

Machine learning of biomarkers and clinical observation to predict eosinophilic chronic rhinosinusitis: a pilot study

Ryan M. Thorwarth, MD¹, Derek W. Scott, BA², Devyani Lal, MD^{1,2}  and Michael J. Marino, MD^{1,2} 

Background: Subtyping chronic rhinosinusitis (CRS) by tissue eosinophilia has prognostic and therapeutic implications, and is difficult to predict using peripheral eosinophil counts or polyp status alone. The objective of this study was to test machine learning for prediction of eosinophilic CRS (eCRS).

Methods: Input variables were defined as peripheral eosinophil count, urinary leukotriene E₄ (uLTE₄) level, and polyp status. The output was diagnosis of eCRS, defined as tissue eosinophil count >10 per high-power field. Patients undergoing surgery for CRS were retrospectively reviewed for complete datasets. Univariate analysis was performed for each input as a predictor of eCRS. Logistic regression and artificial neural network (ANN) machine learning models were developed using random and surgeon-specific training/test datasets.

Results: A total of 80 patients met inclusion criteria. In univariate analysis, area under the receiver operator characteristic curve (AUC) for peripheral eosinophil count and uLTE₄ were 0.738 (95% confidence interval [CI], 0.616 to 0.840) and 0.728 (95% CI, 0.605 to 0.822), respectively. Presence of polyps was 94.1% sensitive, but 51.7% specific. Logistic regression models using random and surgeon

specific datasets resulted in AUC of 0.882 (95% CI, 0.665 to 0.970) and 0.945 (95% CI, 0.755 to 0.995), respectively. ANN models resulted in AUC of 0.918 (95% CI, 0.756 to 0.975) and 0.956 (95% CI, 0.828 to 0.999) using random and surgeon-specific datasets, respectively. Model comparison of logistic regression and ANN was not statistically different. All machine learning models had AUC greater than univariate analyses (all $p < 0.003$).

Conclusion: Machine learning of 3 clinical inputs has the potential to predict eCRS with high sensitivity and specificity in this patient population. Prospective investigation using larger and more diverse populations is warranted. © 2020 ARS-AAOA, LLC.

Key Words:

chronic rhinosinusitis; biomarker; machine learning; eosinophil; leukotriene E₄

How to Cite this Article:

Thorwarth RM, Scott DW, Lal D, Marino MJ. Machine learning of biomarkers and clinical observation to predict eosinophilic chronic rhinosinusitis: a pilot study. *Int Forum Allergy Rhinol.* 2020;00:1-8.

Chronic rhinosinusitis (CRS) is traditionally and most commonly subtyped according to phenotypic classification of patients with (CRSwNP) and without nasal polyps (CRSSNP). Recently, there has been interest in CRS endotypes that more accurately parallel underlying inflammatory mechanisms.¹⁻³ Structured histopathology reports have also been used to identify pathologic features of specimens collected during endoscopic sinus surgery (ESS)

and how this information can be used by clinicians for prognosis.⁴⁻⁶ Nevertheless, clinical trials of targeted biologic therapies for CRS continue to use phenotypic classification as study inclusion criteria.⁷ The widespread use of an endotype approach to disease classification in routine clinical practice also remains limited.

Identifying patients with an eosinophilic CRS (eCRS) subtype has been of particular interest. The subtype is associated with more severe symptoms, higher likelihood of resistance to treatment, and more frequent disease recurrence.^{4,8} Various cutoff values for the number of tissue eosinophils have been used to define eCRS, although there is recognition that identifying this subtype may inform the physician about prognosis, treatment decisions, and clinical monitoring posttreatment.^{4,8,9} Nasal polyps have been associated with eCRS; however, 10% to 50% of CRSwNP patients do not have tissue eosinophilia.^{1,5,10} Conversely, lymphocytic and lymphoplasmacytic inflammatory cell

¹Department of Otorhinolaryngology, Mayo Clinic, Phoenix, AZ; ²Alix School of Medicine, Mayo Clinic, Scottsdale, AZ

Correspondence to: Michael J. Marino, MD, 5777 E Mayo Blvd, Phoenix, AZ 85054; e-mail: marino.michael@mayo.edu

Potential conflict of interest: None provided.

Accepted for oral presentation at the at the Annual ARS Meeting on September 11-12, 2020, in Boston, MA.

Received: 9 March 2020; Revised: 21 May 2020; Accepted: 22 May 2020

DOI: 10.1002/alr.22632

View this article online at wileyonlinelibrary.com.

predominance presented as CRSwNP in up to 30% of cases,¹¹ and tissue eosinophilia was present in 18% of patients with CRSsNP.⁵ To this extent the CRS phenotype does not completely parallel an underlying eosinophilic endotype.

Numerous biomarkers for eCRS have been investigated, including total serum immunoglobulin E (IgE), erythrocyte sedimentation rate (ESR), C-reactive protein (CRP), and serum peripheral eosinophil count. IgE, ESR, and CRP were all generally poor predictors of tissue eosinophilia.¹⁰ Peripheral eosinophil count has been shown to be a fair predictor of tissue eosinophilia by receiver operator characteristic (ROC) analysis, although sensitivity and specificity does not simultaneously exceed 80%.^{4,10} Separately, urinary leukotriene E4 (uLTE4) has been described as a biomarker for aspirin-exacerbated respiratory disease (AERD),^{12,13} and shown to be elevated in CRS patients with comorbid asthma or allergy.¹⁴ Tissue LTE4 has been used to differentiate patients with eosinophilic and non-eosinophilic CRSwNP,^{15,16} and is correlated with interleukin 5 (IL-5) expression.¹⁷ uLTE4, however, has not been specifically studied as a noninvasive biomarker for tissue eosinophilia in CRS patients.

A sensitive and specific noninvasive biomarker strategy for the identification of CRS patients with an eosinophilic endotype would aid prognosis and individualizing treatment. An endotype approach to CRS classification will be increasingly important as targeted biologic therapies become available, and may identify patients that do not meet current criteria for these agents. Potential noninvasive biomarkers have been described for defining CRS subtypes, although achieving simultaneous high sensitivity and specificity for tissue eosinophilia has been challenging. Supervised machine learning is a subset of artificial intelligence that aims to use training input and output data in order to make predictions from new input data. Machine learning is beginning to be used in otolaryngology,^{18,19} and rhinology-specific applications remain limited.^{20,21} The aim of this study was to determine if a machine learning approach could be used to make accurate predictions of tissue eosinophilia from peripheral eosinophil counts, uLTE4 levels, and polyp status. A secondary objective was to describe the predictive ability of uLTE4 for tissue eosinophilia.

Patients and methods

Patients having ESS for the treatment of CRS from July 1, 2011, through November 1, 2019, were retrospectively identified. Input variables were defined as absolute peripheral eosinophil count, uLTE4 level, and notation of the presence of polyps on initial in-office nasal endoscopy. Eosinophil count (cells $\times 10^9/L$) and uLTE4 level (pg/mg creatinine) were reported as continuous variables, whereas polyp status was recorded as a binary categorical variable. The outcome variable was defined as tissue eosinophil count as reported on surgical structured histopathology reports as described by Snidvongs et al.⁴ Patients with

all 3 input variables and the outcome variable of tissue eosinophil count were considered for study inclusion in a pilot investigation of this select population. Exclusion criteria were age <18 years and diagnoses of granulomatosis with polyangiitis, eosinophilic granulomatosis with polyangiitis, and sarcoidosis.

Tissue eosinophil count was completed in an analogous fashion to that reported by Soler et al.²² At the time of surgery, a sample of ethmoidal cavity mucosa was obtained. The sample was fixed in formalin and underwent hematoxylin-eosin staining. The slides were then reviewed by a board-certified pathologist (Matthew Zarka, M.D. Department of Pathology, Mayo Clinic, AZ) with an interest in head and neck pathology. Initial assessment on low power (magnification $\times 40$) was performed in order to identify an area of densest cellular infiltrate. The number of eosinophils averaged over 2 high-power fields (HPFs) (magnification $\times 400$) was then obtained and dichotomized to ≤ 10 or > 10 eosinophils per HPF. Patients with tissue eosinophil counts > 10 per HPF were considered to have an eCRS subtype based on previous publications identifying this as a relevant cutoff point.^{4,22,23}

Demographic and clinical characteristics were recorded from the medical record. Clinical diagnoses of interest included allergic fungal rhinosinusitis (AFRS), allergic rhinitis, asthma, and aspirin-exacerbated respiratory disease (AERD). Active prescriptions for leukotriene receptor antagonists, zileuton, and monoclonal antibodies to IgE, IL-5, and IL-4/13 at the time of laboratory collection for peripheral eosinophil count and uLTE4 were recorded.

Univariate analysis was performed for peripheral eosinophil count, uLTE4 level, and polyp status as a predictor of eCRS. Receiver operator characteristic (ROC) curve and area under the ROC curve (AUC) were used to evaluate the continuous variables. The optimal cut point was selected using Youden's J statistic, and sensitivity and specificity relative to the optimal cut point were calculated. Sensitivity and specificity of polyp status as a predictor of eCRS were also calculated.

The entire study population was then randomly divided into training and test datasets, with two-thirds assigned to the training data and one-third assigned to the test population (Fig. 1). Machine learning models, 1 using binary logistic regression and a second using an artificial neural network (ANN), were developed from the training data. The models were then used within the test data to predict eCRS from the 3 input variables. The AUC was calculated for each model, in addition to the sensitivity, specificity, and accuracy of the model's predictions. A second experiment was conducted where both binary logistic regression and ANN were trained using 1 surgeon's population (accounting for 75% of the total dataset) to make predictions in another surgeon's population (Fig. 1). Again the AUC, sensitivity, specificity, and accuracy were calculated. The nonparametric DeLong test was used to compare AUC values between machine learning models and univariate analyses, to determine if there were statistically significant

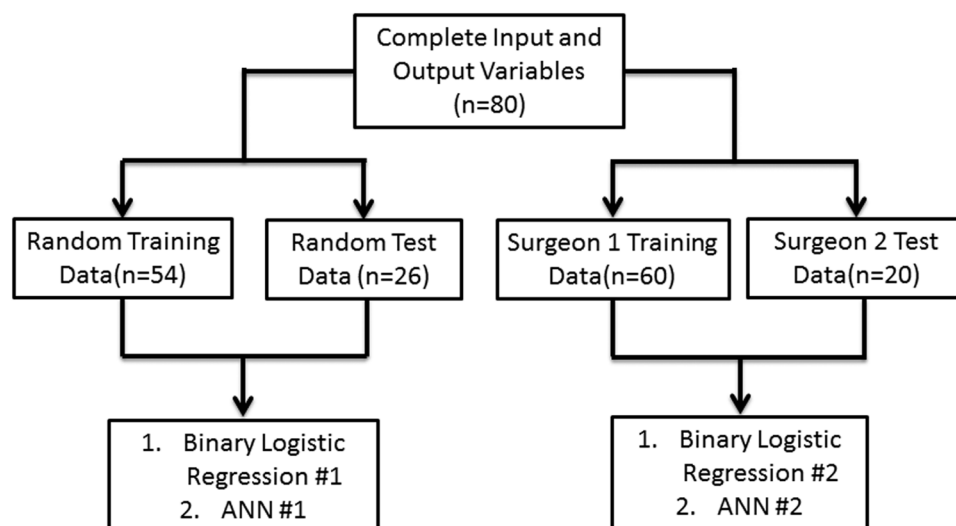


FIGURE 1. Schematic diagram of the study workflow for dividing patients into model development groups.

differences. ANN function parameters were also tested using cross-validation with a one-third holdback method.

Statistical analyses, including development of machine learning models, were performed using JMP Pro, version 14.1.0 (SAS Institute Inc., Cary, NC). Binary logistic regression was performed and the whole model was assessed using the likelihood ratio chi-square statistic. The individual regression coefficients (β) were calculated for each regression parameter and tested using the Wald chi-square statistic. In all instances ANNs were developed with a single layer using sigmoid activation functions, learning rate set to 0.1, and without boosting (Fig. 2). Comparison of 2 continuous variables was performed using Student *t* test, whereas 2 categorical variables were compared with chi-square test. Values of $p < 0.05$ were considered significant. The study was approved by the institutional review board of the Mayo Clinic, Phoenix, AZ.

Results

A total of 761 patients had ESS for treatment of CRS within the study period. Within that dataset, 80 patients

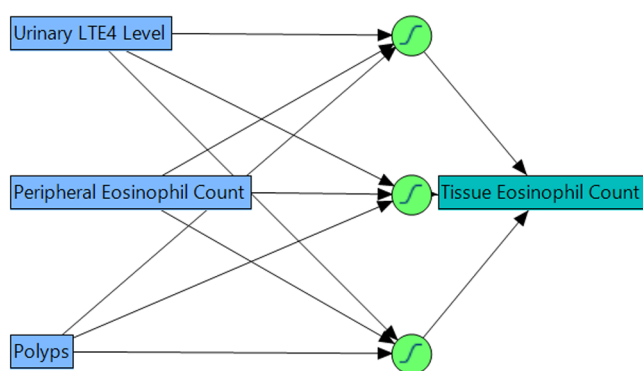


FIGURE 2. Schematic diagram of artificial neural network depicting input variables, single hidden layer, and prediction output.

were identified as having all the input and output variables. A total of 60 patients were under the care of 1 attending surgeon, while the remaining 20 patients were under the care of a second attending surgeon (Fig. 1). All patients were ≥ 18 years of age, and no patient had an exclusion diagnosis. The demographic and clinical characteristics of the entire study population are shown in Table 1.

Univariate ROC analysis of peripheral eosinophil count as a predictor of eCRS was statistically significant ($p < 0.001$) with an AUC of 0.738 (95% confidence interval [CI], 0.616 to 0.840), and the optimal cut point was calculated to be $0.52 \text{ cells} \times 10^9/\text{L}$. ROC analysis of uLTE4 level was also statistically significant ($p < 0.001$) with an AUC of 0.728 (95% CI, 0.605 to 0.822), and the optimal cut point was calculated to be 106 pg/mg creatinine. The sensitivity and specificity of polyp status, in addition to

TABLE 1. Demographic and clinical characteristics of the study population

Characteristic	Value
Age (years), mean $\pm$ SD	48.2 $\pm$ 15.8
Male, n (%)	42 (52.5)
Female, n (%)	38 (47.5)
Nasal polyps, n (%)	62 (77.5)
Asthma, n (%)	59 (73.8)
Allergic rhinitis, n (%)	24 (30.0)
Aspirin-exacerbated respiratory disease, n (%)	21 (26.3)
Peripheral eosinophil count (cells $\times 10^9/\text{L}$), mean $\pm$ SD	0.56 $\pm$ 0.45
Urinary LTE4 level (pg/mg creatinine), mean $\pm$ SD	219.0 $\pm$ 757.1
Tissue eosinophil count > 10 per HPF, n (%)	51 (63.8)

HPF = high-power field; LTE4 = leukotriene E4; SD = standard deviation.

TABLE 2. Predictive ability of univariate and multivariate machine learning models for eCRS

Model	AUC (95% CI)	Sensitivity (95% CI) (%)	Specificity (95% CI) (%)
Univariate peripheral eosinophil count	0.738 (0.616–0.840)	51.0 (37.7–64.1)	86.2 (69.4–94.5)
Univariate urinary LTE4 level	0.728 (0.605–0.822)	56.9 (43.3–69.5)	86.2 (69.4–94.5)
Univariate polyp status	N/A	94.1 (84.1–98.0)	51.7 (34.4–68.6)
Logistic regression (random dataset)	0.882 (0.665–0.970)	82.4 (56.6–96.2)	80.0 (44.4–97.5)
ANN (random dataset)	0.918 (0.756–0.975)	88.2 (63.6–98.5)	80.0 (44.4–97.5)
Logistic regression (surgeon-specific dataset)	0.945 (0.755–0.995)	84.6 (54.6–98.1)	85.7 (42.1–99.6)
ANN (surgeon-specific dataset)	0.956 (0.828–0.999)	92.3 (65.0–99.8)	85.7 (42.1–99.6)

ANN = artificial neural network; AUC = area under the curve receiver operator characteristic curve; CI = confidence interval; eCRS = eosinophilic chronic rhinosinusitis; LTE4 = leukotriene E4; N/A = not applicable.

peripheral eosinophil count and uLTE4 level at the optimal cut point, are shown in Table 2.

Using random training data, the binary logistic regression model had an AUC of 0.882 (95% CI, 0.665 to 0.970) for predicting eCRS in the test dataset, and overall accuracy was 81.5% (Fig. 3A). The logistic regression model was significant by whole model testing ($\chi^2 = 41.5$, $p < 0.001$), and for individual regression parameters (Table 3). The ANN that was developed using the random datasets had an AUC of 0.918 (95% CI, 0.756 to 0.975), with an overall accuracy of 85.2% (Fig. 3B). The AUC for the logistic regression and ANN models using the random datasets were not statistically different ($p = 0.462$), although both models had AUC greater than either univariate peripheral eosinophil count or uLTE4 ROC analysis (all $p < 0.003$, Table 4). Cross-validation using the entire dataset with one-third holdback

and identical ANN function parameters yielded an AUC of 0.882 (95% CI, 0.718 to 0.999) and 85.2% accuracy.

The machine learning algorithms were then trained with the data from 1 surgeon's patient population and tested in the second surgeon's population. Using surgeon specific datasets, the binary logistic regression model had an AUC of 0.945 (95% CI, 0.756 to 0.975) and an overall accuracy of 85.0% (Fig. 4A). The logistic regression model was significant by whole model testing ($\chi^2 = 28.5$, $p < 0.001$), and for individual regression parameters (Table 3). The ANN using the surgeon specific datasets had an AUC of 0.956 (95% CI, 0.828 to 0.999), with an overall accuracy of 90.0% (Fig. 4B). The AUC for the logistic regression and ANN models using the surgeon specific datasets were not statistically different ($p = 0.640$), although all models had AUC greater than either univariate analysis (all

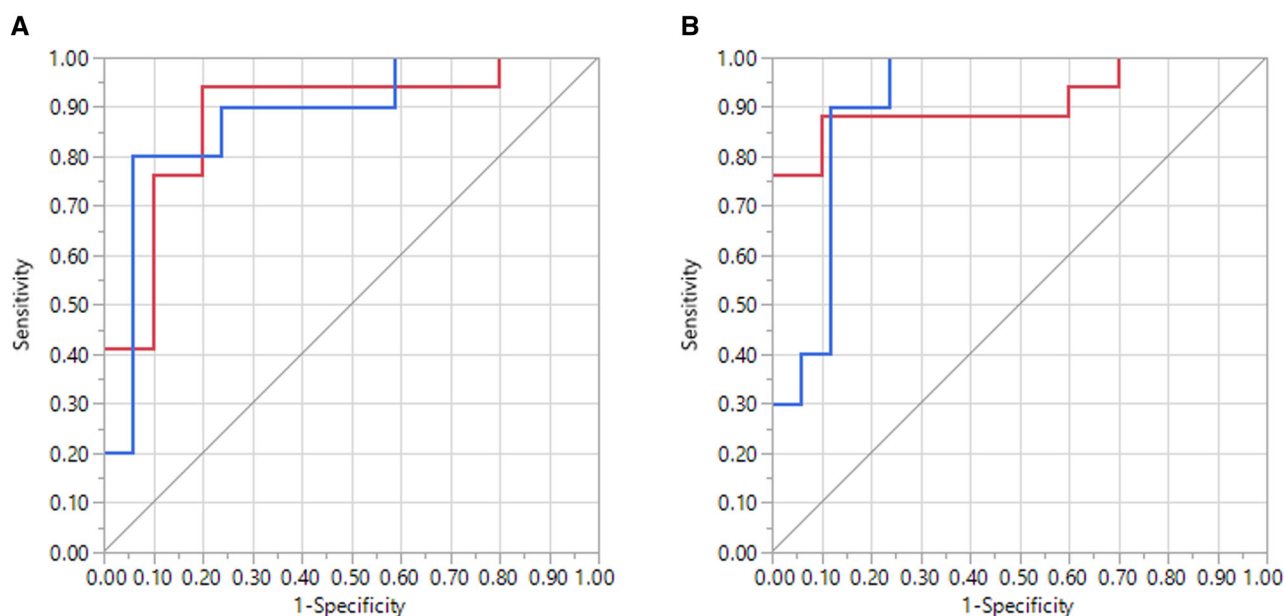


FIGURE 3. ROC curve of binary logistic regression model (A) and artificial neural network (B) for prediction of tissue eosinophilia using random training and test data. Red curves represent prediction of tissue eosinophil count ≥ 10 per high-power field, whereas blue represents inverse curve for the prediction of tissue eosinophil count < 10 per high-power field. ROC = receiver operator characteristic.

TABLE 3. Binary logistic regression model parameters

Parameter	β	SE	Wald chi-square test	<i>p</i>
Model parameter (random dataset)				
Peripheral eosinophil count	2.21	0.99	4.50	0.025
Urinary LTE4	0.01	0.01	5.00	0.026
Polyp status	1.38	0.41	11.26	0.008
Model parameter (surgeon-specific dataset)				
Peripheral eosinophil count	2.02	1.03	3.82	0.021
Urinary LTE4	0.01	0.01	3.70	0.023
Polyp status	1.10	0.43	6.69	0.005

β = regression coefficient; LTE4 = leukotriene E4; SE = standard error.

$p < 0.003$, Table 4). The sensitivity and specificity for all machine learning model predictions of the respective test dataset are shown in Table 2.

The patients in the surgeon specific data used for model testing were more likely to have AFRS ($p = 0.025$) and allergic rhinitis ($p = 0.005$, Table 5). There were no other differences in age, gender, zileuton, and leukotriene receptor antagonist prescriptions, and comorbid disease in the random and surgeon-specific training and test datasets (Table 5). A single patient had been prescribed mepolizumab at the time of laboratory collection, and this patient was included in the random training dataset and the surgeon-specific test dataset. No other patients had been prescribed a relevant monoclonal antibody.

Discussion

The current study represents a preliminary investigation of leveraging machine learning technology for complex interpretation of multiple biomarkers to identify eCRS. The underlying input parameters were investigated individually in a univariate analysis. Peripheral eosinophil count alone resulted in an AUC of 0.738, and was consistent with pre-

vious reports.^{4,10} uLTE4 was also investigated as a novel biomarker for eCRS and had fair predictive value (AUC = 0.728). Presence of polyps was sensitive, but not specific for tissue eosinophilia. The predictive ability of multivariate models using these input variables was good to excellent and reproducible with different training and test datasets, and in cross-validation. The AUC for all machine learning models was also greater than univariate analyses of peripheral eosinophil counts and uLTE4 levels. Neural networks, however, did not have significantly better predictive value than logistic regression in the study population.

Peripheral eosinophil count, uLTE4 level, and nasal polyps have been linked to tissue eosinophilia in CRS, but are incomplete predictors as lone biomarkers.^{4,5,10,12-14} Differentiating eosinophilic and non-eosinophilic CRS noninvasively remains challenging despite recognition that there are important clinical implications for identifying these subtypes. Machine learning is increasingly being used to process multiple input variables to make classification predictions, and has become routine in popular culture. Applications of machine learning technologies in otolaryngology are relatively new, and implementation in clinical practice has been limited.^{18,19} Nevertheless, there

TABLE 4. Values of *p* for comparison of AUC between all predictive models*

Prediction models	Univariate peripheral eosinophil count	Univariate uLTE4 level	Logistic regression (random dataset)	Logistic regression (surgeon-specific dataset)	ANN(random dataset)
Univariate uLTE4 level	0.894				
Logistic regression (random dataset)	0.002	<0.001			
Logistic regression (surgeon-specific dataset)	0.002	0.001	0.480		
ANN (random dataset)	0.002	0.001	0.472	0.462	
ANN (surgeon-specific dataset)	0.002	0.001	0.776	0.640	0.463

*Bold values are significant.

ANN = artificial neural network; AUC = area under the receiver operator characteristic curve; uLTE4 = urinary leukotriene E4.

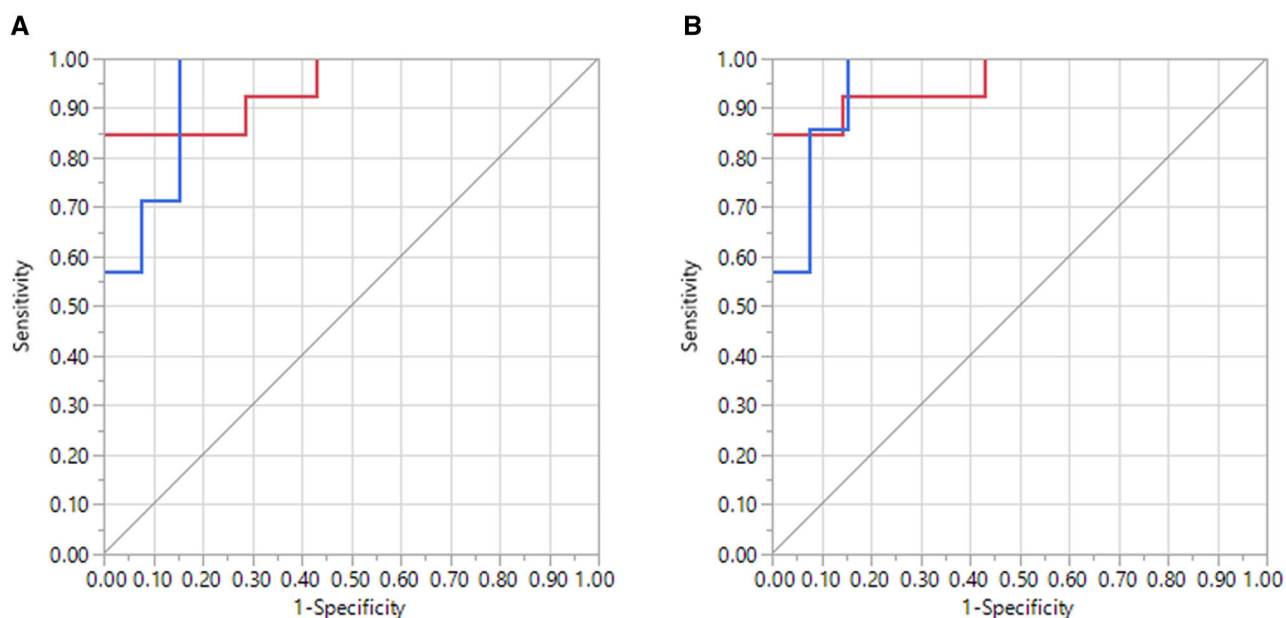


FIGURE 4. ROC curve of binary logistic regression model (A) and artificial neural network (B) for prediction of tissue eosinophilia using surgeon specific training and test data. Red curves represent prediction of tissue eosinophil count ≥ 10 per high-power field, whereas blue represents inverse curve for the prediction of tissue eosinophil count < 10 per high-power field. ROC = receiver operator characteristic.

are interesting possibilities for data processing and decision support of multivariable inputs that would be impractical without artificial intelligence assistance.

Univariate analysis of the input variables is important for understanding the contribution of each model feature, and is the first description of uLTE4 as a marker of tissue eosinophilia to our knowledge. The predictive ability of peripheral eosinophil count determined in this study is similar to previously reported data.^{4,10} The optimal cut point of $0.52 \text{ cells} \times 10^9/\text{L}$ is higher than the values reported by Snidvongs et al.,⁴ and Ho et al.,¹⁰ although this is mostly a function of the statistical method used in this study. The cut point could be lowered in the current dataset, which

would lead to higher sensitivity at the expense of specificity. The metabolism of arachidonic acid into various cysteinyl leukotrienes also has a role in upper airway inflammatory disease including asthma, allergic rhinitis, and CRS. LTE4 represents the stable metabolite of this cascade and can be measured in random urine samples. uLTE4 was a fair predictor of tissue eosinophilia, with a lower optimal cut point than has been suggested for diagnosis of AERD.^{12,13} The lower cut point value for tissue eosinophilia may represent a spectrum of disease between eCRS and AERD. Overall, both peripheral eosinophil count and uLTE4 were fair predictors of tissue eosinophilia, although high sensitivity and specificity could not be achieved simultaneously.

TABLE 5. Comparison of demographic and clinical characteristics by training and test datasets

Characteristic	Random training dataset	Random test dataset	<i>p</i>	Surgeon specific training dataset	Surgeon specific test dataset	<i>p</i>
Age (years), mean $\pm$ SD	46.8 $\pm$ 16.1	51.0 $\pm$ 15.2	0.261	47.2 $\pm$ 15.6	51.3 $\pm$ 16.4	0.337
Male, n (%)	26 (49.1)	16 (59.3)	0.388	33 (55.0)	9 (45.0)	0.438
Nasal polyps, n (%)	42 (79.3)	20 (74.1)	0.601	46 (76.7)	16 (80.0)	0.757
Asthma, n (%)	40 (75.5)	19 (70.4)	0.624	44 (73.3)	15 (75.0)	0.883
Allergic rhinitis, n (%)	16 (30.2)	8 (29.6)	0.959	13 (21.7)	11 (55.0)	0.005
AFRS, n (%)	5 (9.4)	4 (14.8)	0.471	4 (6.7)	5 (25.0)	0.025
AERD, n (%)	14 (26.4)	7 (25.9)	0.963	17 (28.3)	4 (20.0)	0.463
LT receptor antagonists, n (%)	27 (50.9)	12 (44.4)	0.582	28 (46.7)	11 (55.0)	0.519
Zileuton, n (%)	3 (5.7)	2 (7.4)	0.760	2 (3.33)	3 (15.0)	0.097

AERD = aspirin-exacerbated respiratory disease; AFRS = allergic fungal rhinosinusitis; LT = leukotriene; SD = standard deviation.

CRS phenotype was associated with underlying eCRS; however, this remains incomplete. In particular specificity of nasal polyps for eCRS was poor at 51%, and reflects that a sizeable portion (22.6%) of CRSwNP patients that did not have tissue eosinophil counts >10 per HPF. Presence of nasal polyps was sensitive for eCRS in the current dataset, although this may be related to a limitation of the retrospective study population. The percentage of CRSsNP patients in the study (22.5%) is lower than the general population, and is due to using uLTE4 as an input variable. Typically uLTE4 has been used in our practice as part of the diagnostic evaluation for AERD, and as such was more frequently ordered in patients with nasal polyps. A study population that more closely represents the distribution of CRSwNP and CRSsNP in the general population might affect the sensitivity of nasal polyps for eCRS.

There has been some skepticism about the application of machine learning and artificial intelligence to healthcare and clinician decision making,²⁴ and practical uses in otolaryngology have been limited.^{18,19} The reproducibility and opaqueness of machine learning models is a particular area of concern.²⁴ Datasets have the potential to include thousands of variables and large models can be difficult for human interpretation. Supervised machine learning is also dependent on accurate data labeling, and inaccuracies in training datasets can affect predictions of real-world inputs. The advantages of machine learning algorithms, however, are the ability to tackle complex problems involving heterogeneous input data (variable type and distribution) that would otherwise be difficult to approach with human interpretation of traditional statistical methods. In the context of the current study, models are able to make simultaneous decisions about the multiple input variables, which have finer granularity than the optimal cut point values determined in univariate analysis. Limiting the number of input variables also allows for meaningful interpretations of the contributions of model features. Model development in this study was reproducible across varied training and test datasets, as well as with cross-validation. Ultimately, models can be made widely available for interpretation of real world datasets in order to support clinician decision making and potentially allow for a noninvasive strategy to identify eCRS patients.

There are several limitations in the current study. The entire study population was relatively small at 80 patients, and is related to using uLTE4 as an input variable. This raises the possibility of overfitting in model development, although this was mitigated by using only 3 input variables. In general the sample size in each training set was 20 times the number of input variables, and exceeds the 10 times rule that has often been suggested for multivariable model development. Because of these limitations preference was given to objective biomarkers and clinical features, such as polyp status, that are easily assessed by the otolaryngologist. As mentioned previously the limited number of patients with uLTE4 resulted in a study population with a higher percentage of asthma and CRSwNP than the general population. This is a recognized selection bias in the study and may limit the generalizability of the pilot investigation. Prospective study of patients with CRS subtypes and comorbidities representative of the general population would be beneficial in verifying the accuracy of these predictive methods. Furthermore the current dataset was acquired within the same institution and may also limit generalizability. Model testing was performed between surgeon-specific populations as an attempt to demonstrate reproducibility and generalizability. Nevertheless, future study would benefit from larger datasets including additional features and model testing across institutional and geographic sites.

Conclusion

Peripheral eosinophil count, uLTE4, and polyp status can be used simultaneously in multivariable logistic regression and ANN machine learning models for prediction of eCRS with sensitivity and specificity greater than 80% in the study population. uLTE4 is also a fair predictor of eCRS at optimal cut point values lower than those reported for diagnosis of AERD. Machine learning of multiple biomarkers may allow for prediction of CRS endotypes, and offer clinician decision support of disease endotypes. Larger and multi-institutional datasets, that are widely representative of diverse CRS subtypes and comorbidities, will be critical for improving the accuracy and generalizability of these models. 

References

- Tomassen P, Vandeplas G, Van Zele T, et al. Inflammatory endotypes of chronic rhinosinusitis based on cluster analysis of biomarkers. *J Allergy Clin Immunol*. 2016;137:1449-1456 e1444.
- Divekar R, Rank M, Squillace D, Kita H, Lal D. Unsupervised network mapping of commercially available immunoassay yields three distinct chronic rhinosinusitis endotypes. *Int Forum Allergy Rhinol*. 2017;7:373-379.
- Liao B, Liu JX, Li ZY, et al. Multidimensional endotypes of chronic rhinosinusitis and their association with treatment outcomes. *Allergy*. 2018;73:1459-1469.
- Snidvongs K, Lam M, Sacks R, et al. Structured histopathology profiling of chronic rhinosinusitis in routine practice. *Int Forum Allergy Rhinol*. 2012;2:376-385.
- Kuhar HN, Tajudeen BA, Mahdavinia M, Gattuso P, Ghai R, Batra PS. Inflammatory infiltrate and mucosal remodeling in chronic rhinosinusitis with and without polyps: structured histopathologic analysis. *Int Forum Allergy Rhinol*. 2017;7:679-689.
- Marino MJ, Garcia JO, Zarka M, Lal D. A structured histopathology-based analysis of surgical outcomes in chronic rhinosinusitis with and without nasal polyps. *Laryngoscope Invest Otolaryngol*. 2019;4:497-503.
- Bachert C, Han JK, Desrosiers M, et al. Efficacy and safety of dupilumab in patients with severe chronic rhinosinusitis with nasal polyps (LIBERTY NP SINUS-24 and LIBERTY NP SINUS-52): results from two multicentre, randomised, double-blind, placebo-controlled, parallel-group phase 3 trials. *Lancet (North American Edition)*. 2019;394:1638-1650.
- McHugh T, Snidvongs K, Xie M, Banglawala S, Sommer D. High tissue eosinophilia as a marker to predict recurrence for eosinophilic chronic rhinosinusitis: a systematic review and meta-analysis. *Int Forum Allergy Rhinol*. 2018;8:1421-1429.
- Grayson JW, Cavada M, Harvey RJ. Clinically relevant phenotypes in chronic rhinosinusitis. *J Otolaryngol Head Neck Surg*. 2019;48:23.
- Ho J, Hamizan AW, Alvarado R, Rimmer J, Sewell WA, Harvey RJ. Systemic predictors of eosinophilic chronic rhinosinusitis. *Am J Rhinol Allergy*. 2018;32:252-257.
- Marino MJ, Garcia JO, Zarka MA, Lal D. Inflammatory cell predominance and patterns in chronic rhinosinusitis with and without nasal polyposis patients. *Laryngoscope Invest Otolaryngol*. 2019;4:573-577.
- Divekar R, Hagan J, Rank M, et al. Diagnostic utility of urinary LTE4 in asthma, allergic rhinitis, chronic

- rhinosinusitis, nasal polyps, and aspirin sensitivity. *J Allergy Clin Immunol Pract.* 2016;4:665-670.
13. Hagan JB, Laidlaw TM, Divekar R, et al. Urinary leukotriene E4 to determine aspirin intolerance in asthma: a systematic review and meta-analysis. *J Allergy Clin Immunol Pract.* 2017;5:990-997.e991.
 14. Santarelli GD, Lam KK, Han JK. Establishing urinary leukotriene E4 as a diagnostic biomarker for chronic rhinosinusitis with comorbid asthma and atopy. *Otolaryngol Head Neck Surg.* 2019;161:764-769.
 15. Steinke JW, Bradley D, Arango P, et al. Cysteinyl leukotriene expression in chronic hyperplastic sinusitis-nasal polyposis: importance to eosinophilia and asthma. *J Allergy Clin Immunol.* 2003;111:342-349.
 16. Wu X, Hong H, Zuo K, et al. Expression of leukotriene and its receptors in eosinophilic chronic rhinosinusitis with nasal polyps. *Int Forum Allergy Rhinol.* 2016;6:75-81.
 17. Perez-Novo CA, Watelet JB, Claeys C, Van Cauwenberge P, Bachert C. Prostaglandin, leukotriene, and lipoxin balance in chronic rhinosinusitis with and without nasal polyposis. *J Allergy Clin Immunol.* 2005;115:1189-1196.
 18. Crowson MG, Ranisau J, Eskander A, et al. A contemporary review of machine learning in otolaryngology-head and neck surgery. *Laryngoscope.* 2020;130:45-51.
 19. Bur AM, Shew M, New J. Artificial intelligence for the otolaryngologist: a state of the art review. *Otolaryngol Head Neck Surg.* 2019;160:603-611.
 20. Chowdhury NI, Li P, Chandra RK, Turner JH. Baseline mucus cytokines predict 22-item Sino-Nasal Outcome Test results after endoscopic sinus surgery. *Int Forum Allergy Rhinol.* 2020;10:15-22.
 21. Chowdhury NI, Smith TL, Chandra RK, Turner JH. Automated classification of osteomeatal complex inflammation on computed tomography using convolutional neural networks. *Int Forum Allergy Rhinol.* 2019;9:46-52.
 22. Soler ZM, Sauer D, Mace J, Smith TL. Impact of mucosal eosinophilia and nasal polyposis on quality-of-life outcomes after sinus surgery. *Otolaryngol Head Neck Surg.* 2010;142:64-71.
 23. Barham HP, Osborn JL, Snidvongs K, Mrad N, Sacks R, Harvey RJ. Remodeling changes of the upper airway with chronic rhinosinusitis. *Int Forum Allergy Rhinol.* 2015;5:565-572.
 24. Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. *JAMA.* 2020;323(4):305-306. <https://doi.org/10.1001/jama.2019.20866>.