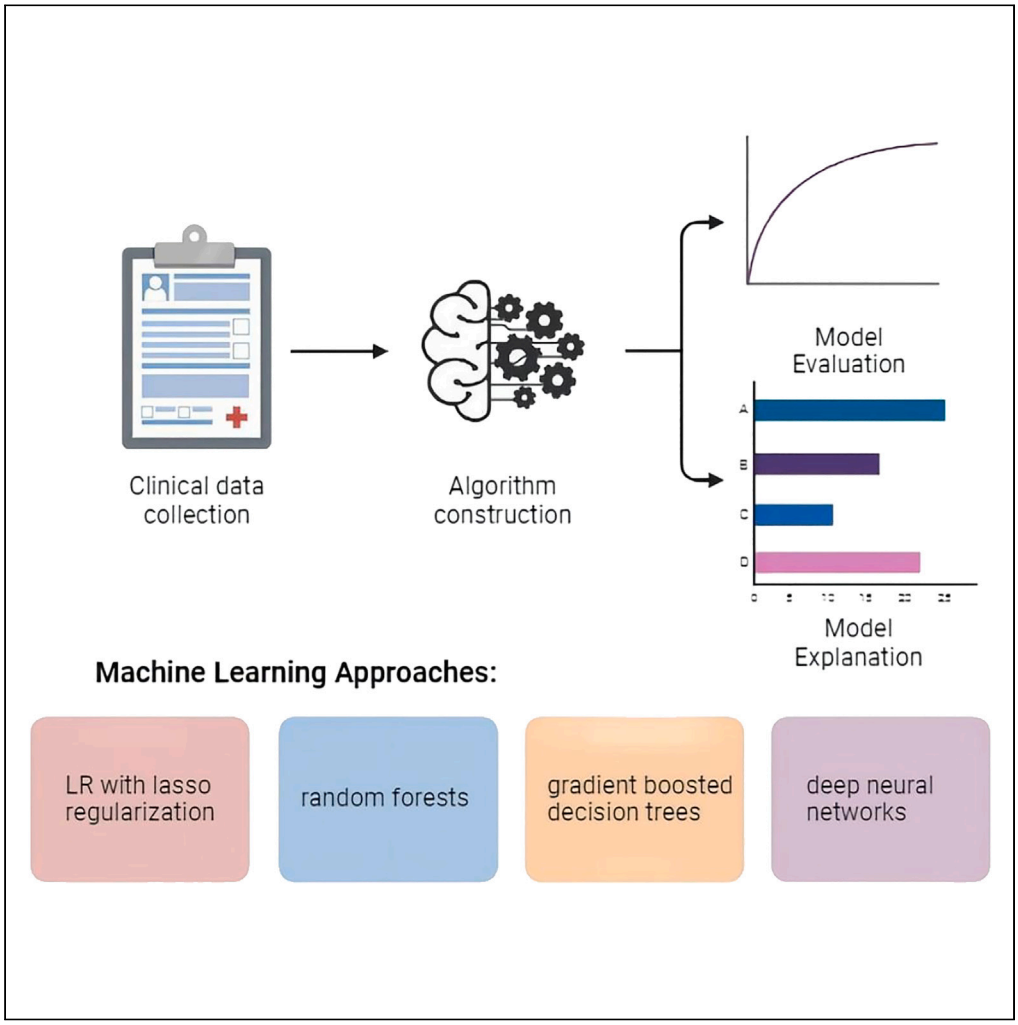
iScience

CellPress
OPEN ACCESS

Article

Predictive modeling for eosinophilic chronic rhinosinusitis: Nomogram and four machine learning approaches



Panhui Xiong, Junliang Chen, Yue Zhang, ..., Dayu Guan, Bowen Zheng, Yucheng Yang

yychxh@163.com

Highlights

We developed nomogram and machine learning models for early diagnosis of ECRS

Model predictions used non-invasive clinical data to predict ECRS

Variable importance, nomogram, and breakdown were conducted for model explanation

The validated models demonstrated good predictive performance and applicability

Xiong et al., iScience 27, 108928
February 16, 2024
https://doi.org/10.1016/j.isci.2024.108928

iScience

CellPress
OPEN ACCESS

Article

Predictive modeling for eosinophilic chronic rhinosinusitis: Nomogram and four machine learning approaches

Panhui Xiong,[1] Junliang Chen,[2] Yue Zhang,[1] Longlan Shu,[1] Yang Shen,[1] Yue Gu,[1] Yijun Liu,[1] Dayu Guan,[1] Bowen Zheng,[1] and Yucheng Yang[1,3,*]

SUMMARY

Eosinophilic chronic rhinosinusitis (ECRS) is a distinct subset of chronic rhinosinusitis characterized by heightened eosinophilic infiltration and increased symptom severity, often resisting standard treatments. Traditional diagnosis requires invasive histological evaluation. This study aims to develop predictive models for ECRS based on patient clinical parameters, eliminating the need for invasive biopsy. Utilizing logistic regression with lasso regularization, random forest (RF), gradient-boosted decision tree (GBDT), and deep neural network (DNN), we trained models on common clinical data. The predictive performance was evaluated using metrics such as area under the curve (AUC) for receiver operator characteristics, decision curves, and feature ranking analysis. In a cohort of 437 eligible patients, the models identified peripheral blood eosinophil ratio, absolute peripheral blood eosinophil, and the ethmoidal/maxillary sinus density ratio (E/M) on computed tomography as crucial predictors for ECRS. This predictive model offers a valuable tool for identifying ECRS without resorting to histological biopsy, enhancing clinical decision-making.

INTRODUCTION

Chronic rhinosinusitis (CRS) has a high global prevalence affecting roughly 5%–12% of the general global population, specifically affecting 8% of individuals in China, 10% in Europe, and 12% in the United States.[1–3] CRS significantly affects quality of life and negatively impacts the global burden of disease. In the United States, the direct healthcare cost of CRS is estimated between 10 and 13 billion dollars and an indirect cost amounting to 20 billion US dollars annually.[4]

Since the publication of EPOS2020, the understanding of the pathophysiology of CRS has increased, and CRS is increasingly recognized as a heterogeneous disease.[1,5] The traditional classification of CRS is based on the presence or absence of nasal polyps, and phenotype while uncovers the complex immunopathophysiology of CRS, namely intrinsic or endotype. Each endotype has a different molecular mechanism, thereby treating it individually with various types of biological agents binding with different biomarkers, which facilitates more accurate diagnosis, treatment, and prevention.

CRS can be classified into two subgroups, ECRS and non-ECRS, based on the number of eosinophils in the nasal polyps or mucosa, as one of the important phenotypes suggestive of type II inflammation, which is likely to complain with more severe symptoms, higher potential for treatment resistance, and more frequent of disease recurrence.[6,7] The recognition of ECRS may inform physicians to make different treatment decisions and take different post-treatment clinical monitoring. For patients requiring surgical treatment, there are some differences in the recommended surgical interventions for ECRS versus non-ECRS. Functional endoscopic sinus surgery (FESS) based on mucosal preservation is more suitable for non-ECRS to some extent, but it is difficult to obtain satisfactory results for ECRS with a large inflammatory burden and polyp obstruction by FESS. The recurrence rate of ECRS is as high as 98.5%.[8] Surgical procedures in patients with ECRS require greater mucosal management to reduce inflammatory burden, such as extended endoscopic sinus surgery and Draf III surgery, which confers a more extensive operating area than non-ECRS surgery.[7,9] The perioperative management of ECRS and non-ECRS is also different. If surgeons can differentiate ECRS and non-ECRS prior to surgery, clinicians are informed to select more appropriate surgical treatment schemes for different types of patients, thus promoting precision treatment. In addition, if patients can be ruled out as surgical patients, these outpatients are likely to be more therapy and follow-up compliant. Additionally, corticosteroids have poor efficacy in non-ECRS, whereas macrolides are not suitable for ECRS treatment.[10] The differences between glucocorticoids and macrolides and biological agents such as interleukin-4 (IL-4), IL-5, and IL-13 can also applied in the clinical treatment of ECRS.[11] Although biopsy is an important step

[1]Department of Otorhinolaryngology, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China
[2]Department of Otorhinolaryngology, Xishui People's Hospital, Xishui County, Zunyi, Guizhou Province 564600, China
[3]Lead contact
*Correspondence: yychxh@163.com
https://doi.org/10.1016/j.isci.2024.108928

Table 1. Demographic and clinical pathological characteristics of patients with chronic rhinosinusitis

Characteristics	Training set (n = 215)	Test set (n = 92)	p value	External validation set (n = 130)
Male, n (%)	151 (70.2)	65 (70.7)	0.941	70 (53.8)
Female, n (%)	64 (29.8)	27 (29.3)	0.943	60 (46.2)
Age (y), mean ± SD	43.6 ± 14.6	43.4 ± 13.3	0.942	40.8 ± 13.9
Overall subjective symptom score (VAS), mean ± SD	5.3 ± 2.5	5.4 ± 2.4	0.587	6.1 ± 1.2
Nasal obstruction, mean ± SD	5.6 ± 1.6	5.9 ± 1.8	0.180	5.3 ± 1.2
Olfactory, mean ± SD	3.5 ± 3.4	4.2 ± 3.7	0.151	1.0 ± 1.9
Purulent nasal discharge, mean ± SD	3.6 ± 2.8	3.8 ± 2.7	0.558	3.8 ± 1.3
Lund-Mackay score (CT), mean ± SD	11.4 ± 4.9	12.6 ± 4.8	0.526	14.3 ± 5.4
E/M, mean ± SD	1.8 ± 1.2	2.0 ± 1.0	0.337	1.7 ± 0.6
polyp, n (%)	132 (61.4)	73 (79.3)	0.328	31 (23.8)
Asthma, n (%)	14 (6.5)	8 (8.6)	0.960	10 (7.7)
AR, n (%)	95 (44.2)	41 (44.6)	0.951	16 (12.3)
Allergy, n (%)	40 (18.6)	21 (22.8)	0.537	19 (14.6)
Lesion extent, n (%)	172 (80.0)	79 (85.9)	0.225	17 (13.1)
Blood eosinophil ratio (%), mean ± SD	4.4 ± 3.3	3.8 ± 3.1	0.160	3.1 ± 3.3
Blood eosinophil number (×109), mean ± SD	0.27 ± 0.21	0.22 ± 0.19	0.121	0.27 ± 0.37
Smoking, n (%)	71 (33.0)	26 (28.3)	0.411	35 (26.9)
Surgical history, n (%)	39 (18.1)	15 (16.3)	0.140	6 (4.6)
Tissue eosinophil count >10 per HPF, n (%)	84 (39.1)	34 (37.0)	0.727	19 (14.6)

for diagnosis, every patient does not need to undergo this invasive and painful procedure that may increase or accumulate unnecessary additional financial burden. Thus, it is only performed for patients who have had single or multiple sinus surgeries and nasal polyps suspected of tumor lesions. However, the assessment of clinical data includes sinus computed tomography (CT), routine blood cultures, and symptomatic assessment. As to the differences in surgical modalities and therapeutic medication between ECRS and non-ECRS, as well as the risk of postoperative asthma and recurrence, a non-invasive clinical diagnosis of ECRS based on routine clinical data is essential to formulate an optimal treatment plan.

Sensitive and specific non-biopsy predictive strategies for identifying patients with ECRS will be indispensable in patients' prognosis and precision treatment. Achieving both high sensitivity and specificity for tissue eosinophilia is challenging, whereas machine learning (ML) offers techniques that are currently used as powerful and reliable as diagnostic tools for the outcome evaluation. The advantage of machine learning methods over standard methods of statistical model building includes its ability to handle complex non-linear relationships between predictor variables and thus can generate more robust predictions.[12] A nomogram is like an ancient calculator, representing the relations between three or more variable quantities using several scales, and similar to a slide rule, which provides a graphical description of a logistic or Cox regression model. It has been used in epidemiology for disease diagnosis prognosis assessment and recurrence prediction.[13–18] Although sophisticated ML methods can provide more accurate and generic predictive models, the convenience and transparent nature of nomograms keep them popular with clinicians.[18]

Both traditional and ML models can be effective tools to improve clinicians' diagnosis and decision-making, providing mechanisms for more accuracy and precision in the treatment of chronic rhinosinusitis. Therefore, this study aimed to explore the independent predictors of ECRS in patients among a broader CRS cohort and to develop a predictive model for identifying patients by using both traditional statistical methods and ML techniques for comparison.

RESULTS

Patient characteristics

The demographic and clinical characteristics of the entire study population are presented in Table 1. In the training set, a total of 307 patients received ESS to treat CRS. Most baseline characteristics were not statistically different and were comparable between the two cohorts. Demographic and clinical characteristics of the independent external validation set are presented in Table 1.

Nomogram: Development and evaluation

The results of univariate and multivariate analyses of clinical predictors of ECRS are displayed in Table 2. The general subjective symptom score (VAS score), presence of allergic rhinitis (AR), E/M, peripheral blood eosinophil ratio (blood EOS ratio), presence of polyps, and

Table 2. Univariate and multivariate analysis of predictors

	Univariate			Multivariate		
Variable	β	OR (95% CI)	p value	β	OR (95% CI)	p value
Overall subjective symptom score (VAS)	0.150	1.162 (1.050–1.285)	0.004	0.167	1.182 (1.029–1.358)	0.018
Olfactory	0.091	1.095 (1.025–1.171)	0.007	−0.062	0.940 (0.850–1.039)	0.223
Purulent nasal discharge	−0.125	0.882 (0.809–0.961)	0.004	−0.210	0.810 (0.720–0.913)	<0.001
Lund-Mackay score (CT)	0.108	1.114 (1.058–1.173)	<0.001	0.036	1.037 (0.963–1.117)	0.340
E/M	0.768	2.156 (1.661–2.798)	<0.001	0.651	1.918 (1.386–2.655)	<0.001
Polyp	1.012	2.752 (1.614–4.693)	<0.001	0.890	2.434 (1.127–5.256)	0.023
AR	0.942	2.564 (1.599–4.111)	<0.001	1.218	3.380 (1.523–7.501)	0.003
Allergy	0.963	2.620 (1.558–4.404)	<0.001	−0.280	0.756 (0.314–1.818)	0.532
Lesion extent	0.867	2.379 (1.218–4.645)	0.011	0.396	1.486 (0.601–3.673)	0.391
Blood EOS ratio	0.378	1.460 (1.317–1.618)	<0.001	0.501	1.650 (1.273–2.139)	<0.001
Blood EOS number	5.313	202.921 (43.985–936.151)	<0.001	−2.475	0.084 (0.002–4.078)	0.211
Asthma	0.730	2.075 (1.021–4.218)	0.044	−0.341	0.711 (0.272–1.857)	0.486
Nasal obstruction	0.000	1.000 (0.872–1.146)	0.999			
Smoking	0.360	1.433 (0.878–2.340)	0.150			
Surgical history	−0.313	0.731 (0.417–1.284)	0.276			
Age	0.000	1.000 (0.984–1.016)	0.961			
Male	0.199	1.220 (0.733–2.030)	0.445			

*Overall subjective symptom score (VAS): patients graded their subjective overall nasal discomfort with the visual analog scale (VAS), marking the intensity of the symptoms on a straight line from 0 to 10 cm.

presence of nasal discharge were identified as independent predictors of clinical diagnosis for ECRS. The final model was built based on the predictors and further visualized as a nomogram by using the "rms" package of the R software (Figure 1). The application of the nomogram is described below. First, we drew an ascending line from the variable axis to the "points" axis to get points for each risk factor. The score of each item for all the variables was added to get the total score. Finally, we produced a downward vertical line from the "Total Score" axis to the "Risk" axis. The corresponding numbers were then expressed as predicted odds of diagnosing ECRS. The beta coefficient of the final logistic regression model is presented in Table 3.

In the training set, the nomogram achieved excellent discrimination with an AUC of 0.867 (95% confidence interval [CI]: 0.816–0.917); the DeLong test comparing AUC between the nomogram and the univariate prediction showed that the nomogram had significantly higher discriminative power (Figure 2A). In terms of the calibration capability, the calibration curve showed good agreement between the predicted and actual risks (Figure 2C). In the internal test set, the nomogram achieved good discrimination with an AUC of 0.867 (95% CI: 0.816–0.917); the DeLong test comparing AUC between the nomogram and the univariate prediction showed that the nomogram also had significantly higher discriminative power (Figure 2B). The nomogram prediction model was in good agreement with the bootstrap correction model (Figure 2D). Furthermore, it showed reasonably acceptable accuracy with a Brier score of 0.157. In external validation, the nomogram showed strong discrimination, accuracy, and calibration capability with an AUC of 0.837 (95% CI: 0.743–0.930).

Machine learning models: Selection and evaluation

The discriminative power of the different models, represented by ROC curves, is shown in Figure 3A. Table 4 presents the generalization ability metrics of the logistic regression model and various ML models in the test set, such as AUC, sensitivity, specificity, etc. There was no statistical difference between logistic regression and the four machine learning models. The discrimination power of the deep neural network model was the lowest (AUC 0.823; 95% CI, 0.738–0.908) with the highest sensitivity (0.886; 95% CI, 0.780–0.991) and negative likelihood ratio (0.905; 95% CI, 0.816–0.994) (Table 4). The RF model presented the highest discriminatory power (AUC 0.879; 95% CI, 0.806–0.952); the lasso regression model showed the highest specificity (0.911; 95% CI, 0.836–0.985) and positive predictive value (0.833; 95% CI, 0.700–0.967); in the meantime, the gradient boosted decision tree model showed the highest sensitivity (0.886; 95% CI, 0.780–0.991). Although LR with lasso regularization had the highest positive likelihood ratio (7.778; 95% CI, 3.278–18.456), the deep neural network displayed the lowest negative likelihood ratio (0.168; 95% CI, 0.066–0.431), which can be selected and used according to different scenarios in clinical work. In the external validation, the logistic regression model and the machine learning model also demonstrated good predictive proficiency, and the ROC curve was shown in Figure 4A. The generalization ability indicators of various ML models in the external validation set are shown in Table 5. Due to the optimal predictive performance of LR using lasso regression, we created a second nomogram, nomogram2, based on LR with a lasso regression model to visualize the model (Figure 5).

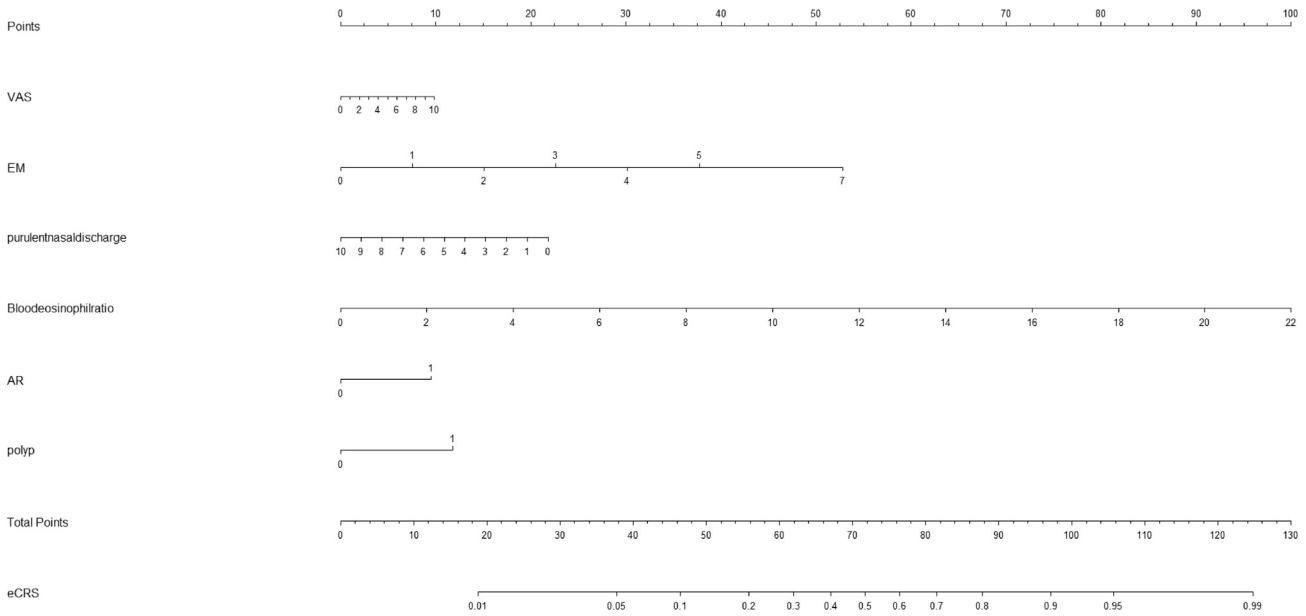


Figure 1. Nomogram of predictive diagnostic eCRS

In the decision curve analysis (Figure 3B), the net benefits of all models were better within the threshold probability range, among which LR with lasso demonstrated the largest net benefit. However, the net benefits of the deep neural network model were lower in most time. In the external validation (Figure 4B), the net gains of all models were good within the threshold probability range, with DNN having the largest net gains and the random forest model showing a low net gain in most cases.

Figure 6 presents the variable importance of each outcome in the gradient-boosting decision tree, with peripheral blood eosinophil ratio, absolute peripheral blood eosinophil, and E/M being important predictors. The importance of these variables was consistent with random forest models (Figure 7). In addition, the model interpretations of the DNN model prediction based on the breakdown method for a single patient in the test set are shown in Figure 8.

DISCUSSION

The clinical predictors of ECRS remain unclear. We assessed the following clinical parameters: patient symptoms, blood tests, and imaging data. The following six valuable predictors that make up the predictive nomogram were able to identify: general subjective symptom score (VAS score), with allergic rhinitis (AR), E/M, blood EOS ratio, with polyps, and accompanied with purulent nasal discharge. The proportion of peripheral blood eosinophils, absolute peripheral blood eosinophils, and E/M were deemed crucial predictors for the diagnosis of ECRS using the ML model. The nomograms and ML models developed in the program both indicated good predictive power and may assist clinicians in diagnosing ECRS in the absence of pathological findings and provide a basis for more precise personalized treatment.

In extreme to non-ECRS patients, the number and proportion of EOS in the peripheral blood of ECRS patients are higher.[19] Additionally, the number and proportion of EOS may be positively correlated with serum immunoglobulin E (IgE), IL-5, and other type II inflammatory factors, which also appear to be the most influential predictor of ECRS in this study.[20] However, peripheral blood eosinophilia is not specific to ECRS or parasitic infection. It may be elevated in other diseases, such as allergic or autoimmune diseases or even medication; so other clinical findings were combined in this study to make the diagnosis more accurate.[21]

Each patient underwent CT scanning, and the CT of the affected sinuses was graded according to the Lund-Mackay staging system. The E/M ratio is the defined threshold ratio of the total ethmoid sinus (E) and total maxillary (M) scores on both sides (E/M ratio). This ratio can reflect the degree of mucosal infiltration in the paranasal sinuses. One study showed that the E/M ratio of paranasal sinus CT was positively correlated with the degree of inflammation and the number of EOS in sinusitis patients. The study also further explored the relationship between the E/M ratio of sinus CT and other sinus-inflammation-related indicators (such as nasal discharge, nasal congestion, etc.), and mentioned that this indicator can be used to distinguish different types of sinusitis, such as CRS with nasal polyps (eCRSwNP).[22] In this study, E/M was either in the nomogram establishment (odds ratio [OR]:1.924; 95% CI, 1.330–2.785) or in the ML models, both appearing as important factors, which also reflected the vital predictive value for ECRS.

Nasal polyps are closely associated with type II inflammation or ECRS,[1] and often refractory CRSwNP has a higher percentage of total inflammatory cells, tissue EOS, eosinophilia aggregates, and Charcot-Leyden crystal (CLC) formation.[23] However, the direct relationship between the presence of nasal polyps and type II inflammation or ECRS remains to be further confirmed. One study showed that intra-type

Table 3. Estimated values and parameters of regression coefficients of nomogram

Parameter	β	SE	Wald chi-square test	OR	95% CI	p value
Overall subjective symptom score (VAS)	0.086	0.078	1.200	1.089	(0.935–1.270)	0.273
E/M	0.655	0.189	12.055	1.924	(1.330–2.785)	<0.001
Purulent nasal discharge	−0.189	0.070	7.295	0.828	(0.721–0.949)	0.007
Blood EOS ratio	0.394	0.073	29.552	1.483	(1.287–1.710)	<0.001
AR	0.823	0.371	4.918	2.278	(1.100–4.714)	0.027
Polyp	1.020	0.423	5.817	2.774	(1.287–1.710)	0.016

*VAS: the visual analog scale (VAS); patients graded their subjective overall nasal discomfort with the VAS scale, marking the intensity of the symptoms on a straight line from 0 to 10 cm.

3 CRS always presents purulent discharge in the sinuses, and IL-17 is the main cytokine produced by Th17 cells and type 3 innate lymphoid cells (ILC3). B lymphocytes and antibodies play an important role in CRS.[24] In the establishment of the column chart of this study, the OR value of dense DDT for predicting ECRS was less than 1 (0.828; 95% CI, 0.721–0.949), suggesting a protective effect on the diagnosis of ECRS.

Recent studies have shown that type II inflammation is of great importance in the occurrence and development of AR. Type II inflammation is an inflammatory response mediated by specific immune cell types, including EOS, CD4 + T-helper 2 cells, and Th2-mediated cytokines, which increase vascular permeability and mucus secretion, leading to upper airway and bronchospasm, airway inflammation, and airway narrowing. All indicated that type II inflammation could be essential to the pathophysiology of allergic rhinitis.[25] However, the causal relationship between AR and ECRS is still controversial. In this study, patients with AR were at higher risk of contracting ECRS compared with patients without allergic rhinitis, which indicated that AR could be a crucial predictor of ECRS in the practice of medicine.

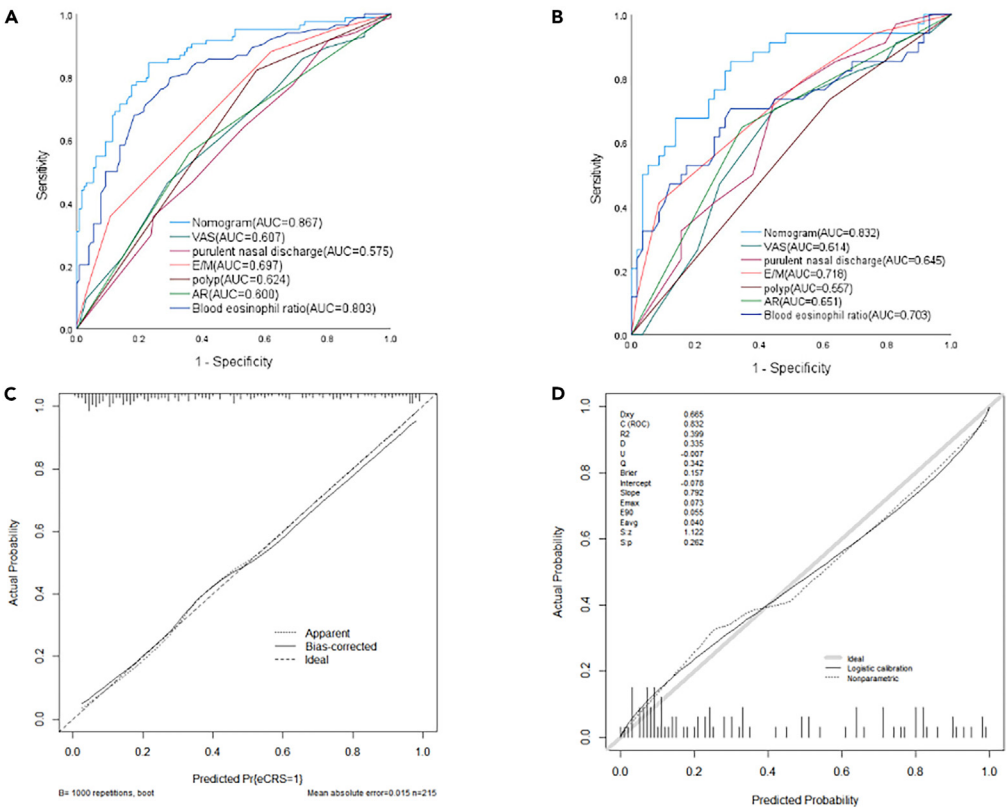


Figure 2. Receiver operating characteristic (ROC) curves and nomogram calibration curves

(A) ROC curve of the nomogram in the training set.

(B) AUC comparison between nomogram and univariate prediction.

(C) Calibration curve of the training set nomogram.

(D) Calibration curves comparing the ideal model, nomogram, and bias-corrected model in the test set.

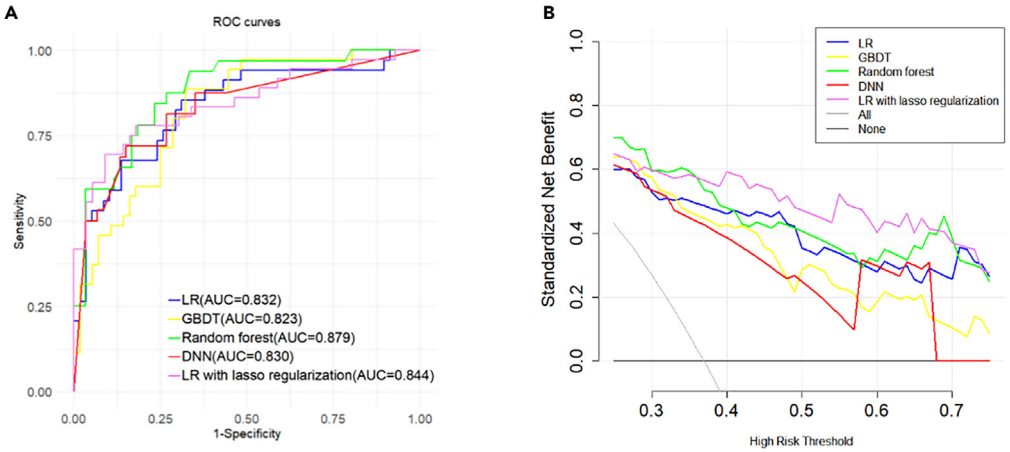


Figure 3. Prediction ability of the LR model and machine learning models for eCRS in the test set

Regarding ML and nomogram, we found that classifiers trained by the LR with the lasso regularization algorithm generated the best predictive power, possibly due to the screening effect of lasso regularization on predictors, which minimizes the potential overfitting of the model. Risk factors identified in the logistic regression model (such as blood EOS ratio, E/M, general subjective symptom score [VAS score], etc.) also ranked high in the characteristic analysis. However, we observed some differences between the results of logistic regression and those of the ML algorithm. In logistic regression analysis, because there may be collinearity between blood EOS ratio and blood EOS count, blood EOS ratio was selected as the most influential factor on the outcome diagnosis in multivariate analysis. To start with ML algorithms, machine learning can calculate more parameters and features to identify key features.[12] Blood EOS ratio and blood EOS counts are regarded as the most important predictors. Secondly, given the limited sample size, only a few variables were included in the logistic regression model to prevent overfitting. As to the same sample size, more variables were displayed in the ML model. Apart from the general subjective symptom global score (VAS score), the presence of allergic rhinitis (AR), E/M, blood EOS ratio, presence of polyps and purulent nasal discharge, blood EOS count, CT score, and olfactory score are also considered as key factors on the prediction of ECRS. Thirdly, unlike machine learning, the nomogram provides us with a visual clinical diagnosis probability on the prediction of ECRS diagnosis in a visual manner. Although the predictive power of a nomogram is not the best, its advantages lie in its interpretability and ease of use in the clinic. Machine learning is increasingly applied to process multiple input variables for classification prediction. Machine learning methods can automatically select useful features from piles of features, incorporate high-order nonlinear interactions between predictors, and have good generalization capabilities that cannot be solved by traditional modeling methods (such as logistic regression models).[26] Although machine learning techniques have been used relatively infrequently in otolaryngology, their implementation in clinical practice is limited.[27] However, with the help of artificial intelligence, data processing and decision support for large samples and multivariate inputs become easier and more comfortable. Finally, in our study, the results of

Table 4. Prediction ability of the LR model and four machine learning models for eCRS in the test set

Outcome and model	AUC (95% CI)	p Value	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR (95% CI)	NLR (95% CI)
Test set								
Logistic regression	0.832(0.741–0.923)	Ref.	0.853(0.734–0.972)	0.690(0.571–0.809)	0.617(0.478–0.756)	0.889(0.797–0.981)	2.748(1.827–4.134)	0.213(0.093–0.488)
Logistic regression with lasso regularization	0.844(0.755–0.933)	0.8851	0.694(0.544–0.845)	0.911(0.836–0.985)	0.833(0.700–0.967)	0.823(0.727–0.918)	7.778(3.278–18.456)	0.336(0.204–0.553)
Random forest	0.879(0.806–0.952)	0.4324	0.844(0.718–0.970)	0.767(0.660–0.874)	0.659(0.513–0.804)	0.902(0.820–0.984)	3.616(2.232–5.857)	0.204(0.090–0.461)
Gradient-boosted decision tree	0.823(0.738–0.908)	0.9765	0.886(0.780–0.991)	0.679(0.556–0.801)	0.633(0.498–0.768)	0.905(0.816–0.994)	2.756(1.849–4.106)	0.168(0.066–0.431)
Deep neural network	0.830(0.739–0.922)	0.8584	0.719(0.563–0.875)	0.850(0.760–0.940)	0.719(0.563–0.875)	0.850(0.760–0.940)	4.792(2.526–9.089)	0.331(0.188–0.582)

*NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value.
*a Comparison of area under the curve of the reference model (conventional triage approaches) with those of each machine learning approach by using a DeLong test.

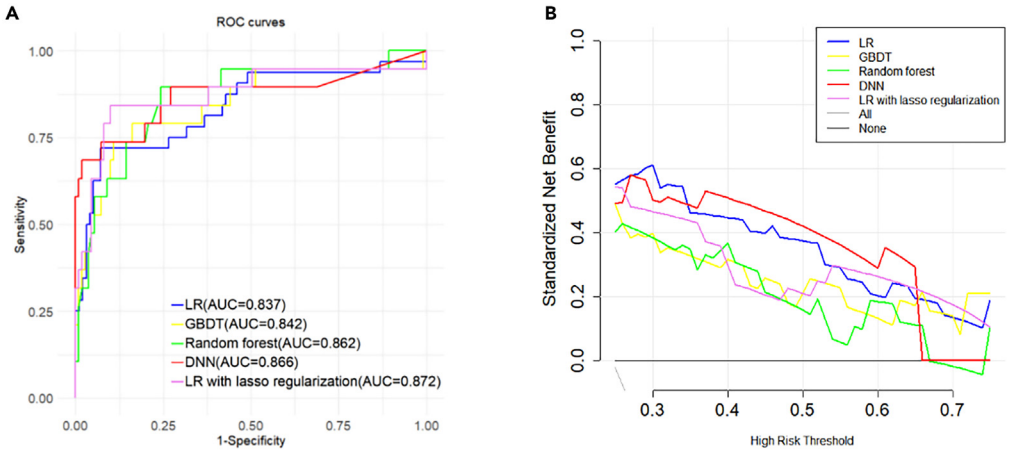


Figure 4. Prediction ability of the LR model and machine learning models for eCRS in the external validation set

external validation suggest that the predictive power of the machine learning model is not significantly higher than that of the nomogram, which may be due to the limited sample size and predictor variables recruited in this study. We believe that this study is a preliminary exploration of ML model prediction of ECRS, and clinical studies with larger scale and more centers should be carried out in the future to further improve and update the ML model. More in-depth research should be carried out on algorithm selection and hyperparameter tuning.

In previous studies, Smith et al. predicted ECRS through eosinophil peroxidase (EPX) analysis of ethmoid and nasal polyp tissue.[28] The study was invasive and unconventional. Thorwarth et al. preliminarily explored the prediction of ECRS through the urinary leukotriene E4 (uLTE4) level, peripheral eosinophil count, and polyp status.[29] Ulte4 in this study was not a routine clinical examination. The study included only 80 patients, with a small sample size and fewer included predictors, with no external validation set. We believe our research had advantages. First, our study included non-invasive common clinical data, including subjective symptoms, nasal endoscopic results, imaging data, etc. Second, we validated our five models in the test set and external validation set. Finally, we also used nomogram, variable importance analysis, and breakdown methods to interpret and visualize the model, so that clinicians can better understand the model. Overall, this study is meaningful for the prediction of ECRS based on clinical parameters. Compared with non-ECRS patients, ECRS patients need repeated corticosteroid therapy and multiple revision surgeries to reduce their symptoms. In the era of precision CRS treatment, surgical approaches often need to be tailored to the underlying type of inflammation. Their perioperative medication choices differ in terms of medication and clinical follow-up for non-surgical patients, as well as differences in medication and follow-up frequency between ECRS patients and non-ECRS patients. Our model shows satisfactory performance and can provide valuable information for accurate diagnosis and treatment. In addition, the combination of nomogram and machine learning models in this study may be more clinically relevant. When patients have limited clinical data or undergo rapid screening, we can choose nomogram1 for ECRS diagnosis. After using nomogram1 for preliminary screening, positive patients can improve the quality of diagnosis by using the LR with lasso regularization model, which has the best predictive performance for ECRS diagnosis, after completing clinical data. It can be seen that after using the joint diagnostic strategy of nomogram1 and the LR with lasso regularization model, we increased the AUC of ECRS to 0.895 (95% CI: 0.804–0.987) in the test set and 0.922 (95% CI: 0.842–1.000) in the external validation set (Figure 9). In recent years, there have been relatively few studies using machine learning to predict ECRS,

Table 5. Prediction ability of the LR model and four machine learning models for ECRS in the external validation set

Outcome and model	AUC (95% CI)	p value	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	PLR (95% CI)	NLR (95% CI)
External validation set								
Logistic regression	0.837(0.743–0.930)	Ref.	0.719(0.563–0.875)	0.929(0.878–0.980)	0.767(0.615–0.918)	0.910(0.854–0.966)	10.063(4.772–21.218)	0.303(0.174–0.528)
Logistic regression with lasso regularization	0.872(0.758–0.987)	0.6371	0.842(0.678–1.000)	0.901(0.845–0.956)	0.593(0.407–0.778)	0.971(0.938–1.003)	8.498(4.693–15.387)	0.175(0.062–0.496)
Random forest	0.862(0.716–0.963)	0.7168	0.895(0.757–1.000)	0.757(0.677–0.837)	0.386(0.242–0.530)	0.977(0.945–1.009)	3.678(2.560–5.286)	0.139(0.037–0.518)
Gradient-boosted decision tree	0.842(0.724–0.960)	0.9443	0.737(0.539–0.935)	0.892(0.834–0.950)	0.538(0.347–0.730)	0.952(0.911–0.993)	6.816(3.748–12.396)	0.295(0.139–0.628)
Deep neural network	0.866(0.748–0.989)	0.6822	0.684(0.475–0.893)	0.982(0.957–1.000)	0.867(0.695–1.039)	0.948(0.907–0.988)	37.974(9.299–155.065)	0.322(0.166–0.624)

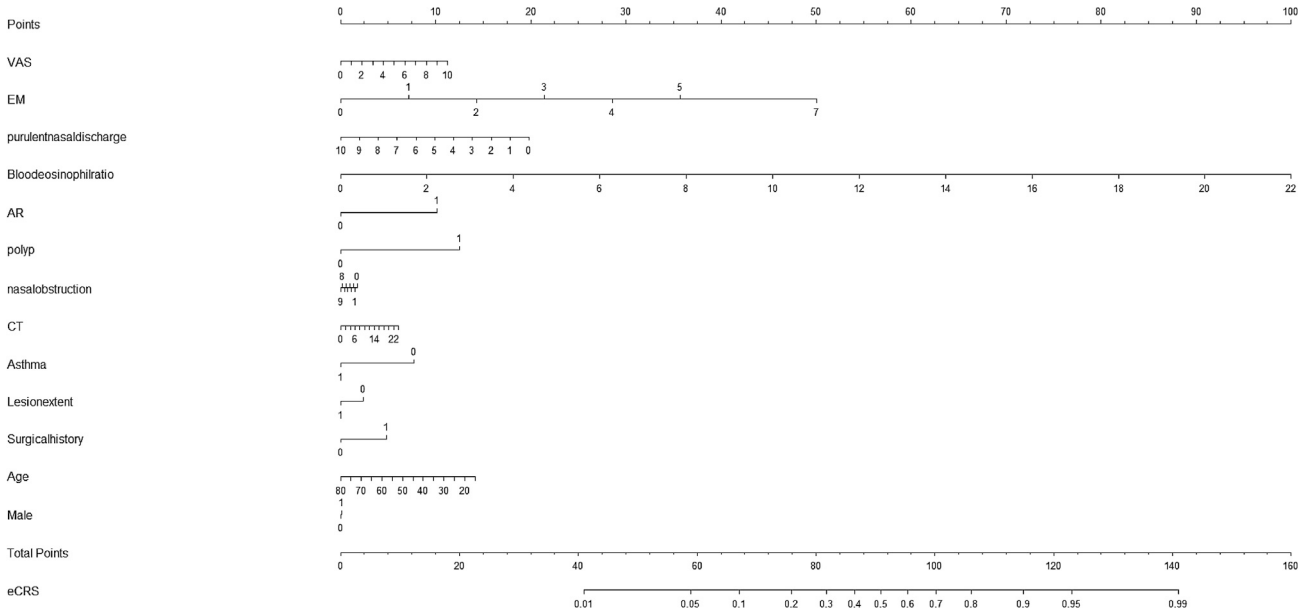


Figure 5. Nomogram2 based on LR with lasso regression model

and the accuracy of prediction can still be further improved. Due to the lack of unified validation standards and procedures, as well as the heterogeneity of patient cohorts, there is currently no model that can be widely adopted in the real world. In the future, under the premise of multiple centers and large samples, further optimization of algorithms and hyperparameter selection can be carried out, and machine learning models can be promoted to clinical use through online network calculators and other methods.

Limitations of the study

For the assessment of purulent discharge, no distinction was made between chronic or acute purulent discharge. The size of the samples limits the predictors included in the study. Furthermore, due to this limitation, objective biomarkers and clinical features such as polyp status and the presence of AR are preferred, which are easily assessed by otolaryngologists. As mentioned earlier, the limited sample size resulted in individual epidemiological characteristics such as asthma and AR percentages being higher in the study population than in the general population. This is a well-recognized selection bias in research and may limit the generalizability of pilot surveys. In addition, due to the limitation of pathological samples, the prediction of inflammatory factor addition was not included in this study. In the future, a variety of inflammatory factors could be added, and the phenotype and endotype judgment of CRS will be more comprehensive and accurate. Therefore, we believe that more potential predictors should be included in prospective studies with larger sample sizes. Future studies should include more sensitive biomarkers for model development.

Conclusion

In summary, we constructed and validated a predictive nomogram and four ML models for predicting potential ECRS patients. The predictive model was developed by applying objective data routinely collected in clinical practice. The clinical nomogram was prepared by using the six best factors for predicting ECRS, including general subjective symptom score (VAS score), allergic rhinitis (AR), E/M, blood EOS ratio, polyps, and nasal discharge. Checking the ML algorithm, we found that blood EOS ratio, blood EOS number, and E/M were key predictors. In this study, both the nomogram and the four machine learning models showed superiority in predicting the diagnosis of ECRS, and we believe that this combination may be more clinically relevant than either the nomogram or the ML model alone. Early prediction of ECRS is important and may greatly influence clinician options, providing insights for more accurate diagnosis and more precise personalized treatment of patients with CRS.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

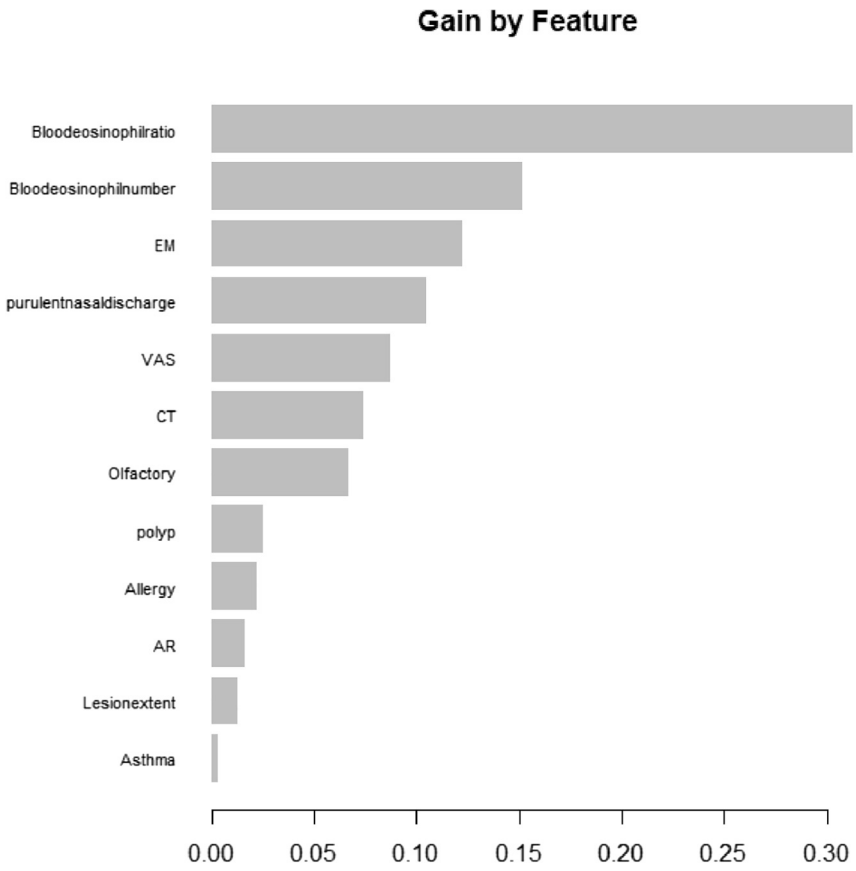


Figure 6. Importance of each predictor in the gradient-boosted decision tree models

- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ethics approval and consent to participate
- METHOD DETAILS

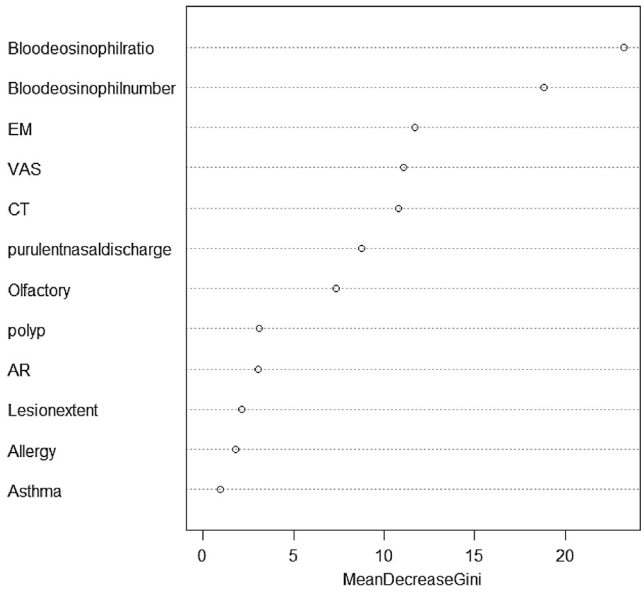


Figure 7. Importance of each predictor in the random forest models

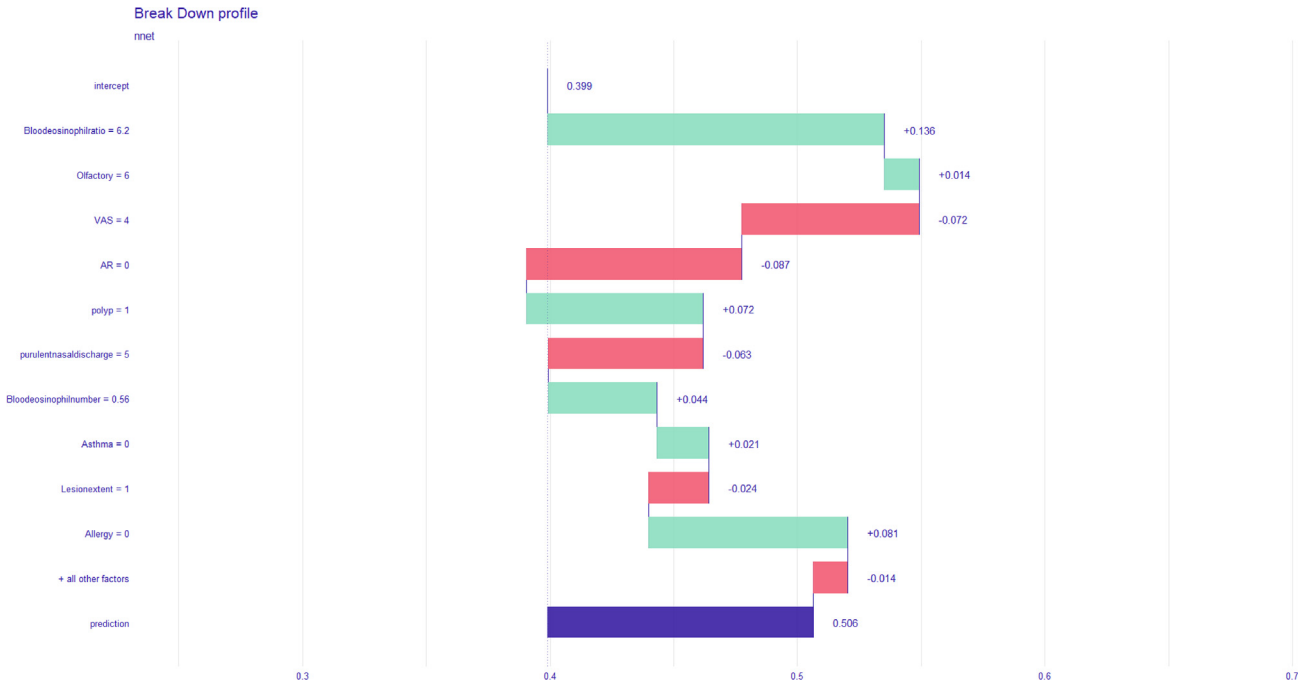


Figure 8. The breakdown result of a single CRS patients from the test set; green bar represents positive prediction and red bar represents negative prediction; numbers near each bar represent how features increase/decrease final prediction value

- Histological analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Nomogram prediction model
  - Machine learning model
  - Model evaluation

ACKNOWLEDGMENTS

We thank Qiuling Shi for guidance in the statistical methods of our study and Xiuqin XU for participating in data collection. This study was supported by the National Natural Science Foundation of China, Grant/Award Number: 81970864; the Chongqing Talents Project,

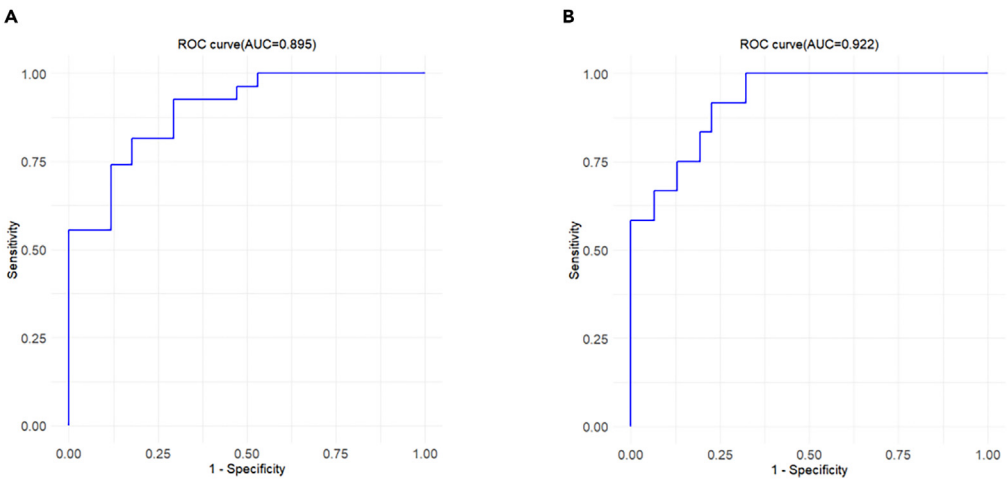


Figure 9. Receiver operating characteristic (ROC) curves of the LR with lasso regularization model after initial screening of nomogram1

(A) ROC curve in the test set.

(B) ROC curve in the external validation set.

Grant/Award Number: cstc2021ycjhbgzxm0080; the Chongqing Middle and Youth Medical High-end Talent Studio Project (Yu Wei (2018) No. 2); the Chongqing Health and Family Planning Commission Research Project (2020jstg004); Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau), Grant/Award Number:2023GDRC003; and China Post-doctoral Science Foundation(2022M720608), Foundation of State Key Laboratory of Ultrasound in Medicine and Engineering (Grant No.2023KFKT001).

AUTHOR CONTRIBUTIONS

P.X. and Y.Y. contributed to the conception and design of the study. P.X. and J.C. collected the clinical data, organized the database, and wrote the first draft of the manuscript. P.X. performed the statistical analysis. Y.Z., L.S., Y.S., Y.G., D.G., B.Z., and Y.L. revised the manuscript. All authors have contributed to the article and approved the submitted version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 4, 2023
Revised: December 4, 2023
Accepted: January 12, 2024
Published: January 17, 2024

REFERENCES

1. Fokkens, W.J., Lund, V.J., Hopkins, C., Hellings, P.W., Kern, R., Reitsma, S., Toppila-Salmi, S., Bernal-Sprekelsen, M., Mullol, J., Alobid, I., et al. (2020). European position paper on rhinosinusitis and nasal polyps 2020. Rhinology *58* (*Suppl S29*), 1–464.
2. DeConde, A.S., and Soler, Z.M. (2016). Chronic rhinosinusitis: Epidemiology and burden of disease. Am. J. Rhinol. Allergy *30*, 134–139.
3. Yim, M.T., and Orlandi, R.R. (2020). Evolving rhinology: understanding the burden of chronic rhinosinusitis today, tomorrow, and beyond. Curr. Allergy Asthma Rep. *20*, 7.
4. Rudmik, L. (2017). Economics of chronic rhinosinusitis. Curr. Allergy Asthma Rep. *17*, 20.
5. Brown, H.J., Tajudeen, B.A., Kuhar, H.N., Gattuso, P., Batra, P.S., and Mahdavinia, M. (2021). Defining the Allergic Endotype of Chronic Rhinosinusitis by Structured Histopathology and Clinical Variables. J. Allergy Clin. Immunol. Pract. *9*, 3797–3804.
6. Naclerio, R., Mullol, J., and Stevens, W.W. (2023). A Decade of Clinical Advances in Chronic Rhinosinusitis: 2012-2022. J. Allergy Clin. Immunol. Pract. *11*, 43–50.
7. McHugh, T., Snidvongs, K., Xie, M., Banglawala, S., and Sommer, D. (2018). High tissue eosinophilia as a marker to predict recurrence for eosinophilic chronic rhinosinusitis: a systematic review and meta-analysis. Int. Forum Allergy Rhinol. *8*, 1421–1429.
8. Cardell, L.O., Stjärne, P., Jonstam, K., and Bachert, C. (2020). Endotypes of chronic rhinosinusitis: impact on management. J. Allergy Clin. Immunol. *145*, 752–756.
9. Bachert, C., Marple, B., Hosemann, W., Cavaliere, C., Wen, W., and Zhang, N. (2020). Endotypes of Chronic Rhinosinusitis with Nasal Polyps: Pathology and Possible Therapeutic Implications. J. Allergy Clin. Immunol. Pract. *8*, 1514–1519.
10. Zeng, M., Wang, H., Liao, B., Wang, H., Long, X.B., Ma, J., Liu, J.X., and Liu, Z. (2021). Clinical and biological markers predict the efficacy of glucocorticoid- and macrolide-based postoperative therapy in patients with chronic rhinosinusitis. Am. J. Rhinol. Allergy *35*, 596–606.
11. (2016). Rhinology/Allergy. Otolaryngol. Head Neck Surg. *155* (1_suppl), P28–P31.
12. Li, D., Kulasegaram, K., and Hodges, B.D. (2019). Why We Needn't Fear the Machines: Opportunities for Medicine in a Machine Learning World. Acad. Med. *94*, 623–625.
13. Grimes, D.A. (2008). The nomogram epidemic: resurgence of a medical relic. Ann. Intern. Med. *149*, 273–275.
14. Dong, D., Fang, M.J., Tang, L., Shan, X.H., Gao, J.B., Giganti, F., Wang, R.P., Chen, X., Wang, X.X., Palumbo, D., et al. (2020). Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. Ann. Oncol. *31*, 912–920.
15. He, Y., Zhu, Z., Chen, Y., Chen, F., Wang, Y., Ouyang, C., Yang, H., Huang, M., Zhuang, X., Mao, R., et al. (2019). Development and validation of a novel diagnostic nomogram to differentiate between intestinal tuberculosis and Crohn's disease: a 6-year Prospective Multicenter Study. Am. J. Gastroenterol. *114*, 490–499.
16. Li, L., Zhang, B., He, B., Gong, Z., and Chen, X. (2020). Critical patients with coronavirus disease 2019: risk factors and outcome nomogram. J. Infect. *80*, e37–e38.
17. Dong, Y.M., Sun, J., Li, Y.X., Chen, Q., Liu, Q.Q., Sun, Z., Pang, R., Chen, F., Xu, B.Y., Manyande, A., et al. (2021). Development and validation of a nomogram for assessing survival in patients with COVID-19 pneumonia. Clin. Infect. Dis. *72*, 652–660.
18. Jeong, S.H., Kim, R.B., Park, S.Y., Park, J., Jung, E.J., Ju, Y.T., Jeong, C.Y., Park, M., Ko, G.H., Song, D.H., et al. (2020). Nomogram for predicting gastric cancer recurrence using biomarker gene expression. Eur. J. Surg. Oncol. *46*, 195–201.
19. Sakuma, Y., Ishitoya, J., Komatsu, M., Shiono, O., Hirama, M., Yamashita, Y., Kaneko, T., Morita, S., and Tsukuda, M. (2011). New clinical diagnostic criteria for eosinophilic chronic rhinosinusitis. Auris Nasus Larynx *38*, 583–588.
20. Seah, J.J., Thong, M., and Wang, D.Y. (2023). The Diagnostic and Prognostic Role of Biomarkers in Chronic Rhinosinusitis. Diagnostics *14;13*, 715.
21. Roufosse, F., and Weller, P.F. (2010). Practical approach to the patient with hypereosinophilia. J. Allergy Clin. Immunol. *126*, 39–44.
22. Meng, Y., Lou, H., Wang, C., and Zhang, L. (2016). Predictive significance of computed tomography in eosinophilic chronic rhinosinusitis with nasal polyps. Int. Forum Allergy Rhinol. *6*, 812–819.
23. Cui, Y., Wang, K., Shi, J., and Sun, Y. (2023). Endotyping Difficult-to-Treat Chronic Rhinosinusitis with Nasal Polyps by Structured Histopathology. Int. Arch. Allergy Immunol. *184*, 1036–1046.
24. Tan, B.K., Peters, A.T., Schleimer, R.P., and Hulse, K.E. (2018). Pathogenic and protective roles of B cells and antibodies in patients with chronic rhinosinusitis. J. Allergy Clin. Immunol. *141*, 1553–1560.
25. Helman, S.N., Barrow, E., Edwards, T., DelGaudio, J.M., Levy, J.M., and Wise, S.K. (2020). The Role of Allergic Rhinitis in Chronic Rhinosinusitis. Immunol. Allergy Clin. North Am. *40*, 201–214.
26. Kuhn, M., and Johnson, K. (2013). Applied Predictive Modeling, *Vol 26* (Springer).
27. Crowson, M.G., Ranisau, J., Eskander, A., Babier, A., Xu, B., Kahmke, R.R., Chen, J.M., and Chan, T.C.Y. (2020). A contemporary review of machine learning in otolaryngology-head and neck surgery. Laryngoscope *130*, 45–51.
28. Smith, K.A., Gill, A.S., Pollard, C.E., Sumsion, J.S., Saffari, H., Ashby, S., Witt, B.L., Shipman, P.A., Gabrielsen, D.A., Yim, M.T., et al. (2023). An eosinophil peroxidase activity assay accurately predicts eosinophilic chronic rhinosinusitis. J. Allergy Clin. Immunol. *152*, 400–407.
29. Thorwarth, R.M., Scott, D.W., Lal, D., and Marino, M.J. (2021). Machine learning of biomarkers and clinical observation to predict eosinophilic chronic rhinosinusitis: a pilot study. Int. Forum Allergy Rhinol. *11*, 8–15.

30. Snidvongs, K., Lam, M., Sacks, R., Earls, P., Kalish, L., Phillips, P.S., Pratt, E., and Harvey, R.J. (2012). Structured histopathology profiling of chronic rhinosinusitis in routine practice. Int. Forum Allergy Rhinol. *2*, 376–385.
31. Barham, H.P., Osborn, J.L., Snidvongs, K., Mrad, N., Sacks, R., and Harvey, R.J. (2015). Remodeling changes of the upper airway with chronic rhinosinusitis. Int. Forum Allergy Rhinol. *5*, 565–572.
32. Soler, Z.M., Sauer, D., Mace, J., and Smith, T.L. (2010). Impact of mucosal eosinophilia and nasal polyposis on quality-of-life outcomes after sinus surgery. Otolaryngol. Head Neck Surg. *142*, 64–71.
33. glmnet: lasso and elastic-net regularized generalized linear models. https://cran.r-project.org/web/packages/glmnet/index.html.
34. ranger:a fast implementation of randomforests. https://cran.r-project.org/web/packages/ranger/index.html.
35. xgboost:extreme gradient boosting. https://cran.r-project.org/web/packages/caret/index.html.
36. Github. R interface to Keras. https://github.com/rstudio/keras/.
37. caret: classification and regression training. https://cran.r-project.org/web/packages/caret/index.html.
38. The Lancet Respiratory Medicine (2018). Opening the black box of machine learning. Lancet Respir. Med. *6*, 801.
39. Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. J. Mach. Learn. Res. *19*, 1–5.
40. Vickers, A.J., and Elkin, E.B. (2006). Decision curve analysis: a novel method for evaluating prediction models. Med. Decis. Making *26*, 565–574.
41. DeLong, E.R., DeLong, D.M., and Clarke-Pearson, D.L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics *44*, 837–845.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Clinical data of patients with chronic sinusitis	This paper	N/A
Software and algorithms		
R (v4.3.0)	R CRAN	https://cran.r-project.org/
caret R package (v4.3.0)	R CRAN	https://cran.r-project.org/web/packages/caret/index.html
randomForest R package (v4.3.0)	R CRAN	https://cran.r-project.org/web/packages/randomForest/index.html
nnet R package (v4.3.0)	R CRAN	https://cran.r-project.org/web/packages/nnet/index.html
DALEX R package (4.3.0)	R CRAN	https://cran.r-project.org/web/packages/DALEX/index.html
rms R package (v4.3.0)	R CRAN	https://cran.r-project.org/web/packages/DALEX/index.html
glmnet R package (v4.3.0)	R CRAN	https://cran.r-project.org/web/packages/DALEX/index.html
ggplot2 R package (v4.3.0)	R CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and data should be directed to and will be fulfilled by the lead contact, Yucheng Yang (yychxh@163.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The complete original data reported in this study cannot be deposited in a public repository because these data are confidential medical records. The dataset used and analysed during the current study is available upon reasonable request, contact Mr. Yucheng Yang (yychxh@163.com).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This multicenter retrospective cohort study was approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University (ethics approval number: [2017]164). All the enrolled patients who provided written informed consent were recruited for this project. The enrollment criteria are as follows: (1) Patients aged >18 years; (2) diagnosis of CRS based on EPOS 2012; (3) patients received ESS treatment during hospitalization. Additionally, patients who met the following criteria were excluded: (1) Patients with incomplete baseline blood routine, tissue specimen, sinus CT, and nasal endoscopy data; (2) aged < 18 years; (3) patients with granulomatous polyangiitis, eosinophilic granulomatous polyangiitis, sarcoidosis, cystic fibrosis, fungal rhinosinusitis, and sinus malignant tumor. All included participants' data were from Chinese patients (n = 307; males = 216, 70.4%; females = 91, 29.6%).

Patients who met the diagnostic criteria of the 2012 European Position Paper on Rhinosinusitis and Nasal Polyposis (EPOS2012) guidelines for CRS underwent Endoscopic Sinus Surgery (ESS) from January 2015 to December 2022.

Ethics approval and consent to participate

This research was approved by the Ethics Committees of the First Affiliated Hospital of Chongqing Medical University (approval notice #[2017] 164) and conformed to the tenets of the Declaration of Helsinki.

METHOD DETAILS

Data were retrieved from medical records including age, sex, smoking history, surgical history(sinus surgery history), allergy history, lesion extent(bilateral CRS or not), asthma, allergic rhinitis, purulent discharge, the absolute value of blood eosinophils, percentage of blood

eosinophils, sinus CT Lund-Mackay score, the threshold ratio of total ethmoid sinus (E) and total maxillary (M) scores for both sides (E/M ratio, determined in a CT scan according to the Lund-Mackay scoring system), VAS overall subjective symptom score, VAS nasal obstruction score, VAS olfactory score.

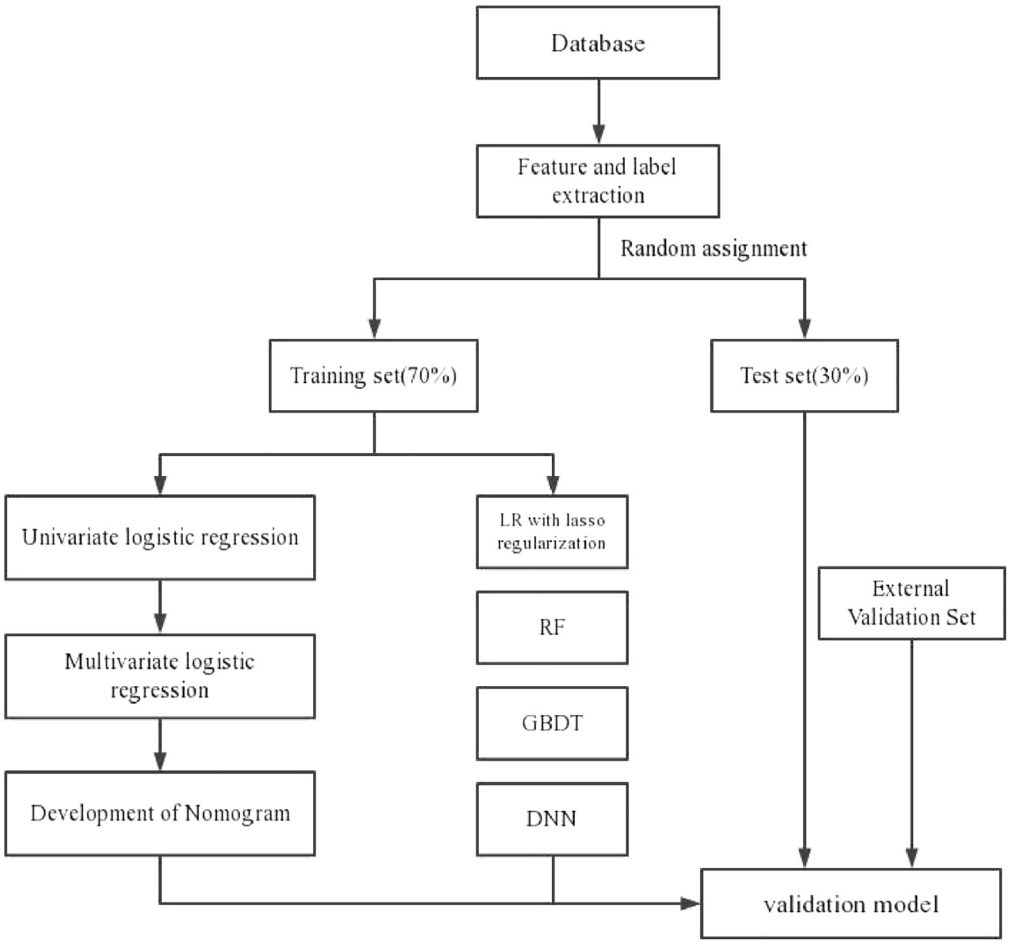
Histological analysis

Mucosal tissue from nasal polyps or ethmoidal polypoid lesions in CRS patients was obtained during surgery. Tissues were immediately fixed in 10% formalin, embedded in paraffin, sectioned, and stained with hematoxylin-eosin. Initial assessment on low power (magnification ×40) was performed in order to identify an area of densest cellular infiltrate. The number of eosinophils averaged over 3 high-power fields (HPFs) (magnification ×400) was then obtained and dichotomized to ≤10 or >10 eosinophils per HPF.[30] Patients with tissue eosinophil count>10/HPF were considered to be termed ECRS subtype, which was identified as a relevant cut-point based on previous publications.[31,32] Histological examination was performed by three separate expert physicians who were blinded to the clinical data.

QUANTIFICATION AND STATISTICAL ANALYSIS

Nomogram prediction model

In the training set (70% random sample), potential predictors were assessed using univariate logistic analysis, and predictors ($p<0.05$) were included in the multivariate logistic regression. Multivariate logistic regression analysis was applied to distinguish independent prognostic factors associated with predicting the ECRS subtype. Next, a nomogram prediction model was established based on independent prognostic factors and validated in both training and test sets. Nomogram development was performed by using the library "rms" in R for MACOS. The predictor variables registered in the nomogram met the requirement of 10 events per variable to reduce bias and variability in the logistic model. The development process of the nomogram is shown in the figure below.



The process of development of nomogram and machine learning methods

Machine learning model

In the training set (70% random sample), potential predictors were assessed using univariate logistic regression. Through the predictors described above ($p < 0.05$), we constructed four machine learning prediction models: (1) logistic regression with lasso regularization (lasso regression), (2) random forests, (3) gradient boosted decision trees, and (4) deep neural networks.[33–36] The development process for machine learning is shown in the figure in "nomogram prediction model" section.

For lasso regression, we chose the regularization parameter (lambda) which allowed the minimum misclassification error rate to penalize large coefficients from small sample sizes. The minimum lambda was calculated by the gillnet package using 10-fold cross-validation. Extending the standard regression model with lasso regularization allowed us to select important predictors (feature selection), which is more understandable and clinically useful. After using lasso regression for feature selection, the selected features were incorporated into the logistics regression to establish the model.

We used the random forest collection method from the decision tree created by using bootstrap samples of training data and random feature selection in tree induction. Random forest is a nonlinear machine learning model that captures more complexity in the data than linear and logistic regression. The feature importance given by random forest represents how well each feature performs in terms of prediction, allowing researchers to prioritize the most important features for their studies.

Gradient boosted decision trees were another ensemble approach-a decision tree addition model estimated by gradient descent. We used a grid search strategy to determine the best combination of hyperparameters by using the "ranger" and "caret" packages.[37] The search range was (100, 300, 500, 1000), (2, 10, 1), and (0.01, 0.03, 0.05, 0.1, 0.3, 0.5) for the number of estimators, maximum depth of the tree, and learning rate, respectively.

A deep neural network is a class of machine learning algorithms consisting of multiple layers of nonlinear networks that process the values of the parameters learned by the unit to make the best prediction of the outcome. A single artificial neuron takes in any number of input values, applies a specific mathematical function, and returns an output value. The function used is usually represented as

$$y\,=\sigma \left(\mathop{\sum }\limits_{i=1}^{n}({w}_{i}{x}_{i})+b\right)$$

where xi represents a single input variable or feature (there are n such inputs), wi represents a learnable weight for that input, b represents a learnable bias term and σ represents a non-linear activation function that takes a single input and returns a single output. To create a network, artificial neurons are arranged in layers, with the output of one layer being the input of the next.

Algorithm development improves the complexity of the model, such as in the ensemble or deep learning models, which further complicates the interpretation of the model. Therefore, opening the 'black box' of Machine Learning Models is crucial since it allows clinicians to easily understand the internal logic of each prediction.[38] In response to this problem, we created nomogram2 to explain the LR with the lasso regression model. For RF and GBDT models, we calculated the variable importance of each output to gain insight into the contribution of each predictor variable to the machine learning model.[26] For the DNN model, we used the breakdown method to explain it. The moDel Agnostic Language for Exploration and eXplanation (DALEX) was a comprehensive packaged algorithms system based on the principle of the Breakdown and helps in calculating local and global feature importance.[39] The above model visualization methods can support clinicians without algorithmic backgrounds to better understand the model.

Model evaluation

Features and characteristics are represented by mean ± standard deviation (Mean ± SD) for continuous variables and as count (percentage) for categorical data. Continuous variables were compared using the t-test and Kruskal-Wallis test, while the Chi-square test compared categorical variables. A two-sided p-value of <0.05 was considered statistically significant. For the nomogram, the receiver operating characteristics (AUC) curves of the models were evaluated for the resolving power, and the calibration curves were measured for the calibration power. Meanwhile, model accuracy, assessed by the Brier score, improved as the score approached 0. For machine learning models, in the test set (30% random sample) and independent external validation set, we calculated (1) AUC (i.e., area under receiver operating characteristic [ROC] curve); (2) prospective predictive outcomes (i.e., sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, negative likelihood ratio); and (3) decision curve analysis.[40] Decision curve analysis was a measure that took different weights and types of misclassifications into account and had a direct clinical interpretation (e.g., a trade-off between under-and over-classification for each model).[40] Specifically, the relative impact of false negatives (false positive results given a threshold probability or clinical preference) was considered to produce a net benefit in each model. The net benefit of each model within the specified outcome threshold probability is graphically displayed as a decision curve. ROC curves were compared by the DeLong test.[41] The value of $p < 0.05$ was considered statistically significant. The above data were collected by IBM SPSS Statistics 27.0 and R software version 4.3.0 (R Foundation for Statistical Computing, Vienna, Austria; www.R-project.org).