Observational Study

Medicine

OPEN

Prediction of renal damage in children with IgA vasculitis based on machine learning

Jinjuan Wang, MD, Huimin Chu, MD, Yueli Pan, MD*

Abstract

This article is objected to explore the value of machine learning algorithm in predicting the risk of renal damage in children with IgA vasculitis by constructing a predictive model and analyzing the related risk factors of IgA vasculitis Nephritis in children. Case data of 288 hospitalized children with IgA vasculitis from November 2018 to October 2021 were collected. The data included 42 indicators such as demographic characteristics, clinical symptoms and laboratory tests, etc. Univariate feature selection was used for feature extraction, and logistic regression, support vector machine (SVM), decision tree and random forest (RF) algorithms were used separately for classification prediction. Lastly, the performance of four algorithms is compared using accuracy rate, recall rate and AUC. The accuracy rate, recall rate and AUC of the established RF model were 0.83, 0.86 and 0.91 respectively, which were higher than 0.74, 0.80 and 0.89 of the logistic regression model; higher than 0.70, 0.80 and 0.89 of SVM model; higher than 0.74, 0.80 and 0.81 of the decision tree model. The top 10 important features provided by RF model are: Persistent purpura ≥4 weeks, Cr, Clinic time, ALB, WBC, TC, Relapse, TG, Recurrent purpura and EB-DNA. The model based on RF algorithm has better performance in the prediction of children with IgA vasculitis renal damage, indicated by better classification accuracy, better classification effect and better generalization performance.

Abbreviations: ALB = blood albumin, APTT = activated partial thromboplastin time, ASO = anti-hemolytic streptococcus antibody, AUC = area under curve, BUN = Blood Urea Nitrogen, Cr = creatinine, CRP = hypersensitive C-reactive protein, Cys-C = Cystatin C, EB-DNA = Epstein-Barr virus DNA, ESR = erythrocyte sedimentation rate, FIB = Fibrin Original, HLD = high-density lipoprotein, IgA = Immunoglobulin A, IgAV = Immunoglobulin A vasculitis, IgAVN = Immunoglobulin A vasculitis Nephritis, IgE = Immunoglobulin E, IgG = Immunoglobulin G, IgM = Immunoglobulin M, L/M = lymph/mononuclear, LDL = low-density lipoprotein, LYMPH# = lymphocyte absolute value, MONO# = monocyte absolute value, Mp-Ab = mycoplasma pneumoniae infection-antibody, N/L = neutrophil/lymph, NEUT# = neutrophil absolute value, PLT = platelet, RBP = retinol-binding protein, RF = random forest, RFE = recursive feature elimination, ROC = receiver operating characteristic curve, SVM = support vector machine, TC = total cholesterol, TG = triglycerides, TT = thrombin time, U/C = Urinary albumin/creatinine ratio, WBC = white blood cell.

Keywords: IgA vasculitis, machine learning, prediction, random forest, renal impairment

1. Introduction

IgA vasculitis (IgAV), also known as Henoch-Schönlein Purpura (HSP), is one of the most common systemic vasculitis in children.[1] The onset of IgAV is rather acute, and the incidence rate of IgAV is about 3/100,000 to 27/100,000.[2,3] The clinical features of IgAV are skin purpura, abdominal pain, joint swelling and pain, gastrointestinal bleeding, hematuria, proteinuria and other renal involvement, among which renal involvement is the key factor affecting the prognosis. It has been reported that about 20% to 80% of children with IgAV have involved to IgA vasculitis Nephritis (IgAVN) in the first or second month after onset, and 1% of patients may even develop end-stage renal failure,[4,5] which would seriously affect the quality of life of the children with the disease. The presence of renal involvement is a key factor in the long-term prognosis for children with IgAV, and the early accurate diagnosis of IgAVN is crucial for the prognosis and individualized treatment of children.

In recent years, big data analysis and mining has attracted increasing attention. Machine learning, as an emerging statistical analysis method, is capable of deep mining and analysis of big data, and has been widely used in disease occurrence, prognosis prediction and other aspects, achieving certain results.[6,7] Therefore, this article aims to build a prediction model of renal damage in children with IgAV based on machine learning algorithm, identify related risk factors, and provide help for the early diagnosis and individualized intervention treatment of children with IgAVN.

The authors have no funding and conflicts of interest to disclose.

The authors have no consent for publication.

All data generated or analyzed during this study are included in this published article [and its supplementary information files].

This is an observational study. The study was approved by the Medical Ethics Committee of Affiliated Hospital of Shandong University of Traditional Chinese Medicine. All methods were carried out in accordance with relevant guidelines and regulations. We confirmed that informed consent was obtained from all patients or their parents.

Shandong University of Traditional Chinese Medicine, Shandong, PR China.

** Correspondence: Yueli Pan, Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan, Shandong 250014, PR China (e-mail: pyl_sd@163.com).*

How to cite this article: Wang J, Chu H, Pan Y. Prediction of renal damage in children with IgA vasculitis based on machine learning. Medicine 2022;101:42(e31135).

Received: 9 June 2022 / Received in final form: 13 September 2022 / Accepted: 14 September 2022

http://dx.doi.org/10.1097/MD.0000000000031135

2. Methods

2.1. Research objects

This article selected children with IgAV who were treated from November 2018 to October 2021 as study subjects. After following up for at least 6 months, the patients were divided into IgAV group and IgAVN group according to urine test results.

2.2. Classification and diagnostic criteria

The classification criteria of IgAV refer to the criteria of IgAV formulated by the EULAR/PRES/PRINTO criteria in 2010.[8] The diagnostic criteria of IgAV is palpable rash(necessary condition) with at least one of the following four clinical symptoms: abdominal pain, arthritis/joint pain, renal involvement, histopathological findings suggesting IgA deposition.

According to "Evidence-based Guideline for Diagnosis and Treatment of Purpura Nephritis (2016)" by the Nephrology group of pediatric branch of Chinese Medical Association, the diagnostic criteria of IgAVN[9] were proposed as: hematuria and/or proteinuria occurred within 6 months of the course of IgAV. Hematuria definition: gross hematuria or hematuria with more than 3 red blood cells/high-power field (HP) under microscope 3 times within 1 week. Proteinuria definition: meet any of the following conditions: 3 routine urine tests within 1 week qualitatively indicate positive urine protein; 24 h quantitative urine protein > 150 mg or Urinary albumin/creatinine ratio (U/C) (mg/mg) > 0.2; Urinary microalbumin was higher than normal for 3 times within 1 week.

2.3. Inclusion criteria

The patient was included if he/she met the diagnostic criteria above and was under 18 years old, and the parents of the patient gave consent.

2.4. Exclusion criteria

Patients are excluded if they had severe heart, liver, brain, immune system or other diseases or consumptive diseases, or do not cooperate with the observer. Patients with kidney diseases below are excluded from the study: primary glomerular disease (glomerulonephritis, primary nephrotic syndrome, etc), secondary glomerular nephropathy (lupus nephritis, HBV-associated glo-Merulonephritis, ANCA-associated systemic vasculitis, etc), hereditary glomerular disease (congenital nephrotic syndrome, Alport syndrome, etc).

2.5. Research factors

42 features were researched as the variables to be predicted:

- Sex
- Season
- Obesity
- Relapse
- D-dimer
- Age ≥ 8
- Clinic time
- Skin purpura
- Platelet (PLT)
- Creatinine (Cr)
- Abdominal pain
- Joint symptoms
- Recurrent purpura
- Triglycerides (TG)
- Cystatin-C (Cys-C)
- Fibrin Original (FIB)
- Thrombin time (TT)
- Blood albumin (ALB)
- Total cholesterol (TC)
- White blood cell (WBC)
- Neutrophil/lymph (N/L)
- Immunoglobulin A (IgA)
- Immunoglobulin E (IgE)
- Immunoglobulin G (IgG)
- Immunoglobulin M (IgM)
- Lymph/mononuclear (L/M)
- Digestive tract hemorrhage
- Blood Urea Nitrogen (BUN)
- Retinol-binding protein (RBP)
- Persistent purpura ≥ 4 weeks
- High-density lipoprotein (HLD)
- Low-density lipoprotein (LDL)
- Epstein-Barr virus DNA (EB-DNA)
- Neutrophil absolute value (NEUT#)
- Monocyte absolute value (MONO#)
- Erythrocyte sedimentation rate (ESR)
- Urinary albumin/creatinine ratio (U/C)
- Lymphocyte absolute value (LYMPH#)
- Hypersensitive c-reactive protein (CRP)
- Activated partial thromboplastin time (APTT)
- Anti-hemolytic streptococcus antibody (ASO)
- Mycoplasma pneumoniae infection–antibody (MP-Ab)

Clinical time refers to the first time to visit the hospital; Persistent purpura usually refers to the rash lasting more than 1 month; Recurrent purpura refers to the recurrence of a typical purpura-like rash in groups (more than 3 times) after the previous rash has completely subsided; Recurrence refers to the recurrence of characteristic manifestations of IgAV in children diagnosed with IgAV at least 1 month after the disappearance of symptoms; Obesity refers to the body mass index of children exceeded the 95th of children of the same age percentile.

2.6. Statistical analysis

All the data in this article are processed and analyzed by Python 3.9 (Python Software Foundation, Beaverton, OR)[. The data with different dimensions may vary greatly, and some of them are negative, which would affect the subsequent feature selection and machine learning. In order to solve the problem caused by dimensional disunity, it is necessary to standardize the data. This article uses the max-min method, which has the advantage of improving the degree of normalization of the model. Missing values may cause confusion in model fitting, resulting the output values unreliable, and features with missing values exceeding 60% contain little availability information, so the features were deleted directly. Because some features in the data have outliers, in order to avoid affecting the overall effect of the model, the remaining missing values were filled with median.[10] Univariate feature selection was used to screen out some possible irrelevant variables. χ^2 test was used for discrete features, and correlation analysis was used for continuous features. Among the results of this feature selection, $P < .05$ was considered statistically significant.

2.7. Machine learning methods

Since the development of machine learning algorithms, the classical classification algorithms mainly include logistic regression, decision tree, support vector machine (SVM), Naive Bayesian, K-nearest neighbor algorithm, and integrated learning algorithms such as AdaBoost, random forest (RF), Gradient Boosting Decision Tree (GBDT) and eXtreme Gradient Boosting (XGBoost) algorithm, etc. This article mainly uses logistic regression, SVM,

decision tree and RF algorithm to conduct modeling analysis. RF is a Bagging algorithm based on decision tree. The traditional decision tree chooses one optimal attribute each time when selecting partition attributes, but RF introduces random attribute selection in the training process of decision tree.[11] When building a decision tree in RF and selecting the partitioning attribute of a node, first a subset of K attributes from the attribute set of the node was selected, and then an optimal attribute from this subset was selected for partition. If k = n (where n is the number of attributes of the current node), the construction of a decision tree in RF is the same as that of a traditional decision tree. If k = 1, an attribute is randomly selected for partitioning. The adopted model evaluation indicators include the precision rate, accuracy rate, recall rate, F1 Score, receiver operating characteristic curve (ROC) and area under curve (AUC).

The whole model construction process of risk prediction and feature analysis of renal damage in children with IgAV based on machine learning algorithm is shown in Figure 1.

3. Results

3.1. Basic information

A total of 303 cases were included in this study, 15 cases were screened out, and 288 cases finally met the requirements. After 6 months of follow-up, 174 cases (60.42%) had no renal damage and 114 cases (39.58%) had renal damage.

3.2. Univariate feature selection of renal damage

Univariate feature selection was carried out on the 288 samples collected, and some possible irrelevant variables were preliminarily screened out. The results of Tables 1 and 2 showed that the 11 features - Obesity, Clinic time, Recurrent purpura, 'Persistent purpura ≥4 weeks, EB-DNA, WBC, ALB, Cr, TG, TC had significant impacts on the occurrence of renal damage ($P < .05$).

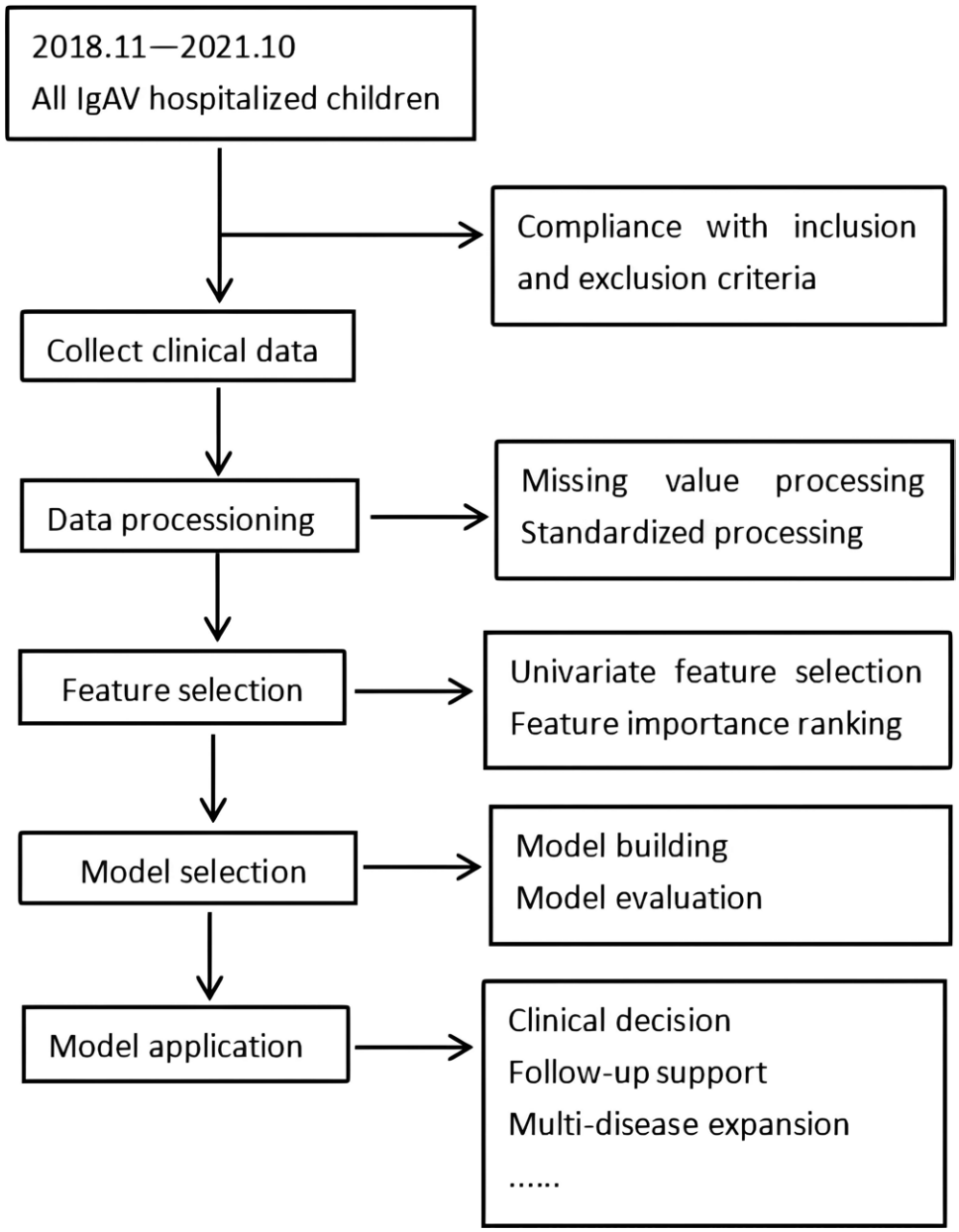

Figure 1. Data processing flow chart.

Table 1

χ^2 and *P* test results for univariate features.

Feature	χ^2	*P*
Persistent purpura ≥4 weeks	47.433757	5.689444e-12
Relapse	37.267170	1.030041e-09
Recurrent purpura	23.280931	1.399810e-06
Clinic time	14.328748	1.535027e-04
Obesity	8.741682	3.110175e-03
Epstein-Barr virus DNA	7.632975	5.731069e-03
Joint symptoms	1.724513	1.891131e-01
Age	0.921619	3.370503e-01
Sex	0.526040	4.682765e-01
Mp-Ab	0.444646	5.048885e-01
Season	0.180479	6.709618e-01
Digestive tract hemorrhage	0.164909	6.846765e-01
Skin purpura	0.149754	6.987708e-01
Abdominal pain	0.130076	7.183537e-01

Mp-Ab = Mycoplasma pneumoniae infection–antibody.

Table 2

Correlation analysis results of univariate features.

Feature	*F*	*P*
Creatinine	16.933456	.000051
Blood albumin	8.532509	.003767
Triglycerides	7.622047	.006139
White blood cell	4.884702	.027888
Total cholesterol	4.064121	.044738
Blood urea nitrogen	3.532561	.061191
Monocyte absolute value	3.296978	.070455
Low-density lipoprotein	2.572281	.109855
APTT	2.504080	.114657
Urinary albumin/creatinine ratio	2.303150	.130216
Thrombin time	2.072313	.151088
Platelet	1.973516	.161160
High-density lipoprotein	1.899490	.169213
Retinol-binding protein	1.815747	.178887
Lymph/mononuclear	1.385707	.240110
Neutrophil/lymph	1.202218	.273801
Fibrin Original	0.824394	.364663
Immunoglobulin E	0.814235	.367631
Cystatin C	0.794239	.373571
Neutrophil absolute value	0.472176	.492545
CRP	0.284482	.594193
ESR	0.282116	.595730
Immunoglobulin M	0.179002	.672550
Immunoglobulin A	0.082517	.774123
Lymphocyte absolute value	0.066034	.797386
D-dimer	0.015206	.901945
ASO	0.011321	.915341
Immunoglobulin G	0.001387	.970321

ASO = anti-hemolytic streptococcus antibody, APTT = activated partial thromboplastin time, CRP = hypersensitive c-reactive protein, ESR = erythrocyte sedimentation rate.

3.3. Ranking of feature importance in RFs

The 11 features obtained from univariate feature selection were ranked according to the features importance provided by the RF model, as shown in Figure 2. By setting appropriate threshold, 10, 8 and 5 features were screened as independent variables for model input. As shown in Figure 3, the AUC of the model is 0.916, 0.880 and 0.886, respectively. Therefore, the performance of the model is best when the feature with the top ten RF importance score is used as the model input. According to Figure 2, the top 10 features are: persistent purpura ≥4 weeks, Cr, Clinic time, ALB, WBC, TC, Relapse, TG, Recurrent purpura, EB-DNA.

3.4. Model construction

In order to optimize the effect of the model, the super parameters need to be tuned during modeling. First, the number of decision trees (N_estimators) and the maximum depth of the decision tree (MAX_depth) are determined by grid search, and then the minimum number of samples needed for splitting (min_samples_split) and the minimum number of samples carried by leaf nodes (min_samples_leaf) are determined. It can be directly seen that the performance of the RF model is better, which has great advantages compared with the linear model, as shown in Table 3.

3.5. Model performance evaluation

Logistic regression, SVM, decision tree and RF algorithm were respectively used to construct the classification prediction model, and the precision rate, accuracy rate, recall rate and F1 Score of each model were calculated. As seen from Table 4 that the precision rate, accuracy rate, recall rate and F1 Score of the RF model are the highest, which are 0.83, 0.87, 0.86 and 0.85 respectively. It shows that RF model has better effect and its stability is better than the other three models.

ROC of the four models was drawn for comparison, as shown in Figure 4. It was found that the AUC of the RF model was 0.912, which was also significantly higher than the other three models, indicating that the classification accuracy of the RF model was greater, the classification effect was better, and it had a great generalization performance. Therefore, it is considered that the fusion based on the RF model has a better performance in the prediction of children's IgAV renal damage.

4. Discussion

Machine learning can effectively learn the characteristics of a large number of data, which provides new research ideas and methods for accurate prediction. Machine learning algorithms include conventional algorithms (K-nearest neighbor, decision tree, SVM, etc) and integrated algorithms (RF, XGBoost, limit tree, etc). In this study, logistic regression, SVM, decision tree and RF algorithm were used to construct the damage prediction model of children's IgAVN. Through the comparison of the precision rate, accuracy rate, recall rate and F1 value of each model, we can see that the RF model has a better effect, with values of 0.83, 0.87, 0.86 and 0.85 respectively, and its stability is better than the other three models.

The ROC of the four models was drawn for comparison, and it was found that the AUC of the RF model was 0.912, which was also significantly higher than that of the other three models, indicating that the classification of the RF model was more correct, had better classification effect, and had good generalization performance. RF is a collection of multiple decision trees, which can make up for the weak generalization ability of decision trees.[11] This method relies on computers to learn all the complex nonlinear interactions between variables by minimizing the errors between the observation and the predicted results.[12] With low computational overhead, it shows strong performance in many practical tasks.

IgA vasculitis is a systemic vasculitis mediated by immune complexes, which is a characteristic self-limited disease. Its pathogenesis is related to genetic, immune and other factors. Renal involvement is the key to determine its prognosis. Clinical judgment of renal involvement in children mainly depends on urine test, renal function test and kidney biopsy. However, due to the relatively high risk and low acceptance of kidney biopsy, and lag time of routine urine test, in recent years, a large number of scholars have devoted themselves to studying the high risk factors of IgAV renal damage and the methods of preventing renal damage. It mainly includes the analysis of the

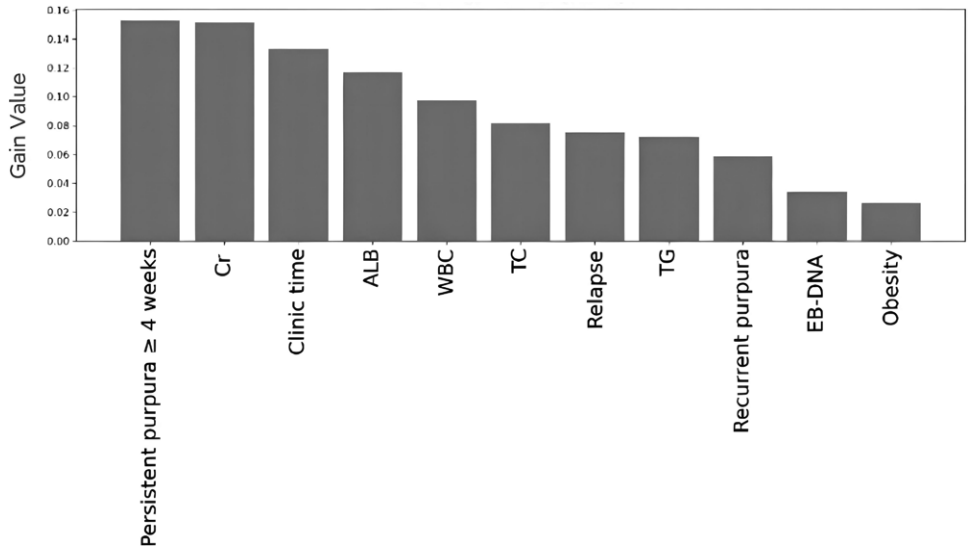

Figure 2. Feature importance of random forest. ALB = blood albumin, Cr = creatinine, EB-DNA = Epstein-Barr virus DNA, TC = total cholesterol, TG = triglycerides, WBC = white blood cell.

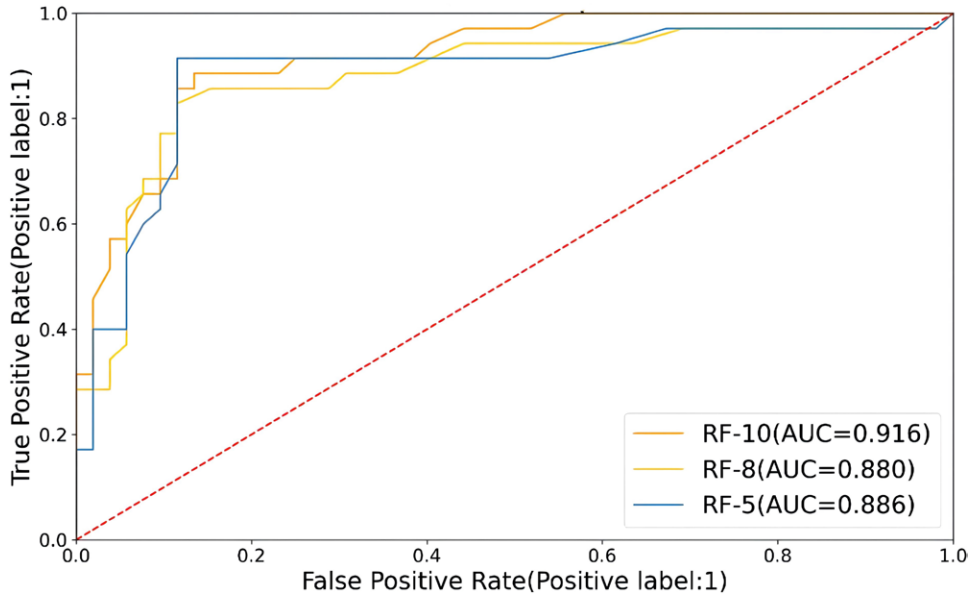

Figure 3. Receiver operating characteristic curve of random forest.

Table 3

Random forest parameter adjustment results.

Parameter name	Parameter adjustment range	Parameter adjustment result	Precision after parameter adjustment	AUC after parameter adjustment
N_estimators	[0, 200]	121	0.83	0.91
MAX_depth	[1, 25]	9		
Min_samples_split	[2, 22]	2		
Min_samples_leaf	[1, 11]	1		

AUC = area under curve.

Table 4

Comparison of model prediction results.

Model	Precision rate	Accuracy rate	Recall rate	F1-Score
Logistic regression	0.74	0.80	0.80	0.77
Support vector machine	0.70	0.78	0.80	0.75
Decision tree	0.74	0.78	0.80	0.73
Random forest	0.83	0.87	0.86	0.85

epidemiological characteristics, clinical manifestations, auxiliary examination, treatment and medication of the disease. This study is ranked according to the feature importance provided by the RF model, the top 10 features are: Persistent purpura ≥4 weeks, Cr, Clinic time, ALB, WBC, TC, Relapse, TG, Recurrent purpura, EB-DNA. These features may be important risk factors associated with IgAVN in children.

Skin purpura is the most common clinical manifestation of IgAV in children.[1] Studies have shown that about 78% to 100%

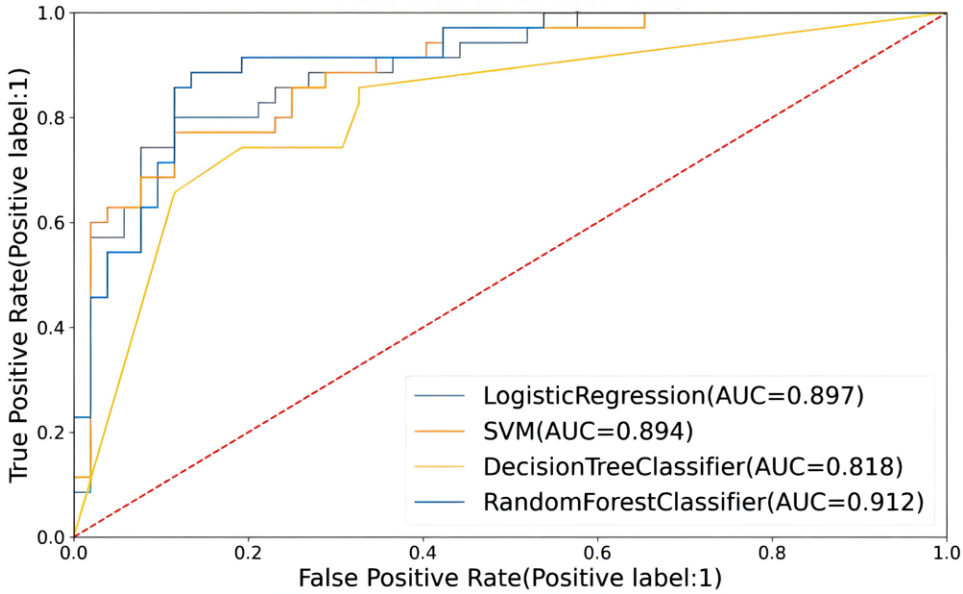

Figure 4. Receiver operating characteristic curve of four models. SVM = support vector machine.

of children are accompanied by skin purpura at the beginning of disease, and the accuracy rate of initial diagnosis is high.[13] Persistent purpura usually refers to the rash lasting more than 1 month. Recurrent purpura refers to the recurrence of a typical purpura-like rash in groups (more than 3 times) after the previous rash has completely subsided. Chan et al[14] found that the risk of renal damage in IgAV children with persistent purpura was 1.22-13.25 times higher than that in non-persistent purpura patients. Rigante et al[15] believed that persistent skin rash for more than one month was an important predictor of renal involvement and disease recurrence in children with IgAV. Ma et al[16] found that the recurrence of rash ≥3 times was a risk factor for renal involvement in children with IgAV. The reason may be the recurrent or persistent skin purpura, indicating that the recurrent and persistent presence of small vasculitis expands the inflammatory cascade reaction of the body, immune complex deposition and complement activation are widely active and persistent, and the renal capillaries are rich leading to renal involvement.

Serum creatinine is an important indicator of renal function, and the increase of serum creatinine caused by the decrease of creatinine clearance is sign of renal insufficiency. AlKhater et al[17] showed that elevated serum creatinine was related to renal damage in children with IgAV. Although in some reports, the average duration of renal disease is about one month after the onset of symptoms, and the risk can last up to six months after the initial symptoms of IgAV appear. This indicates the need to provide adequate follow-up and monitoring of patients to assess renal involvement. The results of this study showed that decreased serum albumin was one of the risk factors for renal involvement. This is related to the damage of the charge barrier of glomerular filtration membrane and the increase of permeability in children with IgAVN, which leads to albuminuria.

In recent years, studies have shown that elevated serum TC is more common in children with IgAV, especially those with renal damage. For example, Xu et al[18] showed that the age, creatinine and TC levels of children with IgAVN were higher than those of children with IgAV. Logistics multivariate analysis showed that TC level was one of the independent risk factors for IgAVN ($P < .05$), which was consistent with the study of Ma et al.[16]

Wang et al[19] studies have showed that patients with an interval of less than 4 days have a higher risk of developing kidney damage and severe kidney disease than patients with an interval of more than 8 days from the onset of symptoms to diagnosis. This risk factor has rarely been reported in previous studies. Thus, IgAV is a self-limited disease in most cases, but for a small number of patients, IgAV may not be self-limited and it will progress to renal involvement or severe renal disease. This finding is similar to the view of Davin et al[20]. These results suggest that early treatment and early diagnosis may be beneficial to children with IgAV. Recurrence refers to the recurrence of characteristic manifestations of IgAV in children diagnosed with IgAV at least 1 month after the disappearance of symptoms. Lei et al[21] defined the interval of recurrence as more than 3 months, including a total of 1002 children, of which 83.6% had one recurrence and 16.4% had more than 2 times of recurrence, and children with recurrence were more likely to have renal damage ($P < .05$).

As studies have shown, infection is the most common cause of IgAV, and about 40% to 70% of children are mainly affected by respiratory tract infection.[22,23] Ma et al[16] showed that the increase of WBC was one of the independent risk factors for IgAVN ($P < .05$). Chang et al[24,25] believed that the mechanism may be tissue damage caused by inflammatory mediators secreted by neutrophils, resulting in swelling and necrosis of renal vascular endothelium, while activated substances such as oxygen free radicals can chemotactic more WBC, aggravate vascular injury and form a vicious circle. EBV belongs to the γ subfamily of Herpesviridae, which is a linear double-strand DNA virus, and human body is its only natural host. It has been reported that viral infection is the etiology of various renal diseases. EBV infection can directly activate cellular and humoral immunity leading to EBV infection-related renal injury, and can also promote the formation of blood antigen-antibody complex, and settle on the renal vascular wall, causing damage to renal function.

This study has the following limitations: the collected cases are one-way retrospective study, the included sample size is limited and has not been externally verified, the results may be biased, and further multicenter large-sample prospective studies are needed for verification; the examination items of children were different, and some index features were omitted due to its absence, and the predictive variables may be left out.

To sum up, this study is based on clinical data, using machine learning algorithm to predict children's IgAVN,

aiming to intervene the possible clinical risk factors, to assist early clinical diagnosis and improve the prognosis of children, and to reduce the damage caused by invasive examination. Prospective intervention experiments can be carried out in the later stage to try to establish an early warning system for renal damage in children with IgAV in hospital, so as to conduct individualized treatment and prevention for patients. The combination of machine learning models and medical big data may provide new ways to predict the risk of children with IgAV.

5. Conclusion

In conclusion, persistent purpura ≥ 4 weeks, Cr, clinical time, ALB, etc are the risk factors of IgAVN, and RF has a good effect in predicting the risk of IgAVN. The RF algorithm performs with a high classification accuracy and prediction accuracy, and its theory and method research have been relatively mature. Therefore, it is recommended to use the RF algorithm in future research and application to predict IgAVN incidence risk.

Acknowledgements

Not applicable.

Author contributions

All authors contributed to the article and approved the submitted version. Jinjuan Wang conducted clinical data analysis and wrote a paper, Huimin Chu helped with clinical data collection, and Yueli Pan revised the paper.
Data curation: Huimin Chu.
Writing – original draft: Jinjuan Wang.
Writing – review & editing: Yueli Pan.

References

[1] Du L, Wang P, Liu C, et al. Multisystemic manifestations of IgA vasculitis. Clin Rheumatol. 2021;40:43–52.
[2] Ruperto N, Ozen S, Pistorio A, et al. EULAR/PRINTO/PRES criteria for Henoch-Schönlein purpura, childhood polyarteritis nodosa, childhood Wegener granulomatosis and childhood Takayasu arteritis: Ankara 2008. Part I: Overall methodology and clinical characterisation [published correction appears in Ann Rheum Dis. 2011 Feb;70(2):397. Paut, Isabelle Kone [corrected to Kone-Paut, Isabella]]. Ann Rheum Dis. 2010;69:790–7.
[3] Piram M, Maldini C, Biscardi S, et al. Incidence of IgA vasculitis in children estimated by four-source capture-recapture analysis: a population-based study. Rheumatology (Oxford). 2017;56:1358–66.
[4] Kawasaki Y, Suyama K, Hashimoto K, et al. Methylprednisolone pulse plus mizoribine in children with Henoch-Schoenlein purpura nephritis. Clin Rheumatol. 2011;30:529–35.
[5] Bluman J, Goldman RD. Henoch-Schönlein purpura in children: limited benefit of corticosteroids. Can Fam Physician. 2014;60:1007–10.
[6] Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19:64.
[7] Ramroach S, Joshi A, John M. Optimisation of cancer classification by machine learning generates an enriched list of candidate drug targets and biomarkers. Mol Omics. 2020;16:113–25.
[8] Ozen S, Pistorio A, Iusan SM, et al. EULAR/PRINTO/PRES criteria for Henoch-Schönlein purpura, childhood polyarteritis nodosa, childhood Wegener granulomatosis and childhood Takayasu arteritis: Ankara 2008. Part II: final classification criteria. Ann Rheum Dis. 2010;69:798–806.
[9]Subspecialty group of renal diseases, the society of pediatrics, Chinese medical association. Zhonghua Er Ke Za Zhi. 2017;55:647–51.
[10] Dong X, Dan X, Yawen A, et al. Identifying sarcopenia in advanced non-small cell lung cancer patients using skeletal muscle CT radiomics and machine learning. Thorac Cancer. 2020;11:2650–9.
[11]Breiman L. Random forests. Mach Learn. 2001;45:5–32.
[12] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform. 2002;35:352–9.
[13] Ozen S. The spectrum of vasculitis in children. Best Pract Res Clin Rheumatol. 2002;16:411–25.
[14] Chan H, Tang YL, Lv XH, et al. Risk factors associated with renal involvement in childhood Henoch-Schönlein purpura: a meta-analysis. PLoS One. 2016;11:e0167346.
[15] Rigante D, Candelli M, Federico G, et al. Predictive factors of renal involvement or relapsing disease in children with Henoch-Schönlein purpura. Rheumatol Int. 2005;25:45–8.
[16] Ma DQ, Li Y, Han ZG, et al. Analysis on kidney injury-related clinical risk factors and evaluation on the therapeutic effects of hemoperfusion in children with Henoch-Schonlein purpura. Eur Rev Med Pharmacol Sci. 2017;21:3894–9.
[17] AlKhater SA, Al Moaigel HM. Clinical spectrum and outcome of immunoglobulin A vasculitis in children: a 10-year clinical study. Int J Clin Pract. 2021;75:e13930.
[18]Xu H, Li W, Mao JH, et al. Association between red blood cell distribution width and Henoch-Schonlein purpura nephritis. Medicine (Baltim). 2017;96:e7091.
[19]Wang K, Sun X, Cao Y, et al. Risk factors for renal involvement and severe kidney disease in 2731 Chinese children with Henoch-Schönlein purpura: a retrospective study. Medicine (Baltim). 2018;97:e12520.
[20] Davin JC, Coppo R. Henoch-Schönlein purpura nephritis in children. Nat Rev Nephrol. 2014;10:563–73.
[21] Lei WT, Tsai PL, Chu SH, et al. Incidence and risk factors for recurrent Henoch-Schönlein purpura in children from a 16-year nationwide database. Pediatr Rheumatol Online J. 2018;16:25.
[22] Kawashima N, Kawada JI, Nishikado Y, et al. Abnormal urinalysis on day 7 in patients with IgA vasculitis (Henoch-Schönlein purpura). Nagoya J Med Sci. 2016;78:359–68.
[23] Xu Y, Wang JJ, Liu FF, et al. Predisposing factors of childhood Henoch-Schönlein purpura in Anhui province, China. J Investig Med. 2019;67:771–8.
[24] Chang H, Cao Y, Lin YI, et al. Association between toll-like receptor 6 expression and auxiliary T cells in the peripheral blood of pediatric patients with allergic purpura. Exp Ther Med. 2015;10:1536–40.
[25] Moretti M, Lava SAG, Zgraggen L, et al. Acute kidney injury in symptomatic primary Epstein-Barr virus infectious mononucleosis: systematic review. J Clin Virol. 2017;91:12–7.