Yang Yu[1]

Received: 12 January 2020 / Accepted: 6 August 2020

Abstract It is difficult to predict RBC consumption accurately. This paper aims to use big data to establish a XGBoost Model to understand the trend of RBC accurately, and forecast the demand in time. XGBoost, which implements machine learning algorithms under the Gradient Boosting framework can provide a parallel tree boosting. The daily RBC usage and inventory (May 2014–September 2017) were investigated, and rules for RBC usage were analysed. All data were divided into training sets and testing sets. A XGBoost Model was established to predict the future RBC demand for durations ranging from a day to a week. In addition, the alert range was added to the predicted value to ensure RBC demand of emergency patients and surgical accidents. The gap between RBC usage and inventory was fluctuant, and had no obvious rule. The maximum residual inventory of a certain blood group was up to 700 units one day, while the minimum was nearly 0 units. Upon comparing MAE (mean absolute error):A:10.69, B:11.19, O:10.93, and AB:5.91, respectively, the XGBoost Model was found to have a predictive advantage over other state-of-the-art approaches. It showed the model could fit the trend of daily RBC usage. An alert range could manage the demand of emergency patients or surgical accidents. The model had been built to predict RBC demand, and the alert range of RBC inventory is designed to increase the safety of inventory management.

Keywords XGBoost model · RBC inventory · Big data · Transfusion prediction

Xiaolin Sun and Zhenhua Xu have contributed equally to this article.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12288-020-01333-5) contains supplementary material, which is available to authorized users.

✉ Deqing Wang
deqingw@vip.sina.com

✉ Yang Yu
yuyangpla301@163.com

Xiaolin Sun
jerrywhtcsun@163.com

Zhenhua Xu
xuzhenhua1983@126.com

Qingqing Yang
yuyangpla301@163.com

Yan Xie
yuyangpla301@163.com

1 Department of Blood Transfusion, Chinese PLA General Hospital, No. 28, Fuxing Rd, Beijing 100853, China

2 HealSci Technology Co., Ltd, 1606, Tower5, 2 Rong Hua South Road, BDA, Beijing 100176, China

Introduction

Blood transfusion has become a common clinical strategy. Although surgical techniques and clinical treatment methods are being constantly updated, they are still impossible to avoid massive blood loss during the perioperative period or blood transfusion support during the treatment of internal medical diseases. RBC, which is the most important blood component, can be targeted for treating anaemia. However, blood resources may face the shortage, or imbalance between blood collection and usage.

In recent years, with the rapid development of the medical industry in China, the imbalance between RBC supply and demand is becoming increasingly prominent. The challenge in RBC inventory management is to find a balance point between the two. Thus far, hospitals in many

Published online: 02 November 2020

regions of our country have experienced seasonal blood shortages, and more than 50 large and medium-sized cities in the country have suffered from insufficient blood supply [1]. Clinicians have to reduce blood consumption or encourage direct donation from families to solve this problem [2]. In addition, the time, amount and population of voluntary blood donors are irregular and unpredictable. Consequently, random replenishment of blood inventory lacking a schedule can easily lead to asymmetries in blood supply and demand. Foreign scholars have also explored this trend of population donation [3].

On the one hand, the amount of blood collection is still the main factor impacting on RBC inventory [4]; on the other hand, RBC demand is also considered a factor to make inventory management more reasonable. Some scholars studied the prediction of blood demand. Mahmood et al. established multivariate risk scores to predict whether patients need intraoperative blood transfusion [5]. Erickson et al. suggested that frozen blood components should be used when there was seasonal shortage of blood stock [6]. However, Cywinski et al. pointed out that regression analysis, logical analysis, classification and regression tree analysis, and other statistical techniques cannot reliably predict whether patients need a large amount of blood during surgery [7].

The application of 'big data' and 'artificial intelligence' technology in the field of blood transfusion provides us a way to solve the problem [8, 9]. According to a Stanford University study, using big data to establish statistical models for quantitatively predicting platelet transfusions can significantly reduce platelet waste [10]. Pendry K proposed that big data is often used to detect transfusion-related complications in the field of blood transfusion, and determine blood schedule for surgery. The information obtained can be further used to monitor patients' blood management and inventory management [11].

XGBoost has also been carried out on the auxiliary diagnosis in the field of intelligent medicine, such as orthopedic auxiliary diagnosis model, accurate prediction model of type 2 diabetes, and surgical prognosis quality score. XGBoost is an integrated learning algorithm. It is an additive model, and the basic model generally chooses the tree model or other models, such as logical regression and so on. XGBoost results in a significant improvement in effect and performance. For samples with missing feature values, XGBoost can automatically learn its splitting direction.

In this paper, we will establish a prospective blood demand-prediction model by sharing clinical data, aiming to predict the patients' demand for RBC accurately, and establish an inventory management system.

Materials and Methods

Data Acquisition

The application of the electronic medical record system can facilitate the tracking of links located in the hospital. The Intelligent Management and Evaluation system of Clinical Blood Transfusion can extract relevant information of the patients by linking with the Hospital Information System (HIS) and Laboratory Information System (LIS) of the PLA general hospital. The data on total RBC transfusion and the RBC inventory were extracted from 1st May, 2014 to 25th September, 2017. The total number was 1243 days. The 'RBC inventory information' was updated at 00:00 am every day, and the total 'RBC transfusion' was observed from intraday 00:00 am to 24:00 pm. The RBC was measured in 'units', which was approximately 150 ml—extracted from 200 ml whole blood product.

The project has been approved by the Medical Ethics Committee of the General Hospital of the Chinese people's Liberation Army and carried out its work in accordance with the relevant requirements of medical ethics.

Model Design Logic

All blood transfusion applications were divided into therapeutic and surgical transfusion purposes. It could be found that blood transfusion for surgery and treatment showed different rules. The transfusion for therapeutic purpose were basically stable at around 100 units/day, even on weekends; while ones for operations fluctuated greatly. About 90% of the applications for operation would be submitted in advance, compared with only 50% for treatment. Therefore, the models for different purposes would be built separately.

The Factors Incorporated by the Algorithm

The algorithm which was based on historical information included the different factors, such as time (including days of the week, month, year), holidays (including number of holidays, previous 1–3 days of holidays, next 1–3 days of holidays), mean transfusion volume of various departments, transfusion purposes (including surgery or therapy), application date, and quantity requested. The factors affecting transfusion were substituted into the model to predict the daily RBC consumption in advance.

1. RBC transfusion statistics: RBC transfusion volume in each department could be regarded as an internal factor. By counting the historical RBC transfusion volume of each department, the trend of RBC usage in various departments could be obtained.

2. Time-factor: considered historical RBC transfusion volume weekly, monthly. yearly by Prophet Model, to manage the trend of transfusion.
3. Holiday-factor: such as approaching holidays, the operation scheme would be arranged in advance or delayed, which would affect the amount of RBC usage.
4. Mean transfusion of departments: The Mean Value and Standard Deviation of RBC transfusion from Monday to Sunday were calculated, because the regularity of RBC consumption per week was similar, and the usages on weekdays and weekends were relatively different.
5. The purpose of applying transfusion, due to the different rules of surgical or therapeutic transfusion, which were involved in the model.
6. Application date, and quantity requested.

Establishment of XGBoost Model

In clinical practice, some blood application data can be obtained in advance, which was different from actual blood consumption. For example, elective surgery would submit the application 1–3 days in advance, and the application volume is often higher than the actual consumption; while for medical patients' transfusion programs, the amount of application and usage were almost consistent.

Results

Rule of RBC Consumption

We obtained data of the RBC usage and inventory each day simultaneously. As shown in Fig. 1, the RBC residual inventory meant the daily RBC inventory minus daily RBC usage, which was relatively different every day. The RBC residual inventory could reach 700 units on a certain day, and it might decrease to 10 units, even to be close to zero inventory on a certain day.

The differences showed fluctuations. The wave valley usually appeared in 7–8 months, 11–12 months, and 1–2 months; the B type fluctuated relatively regularly; the O type had maintained at approximately 250 units, except in August 2016; there were several points when O and AB type were close to zero inventory. A, B and O were even close to 700 units.

The Prophet Model Calculated as Factors

The data were divided into various Time-Series components, and the separated factors are shown in Fig. 2 (A) After the analysis, from the perspective of the trend, the blood consumption showed a downward trend in recent years; (B) In terms of the perspective of the week, the blood consumption in the weekend was significantly less than the weekdays; (C) In terms of the holiday factor, holidays were related to the decrease of RBC consumption; (D) In terms of the seasonal cycle, the RBC consumption in spring and autumn were higher than other seasons.

Analysis of RBC Application Data Involved in the XGBoost Model

From 2014–2018, RBC consumption for operative purpose were less than therapeutic purpose, (34.0% vs 66.0%). For operative application, unused was much more than used, (77.0% vs 23.0%); for therapeutic application, unused was less than used, (34.3% vs 65.7%). Most of the applications were submitted in advance (92.6% of operative application, 69.4% of therapeutic application) (Fig. 3).

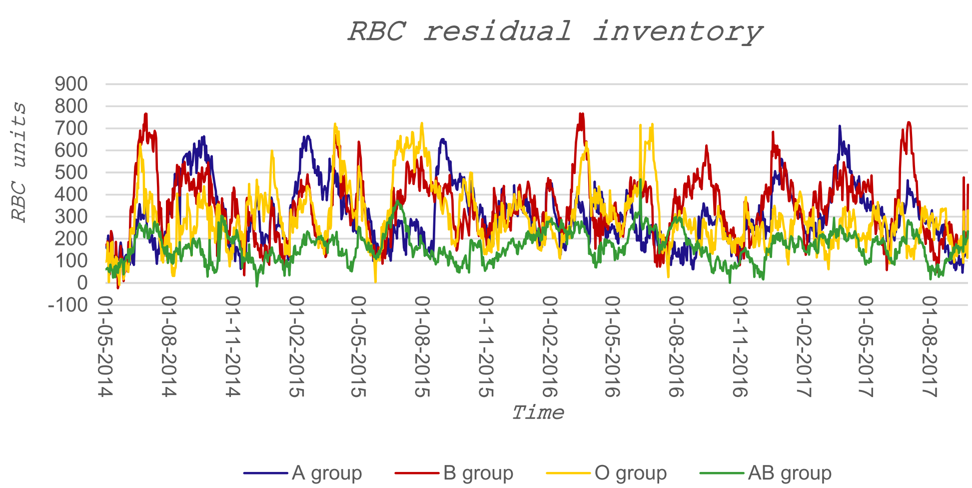

Fig. 1 Daily RBC residual inventory

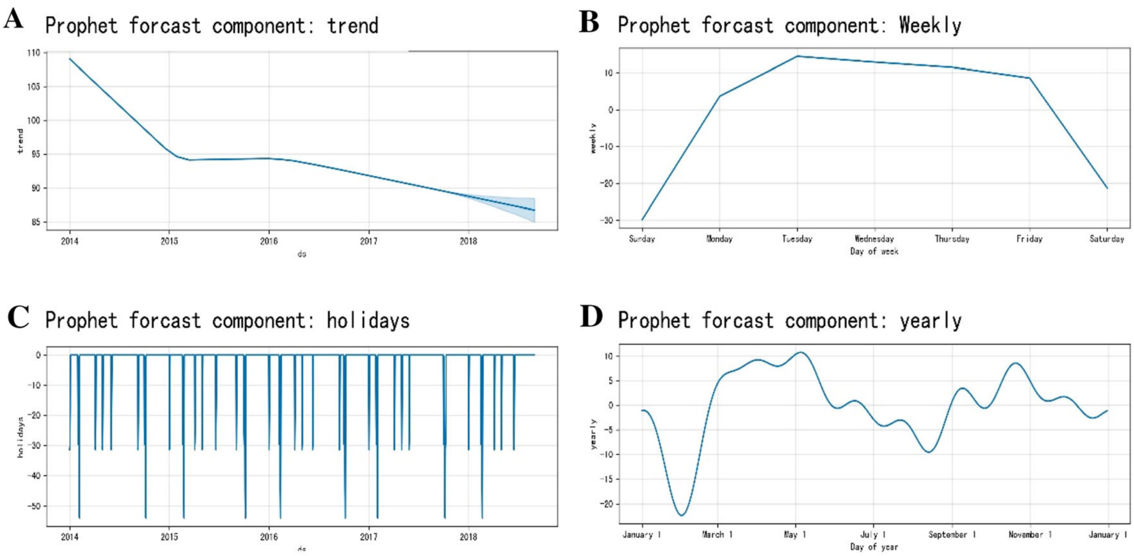

Fig. 2 Analysis of RBC consumption rules by Prophet model

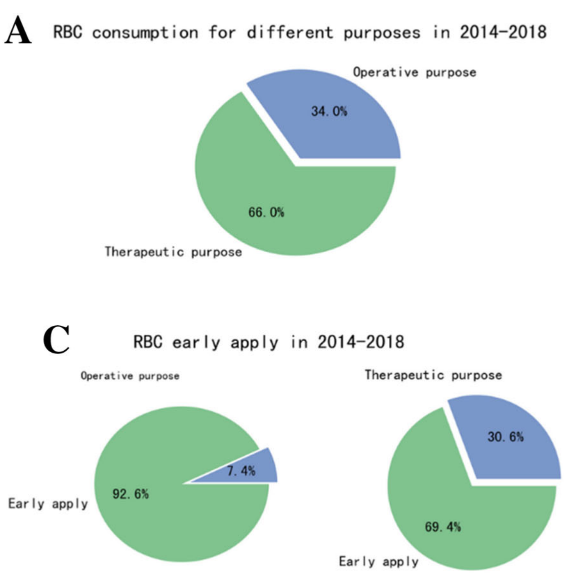

Fig. 3 RBC application data involved in the XGBoost model

Predicting the RBC Demand by Using the XGBoost Model

Using the XGBoost Model, a machine-learning-based prediction method was established, which combined the factors related to transfusion to predict the daily RBC demand. The blue line represented the actual daily RBC consumption, and the orange line represented the predicted value. The MAE values of each blood group predicted by the XGBoost Model were A:10.69, B:11.19, O:10.93, and AB:5.91, respectively, as shown in Fig. 4. From the graph, the predicted volume and actual volume of RBCs were more consistent.

Setting up an Alert Range

To ensure the RBC in stock can meet the demand, a method of setting up an alert range (that is, adding a fixed value to the predicted value) could ensure the RBC inventory meet the RBC demand in most cases. The blue line indicated the actual historical RBC consumption, the red line indicated predicted RBC consumption by the XGBoost Model, and the green line indicated the alert ranges were added to the prediction value. The blood groups were calculated separately. Taking blood type A as an example, as shown in Fig. 5, the green line showed that if 5 units were added to the predicted value, the percentage

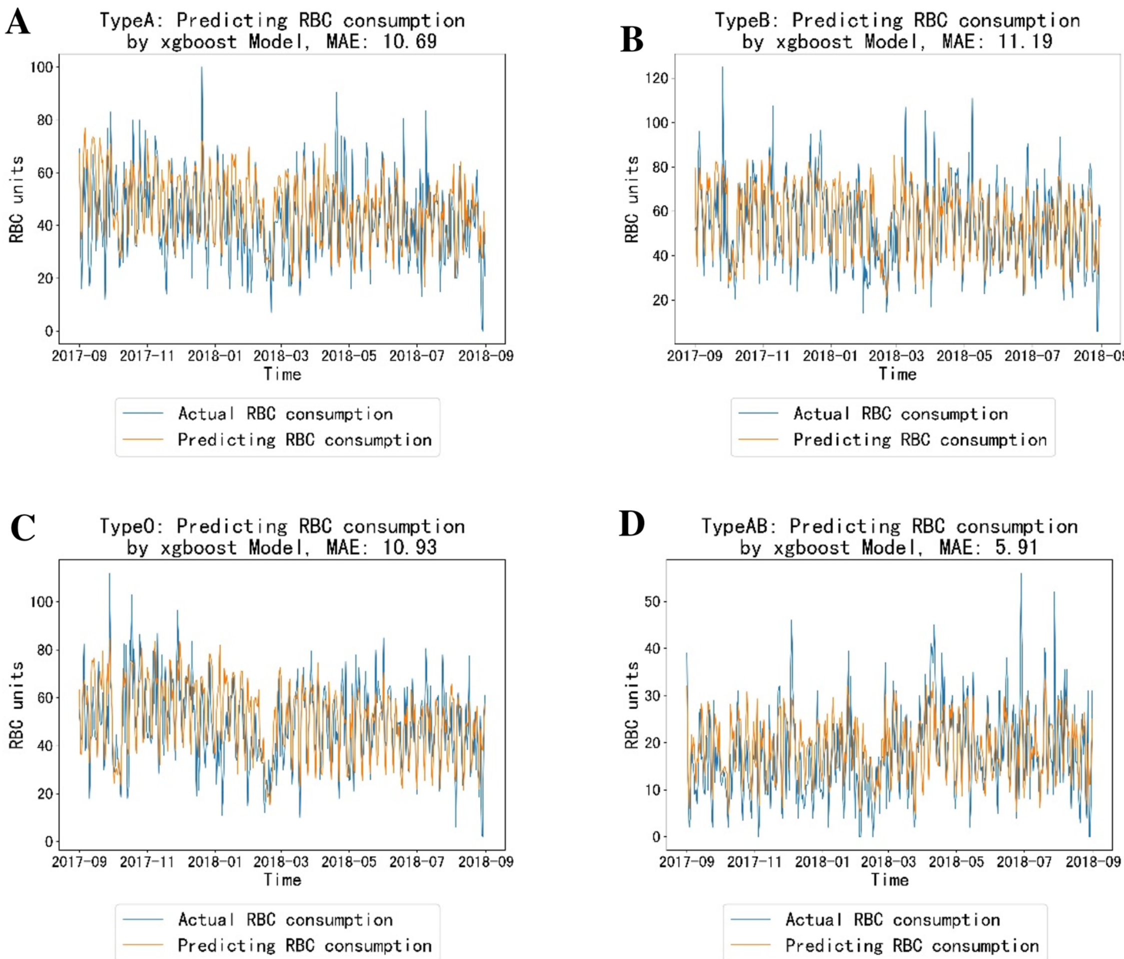

Fig. 4 Predicting the RBC consumption by using the XGBoost Model

of satisfying the accidental consumption is 77.53%; that is, 365 × 77.53% = 282 days. In the case of adding 5 units RBC per day as an alert range, the RBC inventory could meet the accidental requirements for 282 days. An alert range of 10 units was added to the predicted value, and it would satisfy the accidents in the case of 87.67%, and so on.

Other blood groups were predicted and calculated according to the same method. Table 1 lists the correspondence between the percentage that can meet the clinical needs and alert ranges added to the predicted value.

Discussion

Because blood components are scarce medical resources. The RBCs in stock don't only satisfy the selective operations, but also ensure the treatment of anaemic patients in internal departments, perioperative treatment, emergency patients, surgical accidents of obstetric patients [12, 13], etc. It is obvious that a too small stock is risky. Elective surgeries have to be modified or postponed, which prolong the average duration of hospitalization, and hamper treatment of patients. However, from the perspective of health economics, storing a large amount of RBC or setting up a too large alarm range would occupy blood resources, which is unfeasible. The XGBoost Model can predict the RBC demand in the upcoming week, combining the length of the blood storage period to adjust the RBC inventory.

The differences of daily RBC residual inventory were shown in Fig. 1, for a total of 40 months, 1243 days. Such large fluctuations lead to difficulties in RBC management. Excessive RBC inventory might cause resource problems, but the shortage of blood would lead to huge medical risks.

Figure 2 shows the blood transfusion rule. The Prophet model takes advantage of many Bayesian algorithms to make a simple and easy-to-interpret periodic structure. The model can not only predict the future demand, but also find regularity. The blood used for surgery and treatment are different, which cannot be directly substituted. So these relevant factors are added to the historical data to train the

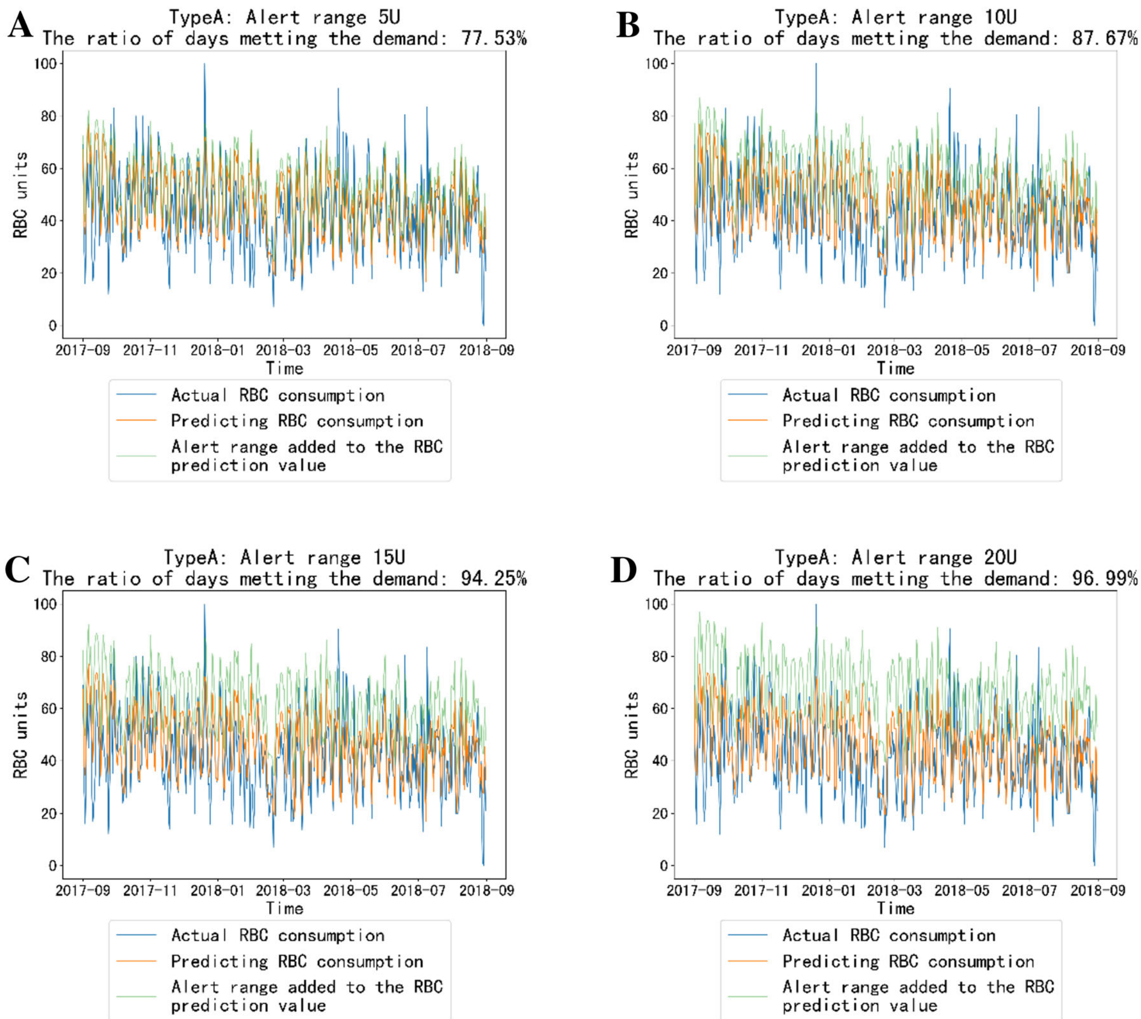

Fig. 5 Relationship between setting up the alert range and meeting the accidental demand-grou

Table1 Adding the alert range values and the percentage that meet the requirements

Set up alert range for RBC	5 units (%)	10 units (%)	15 units (%)	20 units (%)
A type	77.53	87.67	94.25	96.99
B type	72.33	82.74	89.59	93.97
O type	73.15	84.38	91.51	96.16
AB type	83.29	92.6	97.53	99.18

XGBoost Model, which is used to predict the blood demand.

Figure 4 shows the prediction based on the XGBoost Model. The forecast results can better fit the daily blood-flow trend. The MAE values show the degree of closeness between the predicted value and actual value; smaller the MAE value is, closer predicted and actual value are, which means that model performance is better [14]. MAE indicates the proximity of estimated and actual volume. The MAE of AB blood group is the lowest, as seen in Fig. 4, and the proportion of AB blood group in the population is approximately 4.45% [15]. When comparing the MAE values, we also find that the model-dependent prediction results are closer to the actual daily RBC consumption than other methods (Table 1, Supplementary materials), so the prediction results based on the XGBoost Model are better than those state-of-the-art approaches.

Contrary to expectations, the operation transfusion doesn't always comply with the doctors' schedule. In addition, the emergency patients' blood-demand also occupies a certain proportion, such as obstetrical patients, and digestive tract bleeding [16–18], as can be seen from

Fig. 5. Predicting RBC demand cannot cover the peak usage, implying that the model prediction is not ideal for a single day. This may be caused by the limitation of the model itself or the emergency patients and operation accidents. Therefore, if we want to improve the performance of the model, it is necessary to include more specific information of the operations and the patients to predict the incidences of emergency transfusion, such as the baseline haemoglobin level of the patient, existence of organ failure, and historical transfusion volume of the same operation, and other relevant factors [19, 20]. At present, we can ensure safe inventory by adding the alert range, and preparing RBC n units more, to copy with "special events".

But from the reverse point of view, we can't increase the alert range at liberty, which may lead to wastage. Usually, the clinician will submit blood application to the blood bank 3 days in advance. When 5 units/10 units chosen as safety limit, it is necessary to consider the specificity of various operations, and evaluate the possibility of accidental blood consumption (such as Paediatric Cardiac Surgery [21], Children Surgery [22], Critically Ill Surgical [23]. Moreover, after the data of the previous day is uploaded to the model, it will have an impact on the donor recruitment plan in future. Since voluntariness is still the main source of blood, we still have to respond blood demand in a passive manner [24]. For example, 10 units "safety ceiling" RBC are not used in previous day, then the recruitment of blood donors may be reduced in the next day. Of course, autologous blood storage is also a solution to the blood resources. Furuta Y have confirmed non-severe adverse reactions were observed on PAD transfusion at a rate of 0.1% at institution [25].

The model has only a modest improvement, and we may explore more accurate methods or improvements to reduce MAE, which is beneficial for blood transfusion management and treatment.

The establishment of this model is suitable for blood management in the blood transfusion department of tertiary hospitals (The number of beds is more than 500).

Conclusion

Using the machine-learning method to establish the XGBoost Model can facilitate the effective prediction of the RBC demand for the next few days. By setting up the alert range, the security of inventory management can be increased.

Acknowledgement This work was supported by Medical big data research and development project of PLA general hospital (2016MBD-023).

Author Contributions Conceived and designed the experiment: DQW, YY Analysed the data and wrote the paper: XLS, ZHX Managed the data collection: QQY, YNF, YX.

Compliance with Ethical Standards

Conflict of interest The Authors declare no conflicts of interest.

References

1. Ping H, Xing N (2016) Blood shortages and donation in China. Lancet 387(10031):1905–1906
2. Yin YH, Li CQ, Liu Z (2015) Blood donation in China: sustaining efforts and challenges in achieving safety and availability. Transfusion 55(10):2523–2530
3. Volken T et al (2018) Red blood cell use in Switzerland: trends and demographic challenges. Blood Transfus 16(1):73–82
4. Ellingson KD et al (2017) Continued decline in blood collection and transfusion in the United States-2015. Transfusion 57(Suppl 2):1588–1598
5. Mahmood E et al (2018) Multifactorial risk index for prediction of intraoperative blood transfusion in endovascular aneurysm repair. J Vasc Surg 67(3):778–784
6. Erickson ML et al (2008) Management of blood shortages in a tertiary care academic medical center: the Yale-New Haven Hospital frozen blood reserve. Transfusion 48(10):2252–2263
7. Cywinski JB et al (2014) Prediction of intraoperative transfusion requirements during orthotopic liver transplantation and the influence on postoperative patient survival. Anesth Analg 118(2):428–437
8. Hamet P, Tremblay J (2017) Artificial intelligence in medicine. Metabolism 69:S36–S40
9. Tan SS, Gao G, Koch S (2015) Big Data and Analytics in Healthcare. Methods Inf Med 54(6):546–547
10. Guan L et al (2017) Big data modeling to predict platelet usage and minimize wastage in a tertiary care system. Proc Natl Acad Sci U S A 114(43):11368–11373
11. Pendry K (2015) The use of big data in transfusion medicine. Transfus Med 25(3):129–137
12. Butwick AJ, Goodnough LT (2015) Transfusion and coagulation management in major obstetric hemorrhage. Curr Opin Anaesthesiol 28(3):275–284
13. Douglas WG, Uffort E, Denning D (2015) Transfusion and management of surgical patients with hematologic disorders. Surg Clin North Am 95(2):367–377
14. Lau EH et al (2013) Predicting future blood demand from thalassemia major patients in Hong Kong. PLoS ONE 8(12):e81846
15. Ndoula ST et al (2014) Phenotypic and allelic distribution of the ABO and Rhesus (D) blood groups in the Cameroonian population. Int J Immunogenet 41(3):206–210
16. Ng KS et al (2017) Acute lower gastrointestinal haemorrhage: outcomes and risk factors for intervention in 949 emergency cases. Int J Colorectal Dis 32(9):1327–1335
17. Spradbrow J et al (2017) Iron deficiency anemia in the emergency department: over-utilization of red blood cell transfusion and infrequent use of iron supplementation. CJEM 19(3):167–174
18. Compaore GD et al (2014) Readiness of district and regional hospitals in Burkina Faso to provide caesarean section and blood transfusion services: a cross-sectional study. BMC Pregnancy Childbirth 14:158
19. Adjepong KO, Otegbeye F, Adjepong YA (2018) Perioperative Management of Sickle Cell Disease. Mediterr J Hematol Infect Dis 10(1):e2018032

20. Carson JL et al (2016) Transfusion thresholds and other strategies for guiding allogeneic red blood cell transfusion. Cochrane Database Syst Rev 10:CD002042
21. Bishnoi AK et al (2019) Effect of red blood cell storage duration on outcome after paediatric cardiac surgery: a prospective observational study. Heart Lung Circ 28(5):784–791
22. Davies P et al (2007) Calculating the required transfusion volume in children. Transfusion 47(2):212–216
23. Takanishi DM et al (2008) Peripheral blood hematocrit in critically ill surgical patients: an imprecise surrogate of true red blood cell volume. Anesth Analg 106(6):1808–1812
24. Garraud O et al (2016) Voluntariness and blood donation: Proceedings of an ethics seminar held at the National Institute for Blood Transfusion. Transfus Clin Biol 23(3): 168–174
25. Furuta Y et al (2018) Pre-operative autologous blood donation and transfusion-related adverse reactions: a 14-year experience at a university hospital. Transfus Apher Sci 57(5):651–655