Relapse Risk Prediction for Children with Henoch-Schönlein Purpura Based on GA-SVM

Yijun Liu[1], Beihong Wang[2*], Renpu Li[1], Sheng He[1], Haixu Xi[1], Ye Luo[1]

[1]*Key Laboratory of Cloud Computing*
and Intelligent Information Processing of Changzhou City
Jiangsu University of Technology
1801 Zhongwu Road, Changzhou 213001, China
E-mails: yijunliu@vip.sina.com, lrp0109@163.com, hs@jsut.edu.cn,
xihaixu@jsut.edu.cn, luoyejs@hotmail.com

[2]*Department of Pediatrics*
Changzhou No. 2 People's Hospital
29 Xinglong Road, Changzhou 213003, China
E-mail: wang_beihong@sina.com

[*]*Corresponding author*

Received: August 27, 2018 **Accepted: January 08, 2020**

Published: June 30, 2020

Abstract: *The relapse risk prediction for children with Henoch-Schönlein purpura can help pediatricians make an accurate prognosis and offer personalized and appropriate follow-up nursing and relapse control to patients. In this study, we propose a Genetic algorithm-Support vector machine (GA-SVM) learning method combining the support vector machine with the genetic algorithm for parameter optimization to capture the nonlinear mapping from a panel of biomarkers to the relapse risk of HSP children. The GA-SVM prediction model is created by using the dataset of 40 samples in clinical treatment and observation of patients. The inputs of the model consist of 19 biomarkers including gender, age, immunoglobulin M, immunoglobulin G, immunoglobulin A, prothrombin time, etc. The outputs consist of 1 and -1, where 1 indicates high relapse risk and -1 indicates low relapse risk. For comparison, the GS-SVM prediction model based on parameter optimization of grid search is also created. The experimental results show that the GA-SVM prediction model has a high prediction accuracy of 90% and is strong in generalization ability. The GA-SVM model for predicting the relapse risk of HSP children is a promising decision support tool of clinical prognosis, which provides pediatricians with valuable assistance to offer rehabilitation treatment to patients.*

Keywords: *Henoch-Schönlein purpura, Support vector machine, Genetic algorithm, Prediction of relapse risk.*

Introduction

Henoch-Schönlein purpura, abbreviated as HSP, is also known as anaphylactoid purpura. It is one of the most common systemic small vessel vasculitis in the pediatric population [20, 25]. The clinical manifestations of skin rashes, petechiae and other symptoms are commonly seen [15, 24]. Although HSP in children tends to be self-limiting, sometimes renal insufficiency can develop in a few and even renal failure arises in some cases [13, 19]. In order to gain good effects of medical treatments, a prognosis is commonly made for predicting the likely or expected development of a disease, the potential for complications, the likelihood of survival, etc. To help children with HSP make a full recovery, it is necessary for pediatricians to assess and forecast the relapse risk of these HSP patients. The prediction model describing correlation of the relapse risk for HSP children and the influencing factors can be a prognosis

decision support tool to assess the relapse risk and help pediatricians offer personalized and valuable treatments and controls to patients.

Prediction includes classification and regression. The relapse risk prediction of HSP children is a classification problem in which the class set consists of risk ratings. Classification and regression are two tasks of great importance in the area of machine learning and have attracted lots of research interests. However, it is difficult to establish a precise mathematical classification model to capture the disease complexity and the process dynamics due to the reason that the relapse risk of HSP children is influenced by a series of diverse and complicated biomarkers such as immunoglobulin, blood coagulation, complement, etc. Support vector machine, abbreviated as SVM, which is a learning method based on statistical learning theory and structural risk minimization principle, is introduced by Vapnik and co-workers [5, 21]. Compared with neural networks which have been applied in medical and biological areas [12, 14, 26], SVM is expressed in a concise mathematical form due to less parameters and has overcome the disadvantages of local minimum.

The SVM method has gained wide applications in medicine and biology. Sady and Ribeiro [18] used features extracted from symbolic series and time-frequency indices of heart rate variability as inputs in SVM to predict death in patients with Chagas disease. Abut et al. [1] used SVM combined with feature selection to predict VO_{2max} which refers to the highest rate of oxygen consumption an individual can attain during exhaustive exercise. Geng et al. [6] proposed a novel classification method combined SVM with the Adaboost algorithm to handle the imbalance of positive and negative cases in microRNA recognition. Hendel et al. [8] proposed a hybrid approach of self-organizing map and multi-class SVM for mental tasks classification. Hu et al. [11] used SVM for modelling and development of medical information system in web network. Qiu et al. [16] applied SVM to identify and analyse crotonylation sites in histone.

In this study, we apply the genetic algorithm together with the grid search algorithm to parameter optimization of SVM to create prediction models of the relapse risk for HSP children. Firstly, a brief description of methodology is given. Then, the Genetic algorithm-Support vector machine (GA-SVM) model and the Grid-search-Support vector machine (GS-SVM) model for comparison are created. Lastly, we test prediction models and conclude our study with a brief discussion.

Methodology

Support vector machine

A detailed description of SVM can be found in literatures [3, 5, 17, 21]. Here a brief description is given. Consider a linearly separable two-class classification problem, in which X denotes an input space, Y denotes an output space and S denotes a training vectors.

$Y = \{1, -1\}$, $S = \{(x_i, y_i)\}$, $i = 1, 2, \ldots, n$,

where $x_i \in X$ and $y_i \in Y$.

The discriminant function in the high dimensional space that separates two different classes in the training vectors is as follows:

$$h(x) = w^{\mathrm{T}} x + b, \tag{1}$$

where w is the weight vector and b is the bias.

Let consider a non-linear separable two-class classification problem. Given the reasonable classification hyperplane equation, the training vectors can satisfy the requirement of $h(x) \geq 1$, subjected to the constraints given below:

$$y_i\left(w^{\mathrm{T}}x_i+b\right)-1+\xi_i \geq 0, \tag{2}$$

where ξ_i is a positive slack variable. The classification margin is $\frac{2}{\|w\|}$, which reaches the maximum when $\|w\|$ is the minimum value. The optimal hyperplane can be obtained by maximizing the following function:

$$\min L\left(w,\, b,\, a\right)=\frac{1}{2}w^{\mathrm{T}}w-\sum_{i=1}^{n}a_i\left[y_i\left(w^{\mathrm{T}}x_i+b\right)-1+\xi_i\right], \tag{3}$$

where a_i are Lagrange multipliers obtained by solving a quadratic program.

By differentiating L w.r.t. w and b and letting it be zero, and using a mapping function K to transform the training vectors into a higher dimensional space, Eq. (3) is altered as the following dual problem:

$$\max L_D(a)=\sum_{i=1}^{n}a_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_i a_j y_i y_j K\left(x_i, x_j\right),\ (i \neq j) \tag{4}$$

$$\mathrm{s.t.}\sum_{i=1}^{n}a_i y_i=0, \sum_{j=1}^{n}a_j y_j=0,\ 0 \leq a_i, a_j \leq C,\ i, j=1, 2, ..., n,$$

where the constant, $C > 0$, is the penalty parameter or regularization term that controls the penalty degree of the wrongly classified samples and hence maintains the trade-off between the model complexity and the empirical risk of the SVM model. Assuming that the optimal solution is a^*, the optimal classification function is as follows:

$$f\left(x\right)=\mathrm{sgn}\left(\sum_{i=1}^{n}a_i^* y_i K\left(x, x_i\right)+b^*\right), \tag{5}$$

where sgn(·) is the sign function and b^* is the classification threshold. The kernel function K in Eq. (5) makes highly intermeshed overlapping data points easy to be separated linearly in the new space. The commonly used kernel functions of SVM include the linear kernel, the polynomial kernel, the radial basis function and sigmoid function, etc. The RBF, i.e. the radial basis function, is gaining popularity due to promising empirical performance and few parameters to be tuned. In this study, the RBF whose form is shown in Eq. (6) is applied as the kernel function to set up the SVM prediction model:

$$K\left(x, x_i\right)=\exp\left(-\gamma\left|x-x_i\right|^2\right), \tag{6}$$

where γ is the parameter of the kernel. In the SVM method based on RBF, two important parameters of the penalty parameter C in Eq. (4) and the kernel parameter γ in Eq. (6) need to be determined.

Genetic algorithm for parameter optimization in SVM

The choice of the parameters of C and γ is highly application-dependent and it is a task of great importance in the RBF based SVM applications [3]. C determines the trade-off between the empirical risk and the generalization performance. The large value of C tends to decrease the empirical risk but leads to the phenomenon of over learning, and the small value of C gives the slight punishment for the empirical error and thus assures generalizability of the model. The parameter γ affects the sensitivity of the SVM model to the change of data. The small value of γ reduces the anti-noise capability of the model and the large value of γ leads to responding slowness of the model.

The parameter selection in SVM is essentially an optimization problem. Currently the parameter selection methods of SVM include empirical selection, grid search and heuristic search, etc. As an efficient and robust heuristic algorithm, the genetic algorithm has been successfully applied to parameter optimization and feature selection in SVM [2, 3, 7, 17, 22]. In this study the genetic algorithm is used to optimize the SVM parameters of C and γ.

In the genetic algorithm for the problem of parameter optimization in SVM, a population of individuals, i.e. candidate solutions, is evolved toward better individuals. The evolution starts from an initial population consisting of randomly generated individuals. In each generation, every individual is evaluated by the fitness function and the fitter individuals are stochastically selected from the present population. Then the new generation is formed by the genetic operators of crossover and mutation. The process is iterated until the termination condition is satisfied.

Initialization of the population

A certain number of initial individuals are constructed. Every individual is represented in the binary string which is the encoding form of SVM parameters of C and γ.

Fitness function

The fitness function is used to assess the quality of each individual and the fittest individuals are selected to represent the offsprings of the next generation. The fitness function is defined by cross-validation accuracy of the GA-SVM prediction model with the parameters encoded in the individual.

Selection, crossover and mutation

The selection method chosen is the roulette wheel selection method to define which individuals are to be selected for the next generation. The roulette wheel selection is also known as fitness proportionate selection. In this selection method, the candidate solutions with a higher fitness will be more likely to be selected and less likely to be eliminated. The crossover method chosen is the One-Cut-Point crossover, where two individuals exchange part of chromosomes on a randomly picked point between themselves under a certain probability. In the end of each generation individuals perform mutation under a certain probability. The randomly selected gene in the individual is altered.

The GA-SVM method

To predict the relapse risk of HSP children, the GA-SVM model should be obtained. The obtaining process of the GA-SVM prediction model is as follows:

1) The dataset of clinical cases of HSP in children is collected.
2) In the stage of data preprocessing, the data are normalized by [0, 1] and tagged by class labels manually. The relapse risk ratings of HSP children are simply separated into two categories of high risk represented by 1 and low risk represented by -1. Every HSP child's health condition within 6 months after treatment is obtained by the follow-up survey. The clinical records of relapse are labeled as the class of high risk and those of relapse-free are labeled as the class of low risk.
3) The GA-SVM model for predicting the relapse risk of HSP children is created by using the genetic algorithm to optimize the parameters of C and γ in SVM.
4) The GA-SVM model is validated. The optimized GA-SVM model is obtained if the model is proved to be strong in generalization ability. Otherwise the new GA-SVM classifier will continue to be trained.

The established GA-SVM model is used as a decision support tool for assessing the relapse risk of HSP children and providing valuable assistance to the pediatricians.

Model creation

Data set

The dataset of 40 pediatric HSP samples is provided by Department of Pediatrics at Changzhou No. 2 People's Hospital in China. The first 30 samples constitute the training dataset and the remaining constitutes the testing dataset.

The box plot and multi-dimensional visualization of the dataset are shown in Fig. 1 and Fig. 2, respectively.

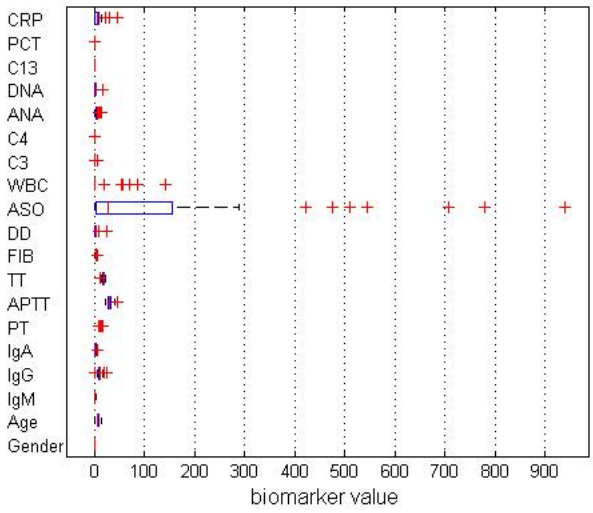


Fig. 1 The box plot of the dataset

A panel of biomarkers as attributes or variables is used to create prediction models. The outputs of the model consist of 1 and -1, where 1 indicates high relapse risk and -1 indicates low relapse risk. Purevdorj et al. [15] explored 12 serum biomarkers for laboratory diagnosis of pediatric HSP. Here the broader 19 biomarkers of HSP children are detected, including gender, age (years), immunoglobulin M (IgM), immunoglobulin G (IgG),

immunoglobulin A (IgA), prothrombin time (PT), activated partial thromboplastin time (APTT), thrombin time (TT), fibrinogen (FIB), D-Dimer (DD), anti-streptolysin O (ASO), white blood cells (WBC), complement component 3 (C3), complement component 4 (C4), antinuclear antibody (ANA), Anti DNA antibody (DNA), Carbon 13 breath test (C13), procalcitonin (PCT), and C-reactive protein (CRP). For the attribute of gender, 0 represents male and 1 represents female.

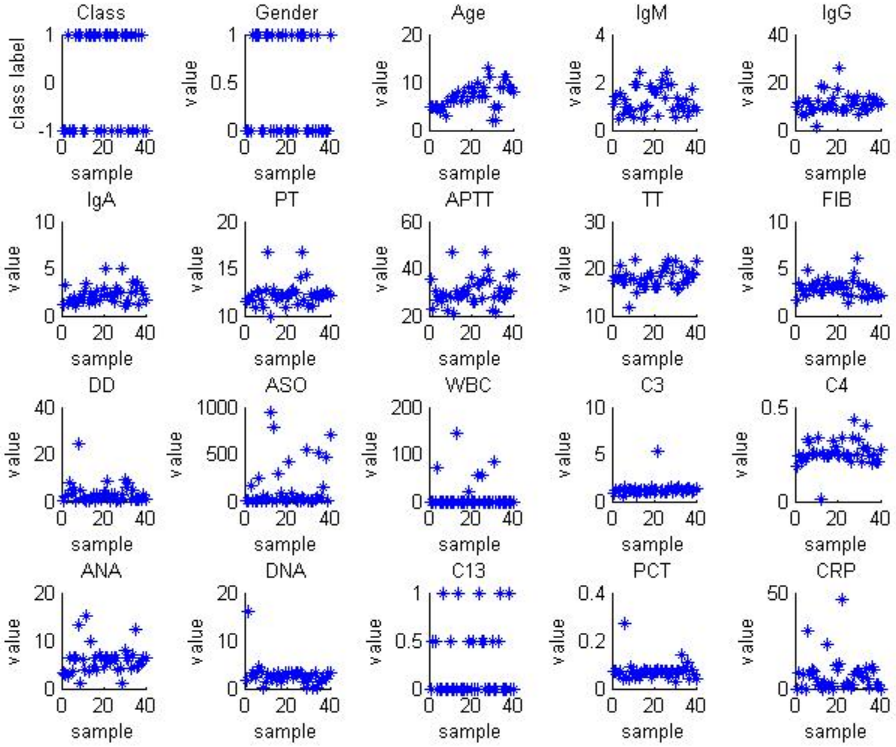


Fig. 2 Multi-dimensional visualization of the dataset

Model Training

GA-SVM model training

The proposed methodology is implemented using Matlab 2014 with an integrated and easy-to-use LIBSVM-FarutoUltimate version package [23]. This package is implemented based on the SVM Toolbox of LIBSVM [4] and the Genetic Algorithm Toolbox of GATBX [10].

The chromosome consists of 40 bits, and the numbers of bits used to represent C and γ values respectively are in the ratio of 1:1. According to [9], the value ranges of C and γ are set to $[2^{-5}, 2^{15}]$ and $[2^{-15}, 2^{3}]$, respectively. The initial population size is set to 100, the termination generation, i.e. number of iterations, is set to 200, and the generation gap is set to 0.9. The roulette wheel selection is used and the crossover probability is set to 0.7. The mutation probability is set to 0.0175 (0.7/40). The fitness of individual is defined as the accuracy of 5-fold cross validation, and the fitness curve is shown in Fig. 3.

When the parameter C is 216.9998 and γ is 6.4289, the best GA-SVM prediction model is obtained and its training accuracy of cross validation is 71.7949%.

GS-SVM model training

For comparison, the parameter optimization method of grid search is also used to establish the GS-SVM prediction model. The value ranges of C and γ are set to $\{2^{-5}, 2^{-4}, \ldots, 2^{14}, 2^{15}\}$ and

$\{2^{-15}, 2^{-14}, \ldots, 2^{2}, 2^{3}\}$, respectively according to [9] and the results of parameter optimization are shown in Fig. 4.

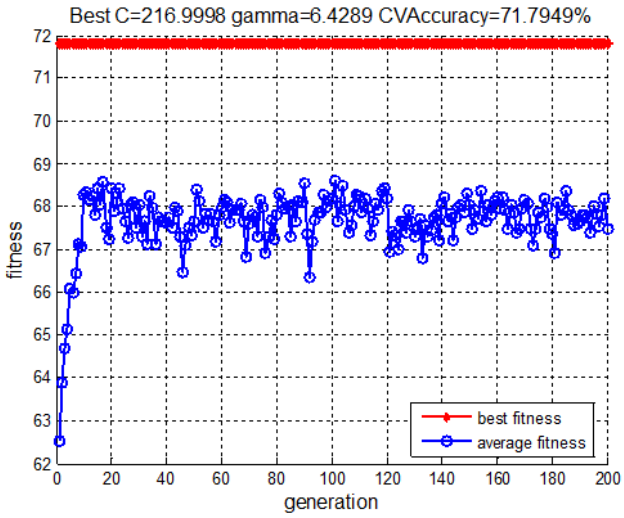


Fig. 3 The fitness curve of parameter optimization for GA-SVM

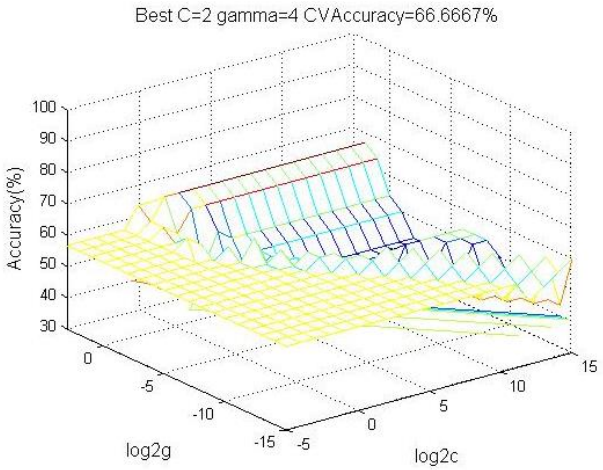


Fig. 4 The 3D visualization of parameter optimization for GS-SVM

The values of 2 and 4 are obtained for the parameters of C and γ respectively, and 66.6667% is obtained for the training accuracy of cross validation.

Model validation and discussion

Model validation

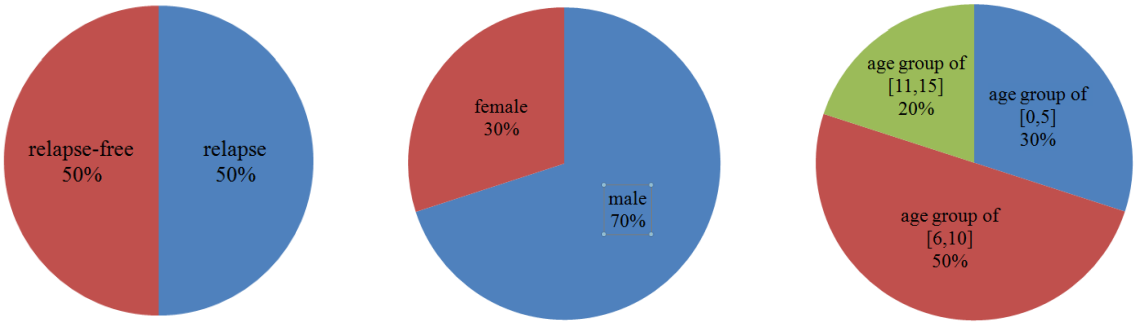
The created GA-SVM model together with the GS-SVM model is validated using the testing dataset of 10 samples as shown in Table 1.

The demographic data for the test subjects is shown in Fig. 5. Patients with HSP under the age of 16 are admitted to our department of pediatrics, and age distribution in the three age groups of [0, 5], [6, 10] and [11, 15] is shown in Fig. 5c.

The well-trained models of GA-SVM and GS-SVM are used to predict the relapse risk for HSP Children. The prediction results of the testing dataset are shown in Table 2.

Table 1. The testing dataset

No.	1	2	3	4	5	6	7	8	9	10
Gender	1	0	0	1	0	0	0	0	0	1
Age	5	2	5	9	9	11	11	9	9	8
IgM	1.35	0.98	1.02	0.93	1.27	0.61	0.8	1.73	0.97	0.85
IgG	9.33	10.69	8.99	8.51	14.22	10.2	13.11	12.76	10.39	10.64
IgA	1.31	1.41	2.4	3.55	2.67	3.61	1.26	2.98	2.29	1.59
PT	12.3	11.1	12	12.5	11.9	12.2	12.1	12.6	12.3	12.2
APTT	28.1	21.5	28.6	31.7	28	27.4	30.5	37	30.5	37.7
TT	15.4	16	20.2	17	19.4	17	18.2	19.2	18.7	21.6
FIB	2.06	3.06	3.29	1.76	2.74	2.18	2.9	2.28	1.94	2.07
DD	6.01	7.76	1.17	4.57	0.7	0.68	0.82	0.93	2.58	1.18
ASO	2	2.3	23.6	41.7	509.3	32.4	143.8	474.2	1	707.6
WBC	85	0	0	0	0	0	0	0	0	0
C3	1.33	1.44	1.2	1.07	1.23	1.27	1.37	1.05	1.22	1.29
C4	0.21	0.28	0.4	0.28	0.24	0.2	0.25	0.21	0.21	0.27
ANA	6.3	6.5	4.3	7	12.2	4.4	4.7	5.7	6.4	6.4
DNA	2.8	0.2	3.3	0.01	1.4	1.71	2.5	1.9	3.2	3.2
C13	0	0	0.5	1	0	0	0	1	0	0
PCT	0.07	0.07	0.14	0.07	0.09	0.11	0.04	0.08	0.06	0.04
CRP	6.5	6.2	10.9	5.2	0.4	11.1	9.9	1.6	2.2	0
Relapse risk class	1	-1	1	-1	1	-1	1	1	-1	-1



a) Relapse and relapse-free b) Gender c) Age (7.80±2.75)

Fig. 5 The demographic data for the test subjects

Table 2. The prediction results of the testing dataset

No.	1	2	3	4	5	6	7	8	9	10
Actual relapse risk class	1	-1	1	-1	1	-1	1	1	-1	-1
Predicted class with GA-SVM	1	-1	1	-1	1	-1	1	1	-1	1
Predicted class with GS-SVM	-1	1	1	1	1	-1	-1	1	-1	1

Discussion

Comparison of the GA-SVM model and the GS-SVM model

The prediction accuracies of the GA-SVM model and the GS-SVM model can be calculated from Table 2, and they are tabulated in Table 3. Compared with the GS-SVM model optimized by the grid search algorithm, the GA-SVM model optimized by the genetic algorithm gains strong generalization ability and has good prediction accuracy. The prediction accuracy of 90% is obtained for the GA-SVM model. It indicates that the relapse risk is largely determined by some biomarkers of HSP children, and the relapse risk prediction model for HSP children built by the GA-SVM algorithm based on available medical information can provide accurate prediction.

Table 3. Comparison of the GA-SVM model and the GS-SVM model

Prediction model	Parameter C	Parameter γ	Training accuracy	Prediction accuracy
GA-SVM model	216.9998	6.4289	71.7949%	90%
GS-SVM model	2	4	66.6667%	50%

The prediction accuracy and the training accuracy of the GA-SVM model are both higher than those of the GS-SVM model. The prediction accuracy the GA-SVM model is 40 percent more than that of the GS-SVM model. It indicates that the genetic algorithm performs better in parameter optimization of the SVM model for the relapse risk prediction. The parameters of C and γ are mutually independent and are in the ranges of $[2^{-5}, 2^{15}]$ and $[2^{-15}, 2^{3}]$ respectively in both of the GA optimization and the GS optimization. However, in the GS optimization the (C, γ) pair is limited to the form of $(2^a, 2^b)$ where a and b are integers, while in the GA optimization the parameters of C and γ are real numbers and are searched parallel and stochastically. Therefore the effective search space of the GA optimization is much larger than that of the GS optimization, and consequently the genetic algorithm has stronger optimization capability than the grid search algorithm.

Although it is possible that C is smaller than γ, generally good results are obtained when C is bigger than γ, e.g. C and γ in [2, 3, 7]. Similarly, the parameter C which is bigger than γ leads to the good model for predicting the relapse risk of HSP children in this study. The reason is that the too small parameter C leads to the too slight penalty and hence results in the under-fitting phenomenon.

Evaluation of the GA-SVM prediction model by using TPR, TNR, FPR and FNR

Four classifier indicators of TPR (True Positive Rate), TNR (True Negative Rate), FPR (False Positive Rate) and FNR (False Negative Rate) are used to evaluate the GA-SVM prediction model, and the indicator values calculated from Table 2 are shown in Fig. 6.

The GA-SVM prediction model has a considerable merit of the high TPR. The TPR is also called sensitivity. The TPR value 100% and the FNR value 0% indicate that all cases with high relapse risk are correctly identified. By using the GA-SVM model with high sensitivity, pediatricians are capable of identifying HSP children with high relapse risk accurately, and then give special attention to them, e.g. arrange more periodic re-examinations for them, to reduce their relapse risk. The HSP children with high relapse risk are guaranteed to get sufficient rehabilitation care and help.

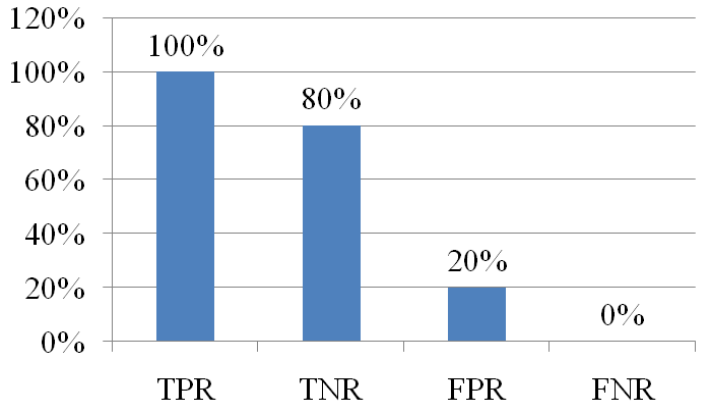


Fig. 6 TPR, TNR, FPR and FNR of the GA-SVM model

The GA-SVM prediction model has the weakness of the relatively low TNR. The TNR is also called specificity. The TNR value 80% indicates that the majority of the cases with low relapse risk, 80%, are identified correctly, and the FPR value 20% indicates that the minority of the cases with low relapse risk, 20%, are wrongly identified as the cases with high relapse risk. Therefore by using the GA-SVM model pediatricians possibly overestimate the relapse risk and then give the patient some unnecessary medical care.

Future directions of the study

There are three future directions of the study. The first direction would be the interpretation of the relapse risk process obtained from the established model with HSP medical experts' knowledge. The second direction would be the further construction of the biomarker index system. Deeper analysis for biomarker indexes is needed to select the most important variables to capture the most relevant information for the relapse risk prediction. The third direction would be the exploration of modeling and computation for the dataset of large-scale samples. More clinical samples of HSP children need to be collected to overcome some disadvantages of small-sample modeling, e.g. the relatively low specificity of the prediction model, in this study.

Conclusion

In this study, a machine learning technique SVM is introduced to the problem of the relapse risk prediction for HSP children in attempt to provide a model with good predictive power. The GA-SVM model is trained and applied to the prediction task. The experimental results show that the genetic algorithm has good optimization ability for the parameters of SVM and the GA-SVM model is impressive in the relapse risk prediction. The GA-SVM model can serve as a decision support tool of medical prognosis to help pediatricians assess the relapse risk and further take personalized relapse control measures and provide specialized care for HSP children.

Acknowledgements

This work is supported by Natural Science Foundation of Jiangsu Province in China (No. BK20161199). Furthermore, we are indebted to the supports and encouragements received from the staff and colleagues.

References

1. Abut F., M. F. Akay, J. George (2016). Developing New VO_{2max} Prediction Models from Maximal, Submaximal and Questionnaire Variables Using Support Vector Machines

Combined with Feature Selection, Computers in Biology and Medicine, 79, 182-192.
2. Almeida B. J. D, R. F. Neves, N. Horta (2018). Combining Support Vector Machine with Genetic Algorithms to Optimize Investments in Forex Markets with High Leverage, Applied Soft Computing, 64, 596-613.
3. Bian X. Q., B. Han, Z. M. Du, J. N. Jaubert, M. J. Li (2016). Integrating Support Vector Regression with Genetic Algorithm for CO_2-oil Minimum Miscibility Pressure (MMP) in Pure and Impure CO_2 Streams, Fuel, 182, 550-557.
4. Chang C. C., C. J. Lin (2011). LIBSVM: A Library for Support Vector Machines, ACM Transactions on Intelligent Systems and Technology, 2(3), 1-27.
5. Cortes C., V. Vapnik (1995). Support-vector Networks, Machine Leaming, 20, 273-297.
6. Geng X., Y. Q. Zhu, Z. Yang (2018). A Novel Classification Method for Class-imbalanced Data and Its Application in microRNA Recognition, International Journal Bioautomation, 22(2), 133-146.
7. Ghorbani M., G. Zargar, H. Jazayeri-Rad (2016). Prediction of Asphaltene Precipitation Using Support Vector Regression Tuned with Genetic Algorithms, Petroleum, 2(3), 301-306.
8. Hendel M., A. Benyettou, F. Hendel (2016). Hybrid Self Organizing Map and Probabilistic Quadratic Loss Multi-class Support Vector Machine for Mental Tasks Classification, Informatics in Medicine Unlocked, 4, 1-9.
9. Hsu C. W., C. C. Chang, C. J. Lin (2016). A Practical Guide to Support Vector Classification, https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf (Access date 25 June 2020)
10. https://codem.group.shef.ac.uk/current/index.php/ga-toolbox (Access date 25 June 2020)
11. Hu C. F., C. C. Ding, L. Dai (2017). Modeling and Development of Medical Information System Based on Support Vector Machine in Web Network, International Journal Bioautomation, 21(4), 283-292.
12. Isah O. R., A. D. Usman, A. M. S. Tekanyi (2017). A Hybrid Model of PSO Algorithm and Artificial Neural Network for Automatic Follicle Classification, International Journal Bioautomation, 21(1), 43-58.
13. Lee K. H., J. H. Park, D. H. Kim, J. Hwang, G. Lee, J. S. Hyun, S. T. Heo, J. H. Choi, M. Kim, M. Kim, S. Kim, M. Eisenhut, A. Kronbichler, J. Shin (2017). Dapsone as a Potential Treatment Option for Henoch-Schönlein Purpura (HSP), Medical Hypotheses, 108, 42-45.
14. Liu Y. J., B. H. Wang, J. L. Tang, M. F. Zhu, D. Chen, H. F. Jiang, X. J. Chen (2015). Prediction of Negative Conversion Days of Childhood Nephrotic Syndrome Based on the Improved Backpropagation Neural Network with Momentum, International Journal Bioautomation, 19(4), 543-554.
15. Purevdorj N., Y. Mu, Y. J. Gu, F. Zheng, R. Wang, J. W. Yu, X. G. Sun (2018). Clinical Significance of the Serum Biomarker Index Detection in Children with Henoch-Schönlein Purpura, Clinical Biochemistry, 52, 167-170.
16. Qiu W. R., B. Q. Sun, H. Tang, J. Huang, H. Lin (2017). Identify and Analysis Crotonylation Sites in Histone by Using Support Vector Machines, Artificial Intelligence in Medicine, 83, 75-81.
17. Raman M. R. G., N. Somu, K. Kirthivasan, R. Liscano, V. S. S. Sriram (2017). An Efficient Intrusion Detection System Based on Hypergraph-Genetic Algorithm for Parameter Optimization and Feature Selection in Support Vector Machine, Knowledge-Based Systems, 134, 1-12.
18. Sady C. C. R., A. L. P. Ribeiro (2016). Symbolic Features and Classification via Support Vector Machine for Predicting Death in Patients with Chagas Disease, Computers in Biology and Medicine, 70, 220-227.

19. Smith G. (2016). Management of Henoch-Schönlein Purpura, Pediatrics and Child Health, 26(8), 339-343.
20. Su Z. T., X. Lv, Y. Liu, J. H. Zhang, J. Y. Guan, Z. T. Gai (2016). Circulating Midkine in Children with Henoch-Schönlein Purpura: Clinical Implications, International Immunopharmacology, 39, 246-250.
21. Vapnik V. (1995). The Nature of Statistical Learning Theory, Springer-Verlag, New York.
22. Vijayanand R., D. Devaraj, B. Kannapiran (2018). Intrusion Detection System for Wireless Mesh Network Using Multiple Support Vector Machine Classifiers with Genetic-algorithm-based feature Selection, Computers & Security, 77, 304-314.
23. Wang X. C., F. Shi, L. Yu, Y. Li (2013). Analysis of 43 Neural Network Cases with MATLAB, Beihang University Press, Beijing.
24. Wang X. C., L. Zhang, Y. Wang, X. M. Liu, H. X. Zhang, Y. Liu, N. Shen, J. J. Yang, Zhongtao Gai (2018). Gut Microbiota Dysbiosis is Associated with Henoch-Schönlein Purpura in Children, International Immunopharmacology, 58, 1-8.
25. Woerner A., C. Rudin, C. Bonetto, C. Santuccio, S. Ozen, R. P. Wise, R. Chandler, J. Bonhoeffer (2017). IgA Vasculitis (Henoch-Schönlein): Case Definition and Guidelines for Data Collection, Analysis, and Presentation of Immunisation Safety Data, Vaccine, 35(11), 1559-1566.
26. Zhang H. Y. (2017). Study on Prediction of Grain Yield Based on Grey Theory and Fuzzy Neutral Network Model, International Journal Bioautomation, 21(4), 331-338.

Assoc. Prof. Yijun Liu, M.Sc.
E-mail: yijunliu@vip.sina.com

Yijun Liu was born in Changzhou, China in June 1978. She got her B.Sc. and M.Sc. degree in Computer Science and Technology from Nanjing University in 2000 and 2003, respectively. She is currently an Associate Professor at School of Computer Engineering of Jiangsu University of Technology in China. Her research interests include data mining and intelligent information system.

Beihong Wang, M.Sc.
E-mail: wang_beihong@sina.com

Beihong Wang was born in Changzhou, China in November 1978. She got her B.Sc. and M.Sc. degree in Pediatrics from Nanjing Medical University in 2001 and 2011, respectively. She is currently an Associate Senior Doctor at Department of Pediatrics of Changzhou No. 2 People's Hospital in China. Her research interests include Henoch-Schönlein purpura in childhood, childhood nephrotic syndrome and biomedical signal processing.

Prof. Renpu Li, Ph.D.
E-mail: lrp0109@163.com

Renpu Li was born in Zibo, China in January 1976. He got B.Sc. degree in Electrical Engineering from Shandong College of Science and Technology in 1997, M.Sc. degree in Computer Application Technology from Tianjin Polytechnic University in 2000, and Ph.D. degree in Management Science and Engineering from Tianjin University in 2003. Currently, he is a Professor at School of Computer Engineering of Jiangsu University of Technology in China. His research interests include data mining and big data application.

Assoc. Prof. Sheng He, M.Sc.
E-mail: hs@jsut.edu.cn

Sheng He was born in Anqing, China in November 1971. He got his B.Sc. degree in Physics from Fuyang Normal University in 1993 and M.Sc. degree in Software Engineering from University of Science and Technology of China in 2005, respectively. Currently, he is an Associate Professor at School of Computer Engineering of Jiangsu University of Technology in China. His research interests include data mining and big data application.

Assoc. Prof. Haixu Xi, M.Sc.
E-mail: xihaixu@jsut.edu.cn

Haixu Xi is an Associate Professor at School of Computer Engineering of Jiangsu University of Technology in China. He received his B.Sc. degree in Physics from Huaibei Normal University in 2003 and M.Sc. degree in Educational Technology from Nanjing Normal University in 2006. His current research interests include intelligent management systems, educational data mining and video intelligence processing.

Ye Luo, M.Sc.
E-mail: luoyejs@hotmail.com

Ye Luo is a lecturer at School of Computer Engineering of Jiangsu University of Technology in China. She received her B.Sc. degree in Computer Education from Nanjing Normal University in 1997 and M.Sc. degree in Computer Application Technology from Soochow University in 2008. Her current research interests include data mining and intelligent information system.