RESEARCH ARTICLE

Blood Science

OPEN

An investigation of long-term outcome of rabbit anti-thymocyte globulin and cyclosporine therapy for pediatric severe aplastic anemia

Lixian Chang[a], Mingchen Yan[b], Jingliao Zhang[a], Binghang Liu[b], Li Zhang[a], Ye Guo[a], Jing Sun[b], Yang Wan[a], Meihui Yi[a], Yang Lan[a], Yuli Cai[a], Yuanyuan Ren[a], Haihui Zheng[b], Aoli Zhang[a], Zhenyu Li[b], Jian Wang[b], Yingrui Li[b,*], Xiaofan Zhu[a,*]

[a]State Key Laboratory of Experimental Hematology, National Clinical Research Center for Blood Diseases, Haihe Laboratory of Cell Ecosystem, Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin, China; [b]Shenzhen Digital Life Institute, Shenzhen, China

Abstract

Children with severe aplastic anemia (SAA) face heterogeneous prognoses after immunosuppressive therapy (IST). There are few models that can predict the long-term outcomes of IST for these patients. The objective of this paper is to develop a more effective prediction model for SAA prognosis based on clinical electronic medical records from 203 children with newly diagnosed SAA. In the early stage, a novel model for long-term outcomes of SAA patients with IST was developed using machine-learning techniques. Among the indicators related to long-term efficacy, white blood cell count, lymphocyte count, absolute reticulocyte count, lymphocyte ratio in bone-marrow smears, C-reactive protein, and the level of IL-6, IL-8 and vitamin B12 in the early stage are strongly correlated with long-term efficacy ($P < .05$). Taken together, we analyzed the long-term outcomes of rabbit anti-thymocyte globulin and cyclosporine therapy for children with SAA through machine-learning techniques, which may shorten the observation period of therapeutic effects and reduce treatment costs and time.

Key Words: Anti-thymocyte globulin; Immunosuppressive therapy; Machine learning; Predictive model; Severe aplastic anemia

1 INTRODUCTION

Patients with severe aplastic anemia (SAA) show the features of pancytopenia and hypocellular bone marrow. The telomere length and the paroxysmal nocturnal hemoglobinuria (PNH) population have been linked to response to immunosuppressive therapy (IST).[1] In the absence of a leukocyte antigen-matched donor, SAA cases are treated by the standard combination of IST with anti-thymocyte globulin (ATG)/porcine anti-lymphocyte globulin/horse anti-lymphocyte globulin and cyclosporine A (CsA). ATG is often used in our center. IST with ATG/CsA, giving a 60% to 70% hematopoietic recovery, a 70% to 80% long-term survival,[2] an 87% to 96% overall survival rate for 2 years and an 84% to 92% rate for 10 years, have been reported.[3] Hence, there is significant heterogeneity of outcome in the IST treatment of SAA; how to identify patients with a poor therapeutic effect from IST in the early stage needs further study and identification. As electronic health records (EHRs) have accumulated, machine-learning methods have been developed for building the models of classification, prediction, and other purposes in a series of disorders and have shown better performance than standard statistical models, especially in the scenario of rich data and abundant indicators. However, no such model yet exists for SAA diagnostic evaluation and ATG response, whether for adult or pediatric patients.

In this study, we hypothesized that a machine-learning–based model using more indicators before and after treatment can resolve clinical heterogeneity and provide superior outcome predictions. We aim to establish an SAA clinical dataset-based model to predict IST efficacy using machine-learning techniques.

2 MATERIALS AND METHODS

2.1 Patients

We analyzed the EHR data from our hospital (January 1, 2000 to September 30, 2016). A total of 203 consecutive, newly diagnosed SAA patients were enrolled and assessed according to criteria previously reported.[4,5] SAA was considered when the bone-marrow cellularity was less than 25%, and severe peripheral blood cytopenia existed in at least 2 of 3 lineages, namely absolute neutrophil count $<0.5\times10^9$/L, platelet count $<20\times10^9$/L, and reticulocyte count $<20\times10^9$/L. As shown in **Fig. 1**, clinical examinations were taken at admission time (D0) and 40 days (D40), 3 months (M3), 6 months (M6), 9 months

* *Address correspondence: Xiaofan Zhu, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, 288 Nanjing Road, Tianjin 300020, China. E-mail address: xfzhu@ihcams.ac.cn (X. Zhu); Yingrui Li, Shenzhen Digital Life Institute, Shenzhen, China. E-mail address: liyr@icarbonx.com (Y. Li).*

Conflict of interest: The authors declare that they have no conflict of interest.

L.C., M.Y., J.Z., B.L. contributed equally to this work.

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS, 2021-I2M-C&T-B-082), National Key Research and Development Program of China (2016YFC0901503) and the National Natural Science Foundation of China (81500156, 81170470).

Blood Science (2023) 5, 180–186

Received December 2, 2022; Accepted March 20, 2023.

http://dx.doi.org/10.1097/BS9.0000000000000157

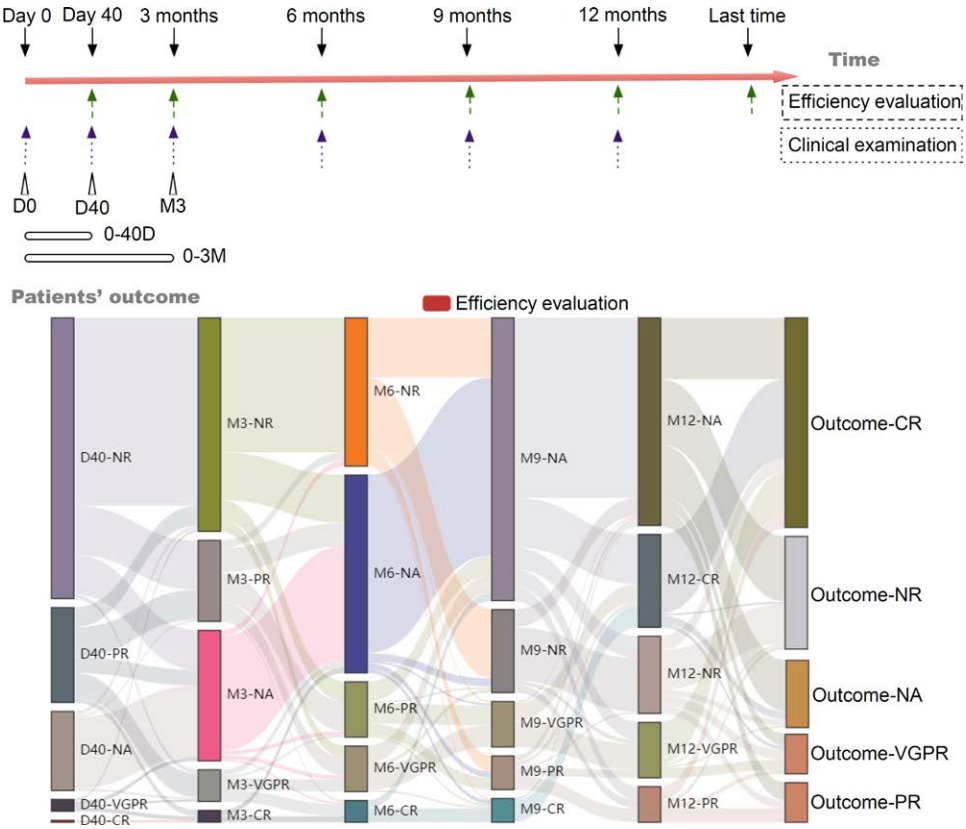

Figure 1. Time set of clinical examinations. CR = complete response, NA = not available, NR = no response, PR = partial response, VGPR = very good partial response.

(M9), and 12 months (M12) after treatment with ATG plus oral CsA. The clinical examinations included routine blood examination and bone-marrow puncture.

As previously published,[6,7] the response definitions included: complete response (CR); very good partial response (PR); PR; and no response (NR). In brief, we defined "complete response" when the absolute neutrophil count was $>1\times10^9$/L, the platelet count $>100\times10^9$/L, and the hemoglobin level >10.0 g/dL. We defined "very good partial response" (VGPR) when the absolute neutrophil count was $>0.5\times10^9$/L, the platelet count $>50\times10^9$/L, and the hemoglobin level >8 g/dL. We defined "partial response" when the absolute neutrophil count was $>0.5\times10^9$/L, the platelet count $>20\times10^9$/L, and the hemoglobin level >8 g/dL. We defined "no response", when the absolute neutrophil count was $<0.5\times10^9$/L, the platelet count $<20\times10^9$/L, and the hemoglobin level <8 g/dL. At the end of each treatment phase, the response evaluation was recorded. The last follow-up was conducted 1 year after the start of therapy. We eliminated 35 patients from the study, who had not completed a 12-month treatment course and therefore had no final efficacy evaluation. Our study has been approved by the Clinical Research Ethics Committee of Blood Diseases Hospital, Chinese Academy of Medical Sciences [NI2020009-EC-1].

2.2 Outcome and prediction window

Based on accepted practice, outcomes were assessed from follow-up data (**Fig. 1**) using criteria as previously reported.[6,7] To establish a 2-class machine-learning model, depending on whether the child was dependent on blood transfusions or not, the CR, VGPR, and PR in the follow-up evaluation were combined with the effective marker as positive and the NR marker as unfavorable. The median follow-up periods for those who fulfilled the eligibility criteria were 48 months. We used the D0 data in our primary analyses to build a baseline model. Then, the model performance was assessed to predict the outcomes during the different phases of treatment. We used the clinical data generated during the treatment course (40 days, 3 months) after the baseline to vary the prediction window.

2.3 Predictors

The 2009 UK Adult SAA Clinical Evaluation Guide provides 7 efficacy-related indicators, including: age; absolute lymphocyte count (LYM); absolute reticulocyte count (ARC); PNH clones; karyotype; the severity of disease; and telomere length. We used the first 5 of them in our datasets as knowledge-based predictors (K and K+ predictor sets, variables used in the prediction models). Also, we built a series of data-driven predictors (D and D+ predictor sets). The method of predictor selection is recursive feature elimination (RFE), which refers to sorting all candidate predictors using random forest (RF) feature importance, removing the least significant predictor at the end, re-ordering the remaining features, and continuously removing the least important until all features were exhausted. The number of features between 10 and 30 corresponding to the maximum area-under-the-curve (AUC) was selected as the data-driven predictors.

As the treatment progressed, additional information, such as the amount of platelet transfusion, became available. We analyzed the difference between several models throughout the treatment course for the performance assessment. For the knowledge-based predictors, in the first step, we updated the predictors' value by subsequence treatment phase as fixed time point sets of predictors (K, eg, using LYM (D40) to replace LYM (D0)). In the second step, we added subsequence treatment phase variables and included information about variable timing as time-series sets of predictors (K+, eg, both LYM (D0) and LYM (D40) are reserved). For the data-driven predictors, in the first step, we updated the candidate predictors

by subsequence treatment phase and used RFE to renew the predictors' collection as fixed time point sets of predictors (D40-D, M3-D). In the second step, we added subsequence treatment phase variables with the variable timing data, then used the RFE to renew the collection of predictors as time-series sets of predictors (D40-D+, M3-D+). Table S1, http://links.lww.com/BS/A57, shows the complete list of predictors. The missingness of each variable in the raw data was recorded in Table S2, http://links.lww.com/BS/A57. Missing laboratory and clinical tests were recorded as the number of individuals and as a percentage of the total. Efficacy assessments (clinical diagnoses) were considered when they were documented. We removed those samples and indicators which have missing values over 30%. Data missingness after cleaning for each variable is shown in Table S3, http://links.lww.com/BS/A57. Then, missing data were inputted with mean, median, or appropriate default values. Ultimately, 23 and 11 variables were identified for the D40-D+, and M3-D+ time-series predictor sets, respectively, 11 for the D0-D, D40-D, and M3-D, and 5 for the K and K+ predictor sets.

2.4 Machine-learning algorithms

A knowledge-based model was first developed as a benchmark model. The predictor variables in K and K+ were involved. The same method was then used to build the model based on D0-D, D40-D, M3-D, D40-D+, and M3-D+ predictor sets. To avoid the bias as a result of classifier selection, we used 4 classifiers with different mathematical principles, namely logistic regression (LR), multilayer perceptron (MLP), support vector machine (SVM), and RF. These methods were used because they have been successfully applied to medical data sets for disease classification, and their characteristics have been previously described in detail.[8] The liblinear solver was used to optimize functions of LR. The MLP contained 2 hidden layers with 50 and 20 neurons, respectively, and each layer led to a reduced neuron number. The radial basis functions were used as kernel parameters of SVM model construction. After extensively searching the parameter space for the RF classifiers, we adjusted the hyperparameters.

Overall, 20 data-driven models (5 sets of predictors and 4 modeling techniques) and 20 knowledge-based models (3 fixed time point data, 2 time-series data, and 4 modeling techniques) were identified for the different analyses. Then, the evaluation efficacy of the final follow-up was explored for the given models.

2.5 Evaluation methodology and metrics

For the cohort, owing to the small number of samples in this cohort and a large number of clinical indicators, we referred to the literature for similar situations and used a conservative holdout method.[9] We randomly chose 75% samples as a training set and the remaining 25% as a testing set. We repeated this 50 times and reported the F1 score value and the sensitivity, specificity, precision and recall of each of the 4 different classifiers. Table S4, http://links.lww.com/BS/A57, presents the 50-fold retention results for each predictor set and each classifier. Smaller cohorts showed higher variance because each sample had a more significant relative impact on classification performance. Table S5, http://links.lww.com/BS/A57, shows the common predictors for all 50 training iterations. Our findings were reported following the guidelines for the biomedical research-related machine-learning predictive models and the Transparent Reporting of Multivariable Predictive Models for Individual Prognosis or Diagnosis (TRIPOD). We used Python 3.6 and R version 3.2 for all statistical analyses and presented the workflow in **Fig. 2**.

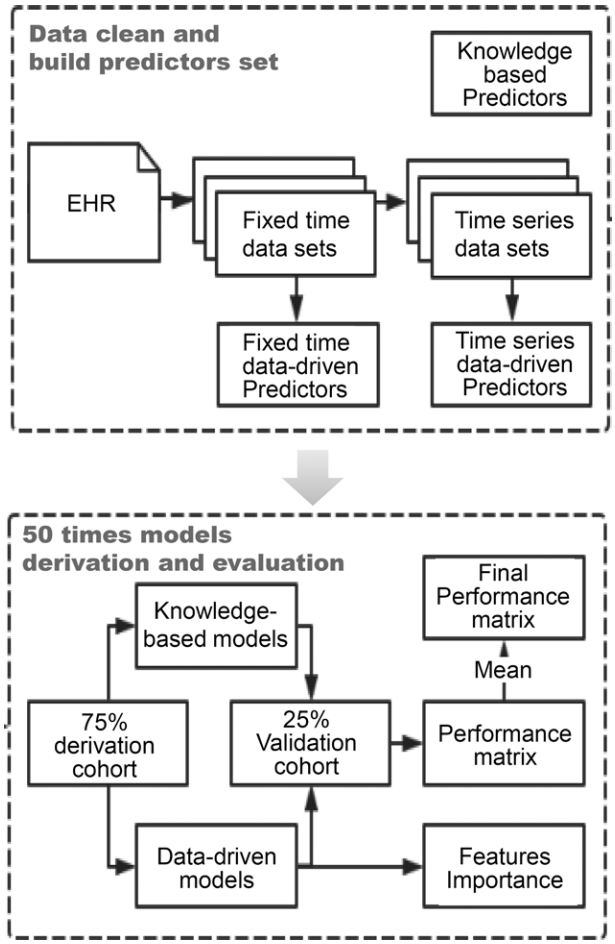

Figure 2. Workflow of data processing in our study. EHR = electronic health record.

3. RESULTS

3.1 Basic statistics

The clinical characteristics of the cohort are presented in Table S2, http://links.lww.com/BS/A57. As the treatment progressed, the condition of some patients improved, so there was no need for a partial examination; this resulted in missing data after 6 months. Thus, only the first 3 months of early treatment data were included in this analysis. As expected, the missing data of each indicator made up less than 30% after the data were cleaned (Table S3, http://links.lww.com/BS/A57).

After cleaning the data, we obtained 185, 161, and 92 patients and 74, 63, and 35 predictors at the 3 fixed time points (D0, D40, and M3). Meanwhile, we took the intersection of the patients at each fixed time point and obtained 2 time-series data sets: the D40+ dataset with 151 patients and 137 features and the M3+ dataset with 76 patients and 169 features. The average age of all patients in the cohort was 7.13 years (median, 6.5; std, 4.103; min, 1; max, 18). There was no significant difference in age or gender between the 2 groups (Table S1, http://links.lww.com/BS/A57).

We performed a statistical test of the examination indicators between the 2 groups with different curative effects. The lymphocyte-related indicators in the bone-marrow examination were significantly different at all times ($P < .01$). In the admission examination before the start of treatment, we observed a significant difference (Tables S1, S6, http://links.lww.com/BS/A57, $P < .01$) for a set of indicators, such as erythrocyte count,

erythroid ratio, morphology in bone-marrow smears, granulocyte ratio, lymphocyte ratio, colony forming unit erythroid, burst forming unit erythroid in bone marrow, absolute LYM, c-reactive protein (CRP), interleukin-2 (IL-2), IL-6, IL-8, and vitamin B12 in peripheral blood. Among these factors, the CRP, vitamin B12, IL-8, and IL-6 were new prognostic factors for IST treatment.

3.2 Predictors related to prognosis and importance evaluation

The AUC value tended to be stable between 10 and 30 predictors as the number of retained predictors increased. The AUC value decreased significantly when it was below 10 predictors in the RFE with RF. So, the predictors set with the most significant AUC value were taken as data-driven predictors (Table S4, http://links.lww.com/BS/A57). Using fixed time predictors (D0, D40, and M3), the maximum AUC was above 0.75. Applying D40+ and M3+ to the same method increased the maximum AUC to over 0.85 and 0.95, respectively (**Fig. 3**). Using time-series predictors (D40+ and M3+), a slightly higher AUC was detected in all models, especially for the M3+ predictors. Predictors from a later time point resulted in higher AUC values for all models.

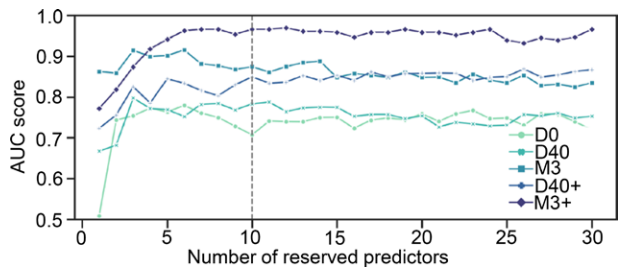

Figure 3. AUC scores of recursive feature elimination. AUC = area-under-the-curve.

We calculated the relative feature importance by analyzing the feature's contribution in each tree in the RF model and then took the average. The calculation method of the contribution was applied using the Gini coefficient. We calculated the relative importance of all data-driven model features, and all considered predictors without RFE selection (Table S7, http://links.lww.com/BS/A57). For instance, when RF was used in the D0 set, the lymphatic system proportion in bone-marrow smears, complement C3, red blood cell count, hemoglobin, and CRP were the top predictors with a higher importance score. With the D40+ set, RF ranked the MONO in the first stage of treatment as a critical predictor, followed by the laboratory test results and the RBC transfusions during treatment. Finally, when M3+ variables were used, both the bone-marrow smear and the treatment approach were shown to be more critical in predicting the outcome, together with biochemical tests and laboratory test results. In contrast, bone-marrow smear results remained among the top few predictors.

3.3 Modeling analysis data of long-term outcome

To develop a clinically applicable tool predicting the long-term response probability of SAA patients benefitting from IST treatment before the start of the treatment process, we used the nomogram to develop a predictive model, based on the clinical covariates (**Fig. 4A**). Predictive factors, including the lymphatic system proportion in bone-marrow smears, C3, RBC, hemoglobin, CRP, serum iron, white blood cell count, CD3+CD8+T cell count, direct bilirubin, folic acid and erythrocyte colony forming units from admission medical examination, were selected using the data-driven RFE method in Table S7, http://links.lww.com/BS/A57. The forest plots are presented in **Fig. 4B**, and calibration curves in **Fig. 4C**. A close link existed between predicted and ideal standard lines, indicating a good predictive ability for our nomogram.

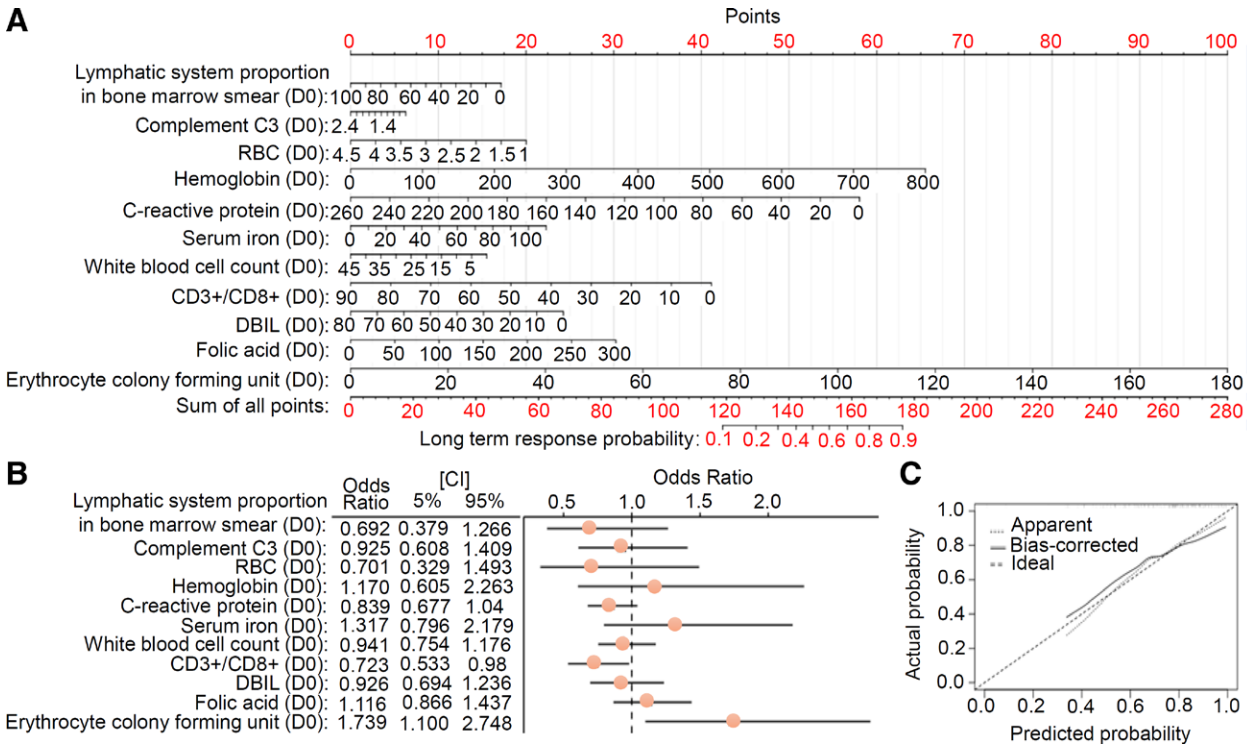

Figure 4. Modeling analysis based on the predictor of admission medical examination. (A) a plotted Nomogram; (B) forest plot data; (C) calibration plot curve. CI = confidence interval.

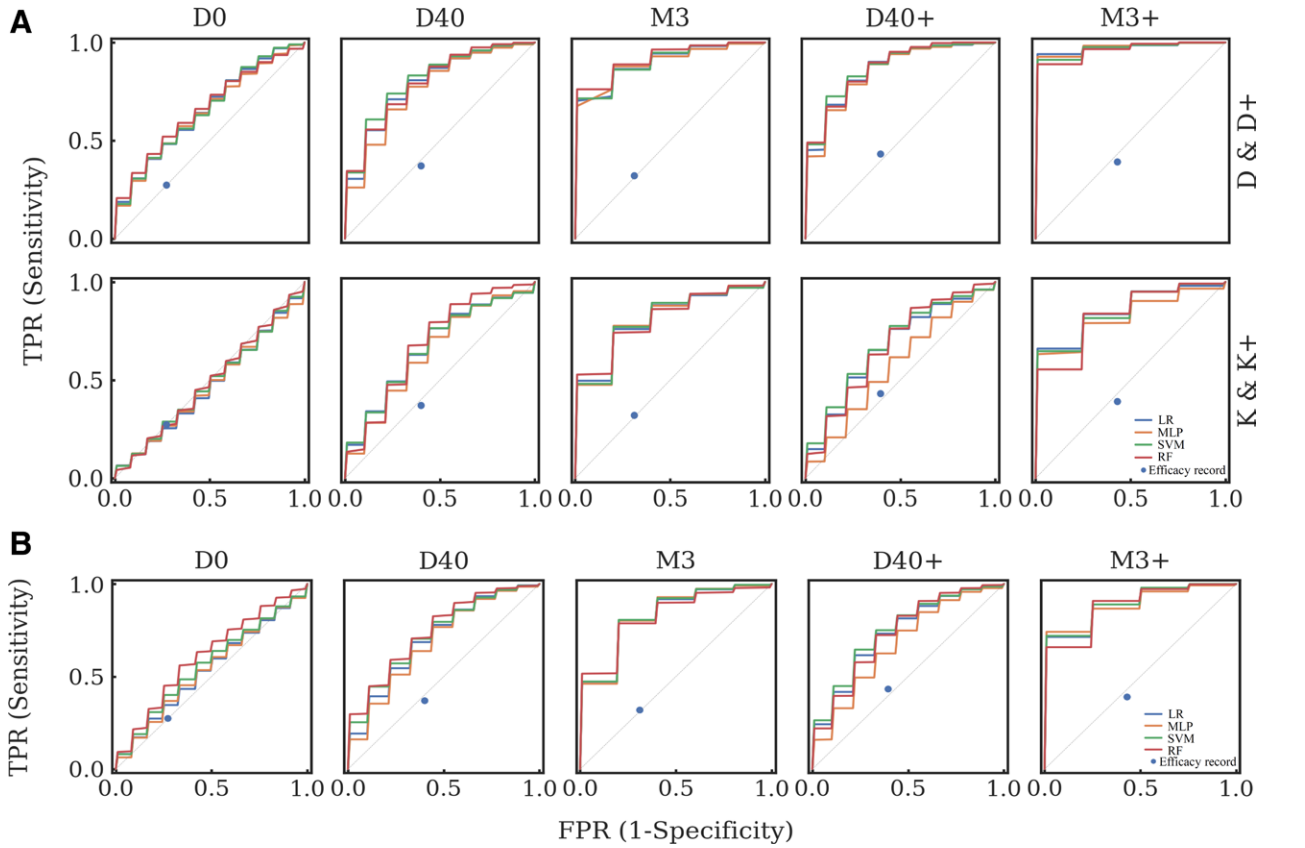

Figure 5. ROC data of the model. (A) the validation set; (B) the K & K+ predictor set with LPR[ms]. LPR[ms] = proportion of lymphocytes in bone-marrow smear. FPR = false positive rate, ROC = receiver operating characteristic, TPR = true positive rate.

3.4 Prognostic models with time-series data-driven indicators

The AUC values from the 50-fold holdout results are given in Tables S4 and S8, http://links.lww.com/BS/A57, and receiver operating characteristic (ROC) curves in **Fig. 5A**. When compared with the knowledge-based models, all data-driven predictor models exhibited an increased value of AUC. Also, all time-series predictors models showed a rise compared with fixed time models' performance. When adding a feature called "Proportion of lymphocytes in bone-marrow smear (LPR[ms])," which is significantly associated with long-term efficacy to knowledge-based predictor sets, we observed the increased AUC and other performance of all models (**Fig. 5B** and Table S9, http://links.lww.com/BS/A57). The performance improvement of the model obtained by adding other indicators proved that there were still other indicators containing valid information that had not been used. A data-driven approach can help us discover more clinical indicators related to long-term outcomes such as LPR[ms].

The best predictive models were built with the M3+ predictors set. Except for the SVM classifier, their AUC values were over 0.96, and the F1-score was over 0.85. LR with the M3+ indicator set had a higher AUC than the RF, but the F1-score and specificity of the RF performed better, and our model (RF with M3+ predictors) showed an AUC of 0.962 using the validation set. The model's excellent accuracy, sensitivity, and precision were detected in derivation/validation sets. Therefore, RF with M3+ predictors was the most valuable for clinical application. Its accuracy, sensitivity, and precision were more significant than 0.9, and the specificity was 0.82.

Using time-series data-driven predictors, the classifiers of the different algorithms had a good performance (AUC ≥ 0.85) on the 40th day of treatment. With the acquisition of new clinical examination data at the 3rd month of treatment, they made a more accurate prediction of long-term efficacy (AUC ≥ 0.95) and better than the current efficiency evaluation method (efficacy record in Table S4, http://links.lww.com/BS/A57, and **Fig. 4**).

4. DISCUSSION

In this article, we have aimed to establish a set of longitudal EMR data-based models to predict the treatment efficiency of SAA cases using machine learning. The model was studied by traditional statistical methods and Local Interpretable Model-Agnostic Explanations (LIME). The significant results included the following: i) a clinically usable long-term response prediction model in advance of treatment; ii) as the course of treatment progressed, a data-driven model of the continuously updated indicators to achieve better prediction performance and provide a reference for the next stage of the treatment plan; iii) based on classic statistics and machine-learning methods, some previously unreported indicators related to the prognosis of children with SAA were discovered, such as CRP, vitamin B12, IL-6, and IL-8.

We comprehensively analyzed the general situation, bone-marrow results, and hematology-related tests in SAA children receiving IST and put forward some prognostic factors in this treatment. Some have been widely reported in adult SAA, and some have been rarely studied in childhood SAA. Some results were similar to those already reported, while others were different. It was found that specific parameters were better predictors of prognosis, such as sex, age, and higher absolute neutrophil count.[10–12] In the adult, the 2009 UK Adult SAA Clinical Evaluation Guide provided 7 efficacy-related indicators, including age, absolute lymphocyte count, ARC, PNH clones, karyotype, the severity of disease, and telomere length. In our study, age was not different between the 2 groups, indicating that age

was not an efficacy indicator for children. However, a short granulocyte telomere for pediatric acquired aplastic anemia patients resistant to IST has been reported.[13] However, only 14 cases were enrolled, and statistical analysis was not performed. The multivariate analyses of some studies of pediatric SAA suggested that the "male" factor may be an efficient predictor of good response.[3]

Nevertheless, there was no significant difference in our statistical result. In this and our previous study, the higher absolute neutrophil count, platelet count, and PNH-clone negativity might be good predictive factors of response.[14] The LYM and ARC were also important, but the "Proportion of lymphocytes in bone-marrow smears (LPR^{ms})" was significantly associated with long-term efficacy and probably more effective. The CD3+CD8+ T cell proportion for bone marrow also served as a predictor of IST response, which was in line with a previous study.[15] Gupta et al[16] detected high levels of IL-6/IL-8 for childhood aplastic anemia patients and a correlation between increased levels and disease severity. Even though no statistical difference was detected for the levels between responders and non-responders,[16] we, by contrast, observed a significant difference. This might be related to the low number of cases in the Gupta study. Compared with the children in the positive group, the serum level of CRP was higher in the opposing group, which indicated that infection of children with SAA before treatment was associated with a treatment effect; IL-6/IL-8 and CRP all reflected the activation status of white blood cells and the occurrence of infection, suggesting that infection was a factor responsible for poor ATG treatment efficacy. In addition, the serum level of vitamin B12 in SAA children increased, although there was a statistical difference of level between positive and negative groups. It may be related to reduced bone-marrow hematopoiesis and surplus hematopoietic materials in patients with aplastic anemia.

Longitudal clinical detection after treatment is essential for the prediction of treatment efficiency. First, time-series predictors were more effective than fixed time predictors based on the AUC values at D40 or M3. More accurate results could be obtained by combining previous data in clinical treatment. Models based on data from a fixed day, no matter whether D0, D40, or M3, never performed better than that from D0&D40 or D0&D40&M3. Second, the most recent data could bring a better predictive model, consistent with other similar research.[4] The effect of prediction at M3 was better than that of 40D. Third, for SAA in children and treatment with ATG, the predictive efficiency had already been high to 0.85 at 40D+ and extended to 0.96 when adding 3rd month data. Overall, models for different post-treatment time points could be combined to improve the clinical treatment plan and, in the end, potentially shorten the treatment efficiency observation period and save time and cost for treatment.

The amount of indicator detected clinically was also essential for prediction. When considering too few indicators, such as fewer than 10 for SAA or with indicators in clinical guidelines, the performance will not be as good or stable as expected. The AUC value of knowledge-based features was smaller than that of the knowledge-based features plus lymphatics, showing the necessity of adding more essential indicators. In our data-driven models, we utilized about 10 indicators and obtained stable results with high performance. It is necessary to mention that the indicators obtained by feature ranking were consistent with those based on the data-driven models, but they were not the same. Markers with high correlation would not be selected as the most predictive indicators owing to their possible correlation with other predictive indicators. Meanwhile, predictive indicators with lower correlation ranking might be related to long-term treatment efficacy. Indicators which correlated highly with treatment response and used in predictive models could bring a greater understanding of the pathology of SAA.

This study has some inherent limitations. First, the number of patients for experimental and ineffective treatments was imbalanced in our dataset. To reduce the over-fitting caused by the imbalance of positive/negative cases and enhance the model's generalization ability, we adopted the SMOTE algorithm. SMOTE was applied to the random over-sampling. Second, all the data came from one hospital, which might introduce bias. So, more tests are needed from other sources. Future work will include testing data from different hospitals and improving models for clinical application through the ensemble-learning method.

In summary, we used machine-learning methods to predict long-term outcomes after rabbit ATG and cyclosporine therapy for childhood SAA. Considering that there are currently no publicly acceptable efficiency prediction models for this condition, we hope that our models can shorten the observation period of therapeutic effects while saving time and treatment costs.

ACKNOWLEDGMENTS

This work was supported by the CAMS Innovation Fund for Medical Sciences (CIFMS, 2021-I2M-C&T-B-082), National Key Research and Development Program of China (2016YFC0901503) and the National Natural Science Foundation of China (81500156, 81170470).

We would like to acknowledge the AiYou Foundation and the patients for participating in the follow-up and those who provided feedback on the written information. We also thank Danni Li, Danhua Qin and Jing Sun from the Shenzhen Digital Life Institute for their suggestions on data analysis.

REFERENCES

[1] Narita A, Muramatsu H, Sekiya Y, et al; Japan Childhood Aplastic Anemia Study Group. Paroxysmal nocturnal hemoglobinuria and telomere length predicts response to immunosuppressive therapy in pediatric aplastic anemia. *Haematologica* 2015;100(12):1546–1552.

[2] Zhang L, Jing L, Zhou K, et al. Rabbit antithymocyte globulin as first-line therapy for severe aplastic anemia. *Exp Hematol* 2015;43(4):286–294.

[3] Jeong DC, Chung NG, Cho B, et al. Long-term outcome after immunosuppressive therapy with horse or rabbit antithymocyte globulin and cyclosporine for severe aplastic anemia in children. *Haematologica* 2014;99(4):664–671.

[4] Camitta BM, Rappeport JM, Parkman R, Nathan DG. Selection of patients for bone marrow transplantation in severe aplastic anemia. *Blood* 1975;45(3):355–363.

[5] Bacigalupo A, Hows J, Gluckman E, et al. Bone marrow transplantation (BMT) versus immunosuppression for the treatment of severe aplastic anaemia (SAA): a report of the EBMT SAA working party. *Br J Haematol* 1988;70(2):177–182.

[6] Townsley DM, Scheinberg P, Winkler T, et al. Eltrombopag added to standard immunosuppression for aplastic anemia. *N Engl J Med* 2017;376(16):1540–1550.

[7] Rogers ZR, Nakano TA, Olson TS, et al. Immunosuppressive therapy for pediatric aplastic anemia: a North American Pediatric Aplastic Anemia Consortium study. *Haematologica* 2019;104(10):1974–1983.

[8] Mohammed A, Podila PSB, Davis RL, Ataga KI, Hankins JS, Kamaleswaran R. Using machine learning to predict early onset acute organ failure in critically ill intensive care unit patients with sickle cell disease: retrospective study. *J Med Internet Res* 2020;22(5):e14693.

[9] Stafford P, Cichacz Z, Woodbury NW, Johnston SA. Immunosignature system for diagnosis of cancer. *Proc Natl Acad Sci U S A* 2014;111(30):E3072–E3080.

[10] Yoshida N, Yagasaki H, Hama A, et al. Predicting response to immunosuppressive therapy in childhood aplastic anemia. *Haematologica* 2011;96(5):771–774.

[11] Chang MH, Kim KH, Kim HS, et al. Predictors of response to immunosuppressive therapy with antithymocyte globulin and cyclosporine and prognostic factors for survival in patients with severe aplastic anemia. *Eur J Haematol* 2010;84(2):154–159.

[12] Scheinberg P, Wu CO, Nunez O, Young NS. Predicting response to immunosuppressive therapy and survival in severe aplastic anaemia. *Br J Haematol* 2009;144(2):206–216.
[13] Tutelman PR, Aubert G, Milner RA, Dalal BI, Schultz KR, Deyell RJ. Paroxysmal nocturnal haemoglobinuria phenotype cells and leucocyte subset telomere length in childhood acquired aplastic anaemia. *Br J Haematol* 2014;164(5):717–721.
[14] Liu L-P, Chen X-J, Yang W-Y, et al. Predicting response to porcine antilymphocyte globulin plus cyclosporine A in children with acquired severe aplastic anemia. *Pediatr Res* 2019;86(3): 360–364.
[15] Zhao F-Y, Xu X-J, Song H, Yang S-L, Tang Y-M. Effectiveness of immunosuppressive therapy for childhood aplastic anemia and its predictive factors. *Zhongguo Dang Dai Er Ke Za Zhi* 2012;14(8): 567–570.
[16] Gupta V, Kumar S, Sonowal R, Singh SK. Interleukin-6 and interleukin-8 levels correlate with the severity of aplastic anemia in children. *J Pediatr Hematol Oncol* 2017;39(3):214–216.