

Prediction of Renal Damage in Children with Henoch-Schönlein Purpura Based on Machine Learning

Jinjuan Wang

Shandong University of Traditional Chinese Medicine

Huimin Chu

Shandong University of Traditional Chinese Medicine

Yueli Pan (✉ pyl_sd@163.com)

Affiliated Hospital of Shandong University of Traditional Chinese Medicine

Research Article

Keywords: Henoch-Schönlein Purpura, Renal Impairment, Random Forest, Machine Learning, Prediction.

Posted Date: January 7th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1202084/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

This article is objected to explore the value of machine learning algorithm in predicting the risk of renal damage in children with Henoch-Schönlein Purpura, and to construct a predictive model of Henoch-Schönlein Purpura Nephritis in children and analyze the related risk factors of Henoch-Schönlein Purpura Nephritis in children.

Methods

Case data of 288 hospitalized children with Henoch-Schönlein Purpura 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, decision tree and random forest algorithm were used respectively for classification prediction. Last, the performance of four algorithms are compared using accuracy rate and recall rate.

Results

The accuracy rate, recall rate and AUC of the established random forest 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 support vector machine model; higher than 0.74, 0.80 and 0.81 of the decision tree model. The top 10 important features provided by random forest model are Persistent purpura \geq 4weeks, Cr, Clinic time, ALB, WBC, TC, TG, Relapse, TG, Recurrent purpura and EB-DNA.

Conclusion

The model based on random forest algorithm has better performance in the prediction of children with allergic purpura renal damage, indicated by better classification accuracy, better classification effect and better generalization performance.

Background

Henoch-Schönlein Purpura (HSP), also known as IgA vasculitis (IgAV), is one of the most common systemic vasculitis in children^[1]. Patients with HSP are often acute onset, and the incidence of HSP are about 3/100,000 ~27/100,000^[2-3]. The clinical features of HSP 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 HSP have involved to Henoch-Schönlein Purpura Nephritis (HSPN) in the first 1 to 2 months 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 HSP, and the early accurate diagnosis of HSPN 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 HSP based on machine learning algorithm, identify related risk factors, and provide help for the early diagnosis and individualized intervention treatment of children with HSPN.

Methods

Research Objects

This article selected 288 children with HSP 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 HSP group and HSPN group according to urine test results.

Diagnostic Criteria

The diagnosis of HSP is based on the 2010 EULAR/PRINTO/PRES diagnostic criteria for HSP^[8]. The diagnosis of HSP is defined as palpable rash 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 HSPN^[9] were proposed as: hematuria and/or proteinuria occurred within 6 months of the course of HSP. 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: 3 routine urine tests within 1 week qualitatively indicate positive urine protein, 24h quantitative urine protein > 150 mg or urine egg/creatinine (mg/mg) > 0.2, Urinary microalbumin was higher than normal for 3 times within 1 week.

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.

Exclusion Criteria

Patients are excluded if they had severe heart, liver, kidney, brain, immune system or other diseases or consumptive diseases, or do not cooperate with the observer.

Research factors

42 features are researched as the variables to be predicted: 'Sex', 'Age \geq 8', 'Clinic time', 'Season', 'Skin purpura', 'Recurrent purpura', 'Persistent purpura \geq 4 weeks', 'Abdominal pain', 'Digestive tract hemorrhage', 'Joint symptoms', 'Relapse', 'EB-DNA', 'MP-Ab', 'WBC', 'NEUT#', 'MONO#', 'N /L', 'L/M', 'PLT', 'CRP', 'ASO', 'ESR', 'ALB', 'BUN', 'Cr', 'U/C', 'Cys-C', 'RBP', 'TG', 'TC', 'HLD', 'LDL', 'APTT', 'FIB', 'TT', 'D-dimer', 'IgA', 'IgG', 'IgM', 'IgE'.

Statistical analysis

All the data in this article are processed and analyzed by Python3.9. 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, and the output values are unreliable, and features with missing values exceeding 60% contain little availability information, so they are 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 are filled with median^[10]. Univariate feature selection was used to screen out some possible irrelevant variables. test was used for discrete features, and correlation analysis was used for continuous features. Among the results of this feature selection, $P < 0.05$ was considered statistically significant.

Machine Learning Methods

Since the development of machine learning algorithms, the classical classification algorithms mainly include logistic regression, decision tree, support vector machine, Naive Bayesian, K-nearest neighbor algorithm, and integrated learning algorithms such as AdaBoost, Random Forest (RF), GBDT and XGBoost algorithm, etc. This article mainly uses Logistic regression, support vector machine, decision tree and random forest algorithm to conduct modeling analysis. Random Forest (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 we randomly select a subset of K attributes from the attribute set of the node, and then select an optimal attribute from this subset 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, and is usually recommended.

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

Results

Basic Information

A total of 288 samples were included in the study. After 6 months of follow-up, 174 cases (60.42%) had no renal damage and 114 cases (39.58%) had renal damage.

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 Table 1 and Table 2 showed that the 11 features of 'Obesity', 'Clinic time', 'Recurrent purpura', 'Persistent purpura \geq 4weeks', 'EB-DNA', 'WBC', 'ALB', 'Cr', 'TG', 'TC' had significant impact on the occurrence of renal damage ($P < 0.05$).

Table 1 Symptom features test results

feature		P
Persistent purpura \geq 4 weeks		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
EB-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

Table 2 Correlation analysis results of clinical features

Feature	F	P
Cr	16.933456	0.000051
ALB	8.532509	0.003767
TG	7.622047	0.006139
WBC	I'm sorry	0.027888
TC	4.064121	0.044738
BUN	3.532561	0.061191
MONO#	3.296978	0.070455
LDL	2.572281	0.109855
APTT	2.504080	0.114657
U/C	2.303150	0.130216
TT	2.072313	0.151088
PLT	1.973516	0.161160
HLD	1.899490	0.169213
RBP	1.815747	0.178887
L/M	1.385707	0.240110
N/L	1.202218	0.273801
FIB	0.824394	0.364663
IgE	0.814235	0.367631
CYS-C	0.794239	0.373571
NEUT#	0.472176	0.492545
CRP	0.284482	0.594193
ESR	0.282116	0.595730
IgM	0.179002	0.672550
IgA	0.082517	0.774123
LYMPH#	0.066034	0.797386
D-dimer	0.015206	0.901945
ASO	0.011321	0.915341
IgG	0.001387	0.970321

Ranking of feature importance in random forests

The 11 features obtained from univariate feature selection are ranked according to the features importance provided by the random forest 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 random forest importance score is used as the model input. According to Figure 2, the top 10 features are: Persistent purpura \geq 4weeks, Cr, Clinic time, ALB, WBC, TC, Relapse, TG, Recurrent purpura, EB-DNA.

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 random forest model is better, which has great advantages compared with the linear model, as shown in Table 3.

Table3 RF 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		

Model performance evaluation

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

Table4 Comparison of Model Prediction Results

Model	Precision	Accuracy	Recall	F1-Score
LogisticRegression	0.74	0.80	0.80	0.77
SVM	0.70	0.78	0.80	0.75
DecisionTree	0.74	0.78	0.80	0.73
RandomForest	0.83	0.87	0.86	0.85

ROC curves of the four models were drawn for comparison, as shown in Figure 4. It was found that the AUC of the random forest model was 0.912, which was also significantly higher than the other three models, indicating that the classification accuracy of the random forest 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 random forest model has a better performance in the prediction of children's HSP renal damage.

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, support vector machine, etc.) and integrated algorithms (random forest, XGBoost, limit tree, etc.). In this study, Logistic regression, support vector machine, decision tree and random forest algorithm were used to construct the damage prediction model of children's HSPN. Through the comparison of the precision rate, accuracy rate, recall rate and F1 value of each model, we can see that the random forest 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 curve of the four models was drawn for comparison, and it was found that the AUC of the random forest model was 0.912, which was also significantly higher than that of the other three models, indicating that the classification of the random forest model was more correct, had better classification effect, and had good generalization performance. Random forest is a collection of multiple decision trees, which can make up for the weak generalization ability of decision trees¹¹. 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.

Henoch-Schönlein Purpura 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 renal 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 HSP renal damage and the

methods of preventing renal damage. It mainly includes the analysis of the epidemiological characteristics, clinical manifestations, auxiliary examination, treatment and medication of the disease. This study is ranked according to the feature importance provided by the random forest model, the top 10 features are: Persistent purpura \geq 4weeks, Cr, Clinic time, ALB, WBC, TC, Relapse, TG, Recurrent purpura, EB-DNA. These features may be important risk factors associated with HSPN in children.

Skin purpura is the most common clinical manifestation of HSP in children^[13]. Studies have shown that about 78% ~ 100% of children are accompanied by skin purpura at the beginning of disease, and the accuracy rate of initial diagnosis is high^[14]. 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 H et al^[15] found that the risk of renal damage in HSP children with persistent purpura was 1.22-13.25 times higher than that in non-persistent purpura patients. Rigante D et al^[16] believed that persistent skin rash for more than one month was an important predictor of renal involvement and disease recurrence in children with HSP. Ma DQ et al^[17] found that the recurrence of rash \geq 3 times was a risk factor for renal involvement in children with HSP. 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, so renal involvement happens.

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. AlKhatir et al^[18] showed that elevated serum creatinine was related to renal damage in children with HSP. Although in some reports, the average duration of renal disease is about one month after the onset of symptoms, the risk can last up to six months after the initial symptoms of HSP appear. In the report of Gupta et al^[19], 57.8% of the patients developed renal symptoms within 4 weeks, 84.4% of the patients developed renal symptoms within 8 weeks, and the remaining patients continued to develop renal symptoms within 6 months after diagnosis. This indicates the need to provide adequate follow-up and monitoring of patients to assess renal involvement. It is generally believed that hypoproteinemia in nephropathy is caused by the loss of a large amount of protein from urine. The results of this study showed that decreased serum albumin was one of the risk factors for renal involvement, which was consistent with the study by Mao et al^[20]. This is related to the damage of the charge barrier of glomerular filtration membrane and the increase of permeability in children with HSPN, which leads to albuminuria.

In recent years, studies have shown that elevated serum total cholesterol is more common in children with HSP, especially those with renal damage. For example, Xu et al^[21] showed that the age, creatinine and TC levels of children with HSPN were higher than those of children with NHSPN. Logistics multivariate analysis showed that TC level was one of the independent risk factors for HSPN ($P < 0.05$), which was consistent with the study of Ma et al^[17].

Wang et al^[22] 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, HSP is a self-limited disease in most cases, but for a small number of patients, HSP 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^[23]. These results suggest that early treatment and early diagnosis may be beneficial to children with HSP. Recurrence refers to the recurrence of characteristic manifestations of HSP in children diagnosed with HSP at least 1 month after the disappearance of symptoms. Lei et al^[24] 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 ($P < 0.05$).

As studies have shown, infection is the most common cause of HSP, and about 40%-70% of children are mainly affected by respiratory tract infection^[25-26]. Ma et al^[17] showed that the increase of WBC was one of the independent risk factors for HSPN ($P < 0.05$). Chang et al^[27-28] 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 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^[29].

This study has the following limitations: (1) 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, (2) 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 HSPN, 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 HSP 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 Henoch-Schönlein Purpura.

Declarations

Ethics approval and consent to participate: 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.

Consent for publication: Not applicable.

Availability of data and materials: All data generated or analysed during this study are included in this published article and they are available from the corresponding author on reasonable request.

Competing interests: The authors declare that they have no competing interests.

Funding: No Funding.

Authors' 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.

Acknowledgements: Not applicable.

References

1. Jennette JC, Falk RJ, Bacon PA, et al. 2012 revised international chapel hill consensus conference nomenclature of vasculitides. *Arthritis Rheum.* 2013, 65 (1):1-11.
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. *Ann Rheum Dis.* 2010, 69 (5):790-797.
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(8):1358-1366.
4. Kawasaki Y, Suyama K, Hashimoto K, et al. Methylprednisolone pulse plus mizoribine in children with Henoch-Schönlein Purpura nephritis. *Clin Rheumatol.* 2011, 30 (4): 529-535.
5. Bluman J, Goldman RD. Henoch-Schönlein Purpura in children: limited benefit of corticosteroids. *Can Fam Physician.* 2014, 60 (11): 1007-1010.
6. Dong JY, Wei WQ, Wu K, et al. The application of machine learning in depression. *Adv Psychol Sci.* 2020, 28 (2): 266-274.
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(2): 113-125.
8. Yang YH, Yu HH, Chiang BL. The diagnosis and classification of Henoch–Schönlein purpura: An updated review. *Autoimmunity Reviews.* 2014, 13 (4-5): 355-358.

9. Nephrology Group, Pediatrics Branch of Chinese Medical Association. Evidence-based Guidelines for diagnosis and Treatment of Purpura Nephritis (2016). *Chinese journal of pediatrics*, 2017, 55 (9): 647-651.
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 (9):2650-2659.
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 (5): 352-359.
13. Du L, Wang P, Liu C, Li S, Yue S, Yang Y. Multisystemic manifestations of IgA vasculitis. *Clin Rheumatol*. 2021, 40 (1):43-52.
14. Ozen S. The spectrum of vasculitis in children. *Clin Rheumatol*. 2002, 16:411-425.
15. 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(11):e0167346.
16. 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 (1):45-8.
17. 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 (17):3894-3899.
18. AlKhatir 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 (4):e13930.
19. Gupta A, Kumar A, Gupta A, et al. 365 profile of Henoch Schonlein purpura (HSP) nephritis: 23 years experience at a tertiary care centre in North India. *Lupus Sci Med*. 2017, 4:A161.
20. Mao Song, Zhang Jianhua. Risk factors for renal damage in Henoch-Schonlein purpura: a meta analysis. *Int J Clin Exp Med*. 2016, 9(2):3607-3613.
21. Xu H, Li W, Mao JH, et al. Association between red blood cell distribution width and Henoch-Schonlein purpura nephritis[J]. *Medicine (Baltimore)*. 2017, 96 (23):e7091
22. 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 (Baltimore)*. 2018, 97 (38):e12520.
23. Davin JC, Coppo R. Henoch-Schönlein purpura nephritis in children. *Nat Rev Nephrol*. 2014, 10:563-73.
24. 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 (1):25.
25. 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 (4):359-368.

26. Xu Y, Wang JJ, Liu FF, et al. Predisposing factors of childhood Henoch-Schönlein purpura in Anhui province, China. *J Investig Med*. 2019,67 (4):771-778.
27. 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 . *ExpTher Med*. 2015,10 (4):1536-1540.
28. Rigante D, Candelli M, Federico G, et al. Predictive factors of renal involve-ment or relapsing disease in children with Henoch-Schonlein purpura.*Rheumatol Int*. 2005, 25 (1):45-48.
29. Moretti M, Lava SAG, Zraggen L, et al. Acute kidney injury in symptomatic primary Epstein-Barr virus infectious mononucleosis: Systematic review. *J Clin Virol*. 2017,91:12-17.

Figures

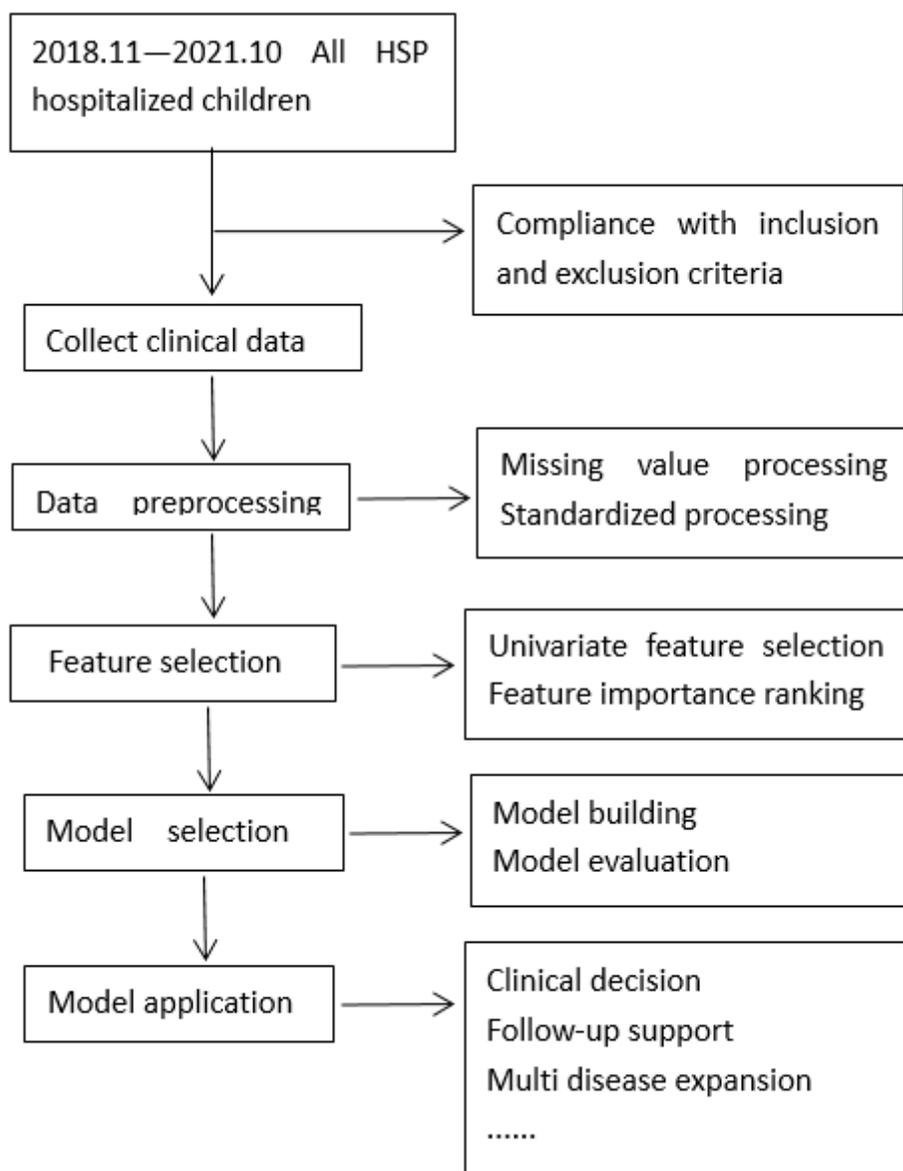


Figure 1

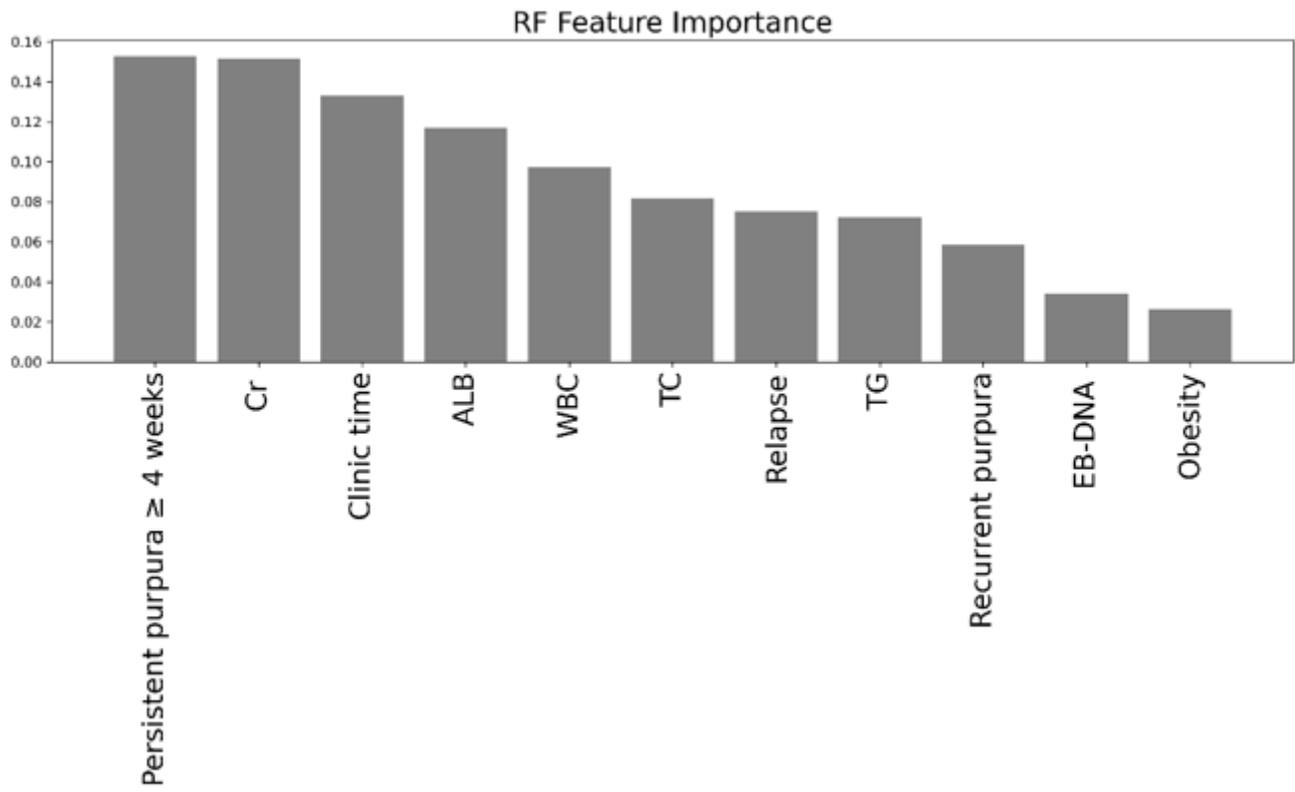


Figure 2

Feature Importance of RF

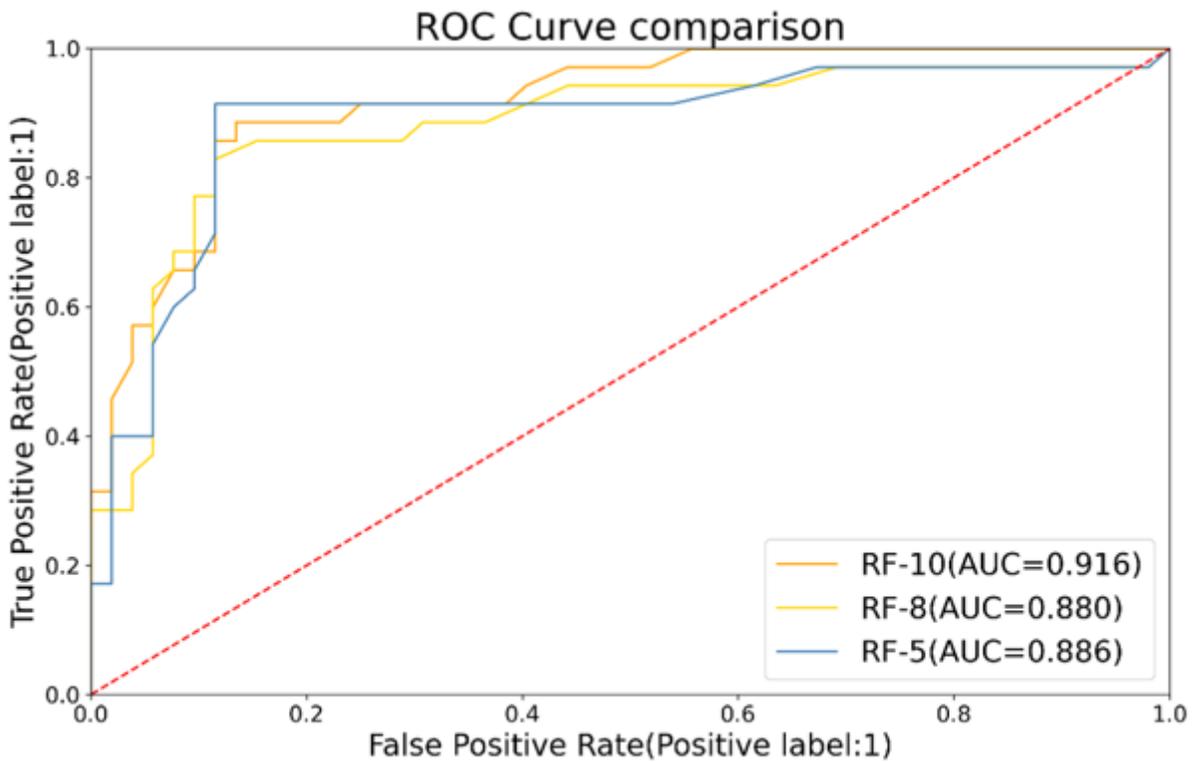


Figure 3

ROC Curve of RF

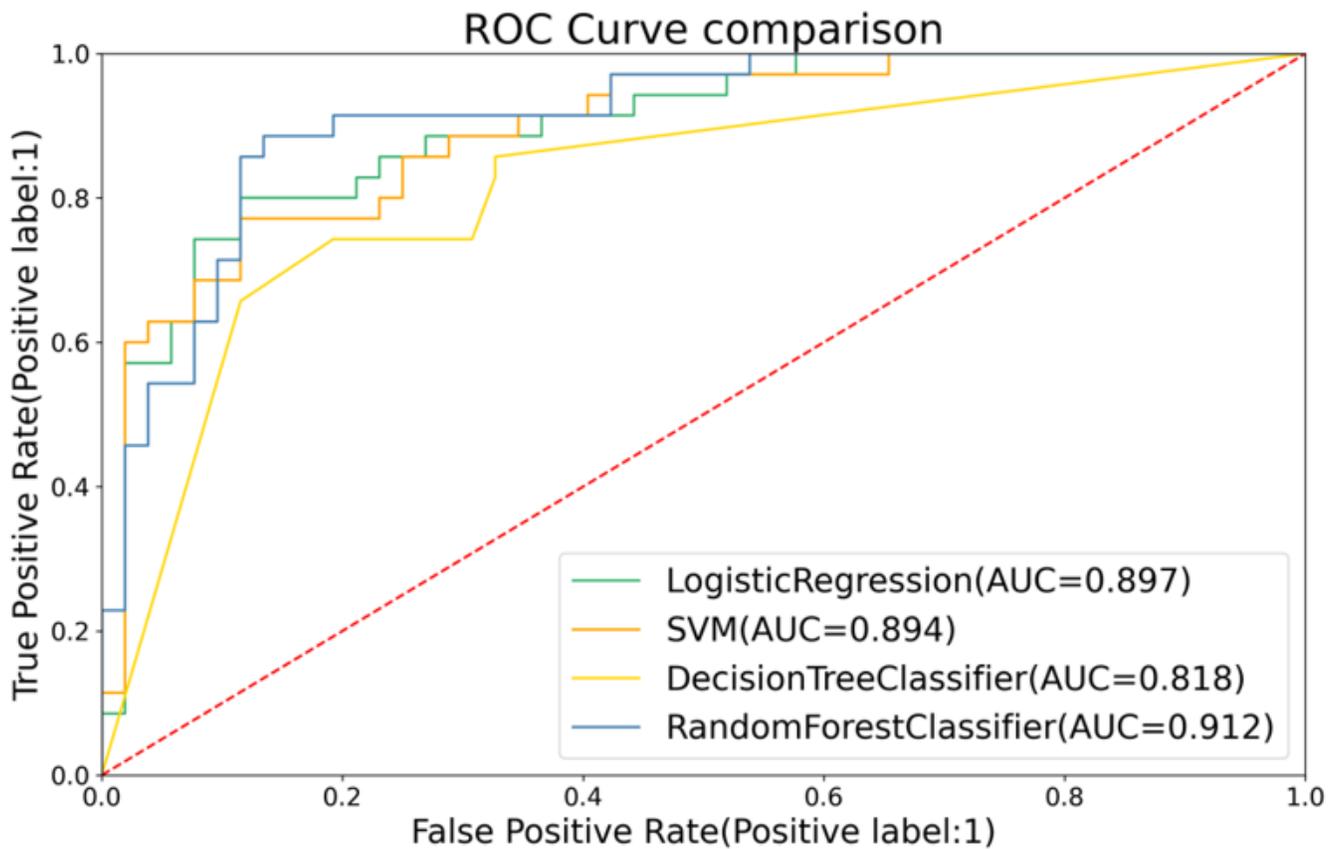


Figure 4

ROC Curve of four Models