Preoperative prediction of complicated appendicitis using machine learning method

Chunbo Kang (chunbo_kang@163.com)
Affiliated Beijing Rehabilitation Hospital of Capital Medical University https://orcid.org/0000-0002-5238-8786

Xubin Li
Affiliated Beijing Rehabilitation Hospital of Capital Medical University

Xiaoqian Chi
Affiliated Beijing Rehabilitation of Capital Medical University

Yabin Yang
Affiliated Beijing Rehabilitation Hospital of Capital Medical University

Haifeng Shan
Affiliated Beijing Rehabilitation Hospital of Capital Medical University

Jie Zhang
Affiliated Beijing Rehabilitation Hospital of Capital Medical University

Qijun Zhang
Southern Medical University Nanfang Hospital

Research article

Keywords: complicated appendicitis, preoperative prediction, logistic regression (LR), support vector machine (SVM), random forest (RF)

Posted Date: May 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-27341/v1

Abstract

BACKGROUND Accurate preoperative prediction of complicated appendicitis (CA) could help selecting optimal treatment and reducing risks of postoperative complications. The study aimed to develop a machine learning model based on clinical symptoms and laboratory data for preoperatively predicting CA.

METHODS 136 patients with clinicopathological diagnosis of acute appendicitis were retrospectively included in the study. The dataset was randomly divided (94: 42) into training and testing set. Predictive models using individual and combined selected clinical and laboratory data features were built separately. Three combined models were constructed using logistic regression (LR), support vector machine (SVM) and random forest (RF) algorithms. The CA prediction performance was evaluated with Receiver Operating Characteristic (ROC) analysis, using the area under the curve (AUC), sensitivity, specificity and accuracy factors.

RESULTS The features of the abdominal pain time, nausea and vomiting, the highest temperature, high sensitivity-CRP (hs-CRP) and procalcitonin (PCT) had significant differences in the CA prediction ($P<0.001$). The ability to predict CA by individual feature was low ($AUC<0.8$). The prediction by combined features was significantly improved. The AUC of the three models (LR, SVM and RF) in the training set and the testing set were 0.805, 0.888, 0.908 and 0.794, 0.895, 0.761, respectively. The SVM-based model showed a better performance for CA prediction. RF had a higher AUC in the training set, but its poor efficiency in the testing set indicated a poor generalization ability.

CONCLUSIONS The SVM machine learning model applying clinical and laboratory data can well predict CA preoperatively which could assist diagnosis in resource limited settings.

1 Background

Acute appendicitis is the most common acute surgical abdominal disease, and its lifetime incidence is approximately 7–9% [1]. Despite the frequency of appendicitis, one of important aspect in modern diagnosis is stratification of simple and complex disease when acute appendicitis has been diagnosed. Histopathologic assays are the gold standard for the diagnosis of acute appendicitis, and the histopathological type of appendicitis determines the clinical management strategies. Acute appendicitis is divided into acute simple appendicitis, acute purulent appendicitis, acute gangrenous or perforated appendicitis, and periappendiceal abscess according to histopathology. The first 2 types of pathology have often been categorized as uncomplicated appendicitis (UA), and the latter two are called complicated appendicitis (CA). Acute appendicitis may present as either UA or CA, and these diseases develop differently according to epidemiology, immunology and histopathology [2, 3, 4].

Accurate differentiation between CA and UA determines the non-operative or operative management and following patient outcomes. For UA patients, it allows for delayed observation and decreased negative appendectomy rate and the risk of wrong diagnosis. While accurate preoperative diagnosis of CA patients

can help optimizing operative treatment, including surgical approach or administration of preoperative and duration of postoperative antibiotics [5]. Improving the differential diagnosis of simple and complex appendicitis can help reduce risk of postoperative complications, shorten recovery time and hospital stay, and avoid the medical burden of hospitals and patients [6, 7, 8].

Among preoperative diagnosis of suspected acute appendicitis, history, clinical manifestations, laboratory examination and medical imaging provide the main diagnostic basis. As popularized imaging tools for acute appendicitis, the sensitivity and specificity were reported to be over 90% for CT and over 80% for ultrasound [9, 10]. The sensitivity of noncontrast-enhanced MRI can also approach 77% [11]. With technological innovations, researchers or clinicians start to pay more attention in timely and accurately identifying pathological types of appendicitis before treatment. Both of CT and ultrasound have been applied to identify complicated appendicitis or predicting its pathological severity [12–14]. CT features including appendix diameter, dependent fluid, appendolithiasis were reported to be associated with appendicitis pathological severity [15] and ultrasound concordance with pathology was reported to be higher for perforated appendicitis as well [16].

However, except for the factors influencing the accuracy for pathology prediction of imaging based methodology (such as the localized inflammation, uncertain appendix position or difficulties in margin measurement), it also exist safety concerns about the use of radiation-based imaging or the availability challenges of skilled clinicians for ultrasound at any time. Therefore, the timely diagnosis of acute appendicitis or even its pathological type through clinical manifestations and laboratory examinations become challenging but necessary. Several scoring systems have been developed to help clinicians with diagnosis of acute appendicitis, including the Alvarado, the Raja Isteri Pengiran Anak Saleha Appendicitis (RIPASA) and the acute inflammatory response (AIR) scoring systems [17–21]. AIR is an index of acute inflammatory responses based on the Alvarado Score, which improves the accuracy of the diagnosis of acute appendicitis. It has also been reported that combined interpretation of white cell counts (WCC) or C-reactive protein (CRP) abnormal results could yield competitive sensitivity as CT in diagnosis of acute appendicitis [22]. However, the latest research results show that these scoring systems still face great challenges in discriminating between UA and CA [17, 21].

It is encouraging that there has been appearing studies involving laboratory factors such as CRP and neutrophil to differentiate simple appendicitis, cellulitis appendicitis and gangrenous appendicitis [23]. In addition, several studies have shown that artificial intelligence (AI) can significantly improve the accuracy and efficiency of prediction and initial diagnosis of acute appendicitis by combining factors derived from physical signs, symptoms and laboratory tests [24–28]. However, only a few studies presented the AI-based predictive models for pathological types of acute appendicitis [29]. The pathological types of acute appendicitis are of great value to clinical treatment decision-making. And AI-assisted tools may enhance the diagnostic confidence in resource limited settings where only conventional clinical methods are accessible. Therefore, based on our previous work [30], we aimed to use machine learning methods to establish an optimized model to preoperatively predict the UA and CA pathological types and enrich the preliminary data for AI-assisted appendicitis diagnosis.

2 Materials And Methods

2.1 Patients

This retrospective study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Aerospace Center Hospital, with the requirement for informed consent waived. During the period of June 2015 to November 2017, 138 patients with acute appendicitis who had underwent surgery and had clinical and laboratory data available, including inflammatory response and pathological diagnosis, were initially included.

The inclusion criteria were as follows: 1) patients had histologically confirmed acute appendicitis including UA and CA; 2) patients had data records including clinical manifestations (RIPASA scoring), laboratory tests before surgery (cell counts in full blood and inflammatory factors), which were listed in Table 1 and Table 2; 3) patients agreed to participate in the study and provided signed informed consent; 4) the preoperational examination indicated no surgical contraindications; and 4) the age range was 18 to 81 years old. The exclusion criteria for the current study were as follows: 1) patients did not meet the inclusion criteria; 2) patients with ileocecal neoplasms. Finally, 2 patients with mucinous adenocarcinoma were excluded, and 136 patients were enrolled in this study. Figure 1 depicts the patient selection process. The 136 patients (UA = 112; CA = 24) were divided at a ratio of ~ 7:3 into training set (n = 94, UA = 78, CA = 16) and validation cohort (n = 42, UA = 34, CA = 8) randomly.

Basic information, including age and sex, physical symptoms, clinicopathological data, and blood assay results before surgery, such as inflammation factors of high-sensitivity C-reactive protein (hs-CRP), procalcitonin (PCT), the lymphocyte subpopulations, were retrospectively extracted from electronic medical records.

2.2 Histopathology

All patients underwent surgical treatment, and all of the surgical specimens were examined by two pathologists. The pathological types of 136 cases of acute appendicitis were as follows: acute simple appendicitis (n = 9), acute purulent appendicitis (n = 103), acute gangrenous or perforated appendicitis (n = 24), and periappendiceal abscess (n = 0). The numbers of CA and UA cases were 24 and 112, respectively.

2.3 Feature selection and CA predictive modeling

In this study, the patients were divided into UA and CA groups according to histopathology. Univariate analysis was used to select the effective features among clinical and laboratory data, which have significant differences between UA and CA groups. CA predictive models based on individual clinical and laboratory data features and models combining clinical and laboratory data features were built separately. The CA prediction probability of individual clinical and laboratory features was identified by univariate logistic regression analysis. To construct predictive model based on combined features, the following three machine learning algorithms with high stability were investigated: logistic regression (LR),

support vector machine (SVM) and random forest (RF). The models were trained and assessed using the repeated ten-fold cross-validation method in the training set, and differentiation performance was evaluated with the testing set.

2.4 Validation of the prediction model

Univariate logistic regression analysis was used to assess the clinical and laboratory features in predicting CA. The diagnostic ability of the single and combined models was studied with Receiver Operating Characteristic (ROC). The CA prediction performance was assessed using the area under the curve (AUC) of ROC curve, sensitivity, specificity and accuracy (ACC). In addition, a nomogram was plotted to better express the predictive effect of logistic regression model. The statistical difference of AUC among the three machine learning models was analyzed. Decision curve analysis (DCA) was conducted to evaluate the clinical usefulness of best preoperative prediction model by quantifying the net benefits at different threshold probabilities in the testing set (31).

2.5 Statistical analysis

Comparisons of proportions and ranks of variables between training and testing set, and between UA and CA groups were performed using the Chi-square test, Fisher's exact test, Kruskal-Wallis H-test, Student's *t*-test or Mann-Whitney *U* test, as appropriate. The clinical and laboratory characteristics were compared using chi square test or Fisher's exact test for the nominal variable, Kruskal-Wallis H-test for the ordinal variable and Mann-Whitney *U* test for the continuous variable with abnormal distribution. Univariate logistic regression analysis was used to present prediction performance of individual clinical or laboratory feature. In addition, ROC curve analyses were performed to determine the AUC, ACC, sensitivity and specificity for each predictive model. The statistical difference of AUC between any two of the machine learning models was analyzed by Delong's test. DCA describe the clinical benefit of the predictive model as the difference between the true-positive and false-positive rates, weighted by the odds of the selected threshold probability of risk.

Statistical analysis was conducted with R software (Version: 3.6.0, https: www.r-project.org). The reported statistical significance levels were all two-sided, and the statistical significance was set at 0.05. The multivariate logistic regression and ROC analysis were performed with the 'stats', 'glmnet' and 'pROC' packages. The construction of the DCA and nomogram diagrams were performed using the 'rms' and 'rmda' packages.

3 Results

3.1 Statistical analysis and feature selection

The clinical and laboratory characteristics of patients in the training and testing sets are shown in Table 1. Demographic, clinical or laboratory features did not significantly differ between the training and testing sets. The statistical analysis of the clinical and laboratory data between UA and CA, including the

inflammatory response of the 136 patients, is presented in Table 2. The results of univariate analysis showed that nausea and vomiting, abdominal pain time, the highest temperature, hs-CRP and PCT were significantly different between UA and CA groups ($P < 0.001$) which associated with the appendicitis pathological type. The performance of these selected clinical and laboratory features in the diagnosis of CA is shown in Table 3. Although compared with other features, hs-CRP has a better efficiency in predicting CA pathology before operation as shown in nomogram in Fig. 2, the AUC of single feature prediction CA is below 0.80.

3.2 Prediction model assessment

By combining clinical and laboratory features, the CA predictive effectiveness of three different machine learning models (LR, SVM and RF) were summarized in Table 4. For logistic regression model, the nomogram in Fig. 2 also indicated that hs-CRP showed more superior than other features in CA prediction and the CA diagnostic probability would reach 80% when the total point was 200.

The ROC analysis results of three combined model (LR, SVM and RF) for the training and testing set are shown in Fig. 3A, 3B and 3C respectively. The AUC of the three models (LR, SVM and RF) in the training and testing sets were 0.805, 0.888, 0.908 and 0.794, 0.895, 0.761, respectively. The ACC of the three models (LR, SVM and RF) in the training and testing sets were 0.862, 0.905, 0.853 and 0.833, 0.805, 0.0.902, respectively. Although there was no significant difference among the three prediction models as shown in Table 5 ($P > 0.05$), the SVM model shows comprehensive advantages by considering accuracy, sensitivity, specificity and AUC (Table 4).

The DCA of the selected SVM model is shown in Fig. 3D. DCA showed that the SVM model had a higher overall net benefit when the threshold probability for a patient was within a range from 0.20 to 0.85. The SVM model showed a better performance in predicting CA pathology among three combined machine learning models.

4 Discussion

Diagnosis of acute appendicitis is a dynamic process that closely correlated with the pathological changes of the disease. Under resource limited conditions, sophisticatedly and fully considering of clinical symptoms, signs and blood biomarkers could potentially help medical practitioners in decision-making for acute appendicitis patients [32]. And the biomarkers of blood inflammation are of great value in the diagnosis and pathological classification of acute appendicitis [20, 21, 33]. AI-assisted diagnosis of acute appendicitis may help clinicians to make better decisions. While there is still a lack of the AI-based predictive models for pathological types of acute appendicitis constructed from general clinical examinations [29]. In the current study, the combined clinical symptoms, signs and laboratory data were used and three models were constructed by LR, SVM and RF algorithms respectively to preoperatively discriminate CA from UA.

In this study, univariate difference analysis was used to select features that are most correlated with the pathological classification of acute appendicitis from data pool including clinical signs and blood indexes, and the effect of each index on the diagnosis results was comprehensively considered. The factor of hs-CRP has a better efficiency in predicting CA pathology before operation as shown in nomogram, with AUC of 0.744(0.606–0.883) and 0.702(0.542–0.862) in the training and testing set respectively. This finding was in accordance with previous report in which C-reactive protein was a good predictor for complicated appendicitis [23]. However, comprehensively considering, the sensitivity and specificity of hs-CRP in the testing set of our study was weaker than this reported study (73.1% sensitivity and 89.5% specificity).

The further prediction model was built based on these selected features to avoid the influence of weak correlation features or possible overfitting on the prediction performance of the joint diagnosis model to a certain extent. Among three combined predictive models constructed respectively by LR, SVM and RF, the SVM model showed a better diagnostic performance. Comparing LR and RF with SVM, it is true that they have no significant difference in CA predictive efficiency. However, SVM is more interpretable, which is probably due to the reasons below.

Firstly, multivariable logistic regression is a classical linear modeling method to derive occurrence probability and influenced by all of the data points, which might be weak to reveal complicated nonlinear relationships and import bias resulted from the less correlated features. In comparison with logistic regression technique, the SVM using nonlinear kernel function can overcome these restrictions [28]. SVM classifier is weighted by data points which have stronger correlation with classification and shows a better generalization ability on the unseen data.

Secondly, SVM may have better classification performance for problems with small samples and high dimension [34]. It could be found that CA subjects occupied minority in our study. The RF model performed better with a 0.908 AUC for the training set but a sharply decreased AUC of 0.761 for the testing set, which might be induced by the small sample size and large depth of decision trees [35]. While the ensemble learning in RF model is based on the accuracy and significance of variables, which may guarantee the accuracy of the model.

The visualization charts derived from the AI-based analysis are easy for clinicians to make quick decisions, combining their clinical experience. However, the current study we have conducted shows several limitations. As can be seen from subject number, a small size of less than 100 was used to build the model through machine learning models. The efficiency of the classifier might not be generalized as a small size may not describe the entire data population. In addition, the performance of each machine learning method closely depends on the dataset properties. One selected algorithm may not always be the best. The choice of algorithm and the prediction procedure should be well interpretable. As the UA and CA distribution is imbalanced along the time line in the current study, splitting the data by period could not be conducted for further model validation. If the research is extended to multiple centers, the generalization and the accuracy of the prediction model might be improved. As the sample size is

enlarged, more features, including that not specific or strongly related to the conventional acute appendicitis diagnosis, should also be considered in the optimizing procedure of CA predicting model. It is helpful to establish the joint diagnosis model without omitting the potential but nonspecific features. And model established upon period splitting should also be considered to validate the feasibility and stability of the AI-based predicting methods.

The current research developed a preoperative prediction method for complicated appendicitis using a machine learning technique. The prediction model constructed by the SVM algorithm showed a better performance than the LR and RF algorithms. The predicted results displayed by the decision curve present good clinical practicability and universality.

5 Abbreviations

CA
Complicated appendicitis;
UA
Uncomplicated appendicitis;
LR
logistic regression;
SVM
support vector machine;
RF
random forest;
ROC
receiver operating characteristic curve;
AUC
area under the curve;
ACC
accuracy;
WBC
white blood cell;
NE
neutrophil;
NLR
blood neutrophil-to-lymphocyte ratio;
hs-CRP
high-sensitivity C-reactive protein;
PCT
procalcitonin;
RIPASA
Raja Isteri Pengiran Anak Saleha Appendicitis;

AIR
appendicitis inflammatory response.

Declarations

1. **Institutional addresses:**

Chun-Bo Kang, Xu-Bing Li, Xiao-Qian Chi, Ya-Bing Yang, Hai-Feng Shan, Jie Zhang, Qi-Jun Zhang, Department of General Surgery, Affiliated Beijing Rehabilitation Hospital of Capital Medical University, Beijing 100144, China

2. **Corresponding author:**

Chun-Bo Kang, Department of General Surgery, Affiliated Beijing Rehabilitation Hospital of Capital Medical University, Xixia Village, Eight Places in Shijingshan District, Beijing 100144, China.

Email: chunbo_kang@163.com

Telephone: 0086-13611104088

Fax: 0086-10-56981478

3. **Authors' contributions:**

Kang C. B. and Li X. B. designed the overall study; Kang C. B. and Li X. B. prepared the data; Chu X. Q. carried out the investigation; Kang C. B. and Yang Y. B. designed the methodology; Kang C. B., Shan H. F. and Zhang Q. J. performed the data analysis; and Kang C. B. and Li X. B. cowrote the manuscript.

4. **Ethics approval and consent to participate**

This study was reviewed and approved by the Ethics Committee of Aerospace Center Hospital. All patients signed informed consent before the operation.

5. **Consent for publication**

All patients or their caregivers signed a consent form giving permission to use their clinical data for research.

6. **Availability of data and materials**

The core data has been included in the manuscript. The datasets generated for this study are available on request to the corresponding author.

7. **Competing interests statement**

The authors declare that they have no competing interests.

8. **Funding**

The study is supported by the Funding Project of Aerospace Center Hospital [grant number: No. YN201429].

9. **Acknowledgments**

To the General Hospital of Aeronautics and Astronautics and Aerospace Center Hospital for providing strong support for the technology, funds and services.

References

1. Anderson JE, Bickler SW, Chang DC, Talamini MA. Examining a Common Disease with Unknown Etiology: Trends in Epidemiology and Surgical Management of Appendicitis in California, 1995–2009. World Journal of Surgery. 2012;36:2787–94.
2. Livingston EH, Woodward WA, Sarosi GA, Haley RW. Disconnect Between Incidence of Nonperforated and Perforated Appendicitis. Annals of Surgery. 2007;245:886–92.
3. Andersson RE. The Natural History and Traditional Management of Appendicitis Revisited: Spontaneous Resolution and Predominance of Prehospital Perforations Imply That a Correct Diagnosis is More Important Than an Early Diagnosis. World Journal of Surgery. 2007;31:86–92.
4. Rubér M, Andersson M, Petersson BF, Olaison G, Andersson RE, Ekerfelt C. Systemic Th17-like cytokine pattern in gangrenous appendicitis but not in phlegmonous appendicitis. Surgery. 2010;147:366–72.
5. Bhangu A, Søreide K, Di Saverio S, Assarsson JH, Drake FT. Acute appendicitis: modern understanding of pathogenesis, diagnosis, and management. The Lancet. 2015;386:1278–87.
6. Carr NJ. The pathology of acute appendicitis. Annals of Diagnostic Pathology. 2000;4:46–58.
7. Hall NJ, Eaton S, Abbo O, Arnaud AP, Beaudin M, Brindle M, Bütter A, Davies D, Jancelewicz T, Johnson K, Keijzer R, Lapidus-Krol E, Offringa M, Piché N, Rintala R, Skarsgard E, Svensson JF, Ungar WJ, Wester T, Willan AR, Zani A, St Peter SD, Pierro A. Appendectomy versus non-operative treatment for acute uncomplicated appendicitis in children: study protocol for a multicentre, open-label, non-inferiority, randomised controlled trial. *BMJ Paediatrics. Open*. 2017; 1(1).
8. Talan DA, Saltzman DJ, Mower WR, Krishnadasan A, Jude CM, Amii R, DeUgarte DA, Wu JX, Pathmarajah K, Morim A, Moran GJ, Olive View–UCLA Appendicitis Study Group. Antibiotics-first versus surgery for appendicitis: a US pilot randomized controlled trial allowing outpatient antibiotic management. Annals of emergency medicine. 2017;70:1–11.
9. Kotagal M, Richards MK, Flum DR, Acierno SP, Weinsheimer RL, Goldin AB. Use and accuracy of diagnostic imaging in the evaluation of pediatric appendicitis. Journal of pediatric surgery. 2015;50(4):642–6.

10. Terasawa T, Blackmore CC, Bent S, Kohlwes RJ. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. Ann Intern Med. 2004;141:537–46.
11. Harringa JB, Bracken RL, Davis JC, Mao L, Kitchin DR, Robbins JB, Ziemlewicz TJ, Pickhardt PJ, Reeder SB, Repplinger MD. Prospective evaluation of MRI compared with CT for the etiology of abdominal pain in emergency department patients with concern for appendicitis. J Magn Reson Imaging. 2019;50(5):1651–8.
12. Kitaoka K, Saito K, Tokuuye K. Important CT findings for prediction of severe appendicitis: involvement of retroperitoneal space. World journal of emergency surgery. 2014; 9(62): https://doi.org/10.1186/1749-7922-9-62.
13. Hansen AJ, Young SW, De Petris G, Tessier DJ, Hernandez JL, Johnson DJ. Histologic Severity of Appendicitis Can Be Predicted by Computed Tomography. Archives of surgery. 2004;139(12):1304–8.
14. Gonzalez DO, Lawrence AE, Cooper JN, Sola RJ, Garvey E, Weber BC, St Peter SD, Ostlie DJ, Kohler JE, Leys CM, Deans KJ, Minneci PC. Can ultrasound reliably identify complicated appendicitis in children? J Surg Res. 2018;229:76–81.
15. Hansen AJ, Young SW, De Petris G, Tessier DJ, Hernandez JL, Johnson DJ. Histologic severity of appendicitis can be predicted by computed tomography. Archives of surgery. 2004;139(12):1304–8.
16. Kotagal M, Richards MK, Flum DR, Acierno SP, Weinsheimer RL, Goldin AB. Use and accuracy of diagnostic imaging in the evaluation of pediatric appendicitis. Journal of pediatric surgery. 2015;50(4):642–6.
17. Unal Y. A new and early marker in the diagnosis of acute complicated appendicitis: immature granulocyes. Turkish Journal of Trauma Emergency Surgery. 2018;24:434–9.
18. Petroianu A. Diagnosis of acute appendicitis. International Journal of Surgery. 2012;10:115–9.
19. Resende F, Almeida AB, Maia JC, Melo RB. Challenges in uncomplicated acute appendicitis. Journal of Acute Disease. 2016;5:109–13.
20. Karami MY, Niakan H, Zadebagheri N, Mardani P, Shayan Z, Deilami I. Which one is Better? Comparison of the acute inflammatory response, Raja isteri Pengiran anak saleha appendicitis and Alvarado scoring systems. Annals of Coloproctology. 2017;33:277–231.
21. Acharya A, Markar SR, Ni M, Hanna GB. Biomarkers of acute appendicitis: systematic review and cost-benefit teade-off anaylsis. Surgical Endoscopy. 2017;31:1022–31.
22. Chen KC, Arad A, Chen KC, Storrar J, Christy AG. The clinical value of pathology tests and imaging study in the diagnosis of acute appendicitis. Postgraduate medicine journal. 2016;92(1092):611–9.
23. Qi FQ, Zhang B. Clinical significance of C-reactive protein levels in the determination of pathological type of acute appendicitis. International Journal of Clinical Experimental Medicine. 2015;8(8):13887–90.
24. Park SY, Kim SM. Acute appendicitis diagnosis using artificial neural networks. Technol Health Care. 2015;23:559–65.

25. Hsieh CH, Lu RH, Lee NH, Chiu WT, Hsu MH, Li YC. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. Surgery. 2010;149:87–93.
26. Sakai S, Kobayashi K, Nakamura J, Toyabe S, Akazawa K. Accuracy in the Diagnostic Prediction of Acute Appendicitis Based on the Bayesian Network Model. Methods of Information in Medicine. 2007;46:723–6.
27. Prabhudesai SG, Gould S, Rekhraj S, Tekkis PP, Glazer G, Ziprin P. Artificial Neural Networks: Useful Aid in Diagnosing Acute Appendicitis. World Journal of Surgery. 2008;32:305–9.
28. Shahmoradi L, Safdari R, Mirhosseini MM, Arji G, Jannat B, Abdar M. Predicting risk of acute appendicitis: A comparison of artificial neural network and logistic regression models. Acta Medica Iranica. 2018;56:784–95.
29. Josephine R, Alessandro R, Natalie K, Maximiliane IM, Jim K, Martina S, Marc R. Diagnosis and classification of pediatric acute appendicitis by artificial intelligence methods: An investigator-independent approach. PloS one. 2019;14:e0222030.
30. Kang CB, Li WQ, Zheng JW, Li XW, Lin DP, Chen XF, Wang DZ, Yao N, Liu XK, Qu J. Preoperative assessment of complicated appendicitis through stress reaction and clinical manifestations. Medicine. 2019;98:23–8.
31. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC medical informatics decision making. 2008;8:53–69.
32. Alvarado A. How to improve the clinical diagnosis of acute appendicitis in resource limited settings. World Journal of Emergency Surgery. 2016;11:16–9.
33. Yuki I, Toshiyuki I, Yuji T, Takahisa S, Satoshi I, Takashi U. Validity of predictive factors of acute complicated appendicitis. World Journal of Emergency Surgery. 2016;11:48–52.
34. You H, Ma Z, Tang Y, Wang Y, Yan J, Ni M, Cen K, Huang Q. Comparison of ANN (MLP), ANFIS, SVM, and RF models for the online classification of heating value of burning municipal solid waste in circulating fluidized bed incinerators. *Waste Management*. 2017; 68: S0956053 × 17301903.
35. Song Y, Gao S, Tan W, Qiu Z, Zhou H, Zhao Y. Multiple Machine Learnings Revealed Similar Predictive Accuracy for Prognosis of PNETs from the Surveillance, Epidemiology, and End Result Database. Journal of Cancer. 2018;9:3971–8.

Tables

Table 1 Statistical difference analysis of clinical and laboratory features between training and testing sets

	Name	Sample	Training	Testing	P-value
Clinical	Sex				0.288
	M	74	54(57.45%)	20(47.62%)	
	F	62	40(42.55%)	22(52.38%)	
	Age (years, Mean±SD)	136	37.00(28.00, 57.00)	35.50(27.00, 54.05)	0.649
	Abdominal pain score (1-10)				0.943
	3	6	2(4.76 %)	4(4.26 %)	
	4	10	3(7.14 %)	7(7.45 %)	
	5	16	5(11.90 %)	11(11.70 %)	
	6	27	7(16.67 %)	20(21.28 %)	
	7	37	13(30.95 %)	24(25.53 %)	
	8	16	5(11.90 %)	11(11.70 %)	
	9	20	6(14.29 %)	14(14.89 %)	
	10	4	1(2.38 %)	3(3.19 %)	
	Abdominal pain time (hours, Mean±SD)	136	24.00(12.00, 48.00)	24.00(14.00, 48.00)	0.266
	Tenderness range (1-10)				0.472
	1	29	17(18.09%)	12(28.57%)	
	2	18	11(11.70%)	7(16.67%)	
	3	53	38(40.43%)	15(35.71%)	
	4	5	3(3.19%)	2(4.76%)	
	5	1	1(1.06%)	0(0.00%)	
	6	1	0(0.00%)	1(2.38%)	
	7	2	2(2.13%)	0(0.00%)	
	8	5	4(4.26%)	1(2.38%)	
	9	22	18(19.15%)	4(9.52%)	
	Nausea and vomiting (0-2)				0.953
	0	38	27(28.72%)	11(26.19%)	
	1	66	45(47.87%)	21(50.00%)	
	2	32	22(23.40%)	10(23.81%)	
	The highest temperature, (°C, Mean±SD)	136	37.50(37.00, 38.10)	37.40(36.80, 37.90)	0.264
Labs	WBC‡ (10^9) (Mean±SD)	136	13.88(11.40, 16.26)	13.69(10.68, 16.47)	0.71
and AIR†	NE§% (Mean±SD)	136	86.80(81.40, 89.60)	86.50(79.98, 89.51)	0.619
	CD3+ (Mean±SD)	136	65.70(59.44, 71.01)	66.80(61.99, 71.31)	0.522
	CD4+ (Mean±SD)	136	33.55(27.79, 39.80)	38.65(28.00, 42.92)	0.054
	CD8+ (Mean±SD)	136	27.20(21.50, 32.41)	26.00(22.27, 32.04)	0.59
	CD19+ (Mean±SD)	136	16.60(12.90, 19.08)	16.80(12.37, 18.20)	0.912
	CD16+56 (Mean±SD)	136	13.75(8.70, 23.60)	14.05(8.47, 17.65)	0.451
	Total T cells (Mean±SD)	136	841.00(542.80, 1158.65)	907.50(609.05, 1122.70)	0.679
	T helper cells (Mean±SD)	136	369.00(263.65, 635.00)	494.00(279.40, 734.90)	0.399
	Inhibiting T cells	136	332.50(209.90, 474.55)	375.50(224.35, 449.20)	0.949
	B cells (Mean±SD)	136	218.50(144.30, 282.10)	194.00(136.25, 299.15)	0.772
	NK cells (Mean±SD)	136	174.50(111.00, 259.00)	173.50(84.00, 260.25)	0.582
	CD4/CD8 (Mean±SD)	136	1.27(0.93, 1.69)	1.46(1.07, 1.95)	0.089
	hs-CRP¶, mg/L (Mean±SD)	136	63.94(15.59, 77.89)	52.53(13.61, 98.41)	0.989
	PCT††, ng/mL (Mean±SD)	136	0.17(0.03, 2.77)	0.08(0.02, 0.89)	0.167

Note: SD, standard deviation; P-value < 0.05 indicated statistical significance.

Abbreviation: AIR†, acute inflammatory response; WBC‡, white blood cell; NE§, neutrophil; hs-CRP¶, high-sensitivity C-reactive protein; PCT††, procalcitonin.

Table 2 Statistical difference analysis of clinical and laboratory features between UA and CA groups

	Name	UA	CA	P value
Clinical	Sex			0.577
	M	61(54.5%)	13(54.2%)	
	F	51(45.5%)	11(45.8%)	
	Age (years, Mean±SD)	41.81±1.65	40.54±3.13	0.984
	Abdominal pain score (1-10)			0.415
	3	6(5.4%)		
	4	8(7.1%)	2(8.3%)	
	5	13(11.6%)	3(12.5%)	
	6	22(19.6%)	5(20.8%)	
	7	31(27.7%)	6(25%)	
	8	13(11.6%)	3(12.5%)	
	9	17(15.2%)	3(12.5%)	
	10	2(1.8%)	2(8.3%)	
	Abdominal pain time, h			
	Mean±SD	24.79±1.44	36.00±2.91	0.001**
	Tenderness range (1-10)			0.135
	1	23(20.6%)	4(16.7%)	
	2	15(13.4%)	3(12.5%)	
	3	49(43.7%)	5(20.8%)	
	4	3(2.7%)	2(8.3)	
	5	1(1%)	1(4.2%)	
	6-8	8(7%)		
	9	13(11.6%)	9(37.5%)	
	Nausea and vomiting (0-2)			0.009**
	0	31(27.7%)	7(29.2%)	
	1	60(53.6%)	6(25%)	
	2	21(18.7%)	11(45.8%)	
	The highest temperature (°C, Mean±SD)	37.44±0.08	38.04±0.19	0.002**
Labs and AIR †	WBC‡ (10^9) (Mean±SD)	13.64±0.40	15.26±0.68	0.081
	NE§% (Mean±SD)	84.40±0.69	87.09±0.94	0.102
	CD3+ (Mean±SD)	65.71±0.80	64.37±2.48	0.685
	CD4+ (Mean±SD)	34.83±0.77	31.70±1.48	0.062
	CD8+ (Mean±SD)	26.95±0.75	31.34±2.37	0.155
	CD19+ (Mean±SD)	17.41±0.68	17.98±1.50	0.760
	CD16+56 (Mean±SD)	15.26±0.79	13.72±1.91	0.333
	Total T cells (Mean±SD)	916.01±47.49	835.49±85.40	0.522
	T helper cells (Mean±SD)	496.63±27.92	396.52±43.66	0.136
	Inhibiting T cells	366.64±20.54	396.88±51.50	0.26
	B cells (Mean±SD)	230.84±12.12	203.17±18.38	0.311

NK cells (Mean±SD)	206.29±14.28	166.75±19.88	0.376
CD4/CD8 (Mean±SD)	1.47±0.07	1.22±0.132	0.058
hs-CRP¶, mg/L (Mean±SD)	54.62±4.91	103.85±14.79	0.000**
PCT††, ng/mL (Mean±SD)	1.36±0.43	4.46±1.88	0.006**

Note: UA, uncomplicated appendicitis; CA, complicated appendicitis; SD, standard deviation; P-value < 0.05 indicated statistical significance; ** $P < 0.05$.

Abbreviation: AIR†, acute inflammatory response; WBC‡, white blood cell; NE§, neutrophil; hs-CRP¶, high-sensitivity C-reactive protein; PCT††, procalcitonin.

Table 3 Performance of clinical and laboratory examinations in the diagnosis of complicated appendicitis

	Accuracy		Sensitivity		Specification		AUC§ (95% CI)	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing
Abdominal pain time (h)	0.723	0.524	0.75	1.00	0.718	0.412	0.728(0.590-0.865)	0.673(0.510-0.836)
Nausea andvomiting	0.787	0.667	0.500	0.375	0.846	0.735	0.608(0.428-0.788)	0.555(0.329-0.781)
The highest temperature (°C)	0.766	0.857	0.688	0.500	0.782	0.941	0.700(0.537-0.863)	0.697(0.460-0.933)
hs-CRP†, mg/L	0.734	0.619	0.688	1.00	0.744	0.529	0.744(0.606-0.883)	0.702(0.542-0.862)
PCT‡, ng/mL	0.755	0.833	0.625	0.375	0.782	0.941	0.682(0.532-0.832)	0.665(0.432-0.899)

Abbreviation: hs-CRP†, high-sensitivity C-reactive protein; PCT‡, procalcitonin; AUC§, area under the curve.

Table 4 Comparison of the predictive performance of three different machine learning models

Prediction Model	LR§		SVM‡		RF¶	
	Training	Testing	Training	Testing	Training	Testing
Accuracy	0.862	0.833	0.905	0.805	0.853	0.902
Sensitivity	0.625	0.750	0.974	0.912	0.882	0.571
Specificity	0.910	0.853	0.706	0.857	0.846	0.971
AUC†	0.805	0.794	0.888	0.895	0.908	0.761
(95% CI)	(0.678-0.932)	(0.600-0.988)	(0.794-0.982)	(0.746-1.00)	(0.837-0.979)	(0.525-0.996)

Abbreviation: AUC†, area under the curve; SVM‡, support vector machine; LR§, logistic regression; RF¶, random forest.

Table 5 Comparison of significant differences among three complicated appendicitis predictive models

	Training Set (P value)		Testing Set (P value)	
	Logistic Regression	SVM	Logistic Regression	SVM
SVM†	0.3		0.4	
Random Forest	0.2	0.6	0.8	0.1

Abbreviation: SVM†, support vector machine.

Figures

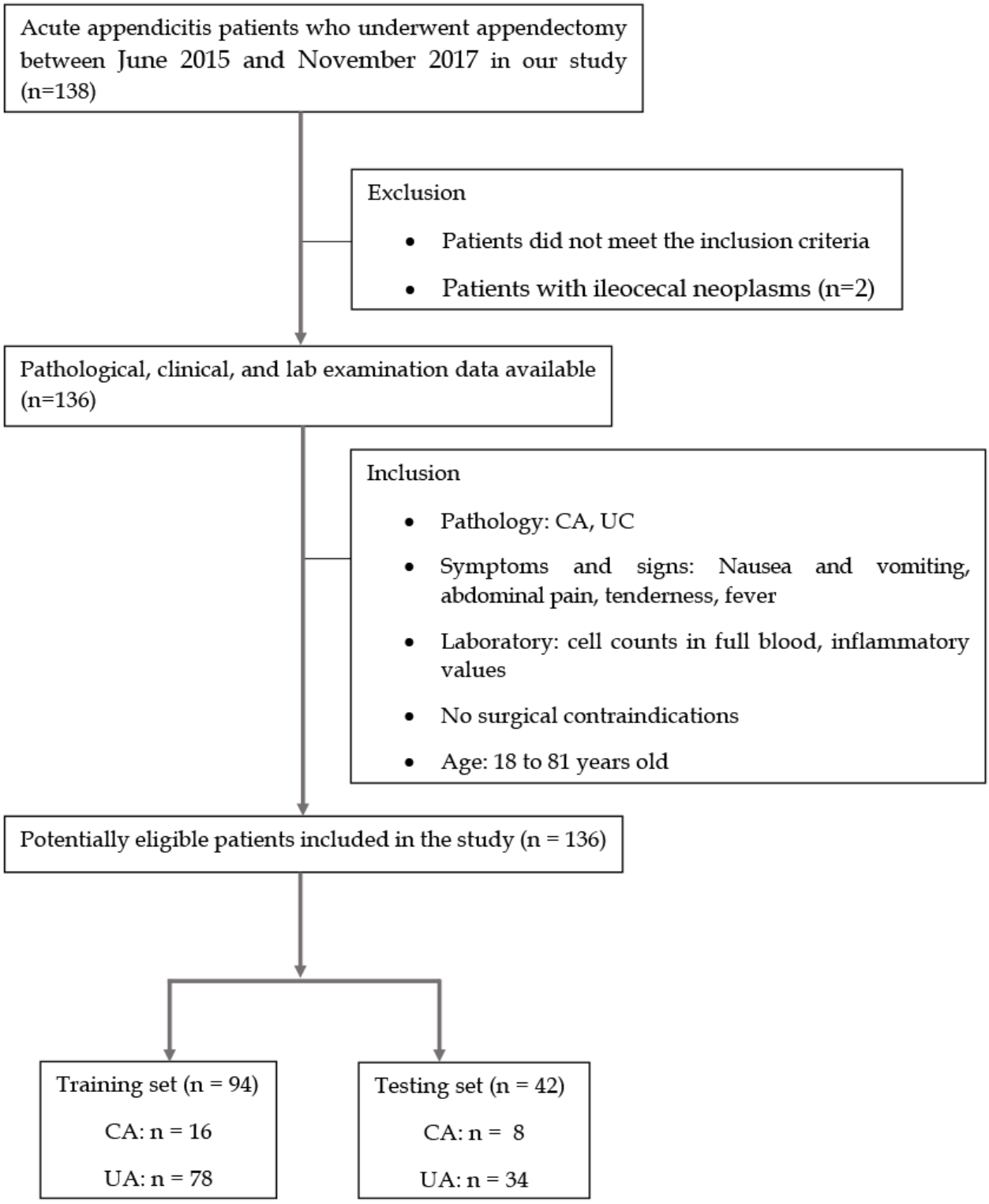

Figure 1

Flow chart of the patient selection and exclusion process

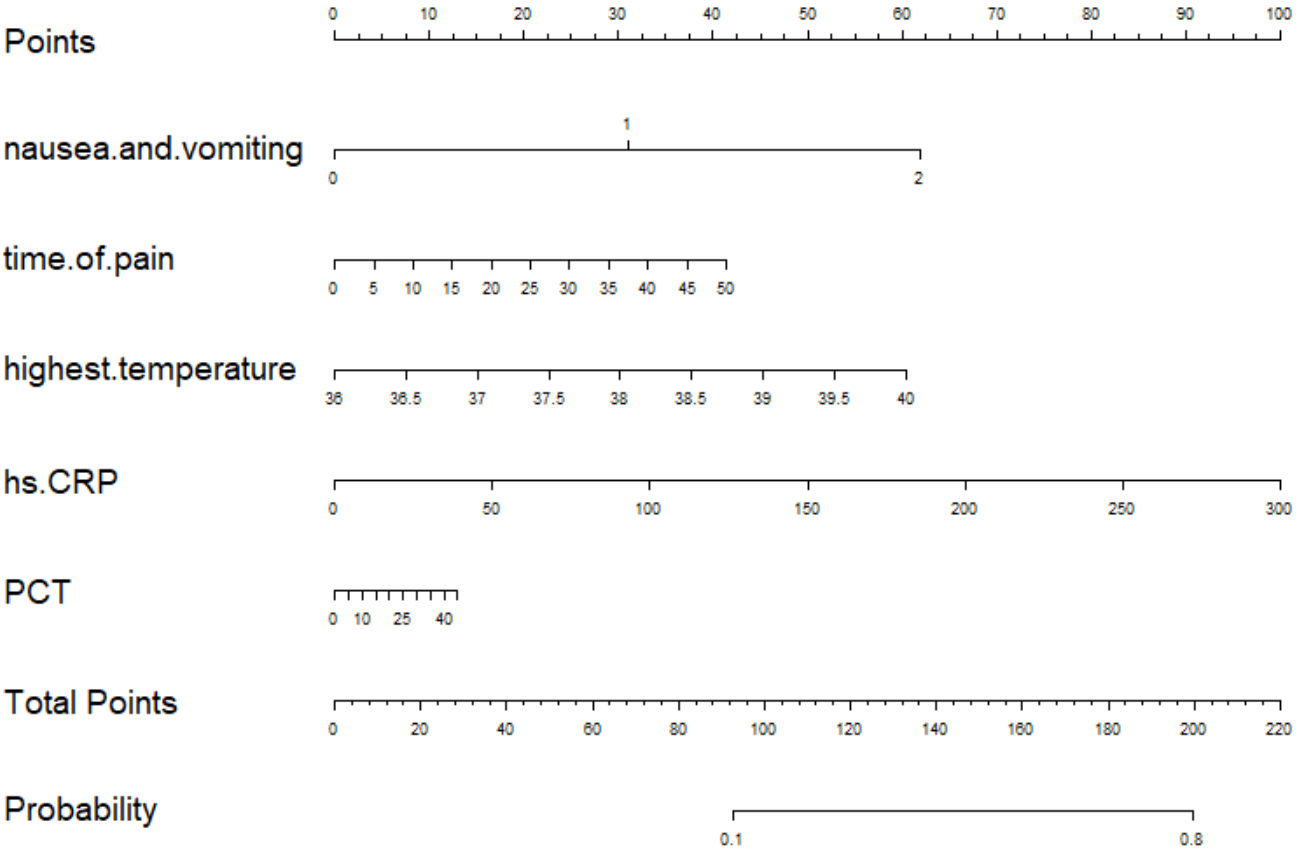

Figure 2

Nomogram for the prediction of CA based on clinical and laboratory features Abbreviation: PCT, procalcitonin; hs-CRP, high-sensitivity C-reactive protein.

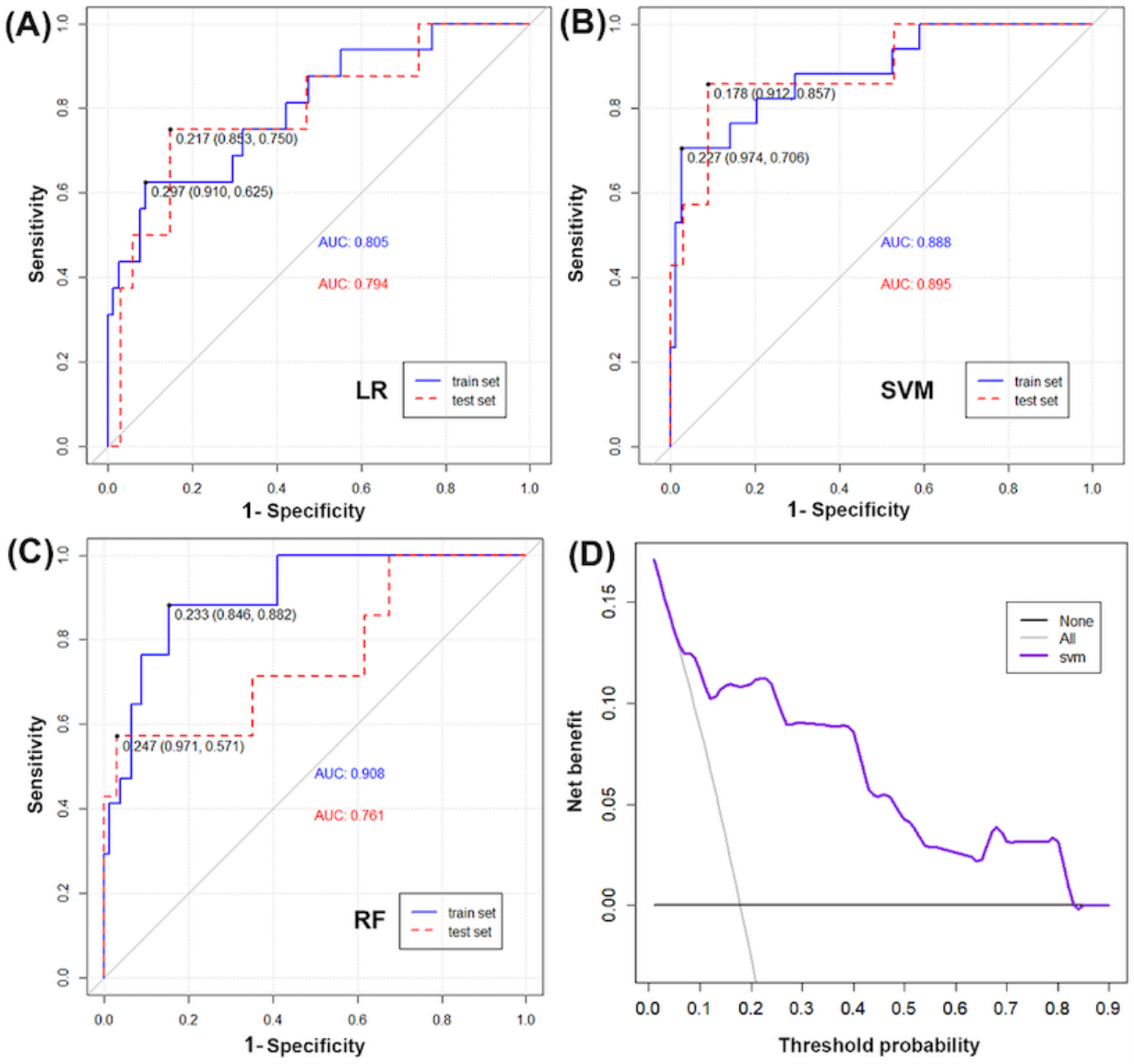

Figure 3

ROC curves of pathological type prediction in the training and testing sets based on three combined models and decision curve for SVM-based model. (A) ROC curves of CA prediction in the training and testing sets using LR algorithm. (B) ROC curves of CA prediction in the training and testing sets using SVM algorithm. (C) ROC curves of CA prediction in the training and testing sets using RF algorithm. (D) Decision curve obtained from the SVM-based model. Abbreviation: ROC, receiver operating characteristic curve; AUC, area under the curve; LR, logistic regression; SVM, support vector machine; RF, random forest.