Predicting Olfactory Loss In Chronic Rhinosinusitis Using Machine Learning

Vijay R. Ramakrishnan, MD[1], Jaron Arbet, PhD[2], Jess C. Mace, MPH, CCRP[3], Zachary M. Soler, MD, MSc[4], Timothy L. Smith, MD, MPH[3]

Department of Otolaryngology-Head and Neck Surgery, University of Colorado, Aurora, CO.[1]

Department of Biostatistics, University of Colorado, Aurora, CO.[2]

Department of Otolaryngology-Head and Neck Surgery, Oregon Health Sciences University, Portland, OR.[3]

Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, Charleston, SC.[4]

Corresponding author:
Vijay R. Ramakrishnan, MD
Department of Otolaryngology, University of Colorado
12631 E 17th Ave, B205
Aurora, CO. 80045
Phone: 303-724-1965
Fax: 303-724-1961
E-mail: Vijay.Ramakrishnan@ucdenver.edu

Running Title: Machine Learning and CRS Olfactory Loss

Financial Disclosures: V.R.R., J.C.M., Z.M.S., and T.L.S. are supported by grants for this investigation from the National Institute on Deafness and Other Communication Disorders (NIDCD), one of the National Institutes of Health, Bethesda, MD., USA [R01 DC005805 (T.L.S.) and K23 DC014747 (V.R.R.)]. Public clinical trial registration (www.clinicaltrials.gov) ID# NCT01332136. JA is supported by NIH/NCATS Colorado CTSA Grant Number UL1 TR002535. Contents are the authors' sole responsibility and do not necessarily represent official NIH views.

Conflicts of Interest: None related to this study

ABSTRACT

Objective

Compare machine learning (ML) based predictive analytics methods compared to traditional logistic regression in classification of olfactory dysfunction in chronic rhinosinusitis (CRS-OD), and identify predictors within a large multi-institutional cohort of refractory CRS patients.

Methods

Adult CRS patients enrolled in a prospective, multi-institutional, observational cohort study were assessed for baseline CRS-OD using a smell identification test (SIT) or brief SIT (bSIT). Four different ML methods were compared to traditional logistic regression for classification of CRS normosmics versus CRS-OD.

Results

Data were collected for 611 study participants who met inclusion criteria between April 2011 and July 2015. 34% of enrolled patients demonstrated olfactory loss on objective testing. Differences between CRS normosmics and those with smell loss included objective disease measures (CT and endoscopy scores), age, sex, prior surgeries, socioeconomic status, steroid use, polyp presence, asthma, and aspirin sensitivity. Most ML methods outperformed traditional logistic regression in terms of predictive ability. Top predictors include known factors reported in the literature, as well as several socioeconomic factors.

Conclusion

Olfactory dysfunction is a variable phenomenon within a large multicenter cohort of refractory CRS patients. ML methods outperform traditional logistic regression in classification of normosmia versus smell loss in CRS, and are able to include numerous risk factors into prediction models. These results carry implications for basic science and clinical research in hyposmia secondary to sinonasal disease, the most common cause of persistent olfactory loss in the general population.

INTRODUCTION

Olfactory loss is common, affecting up to 25% of the adult population, with chronic rhinosinusitis (CRS) being a major cause of persistent olfactory loss [1]. Smell loss is considered a hallmark of CRS, particularly in those with nasal polyps, and is one of the cardinal diagnostic symptoms for CRS [2-4]. Curiously, this symptom does not affect all CRS patients. Just as CRS is a heterogeneous disease [5], previous literature has estimated a 20-80% prevalence of olfactory loss in CRS cohorts, based on the sensitivity of olfactory assessment used and the proportion of enrolled subjects with nasal polyps (CRSwNP). A recent study using well-defined CRS diagnostic criteria and objective testing using Sniffin' Sticks assessment showed a 44% prevalence of smell loss in CRSsNP subjects and 58% in CRSwNP [6]. Risk factors associated with CRS-OD identified in this study included disease severity score on computed tomography (CT) the presence of comorbid conditions such as diabetes, allergy, and asthma. Adding a healthy control group for comparison, Schlosser et al. noted CRS-OD to affect CRSwNP patients more than CRSsNP, and identified age as an additional risk factor [7]. Given that most existing studies consist of single-institution cohorts, with subjects undergoing variable treatment regimens, and different assessments of smell function, a systematic review and meta-analysis was performed by Kohli et al [8] to clarify the presence of OD in CRS and identify consistent risk factors. The authors included 47 studies in the analysis, and identified age, CT score, and presence of nasal polyps as the three most consistent risk factors for CRS-OD in the published literature. Similarly, olfactory response to CRS treatment is highly variable, with resolution of smell disturbance after standard therapies occurring in only approximately 40% of refractory CRS patients [9].

Olfactory dysfunction in CRS is potentially complex and the interplay among possible mechanisms is poorly understood. The simple model of conductive loss due to impaired odorant transport to the olfactory cleft in CRS may explain a portion of the disease for some patients. Additional contributions of local and systemic inflammation likely contribute to a varying degree in some individuals, with underlying factors such as age and presence of comorbid diabetes influencing the capacity for olfactory sensory neuron regeneration. Given the complex and multifactorial nature of CRS-OD, compounded by inherent challenges in human cohort studies, it is no surprise that putative predictors and their relative importance have been inconsistently identified across studies. Beyond the risk of sampling bias found in most single institution studies, existing publications attempt to identify predictor variables for CRS-OD generally through univariate regression models. But, these analyses cannot account for all possible predictor variables or interactions among variables, and therefore, may offer limited conclusions. Additionally, attempting multivariable regressions that were not developed for predictive performance or assessed as such, cannot be relied upon [10].

Traditional statistical approaches can model how system variables relate to one another and generate corresponding metrics of statistical significance. However, novel data analytics approaches including machine learning (ML) methods have shown improved classification accuracy and can be utilized to predict future outcomes, rather than attempting to interpret individual components within a system. ML approaches are particularly useful when there is a role for including complex interactions that may otherwise be ignored or dismissed as noise when using traditional statistics [11], such as in the case of CRS-OD.

Understanding the heterogeneity of CRS-OD and grasping multi-dimensional risk factors for classification prediction is foundational for future clinical care and research. The objective of this study was to test different ML models used for predictive analytics against traditional regression

modeling for classification accuracy of OD in CRS patients. A secondary goal was to explore top predictors in high-performing models to identify common predictor variables of high importance that may help explain the heterogeneity of OD found among CRS patients.

MATERIALS and METHODS

Study Population

Adult patients (≥18 years old) were recruited into a prospective, multi-center, observational cohort study developed to evaluate treatment outcomes of endoscopic sinus surgery (ESS), for which multiple reports have been previously published [12-15]. All study patients were diagnosed with medically refractory CRS as defined by both the American Academy of Otolaryngology-Head and Neck Surgery 2007 guidelines and the European Position Paper on Rhinosinusitis and Nasal Polyps 2012 [3,4]. The Institutional Review Board (IRB) at each performance site governed all study protocols, informed consent documentation, and data safety monitoring. Performance sites consisted of tertiary care sinus surgery centers located within academic hospital systems in North America including: Oregon Health & Science University (OHSU; Portland, OR; eIRB#7198), Stanford University (Palo Alto, CA; IRB#4947), the Medical University of South Carolina (Charleston, SC; IRB#12409), and the University of Calgary (Calgary, Alberta, Canada; IRB#E-24208), with central study coordination at OHSU.

At original study enrollment, participants were screened for demographics, social and medical history including factors listed in **Table 1**. The diagnoses of asthma, allergy, and aspirin sensitivity were made by presence of classic symptoms and self-report of prior testing. Subjects with comorbid cystic fibrosis, ciliary dyskinesia, or autoimmune disease were not excluded from final analysis. Participants were followed postoperatively through the standard of care for up to 18-months and asked to complete postoperative, study evaluations at regular 6-month intervals, either during physician-directed appointments or follow-up study mailings utilizing the postal service and self-addressed return envelopes.

Objective Measures of Disease Severity

Diagnostic measures of disease severity were collected as standard of care, and were also used for investigational purposes. These well-established objective measures have been recommended for clinical study [16], and include Lund-Kennedy nasal endoscopy scoring [17] and high-resolution computed tomography (CT) radiographic imaging graded using the Lund-Mackay scoring system [18]. Olfactory function was using the brief Smell Identification Test (bSIT) in the years 2011-2013, and the Smell Identification Tests (SIT) in the years 2013-2015 (Sensonics Inc., Haddon Heights, NJ). Higher SIT total scores reflect a better sense of smell, and categorical ratings of "smell loss" (anosmia + hyposmia) and "normosmia" were assigned based on established scales [19-21].

Patient Reported Outcome Measures

Other patient reported outcome measures (PROMs) were documented but were not used in the predictive analytics given the potential overlap between particular symptom ratings and objective smell loss. Study PROMs included the 22-item SinoNasal Outcome Test (SNOT-22; ©2006, Washington University, St. Louis, MO) [22,23], the Rhinosinusitis Disability Index (RSDI) [24], and the questionnaire on olfactory dysfunction (QOD-NS) [25].

Data Management and Statistical Analyses

Protected health information was previously removed and study data was safeguarded using unique study identification number assignment for each participant. Study data was securely transferred from a HIPAA compliant, relational database (Access, Microsoft Corp, Redmond, WA.) to the University of Colorado Department of Otolaryngology-Head and Neck Surgery password protected research server, according to specifications of a Data Use Agreement between OHSU and the University of Colorado.

Study data were evaluated descriptively while continuous and ordinal variables were evaluated for assumptions of normality. Comparing variables between normosmics and those with smell loss, Fisher's exact test was used for categorical variables and Welch's two sample t-test was used for continuous variables. Statistical comparisons were reported using a type-I error probability at the 0.050 level of significance. All statistical and data analyses were completed using the R software program [26].

Machine Learning Approaches to Classification

Several machine learning predictive analytics models were applied, in order to compare classification accuracy to tradition *logistic regression*, which was performed here with backwards step-wise AIC variable selection. *Random forests* was chosen as it allows for non-linear and interaction effects; conditional inference trees were used with permutation-based variable importance scores for unbiased variable selection [27]. Similar to traditional logistic regression, least absolute shrinkage and selection operator (*LASSO*), assumes only linear and additive affects (without interaction) where regression coefficients are shrunk toward zero to perform variable selection. In this model, unimportant variables are given coefficients equal to zero, effectively removing them from the model. Along these lines, multivariate adaptive regression splines (*MARS*) was also applied in a similar fashion to traditional logistic regression, but it uses stepwise methods for variable selection and allows for nonlinear and interaction effects. Lastly, the *Support Vector Machine* approach was applied with a radial basis kernel.

To evaluate classification accuracy, repeated 10-fold cross validation (5 repeats) was used to tune and evaluate each model area. To deal with the fact that the two classes are unbalanced (smell loss vs normosmia), down sampling was used within the cross-validation procedure, as implemented in the caret R package. Up sampling was also applied, but produced equal classification accuracy in terms of AUC. The method of "surrogate splits" [28] was used to handle missing values in Random Forest. The missForest R package [29] was used to impute missing values for all other models.

To identify the important predictor variables included in each classification algorithm, permutation-based variable importance scores were used to rank variables from most to least important in Random Forests, while the absolute value of the standardized regression coefficient was used for LASSO variable rankinds. The IML R package [30] was used to calculate variable importance scores for SVM using a permutation method. All variable importance scores were scaled from 0 to 100, where the most important variable for the model has a score of 100. Of note, for a given categorical variable, RF and LASSO calculate separate variable importance scores for each category, whereas the approach used for SVM only calculates an overall score.

RESULTS

Final Cohort Characteristics

A total of 611 study participants met all inclusion criteria and were prospectively enrolled between April 2011 and July 2015. Demographics and comorbid disease characteristics,

examined by olfactory classification, are described in **Table 1**. The overall mean age for the final cohort was 51.3 years (38.3, 61.7) years with slight female preponderance (53% female). Approximately one-third of the cohort exhibited nasal polyps (35%).

Study participants were stratified by SIT category. Statistically significant differences were observed between the normosmic and smell loss cohorts in terms of objective disease measures (CT and endoscopy scores), age, sex, previous surgery, socioeconomic status, steroid use, asthma/COPD, and aspirin sensitivity. Local conditions including nasal polyps, allergic fungal rhinosinusitis, and deviated septum were also significant factors.

Smell loss vs Normosmia Classification Accuracy

A traditional logistic regression model for classification of baseline olfactory dysfunction demonstrated on these criteria demonstrated acceptable accuracy in the context of CRS-OD (AUC 0.707). Three of the four machine learning models outperformed the traditional statistical approach (**Table 2**). The SVM approach was the most accurate, with a relatively good classification accuracy (AUC 0.754) and much higher sensitivity than other the methods. LASSO and Random Forest also outperformed logistic regression, whereas MARS performed worse than all of the approaches.

Machine Learning Predictions: Top variables

ML algorithms are adept are recognizing patterns within complex data, but the lack of *a priori* framework results in the inability to understand *why* the algorithm made its decisions. As we see in the variable importance graphs (**Figs 1, 2, Suppl Fig 1**), the ML models each arrived at somewhat different relative importances of a given predictor variable. SVM, the top predictor model, included the most number of variables in its algorithm with nearly all inputs having significant importance (**Fig 1**). Relative importance of predictor variables in Random Forests and LASSO are shown in **Fig 2** (**and Suppl Fig 1**) and in these models, fewer variables were important and the most weight was placed on the presence of nasal polyps and the nasal endoscopy disease severity score.

Top predictors across the approaches contain many factors associated with olfactory function in previous studies, including steroid use, prior surgery, nasal polyps, aspirin exacerbated respiratory disease, age, sex, and smoking history. Inflammatory disease severity, as measured by CT and nasal endoscopy scores, are not surprisingly very important predictors of CRS-OD.

Interestingly, socioeconomic factors such as insurance status, household income, race, were among top predictors across the three models displayed. Comorbid medical conditions including asthma, diabetes, septal deviation, alcohol use, and sleep apnea, are also of moderate importance. Notably, these factors can be influenced by unrelated medical care, suggesting that other approaches to improve general health and well-being may be considered as part of the holistic approach to CRS-OD management. Although the presence of nasal polyps, asthma, and AERD, are significantly associated with CRS-OD here and in prior literature, the presence of allergic rhinitis by history or by skin testing appeared to carry relatively minimal weight in the CRS-OD classifications.

DISCUSSION

CRS is a heterogeneous process in terms of presentation, subtypes, natural history and responses to treatment [2]. In terms of CRS-OD, current data from a multi-institutional cohort

251 studying olfactory loss in CRS with a sensitive olfactory test (Sniffin' Sticks) demonstrates that
252 nearly 70% of the CRS population has significant olfactory loss, with 20% exhibiting complete
253 anosmia [7]. The pathophysiology of CRS-OD is ascribed to both sensorineural and conductive
254 effects of local tissue inflammation, on top of the myriad factors that can affect olfaction in the
255 general population [1,31]. In previous study, factors associated with worse baseline olfactory
256 function in CRS patients included nasal polyposis, asthma, age, smoking, and eosinophilia [32].
257 Other works have shown CRS-OD response treatment in ~40% of patients, with risk factors such
258 as olfactory dysfunction severity, nasal polyps, female gender, high socio-economic status and
259 non-smoking associating with better quality of life results. In multivariate regression, only nasal
260 polyps and degree of baseline olfactory dysfunction maintained statistical significance,
261 highlighting the importance of sample size and analytical methods to glean useful information
262 from a prospectively collected human dataset [10].
263 Here, we utilized ML-based data science approaches to interrogate a large multicenter
264 CRS cohort to test whether novel predictive analytics approaches could outperform the gold-
265 standard traditional statistical method of logistic regression. ML is defined as the use of
266 algorithms to break down data, learn from it, and then make a determination or prediction about
267 some aspect. In doing so, ML provides a novel approach to uncover previously unrecognized
268 patterns among CRS-OD patients and thereby offers numerous advantages over regression
269 analyses, which have been traditionally employed in studies of CRS disease severity and
270 outcomes [33]. Whereas traditional statistical methods focus on modeling how system variables
271 relate to one another and what statistical inference (e.g. significance in p-values) can be made,
272 the goal of machine learning (ML) is not interpretation of individual components but prediction
273 of future outcomesTop predictorsIn this study, different supervised ML algorithms were used
274 map features of interest to the outcome or "label" of olfactory loss. We observed improved
275 prediction accuracy with most ML approaches, measured by AUC and sensitivity/specificity.
276 This advantage may be attributed to the ability of many ML models to handle high-dimensional
277 data in which more features exist than observations, whereas traditional statistical modeling with
278 a large number of potential features requires some form of dimension reduction or variable
279 selection. As a result, we believe that machine learning and artificial intelligence data analytics
280 are well-suited to prime research and eventually for precision medicine in olfactory disorders,
281 given the idiosyncrasies, nuances, and numerous possible predictor variables.
282 There are some limitations of the current study that should be addressed, the first is the
283 potential for sampling bias. Two major biases in healthcare ML are gender and race [34]. Details
284 of this cohort have been extensively published and include even distributions by age and sex.
285 Racial minorities and low socioeconomic status groups are underrepresented as a function of the
286 geography and practice referral patterns, the general lack of accessible and affordable healthcare
287 and racial disparities in the U.S. healthcare system [35]. A related concern is that models created
288 from our cohort data may "overfit" if new or unseen clinical data are applied. Certainly, these
289 approaches should be further validated but nonetheless demonstrate the important utility of novel
290 data science approaches to uncover unrecognized patterns and predictors from "noisy" human
291 data in a noisy disease process. There are likely potentially unrecognized but important data
292 inputs (i.e., molecular endotyping data) that are not yet clinically available. Serum markers, such
293 as peripheral eosinophilia and serum IgE, have been proposed for CRS in the past but have
294 generally been disappointing and do not appear to correlate with tissue levels [36,37] Some our
295 clinical data input fields may associate phenotypes with likely molecular endotypes, but such

overlap is not universal [5]. It is an exciting direction for the near future to include molecular based biomarkers to the input data, and test additional accuracy gained.

CONCLUSION

Here we demonstrate the importance of novel predictive analytics approaches in study of clinical olfactory disorders. Although founded in data science principles, and capable of recognizing patterns within noisy data, it is challenging to understand *why* the models make predictions with comparable accuracy despite use of some different predictor variable. Nonetheless, consistently observed predictors include some weighted interest in SES and other potentially modifiable conditions such as asthma/AERD. Such approaches may have value for both clinical counseling and guidance for future basic science and translational research.

ACKNOWLEDGMENTS

The authors would like to thank Drs. Miranda Kroehl, PhD, John Rice, PhD, and Krithika Suresh, PhD, for biostatistical expertise in machine learning and prediction models, as well as data management considerations. We would also like to thank Drs. Peter Hwang, MD, and Luke Rudmik, MD, for their contributions in original study enrollment and ongoing collaboration.

313 **REFERENCES**

314

315 1. Hummel T, Whitcroft KL, Andrews P et al. Position Paper on Olfactory Dysfunction.
316 Rhinology 2016 31;56(1):1-30.

317 2. Orlandi RR, Kingdom TT, Hwang PH, et al. International Consensus Statement on Allergy
318 and Rhinology: Rhinosinusitis. Int Forum Allergy Rhinol. 2016;6 Suppl 1:S22-209. Section X.B.

319 3. Fokkens WJ, Lund VJ, Mullol J, et al. European position paper on rhinosinusitis and nasal
320 polyps 2012. Rhinol. Suppl 2012;23: 1-298.

321 4. Rosenfeld RM, Piccirillo JF, Chandrasekhar SS, et al. Clinical practice guideline (update):
322 adult sinusitis. Otolaryngol Head Neck Surg. 2015;152(2 Suppl):S1-39.

323 5. Akdis CA, Bachert C, Cingi C, et al. Endotypes and phenotypes of chronic rhinosinusitis: a
324 PRACTALL document of the European Academy of Allergy and Clinical Immunology and the
325 American Academy of Allergy, Asthma & Immunology. J Allergy Clin Immunol
326 2013;131(6):1479-90.

327 6. Soler ZM, Kohli P, Storck KA, Schlosser RJ. Olfactory Impairment in Chronic Rhinosinusitis
328 Using Threshold, Discrimination, and Identification Scores. Chem Senses 2016;41(9):713-719.

329 7. Schlosser RJ, Smith TL, Mace JC, et al. Factors driving olfactory loss in patients with chronic
330 rhinosinusitis: a case control study. Int Forum Allergy Rhinol 2020;10(1):7-14.

331 8. Kohli P, Naik AN, Harruff EE, et al. The prevalence of olfactory dysfunction in chronic
332 rhinosinusitis. Laryngoscope 2017; 127(2):309-320.

333 9. DeConde AS, Mace JC, Alt JA, Schlosser RJ, Smith TL, Soler ZM. Comparative effectiveness
334 of medical and surgical therapy on olfaction in chronic rhinosinusitis: a prospective, multi-
335 institutional study. Int Forum Allergy Rhinol. 2014;4(9):725-33.

336 10. Katotomichelakis M, Simopolous E, Tripsianis G, et al. Predictors of quality of life
337 outcomes in chronic rhinosinusitis after sinus surgery. Eur Arch Otorhinolaryngol 2014;
338 271(4):733-41.

339 11. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods
340 2018;15(4):233-234.

341 12. Smith TL, Kern R, Palmer JN, et al. Medical therapy vs surgery for chronic rhinosinusitis: a
342 prospective, multi-institutional study with 1-year follow-up. Int Forum Allergy Rhinol
343 2013;3(1):4-9.

344 13. DeConde AS, Mace JC, Alt JA, et al. Longitudinal improvement and stability of the SNOT-
345 22 survey in the evaluation of surgical management for chronic rhinosinusitis. Int Forum Allergy
346 Rhinol. 2015; 5(3):233-239.

347 14. Soler ZM, Smith TL, Alt JA, et al. Olfactory-specific quality of life outcomes after
348 endoscopic sinus surgery. Int Forum Allergy Rhinol. 2016; 6(4): 407-413.

349 15. Ramakrishnan VR, Mace JC, Soler ZM, Smith TL. Examination of high-antibiotic users in a
350 multi-institutional cohort of chronic rhinosinusitis patients. Int Forum Allergy Rhinol.
351 2017;7(4):343-51.

352 16. Hopkins C, Hettige R, Soni-Jaiswal A, et al. Chronic rhinosinusitis outcome measures
353 (CHROME), developing a core outcome set for trials of interventions in chronic rhinosinusitis.
354 Rhinology 2018;56(1):22-32.

355 17. Lund VJ and Kennedy DW. Quantification for staging sinusitis. The Staging and Therapy
356 Group. Ann Otol Rhinol Laryngol Suppl. 1995;167:17-21.

357 18. Lund VJ and Mackay IS. Staging in rhinosinusitis. Rhinology. 1992;31(4):183-4.

358 19. Doty RL. The Smell Identification Test Administration Manual. 3. Haddon Heights, NJ:
359 Sensonics, Inc; 1995.

360 20. Doty RL. The Brief Smell Identification Test Administration Manual. Haddon Heights, NJ:
361 Sensonics, Inc; 2001.

362 21. El Rassi E, Mace JC, Steele TO, et al. Sensitivity analysis and diagnostic accuracy of the
363 Brief Smell Identification Test in patients with rhinosinusitis. Int Forum Allergy Rhinol.
364 2016;6(3):287-92.

365 22. Piccirillo JF, Merritt MG, Jr, Richards ML. Psychometric and clinimetric validity of the 20-
366 Item Sino-Nasal Outcome Test (SNOT-20) Otolaryngol Head Neck Surg. 2002;126(1):41–47.

367 23. Hopkins C, Gillett S, Slack R, et al. Psychometric validity of the 22-item Sinonasal Outcome
368 Test. Clin Otolaryngol. 2009;34(5):447–454.

369 24. Senior BA, Glaze C, Benninger MS. Use of the rhinosinusitis disability index (RSDI) in
370 rhinologic disease. Am J Rhinol. 2001; 15(1):15-20

371 25. Simopoulos E, Katotomichelakis M, Gouveris H, Tripsianis G, Livaditis M, Danielides V.
372 Olfaction-associated quality of life in chronic rhinosinusitis: adaptation and validation of an
373 olfaction-specific questionnaire. Laryngoscope 2012;122:1450–1454.

374 26. R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna,
375 Austria: R Foundation for Statistical Computing.

376 27. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance
377 measures: Illustrations, sources and a solution. BMC bioinformatics 2007;8(1):25.

378 28. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: Rationale, application,
379 and characteristics of classification and regression trees, bagging, and random forests.
380 Psychological methods 2009; 14:323.

381 29. Stekhoven DJ and Bühlmann, P. MissForest—non-parametric missing value imputation for
382 mixed-type data. Bioinformatics 2012;28(1):112-118.

383 30. Molnar C, Casalicchio G, Bischl B. iml: An R package for interpretable machine learning. J
384 Open Source Software 2018;3(26), 786.

385 31. Rombaux P, Huart C, Levie P, Cingi C, Hummel T. Olfaction in Chronic Rhinosinusitis.
386 Curr Allergy Asthma Rep 2016;16(5):41.

387 32. Litvack JR, Fong K, Mace J, et al. Predictors of olfactory dysfunction in patients with
388 chronic rhinosinusitis. Laryngoscope 2008;118:2225–30.

389 33. Michie D, Spiegelhalter D, Taylor C. Machine Learning, Neural, and Statistical
390 Classification. New York, NY: Ellis Horwood; 1994.

391 34. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med
392 2019;380(14):1347-1358.

393 35. 2018 National Healthcare Quality and Disparities Report. Content last reviewed April 2020.
394 Agency for Healthcare Research and Quality, Rockville, MD.
395 https://www.ahrq.gov/research/findings/nhqrdr/nhqdr18/index.html

396 36. Gitomer SA, Fountain CR, Kingdom TT, et al. Clinical Examination of Tissue Eosinophilia
397 in Patients with Chronic Rhinosinusitis and Nasal Polyposis. Otolaryngol Head Neck Surg
398 2016;155(1):173-8.

399 37. Settipane GA. Nasal polyps and immunoglobulin E (IgE). Allergy Asthma Proc
400 1996;17(5):269-73.

401

401 **Table 1. Study population.** Collected demographics and clinical metadata used as inputs
402 for classification models are shown here for the overall cohort, and those with and without
403 olfactory dysfunction. Data sorted from smallest to largest q-value (multiple testing adjusted p-
404 value).
405
406

Characteristic	Overall, N = 611	N	Yes, N = 203[1]	No, N = 408[1]	p-value[2]	q-value[3]
CRS with POLYPS		611			**<0.001**	**<0.001**
No	399 (65%)		86 (42%)	313 (77%)		
Yes	212 (35%)		117 (58%)	95 (23%)		
TOTAL ENDOSCOPY SCORE	6.0 (4.0, 8.5)	607	8.0 (4.0, 11.0)	4.0 (2.0, 7.0)	**<0.001**	**<0.001**
TOTAL CT SCORE	12.0 (7.0, 16.0)	582	15.0 (10.0, 19.0)	11.0 (6.0, 14.0)	**<0.001**	**<0.001**
AERD		611			**<0.001**	**<0.001**
No	567 (93%)		173 (85%)	394 (97%)		
Yes	44 (7.2%)		30 (15%)	14 (3.4%)		
AGE	51.0 (38.3, 61.7)	611	56.6 (44.6, 65.4)	49.1 (36.9, 59.9)	**<0.001**	**<0.001**
PRIOR SURGERY		610			**<0.001**	**<0.001**
0	278 (46%)		66 (33%)	212 (52%)		
1	153 (25%)		56 (28%)	97 (24%)		
2	90 (15%)		34 (17%)	56 (14%)		
3	40 (6.6%)		24 (12%)	16 (3.9%)		
4+	49 (8.0%)		23 (11%)	26 (6.4%)		
INSURANCE		610			**<0.001**	**<0.001**
Employer provided	373 (61%)		112 (55%)	261 (64%)		
Medicaid	24 (3.9%)		4 (2.0%)	20 (4.9%)		
Medicare	129 (21%)		63 (31%)	66 (16%)		
None	6 (1.0%)		5 (2.5%)	1 (0.2%)		
Private	62 (10%)		13 (6.4%)	49 (12%)		
State Assisted	8 (1.3%)		4 (2.0%)	4 (1.0%)		
VA Benefits	3 (0.5%)		0 (0%)	3 (0.7%)		
VA Benefits / Tricare	5 (0.8%)		2 (1.0%)	3 (0.7%)		
ASTHMA		611			**<0.001**	**<0.001**
No	397 (65%)		110 (54%)	287 (70%)		
Yes	214 (35%)		93 (46%)	121 (30%)		

Characteristic	Overall, N = 611	N	Yes, N = 203[1]	No, N = 408[1]	p-value[2]	q-value[3]
SEPTAL DEVIATION		611			**<0.001**	**0.001**
No	408 (67%)		155 (76%)	253 (62%)		
Yes	203 (33%)		48 (24%)	155 (38%)		
RECURRENT ACUTE SINUSITIS		611			**<0.001**	**0.003**
No	571 (93%)		199 (98%)	372 (91%)		
Yes	40 (6.5%)		4 (2.0%)	36 (8.8%)		
STEROID USE		611			**0.003**	**0.009**
No	555 (91%)		174 (86%)	381 (93%)		
Yes	56 (9.2%)		29 (14%)	27 (6.6%)		
SEX		611			**0.013**	**0.037**
Female	324 (53%)		93 (46%)	231 (57%)		
Male	287 (47%)		110 (54%)	177 (43%)		
AFRS		611			**0.019**	**0.049**
No	593 (97%)		192 (95%)	401 (98%)		
Yes	18 (2.9%)		11 (5.4%)	7 (1.7%)		
COPD		611			**0.020**	**0.049**
No	579 (95%)		186 (92%)	393 (96%)		
Yes	32 (5.2%)		17 (8.4%)	15 (3.7%)		
HOUSEHOLD INCOME		595			**0.027**	0.064
0-25000	96 (16%)		34 (17%)	62 (16%)		
26000-50000	102 (17%)		47 (24%)	55 (14%)		
51000-75000	120 (20%)		37 (19%)	83 (21%)		
76000-100000	107 (18%)		28 (14%)	79 (20%)		
100000+	170 (29%)		51 (26%)	119 (30%)		
Inferior Turbinate Hypertrophy		611			0.091	0.200
No	520 (85%)		180 (89%)	340 (83%)		
Yes	91 (15%)		23 (11%)	68 (17%)		
CF or ciliary dysfunction		611			0.117	0.241
No	588 (96%)		199 (98%)	389 (95%)		
Yes	23 (3.8%)		4 (2.0%)	19 (4.7%)		
ALLERGY by HISTORY		611			0.143	0.277
No	480 (79%)		152 (75%)	328 (80%)		
Yes	131 (21%)		51 (25%)	80 (20%)		

Characteristic	Overall, N = 611	N	Yes, N = 203[1]	No, N = 408[1]	p-value[2]	q-value[3]
SITE		611			0.189	0.349
#1	222 (36%)		84 (41%)	138 (34%)		
#2	253 (41%)		77 (38%)	176 (43%)		
#3	136 (22%)		42 (21%)	94 (23%)		
OSA by HISTORY		611			0.215	0.376
No	584 (96%)		191 (94%)	393 (96%)		
Yes	27 (4.4%)		12 (5.9%)	15 (3.7%)		
EDUCATION (Yrs)	16.0 (13.0, 17.0)	608	15.0 (13.0, 16.0)	16.0 (13.0, 17.0)	0.235	0.392
RACE		610			0.247	0.393
African American	49 (8.0%)		18 (8.9%)	31 (7.6%)		
American Indian/Alaska Native	5 (0.8%)		2 (1.0%)	3 (0.7%)		
Asian	24 (3.9%)		10 (4.9%)	14 (3.4%)		
Native Hawaiian/Pacific Islander	1 (0.2%)		1 (0.5%)	0 (0%)		
Other	25 (4.1%)		12 (5.9%)	13 (3.2%)		
White	506 (83%)		160 (79%)	346 (85%)		
GERD		611			0.303	0.460
No	570 (93%)		186 (92%)	384 (94%)		
Yes	41 (6.7%)		17 (8.4%)	24 (5.9%)		
ETHNICITY		611			0.352	0.493
Hispanic / Latino	21 (3.4%)		9 (4.4%)	12 (2.9%)		
Non-Hispanic / Latino	590 (97%)		194 (96%)	396 (97%)		
ALCOHOL Use		608			0.344	0.493
No	313 (51%)		109 (54%)	204 (50%)		
Yes	295 (49%)		92 (46%)	203 (50%)		
ALLERGY_TESTING		611			0.382	0.510
No	362 (59%)		115 (57%)	247 (61%)		
Yes	249 (41%)		88 (43%)	161 (39%)		
DEPRESSION		611			0.403	0.510
No	518 (85%)		176 (87%)	342 (84%)		
Yes	93 (15%)		27 (13%)	66 (16%)		
FIBROMYALGIA		611			0.408	0.510
No	584 (96%)		192 (95%)	392 (96%)		

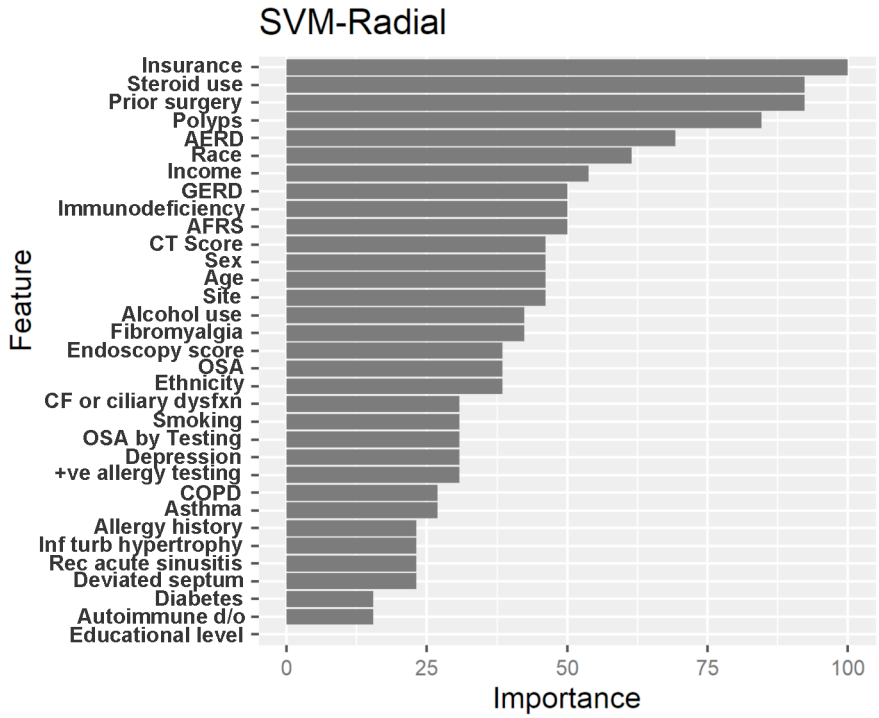
Characteristic	Overall, N = 611	N	Yes, N = 203[1]	No, N = 408[1]	p-value[2]	q-value[3]
Yes	27 (4.4%)		11 (5.4%)	16 (3.9%)		
OSA BY TESTING		611			0.450	0.543
No	557 (91%)		188 (93%)	369 (90%)		
Yes	54 (8.8%)		15 (7.4%)	39 (9.6%)		
ALCOHOL USE (#)	0.0 (0.0, 36.0)	608	0.0 (0.0, 24.0)	0.0 (0.0, 36.0)	0.473	0.552
DIABETES		609			0.536	0.605
No	558 (92%)		183 (91%)	375 (92%)		
Yes	51 (8.4%)		19 (9.4%)	32 (7.9%)		
AUTOIMMUNE DISEASE		610			0.589	0.645
No	573 (94%)		188 (93%)	385 (94%)		
Yes	37 (6.1%)		14 (6.9%)	23 (5.6%)		
IMMUNODEFICIENCY		611			0.626	0.664
No	592 (97%)		198 (98%)	394 (97%)		
Yes	19 (3.1%)		5 (2.5%)	14 (3.4%)		
SMOKER (current or former)		607			1.000	1.000
No	574 (95%)		190 (95%)	384 (95%)		
Yes	33 (5.4%)		11 (5.5%)	22 (5.4%)		

[1]Statistics presented: n (%); median (IQR)
[2]Statistical tests performed: Fisher's exact test; Wilcoxon rank-sum test
[3]False discovery rate correction for multiple testing

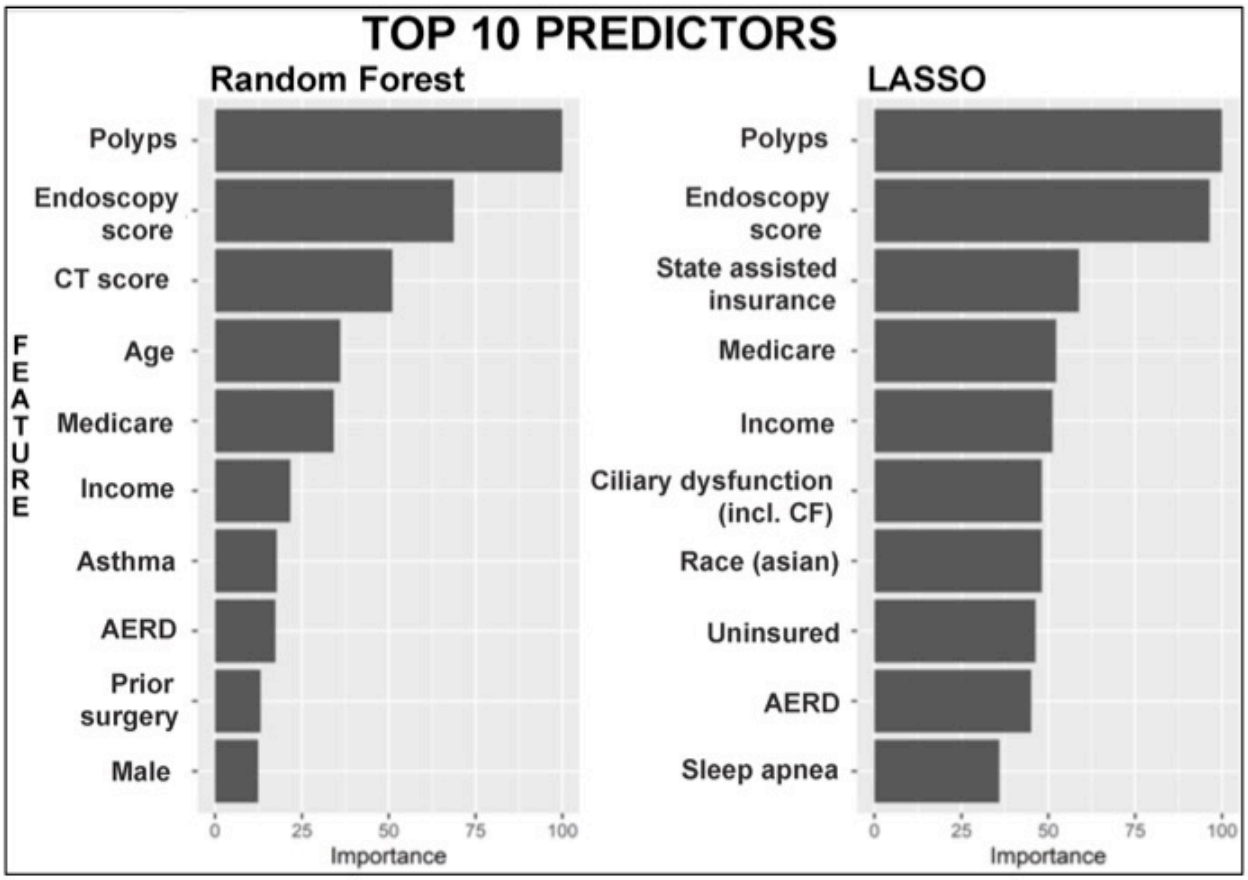
Table 2. **Comparing classification accuracy.** Three ML approaches outperform traditional logistic regression. *AUC = area under receiver operating characteric curve; SVM-radial = support vector machine with a radial basis kernel; Log Reg-Step = stepwise logistic regression.*

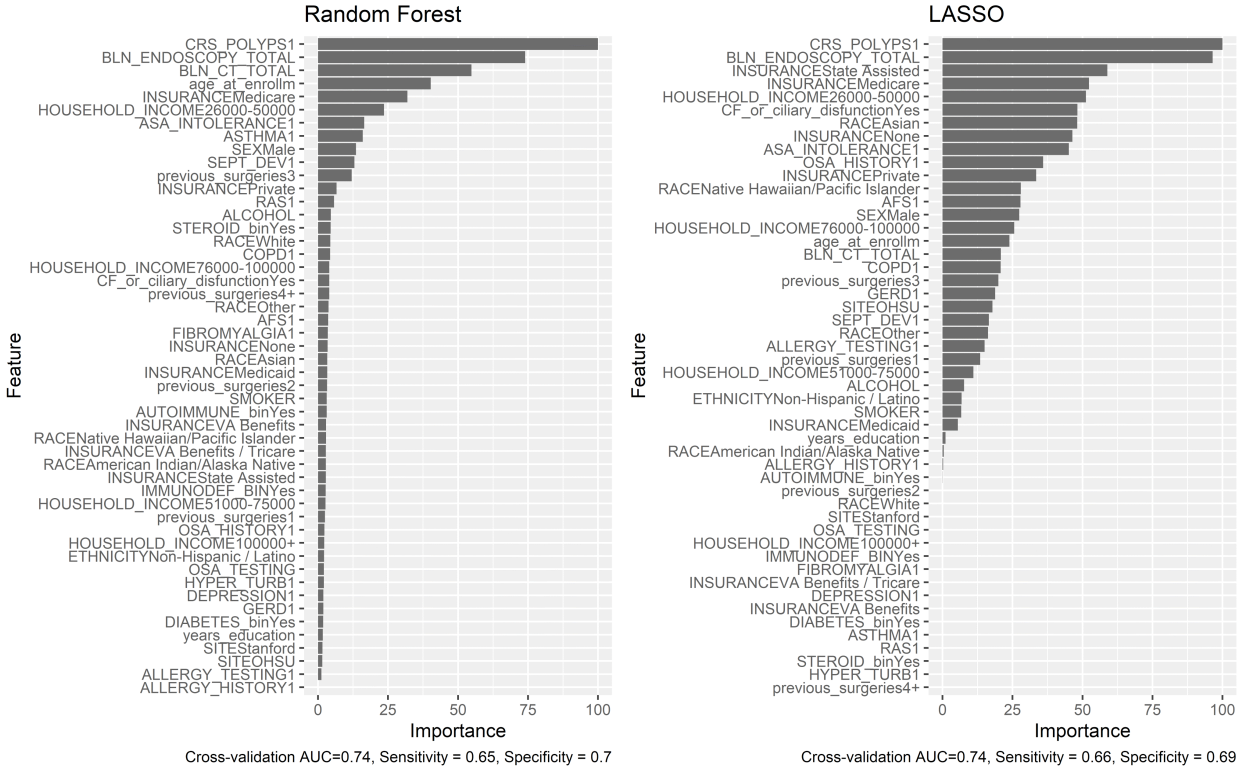
Model	AUC	Sensitivity	Specificity
SVM-Radial	0.754	0.726	0.665
Random Forest	0.744	0.649	0.696
LASSO	0.737	0.659	0.689
Log Reg-Step	0.707	0.635	0.676
MARS	0.682	0.604	0.667

Figure 1. **Variable importance display for most accurate classification model (Support Vector Machine with a radial basis kernel).** Note the inclusion of many predictor variables – 32 are included with > 10% variable importance, suggesting significant interaction between predictors.



424 **Figure 2**. **Top 10 Predictors and Variable importance for Random Forest and LASSO**
425 **models.** *AERD = aspirin exacerbated respiratory disease; CF = cystic fibrosis.*
426



427
428

Supplementary Figure 1. **Complete list of predictors for Random Forest and LASSO models.** Notably, the LASSO model incorporates significantly more features of interest than the Random Forest algorithm (~25 features with >10% importance in the prediction algorithm).