

Predicting Olfactory Loss In Chronic Rhinosinusitis Using Machine Learning

Vijay R. Ramakrishnan, MD¹, Jaron Arbet, PhD², Jess C. Mace, MPH, CCRP³, Krithika Suresh, PhD², Stephanie Shintani Smith, MD, MS,⁴ Zachary M. Soler, MD, MSc⁵, Timothy L. Smith, MD, MPH³

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

Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado-Denver Anschutz Medical Campus, Aurora, CO.²

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

Department of Otolaryngology-Head and Neck Surgery, Northwestern University, Chicago, IL.⁴

Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, Charleston, SC.⁵

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

Word count: abstract 228; main document 3217

Financial Disclosures: This study was supported in part by a grant from the Ludeman Family Center for Women's Health Research at the University of Colorado Anschutz Medical Campus (V.R.R.). V.R.R., J.C.M., Z.M.S., T.L.S., and S.S.S. are supported by grants for this investigation from the National Institute on Deafness and Other Communication Disorders (NIDCD) and the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health, Bethesda, MD., USA [R01 DC005805 (T.L.S. and Z.M.S.), K23 DC014747 (V.R.R.), 1P01AI145818-01 (S.S.S.)]. Public clinical trial registration (www.clinicaltrials.gov) ID# NCT01332136. Contents are the authors' sole responsibility and do not necessarily represent official NIH views.

Conflicts of Interest: None related to this study

41 **ABSTRACT**

42 **Objective**

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

46

47 **Methods**

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

52

53 **Results**

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

61

62 **Conclusion**

63 Olfactory dysfunction is a variable phenomenon in CRS patients. ML methods perform well
64 compared to traditional logistic regression in classification of normosmia versus smell loss in
65 CRS, and are able to include numerous risk factors into prediction models. Several actionable
66 features were identified as risk factors for CRS-OD. These results suggest that ML methods may
67 be useful for current understanding and future study of hyposmia secondary to sinonasal disease,
68 the most common cause of persistent olfactory loss in the general population.

69

70 INTRODUCTION

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

91 Olfactory dysfunction in CRS is potentially complex and the interplay among possible
92 mechanisms is poorly understood. The simple model of conductive loss due to impaired odorant
93 transport to the olfactory cleft in CRS may explain a portion of the disease for some patients.
94 Local and systemic inflammation likely contribute to varying degrees in particular individuals,
95 with underlying factors such as age and presence of comorbid diabetes influencing the capacity
96 for olfactory sensory neuron regeneration. Given the intricate and multifactorial nature of CRS-
97 OD, compounded by inherent challenges of human cohort studies, it is no surprise that putative
98 predictors and their relative importance have been inconsistently characterized across studies.
99 Existing single-institution studies carry a significant risk of sampling bias and their use of
100 univariate regression models are insufficient to identify predictor variables for CRS-OD. These
101 analyses cannot account for all possible predictor variables or interactions among variables, and
102 therefore, may offer limited insights. Additional attempts at categorizing outcomes with
103 multivariable regression approaches that were not developed for predictive performance, or
104 assessed as such, cannot be relied upon [10].

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

112 Understanding the heterogeneity of CRS-OD and interpreting multi-dimensional risk factors
113 for outcomes prediction is foundational for future clinical care and research. The objective of this
114 study was to test different ML-based predictive analytics approaches against traditional
115 regression modeling for classification accuracy of OD in CRS patients. A secondary goal was to

116 explore top predictors in high-performing models to identify common predictor variables of
117 importance that may help explain the heterogeneity of OD found among CRS patients.

118

119 **MATERIALS and METHODS**

120 *Study Population*

121 Adult patients (≥ 18 years old) were recruited into a prospective, multi-center,
122 observational cohort study, for which multiple reports have been previously published [12-15].
123 All study patients were diagnosed with medically refractory CRS as defined by the American
124 Academy of Otolaryngology-Head and Neck Surgery 2007 guidelines [4]. Performance sites
125 consisted of tertiary care sinus centers located within academic hospital systems in North
126 America including: Oregon Health & Science University (OHSU; Portland, OR; eIRB#7198),
127 Stanford University (Palo Alto, CA; IRB#4947), and the Medical University of South Carolina
128 (Charleston, SC; IRB#12409), with central study coordination at OHSU. The Institutional
129 Review Board (IRB) at each performance site governed all study protocols, informed consent
130 documentation, and data safety monitoring, in accordance with the Declaration of Helsinki for
131 Medical Research Involving Human Subjects.

132 At original study enrollment, participants were screened for demographics, social and
133 medical history (**Table 1**). The diagnoses of asthma, allergy, and aspirin-exacerbated respiratory
134 disease (AERD) were made by evaluation of the medical record, presence of classic symptoms
135 and self-report of prior testing. Subjects with comorbid cystic fibrosis, ciliary dyskinesia, or
136 autoimmune disease were included in the analysis. Patient reported outcome measures (PROMs)
137 were collected during study enrollment, but were not used in the predictive analytics given the
138 overlap between particular symptom ratings and objective smell loss [22-25]. Given the
139 exploratory, retrospective nature of this study, all available subjects were included and sample
140 size calculations were not conducted.

141

142 *Objective Measures of Disease Severity*

143 Diagnostic measures of disease severity were collected as standard of care, and were also
144 used for investigational purposes. These well-established objective measures have been
145 recommended for clinical study [16], and include Lund-Kennedy nasal endoscopy scoring [17]
146 and high-resolution CT imaging graded using the Lund-Mackay scoring system [18].

147 The main outcome of interest in the current study is olfactory dysfunction as measured by
148 psychophysical testing. Olfactory function was initially assessed using the brief Smell
149 Identification Test (bSIT) in the years 2011-2013, and subsequently using the 40-item Smell
150 Identification Test (SIT) in the years 2013-2015 (Sensonics Inc., Haddon Heights, NJ). Higher
151 SIT total scores reflect a better sense of smell, and categorical ratings of “smell loss” (anosmia +
152 hyposmia) and “normosmia” were assigned based on established scales [19-21].

153

154 *Data Management and Statistical Analyses*

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

161

162 Study data were evaluated descriptively using count (%) for categorical variables and
163 median and interquartile range (IQR; 25th quantile, 75th quantile) for numeric variables.
164 Comparing variables between normosmics and those with smell loss, Fisher's exact test was used
165 for categorical variables and Wilcoxon rank sum test was used for numeric variables. Both the
166 original p-values and Benjamini-Hochberg multiple testing adjusted p-values ("q-values") are
167 reported [26]. Statistical comparisons with q-values < 0.05 were regarded as statistically
168 significant. All statistical and data analyses were completed using the R software program [27].
169 Machine learning models were compared to simpler logistic regression model with backwards
170 step-wise variable selection using Akaike information criteria (AIC). The step-wise variable
171 selection was re-performed during each iteration of the cross-validation procedure (details later),
172 in order to prevent biased estimates of classification accuracy [28].

173 174 ***Machine Learning Approaches to Classification***

175 Four different machine learning predictive analytics models were applied, in order to
176 compare classification accuracy to logistic regression, and code was made publicly available at
177 <https://github.com/vijayramakrishnan/chemsenses-CRS-olfactory-prediction.git>. All classifiers
178 were fit using the caret R package [29], which provides a unified framework for tuning and
179 evaluating classification accuracy. When including a categorical predictor in a model, K-1
180 dummy variables were used, where K is the total number of categories. *Random forests (RF)*
181 was chosen as it allows for non-linear and interaction effects, and can handle a large number of
182 potential predictors [30]. Although the original random forest algorithm is biased towards
183 favoring features that have more possible cut-points (e.g. categorical variables with more
184 categories), importantly, we used a special type of random forests for unbiased feature selection:
185 an ensemble of "conditional inference trees" with permutation-based variable importance scores
186 [31-33]. Similar to traditional logistic regression, least absolute shrinkage and selection operator
187 (*LASSO*) assumes only linear and additive effects (without interaction) where regression
188 coefficients are shrunk toward zero to perform variable selection [34,35]. In this model,
189 unimportant variables are given coefficients equal to zero, effectively removing them from the
190 model. Along these lines, multivariate adaptive regression splines (*MARS*) was also applied in a
191 similar fashion to traditional logistic regression, but it uses stepwise methods for variable
192 selection and allows for nonlinear and interaction effects [36]. Lastly, the *Support Vector*
193 *Machine (SVM)* approach was applied with a radial basis kernel, which allows for non-linear and
194 interaction effects [37]. Although many machine learning methods exist, these four methods
195 were selected since RF and SVM with a Gaussian/radial basis kernel have been shown to
196 perform well across a wide variety of empirical applications, while LASSO and MARS were
197 chosen as simpler, easier to interpret methods [38-40]. Given that the primary goal is to compare
198 how well different models classify a binary olfactory dysfunction outcome, unsupervised
199 machine learning methods (which do not use an outcome) were not considered appropriate for
200 this study.

201 Repeated 10-fold cross validation (repeated 5 times [41]) was used to tune and internally
202 validate the classification accuracy of each model. The following measures of accuracy were
203 considered: area under receiver operating characteristic curve (AUC), sensitivity (proportion of
204 subjects with olfactory dysfunction that were correctly classified), and specificity (proportion of
205 subjects without olfactory dysfunction that were correctly classified). To deal with the fact that
206 the two classes are unbalanced (dysosmia vs normosmia), under-sampling was used within the
207 cross-validation procedure, as implemented in the caret R package [29,42]. Over-sampling was

208 also considered, but produced nearly equal classification accuracy in terms of AUC (results not
209 shown). The method of “surrogate splits” [33] was used to handle missing values in RF. The
210 missForest R package [43] was used to impute missing values for all other models, while
211 excluding the outcome in order to prevent over-fitting. Cross-validated AUC was compared
212 between models using a corrected resampled t-test [44], which corrects for the overlap that
213 occurs during the resampling process, and the Benjamini-Hochberg method was used to adjust
214 for multiple testing.

215 To identify the important predictor variables included in each classification algorithm,
216 permutation-based variable importance scores were used to rank variables from most to least
217 important in RF, while the absolute value of the standardized regression coefficient was used for
218 LASSO. The IML R package [45] was used to calculate variable importance scores for SVM
219 using a permutation method. All variable importance scores were scaled from 0 to 100, where
220 the most important variable for the model has a score of 100. Of note, for a categorical variable
221 with more than two categories, RF and LASSO calculate separate variable importance scores for
222 each category, whereas the approach used for SVM only calculates an overall score.

223

224 **RESULTS**

225 *Final Cohort Characteristics*

226 A total of 611 study participants met all inclusion criteria and were prospectively enrolled
227 between April 2011 and July 2015. Demographics and comorbid disease characteristics,
228 examined by olfactory classification, are described in **Table 1**. The overall median age for the
229 final cohort was 51 years (IQR= 38.3, 61.7) years with slight female preponderance (53%
230 female). Approximately one-third of the cohort exhibited nasal polyps (35%).

231 Study participants were stratified by olfactory testing category. Statistically significant
232 differences ($q < 0.05$) were observed between the normosmic and hyposmic cohorts in terms of
233 objective disease measures (CT and endoscopy scores), age, sex, previous surgery,
234 socioeconomic status, corticosteroid dependence, asthma/COPD, and AERD. Sinonasal
235 conditions including nasal polyps, allergic fungal rhinosinusitis, and deviated septum were also
236 significant factors.

237

238 *Smell loss vs Normosmia Classification Accuracy*

239 **Figure 1** illustrates a comparison of the different models based on AUC, sensitivity, and
240 specificity. A logistic regression model for classification of baseline olfactory dysfunction
241 demonstrated acceptable accuracy in the context of CRS (AUC 0.707 ± 0.07). Three of the four
242 machine learning models performed favorably compared with the traditional logistic regression
243 model. The SVM approach was the most accurate, with a relatively good classification accuracy
244 (AUC 0.754 ± 0.05) and much higher sensitivity than other the methods. LASSO and Random
245 Forest also outperformed logistic regression, whereas MARS had the lowest AUC. The corrected
246 resampled t-test with Benjamini-Hochberg adjustment (**Figure 1c**) indicated that SVM had a
247 significantly higher mean AUC than MARS ($q = 0.02$), but was not significantly different than the
248 Random Forest, LASSO, or logistic regression. LASSO had a significantly higher mean AUC
249 than MARS ($q = 0.012$) and logistic regression ($q = 0.012$).

250

251 *Machine Learning Predictions: Top variables*

252 ML algorithms are adept at recognizing patterns within complex data, but the lack of a
253 *priori* framework results in the clinical challenge of understanding *why* the algorithm made its

254 decisions. The ML models carried some different relative importances of predictor variables.
255 SVM, the model with the largest AUC, considered many variables to be influential, with 20
256 variables having variable importance scores >25 (**Fig 2**). In Random Forest and LASSO models,
257 fewer variables were important and the most weight was placed on the presence of nasal polyps
258 and the nasal endoscopy severity score (**Fig 3 and Suppl Fig 1**).

259 Top predictors across the approaches include corticosteroid dependence, prior surgery,
260 nasal polyps, AERD, age, sex, and smoking history. Inflammatory disease severity, as measured
261 by CT and nasal endoscopy scores, were also strong predictors of CRS-OD.

262 Socioeconomic factors such as insurance status, household income, and race are among
263 top predictors across the three models displayed. Age was a less important variable in all models.
264 Comorbid medical conditions including asthma, diabetes, septal deviation, alcohol use, and sleep
265 apnea, are also of moderate importance. Although the presence of nasal polyps, asthma, and
266 AERD, are significantly associated with CRS-OD, the presence of allergic rhinitis by history or
267 by skin testing appears to carry relatively minimal weight in the CRS-OD classifications.

268

269 **DISCUSSION**

270 CRS is a heterogeneous process in terms of presentation, subtypes, natural history and
271 response to treatments [2]. Current data from a multi-institutional CRS cohort studying olfactory
272 loss in CRS with a sensitive olfactory test (Sniffin' Sticks) demonstrates that nearly 70% of the
273 CRS population has significant olfactory loss, with 20% exhibiting complete anosmia [7]. The
274 pathophysiology of OD in CRS could result from both sensorineural and conductive effects of
275 local tissue inflammation, on top of the myriad factors that can affect olfaction in the general
276 population [1,46]. In previous studies, factors associated with worse baseline olfactory function
277 in CRS patients included nasal polyposis, asthma, age, smoking, and eosinophilia [47]. Other
278 works have shown CRS-OD response treatment in ~40% of patients, with risk factors such as
279 olfactory dysfunction severity, nasal polyps, female gender, high socioeconomic status and non-
280 smoking associating with better quality of life results. In multivariate regression, only nasal
281 polyps and degree of baseline olfactory dysfunction maintained statistical significance,
282 highlighting the need for robust sample size and improved analytical methods to glean useful
283 information from human study [10].

284 Here, we utilized ML-based data science approaches in addition to logistic regression in
285 predictive modeling of CRS-OD classification. Top predictors across the approaches contain
286 many variables associated with olfactory function in previous studies including nasal polyps,
287 prior surgery, AERD, age, sex, corticosteroid dependence, and smoking history. Inflammatory
288 disease severity, as measured by CT and nasal endoscopy scores, are also apparently important
289 predictors of CRS-OD. Socioeconomic factors such as insurance status, household income, and
290 race were among the important variables across the three top performing models. Since age was
291 less important in all models, insurance status appears to be an independent predictor (i.e.,
292 Medicare insurance is not simply a proxy for advanced age). Comorbid medical conditions
293 including asthma, diabetes, septal deviation, alcohol use, and sleep apnea, are also of moderate
294 importance. Notably, these factors can be influenced by medical care and lifestyle modification,
295 suggesting that other approaches to improve general health and well-being may be considered as
296 part of a holistic approach to CRS-OD management. Although the presence of nasal polyps,
297 asthma, and AERD, are significantly associated with CRS-OD here and in prior literature, the
298 presence of allergic rhinitis by history or by skin testing appeared to carry relatively minimal

299 weight in the CRS-OD classifications. This observation suggests that evaluation and treatment
300 for allergic rhinitis may not necessarily be expected to benefit olfactory function in CRS.

301 We observed favorable prediction accuracy with most ML approaches compared to
302 traditional regression modeling. ML is defined as the use of algorithms to break down data, learn
303 from it, and then make a determination or prediction about some aspect. Whereas traditional
304 statistical methods focus on modeling how system variables relate to one another and what
305 statistical inference (e.g. significance in p-values) can be made, the goal of machine learning
306 (ML) is not interpretation of individual components but prediction of future outcomes. In doing
307 so, ML provides a novel approach to uncover previously unrecognized patterns among CRS-OD
308 patients and thereby offers numerous advantages over regression analyses, which have been
309 traditionally employed in studies of CRS disease severity and outcomes [48]. As more complex
310 and multi-dimensional data are adding in “deep phenotyping” approaches to CRS, machine
311 learning approaches may be well-suited to include the expansive amounts of added data (i.e.,
312 multi-‘omics). Further external validation using appropriate cohorts will certainly be of value.

313 In this study, four different supervised ML algorithms were used map features of interest
314 to the outcome or “label” of olfactory loss, and three outperformed traditional regression. In
315 particular, the SVM model had improved accuracy that was driven largely by improved
316 sensitivity. The classification accuracy may be attributed to the ability of these ML models to
317 handle high-dimensional data in which more features exist than observations, and include non-
318 linear and interaction effects. In contrast, traditional statistical modeling with a large number of
319 potential features requires some form of dimension reduction or variable selection, and
320 exploration of interactions and non-linear effects is challenging with more than a small number
321 of predictors. As a result, we believe that ML and artificial intelligence data analytics are well-
322 suited to prime research and eventually for the application of precision medicine in olfactory
323 disorders, given the idiosyncrasies, nuances, and numerous possible predictor variables.

324 There are some limitations of the current study that should be addressed with regard to
325 ML. Besides the four ML models considered in this study, many other ML methods exist and
326 could be evaluated in future studies. Adequate enrollment by sex and race are common sampling
327 issues in healthcare ML [49,50]. Details of this cohort have been extensively published and
328 include even distributions by age and sex. Racial/ethnic minorities and low socioeconomic status
329 groups are underrepresented perhaps as a function of geography and referral patterns of the
330 enrollment sites, but also result from the general lack of accessible and affordable healthcare and
331 racial disparities inherent within the U.S. healthcare system [51,52]. A related concern is that
332 models created from our cohort data may “overfit” when new or unseen clinical data are applied.
333 Certainly, these approaches should be further validated in independent studies, but nonetheless
334 demonstrate the utility of data science approaches to uncover patterns within expansive human
335 data in a noisy disease process. To incorporate a large enough patient dataset across three sites
336 over several years of prospective data collection, psychophysical testing included classification
337 of smell loss by either SIT or bSIT. Although the SIT may be more sensitive despite both of
338 these tests being extensively validated in the literature, our prior analysis in a CRS cohort has
339 demonstrated excellent diagnostic accuracy of the bSIT (AUC=0.873; 95% CI: 0.819, 0.927)
340 [21]. Additionally, there may be important data inputs (i.e., molecular endotyping data) that are
341 not yet clinically available. Serum markers, such as peripheral eosinophilia and serum IgE, have
342 been proposed for CRS in the past but have generally been disappointing and do not appear to
343 correlate with tissue levels [53,54]. Some of the clinical data input fields may associate
344 phenotypes with likely molecular endotypes, but such overlap is not universal [5]. Future goals

345 include testing whether additional accuracy is gained with the addition of molecular based
346 biomarkers to the input data.

347

348 **CONCLUSION**

349 Here we demonstrate the utility of novel predictive analytics approaches in the study of
350 clinical olfactory disorders. Founded in data science principles rather than traditional statistics,
351 ML models performed favorably compared to logistic regression. Consistently observed
352 predictors across models include a weighted interest in socioeconomic status, and other
353 potentially modifiable conditions such as asthma/AERD. Further studies will be important to
354 reproduce and validate these findings, and might explore why certain novel predictor variables
355 appear to be important. Such approaches may have value for both clinical counseling and
356 guidance for future basic science and translational research.

357

358 **ACKNOWLEDGMENTS**

359 The authors would like to thank Drs. Miranda Kroehl, PhD, John Rice, PhD, and Debashis
360 Ghosh, PhD, for biostatistical expertise in discussions regarding validation of machine learning
361 and prediction models. We would also like to thank Drs. Peter Hwang, MD, Rod Schlosser, MD,
362 and Luke Rudmik, MD, for their contributions in original study enrollment and ongoing
363 collaboration.

364 **REFERENCES**

365

- 366 1. Hummel T, Whitcroft KL, Andrews P et al. Position Paper on Olfactory Dysfunction.
367 *Rhinology* 2016 31;56(1):1-30.
- 368 2. Orlandi RR, Kingdom TT, Hwang PH, et al. International Consensus Statement on Allergy
369 and Rhinology: Rhinosinusitis. *Int Forum Allergy Rhinol.* 2016;6 Suppl 1:S22-209. Section X.B.
- 370 3. Fokkens WJ, Lund VJ, Mullol J, et al. European position paper on rhinosinusitis and nasal
371 polyps 2012. *Rhinol. Suppl* 2012;23: 1-298.
- 372 4. Rosenfeld RM, Piccirillo JF, Chandrasekhar SS, et al. Clinical practice guideline (update):
373 adult sinusitis. *Otolaryngol Head Neck Surg.* 2015;152(2 Suppl):S1-39.
- 374 5. Akdis CA, Bachert C, Cingi C, et al. Endotypes and phenotypes of chronic rhinosinusitis: a
375 PRACTALL document of the European Academy of Allergy and Clinical Immunology and the
376 American Academy of Allergy, Asthma & Immunology. *J Allergy Clin Immunol*
377 2013;131(6):1479-90.
- 378 6. Soler ZM, Kohli P, Storck KA, Schlosser RJ. Olfactory Impairment in Chronic Rhinosinusitis
379 Using Threshold, Discrimination, and Identification Scores. *Chem Senses* 2016;41(9):713-719.
- 380 7. Schlosser RJ, Smith TL, Mace JC, et al. Factors driving olfactory loss in patients with chronic
381 rhinosinusitis: a case control study. *Int Forum Allergy Rhinol* 2020;10(1):7-14.
- 382 8. Kohli P, Naik AN, Harruff EE, et al. The prevalence of olfactory dysfunction in chronic
383 rhinosinusitis. *Laryngoscope* 2017; 127(2):309-320.
- 384 9. DeConde AS, Mace JC, Alt JA, Schlosser RJ, Smith TL, Soler ZM. Comparative effectiveness
385 of medical and surgical therapy on olfaction in chronic rhinosinusitis: a prospective, multi-
386 institutional study. *Int Forum Allergy Rhinol.* 2014;4(9):725-33.
- 387 10. Katotomichelakis M, Simopolous E, Tripsianis G, et al. Predictors of quality of life
388 outcomes in chronic rhinosinusitis after sinus surgery. *Eur Arch Otorhinolaryngol* 2014;
389 271(4):733-41.
- 390 11. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*
391 2018;15(4):233-234.
- 392 12. Smith TL, Kern R, Palmer JN, et al. Medical therapy vs surgery for chronic rhinosinusitis: a
393 prospective, multi-institutional study with 1-year follow-up. *Int Forum Allergy Rhinol*
394 2013;3(1):4-9.
- 395 13. DeConde AS, Mace JC, Alt JA, et al. Longitudinal improvement and stability of the SNOT-
396 22 survey in the evaluation of surgical management for chronic rhinosinusitis. *Int Forum Allergy*
397 *Rhinol.* 2015; 5(3):233-239.
- 398 14. Soler ZM, Smith TL, Alt JA, et al. Olfactory-specific quality of life outcomes after
399 endoscopic sinus surgery. *Int Forum Allergy Rhinol.* 2016; 6(4): 407-413.
- 400 15. Ramakrishnan VR, Mace JC, Soler ZM, Smith TL. Examination of high-antibiotic users in a
401 multi-institutional cohort of chronic rhinosinusitis patients. *Int Forum Allergy Rhinol.*
402 2017;7(4):343-51.

- 403 16. Hopkins C, Hettige R, Soni-Jaiswal A, et al. Chronic rhinosinusitis outcome measures
404 (CHROME), developing a core outcome set for trials of interventions in chronic rhinosinusitis.
405 *Rhinology* 2018;56(1):22-32.
- 406 17. Lund VJ and Kennedy DW. Quantification for staging sinusitis. The Staging and Therapy
407 Group. *Ann Otol Rhinol Laryngol Suppl.* 1995;167:17-21.
- 408 18. Lund VJ and Mackay IS. Staging in rhinosinusitis. *Rhinology.* 1992;31(4):183-4.
- 409 19. Doty RL. The Smell Identification Test Administration Manual. 3. Haddon Heights, NJ:
410 Sensonics, Inc; 1995.
- 411 20. Doty RL. The Brief Smell Identification Test Administration Manual. Haddon Heights, NJ:
412 Sensonics, Inc; 2001.
- 413 21. El Rassi E, Mace JC, Steele TO, et al. Sensitivity analysis and diagnostic accuracy of the
414 Brief Smell Identification Test in patients with rhinosinusitis. *Int Forum Allergy Rhinol.*
415 2016;6(3):287-92.
- 416 22. Piccirillo JF, Merritt MG, Jr, Richards ML. Psychometric and clinimetric validity of the 20-
417 Item Sino-Nasal Outcome Test (SNOT-20) *Otolaryngol Head Neck Surg.* 2002;126(1):41-47.
- 418 23. Hopkins C, Gillett S, Slack R, et al. Psychometric validity of the 22-item Sinonasal Outcome
419 Test. *Clin Otolaryngol.* 2009;34(5):447-454.
- 420 24. Senior BA, Glaze C, Benninger MS. Use of the rhinosinusitis disability index (RSDI) in
421 rhinologic disease. *Am J Rhinol.* 2001; 15(1):15-20
- 422 25. Simopoulos E, Katotomichelakis M, Gouveris H, Tripsianis G, Livaditis M, Danielides V.
423 Olfaction-associated quality of life in chronic rhinosinusitis: adaptation and validation of an
424 olfaction-specific questionnaire. *Laryngoscope* 2012;122:1450-1454.
- 425 26. Benjamini Y and Hochberg Y. Controlling the false discovery rate: a practical and powerful
426 approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*
427 1995;57(1):289-300.
- 428 27. R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna,
429 Austria: R Foundation for Statistical Computing.
- 430 28. Hastie T, Tibshirani R, and Friedman J. The wrong and right way to do cross-
431 validation. *Elements of Statistical Learning, Data Mining, Inference, Prediction*, Springer, NY
432 2009: 245-247.
- 433 29. Kuhn, M. Building predictive models in R using the caret package. *Journal of statistical*
434 *software* 2008; 28(5):1-26.
- 435 30. Boulesteix AL, Janitza S, Kruppa J, and König IR. Overview of random forest methodology
436 and practical guidance with emphasis on computational biology and bioinformatics. *Wiley*
437 *Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2012;2(6):493-507.
- 438 31. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance
439 measures: Illustrations, sources and a solution. *BMC bioinformatics* 2007;8(1):25.
- 440 32. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference
441 framework. *Journal of Computational and Graphical statistics* 2006;15(3):651-674.

- 442 33. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application,
443 and characteristics of classification and regression trees, bagging, and random
444 forests. *Psychological methods* 2009; 14(4):323.
- 445 34. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal*
446 *Statistical Society: Series B (Methodological)* 1996;58(1):267-288.
- 447 35. Friedman J, Hastie T, and Tibshirani R. Regularization paths for generalized linear models
448 via coordinate descent. *Journal of statistical software* 2010;33(1):1.
- 449 36. Friedman JH. Multivariate adaptive regression splines. *Annals of Statistics* 1991;1-67.
- 450 37. Cristianini N and Shawe-Taylor J. An introduction to support vector machines and other
451 kernel-based learning methods. Cambridge university press, 2000.
- 452 38. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: data mining,
453 inference, and prediction. Springer Science & Business Media 2009.
- 454 39. Lötsch J, Kringel D, Hummel T. Machine learning in human olfactory research. *Chemical*
455 *senses* 2019;44(1):11-22.
- 456 40. Fernández-Delgado M, Cernadas E, Barro S, Amorim, D. Do we need hundreds of classifiers
457 to solve real world classification problems? *The journal of machine learning research*,
458 2014;15(1): 3133-3181.
- 459 41. Kim JH. Estimating classification error rate: Repeated cross-validation, repeated hold-out
460 and bootstrap. *Computational statistics & data analysis* 2009;53(11):3735-3745.
- 461 42. He H and Garcia EA. Learning from imbalanced data. *IEEE Transactions on knowledge and*
462 *data engineering* 2009;21(9):1263-1284.
- 463 43. Stekhoven DJ and Bühlmann, P. MissForest—non-parametric missing value imputation for
464 mixed-type data. *Bioinformatics* 2012;28(1):112-118.
- 465 44. Nadeau C and Bengio Y. Inference for the generalization error. In *Machine Learning*
466 2003;52:239– 281.
- 467 45. Molnar C, Casalicchio G, Bischl B. iml: An R package for interpretable machine learning. *J*
468 *Open Source Software* 2018;3(26), 786.
- 469 46. Rombaux P, Huart C, Levie P, Cingi C, Hummel T. Olfaction in Chronic Rhinosinusitis.
470 *Curr Allergy Asthma Rep* 2016;16(5):41.
- 471 47. Litvack JR, Fong K, Mace J, et al. Predictors of olfactory dysfunction in patients with
472 chronic rhinosinusitis. *Laryngoscope* 2008;118:2225–30.
- 473 48. Michie D, Spiegelhalter D, Taylor C. *Machine Learning, Neural, and Statistical*
474 *Classification*. New York, NY: Ellis Horwood; 1994.
- 475 49. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*
476 2019;380(14):1347-1358.
- 477 50. Cuttillo CM, Sharma KR, Foschini L, et al; MI in Healthcare Workshop Working Group.
478 Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and
479 transparency. *NPJ Digit Med* 2020;3:47.

- 480 51. 2018 National Healthcare Quality and Disparities Report. Content last reviewed April 2020.
481 Agency for Healthcare Research and Quality, Rockville, MD.
482 <https://www.ahrq.gov/research/findings/nhqrdr/nhqrdr18/index.html>
- 483 52. Soler ZM, Mace JC, Litvack JR, Smith TL. Chronic rhinosinusitis, race, and ethnicity. Am J
484 Rhinol Allergy 2012 ;26(2):110-6.
- 485 53. Gitomer SA, Fountain CR, Kingdom TT, et al. Clinical Examination of Tissue Eosinophilia
486 in Patients with Chronic Rhinosinusitis and Nasal Polyposis. Otolaryngol Head Neck Surg
487 2016;155(1):173-8.
- 488 54. Settiple GA. Nasal polyps and immunoglobulin E (IgE). Allergy Asthma Proc
489 1996;17(5):269-73.
- 490

491 **TABLE AND FIGURE LEGEND**

492

493 **Table 1. Study population.** Collected demographics and clinical metadata used as inputs for
 494 classification models are shown here for the overall cohort, and those with and without olfactory
 495 dysfunction. Count (%) and Fisher’s exact test were used for categorical variables, while median
 496 (25th quantile, 75th quantile) and Wilcoxon rank sum test were used for numeric variables. Data
 497 sorted from smallest to largest q-value (multiple testing adjusted p-value).

498

Characteristic	Overall, n = 611	n	Olfactory Dysfunction		p- value	q- value
			Yes, n = 203	No, n = 408		
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

Characteristic	Overall, n = 611	n	Olfactory Dysfunction		p- value	q- value
			Yes, n = 203	No, n = 408		
No	397 (65%)		110 (54%)	287 (70%)		
Yes	214 (35%)		93 (46%)	121 (30%)		
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 DEPENDENCE		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%)		

Characteristic	Overall, n = 611	n	Olfactory Dysfunction		p- value	q- value
			Yes, n = 203	No, n = 408		
ALLERGY by HISTORY		611			0.143	0.277
No	480 (79%)		152 (75%)	328 (80%)		
Yes	131 (21%)		51 (25%)	80 (20%)		
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

Characteristic	Overall, n = 611	n	Olfactory Dysfunction		p- value	q- value
			Yes, n = 203	No, n = 408		
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%)		
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%)		

499

500

501

502

503 **Figure 1. Comparing AUC between methods. A and B) Classification accuracy for the**
504 **different models.** The sample mean and standard deviation of the performance metrics is
505 reported across all 50 cross-validation resampling estimates. **C) Statistical comparison of**
506 **methods.** Upper triangle depicts the mean difference in AUC (d) between the model on the y-
507 axis and the model on the x-axis. Lower triangle depicts the q-value (multiple testing adjusted p-
508 value) for the corrected resampled t-test. Blue values indicated statistically significant
509 differences in mean AUC between the corresponding models on the x-axis and y-axis. *AUC =*
510 *area under receiver operating characteristic curve. SVM-Radial = support vector machine with*
511 *a radial basis kernel; RF = random forest; Log Reg-Step = logistic regression with step-wise*
512 *variable selection.*

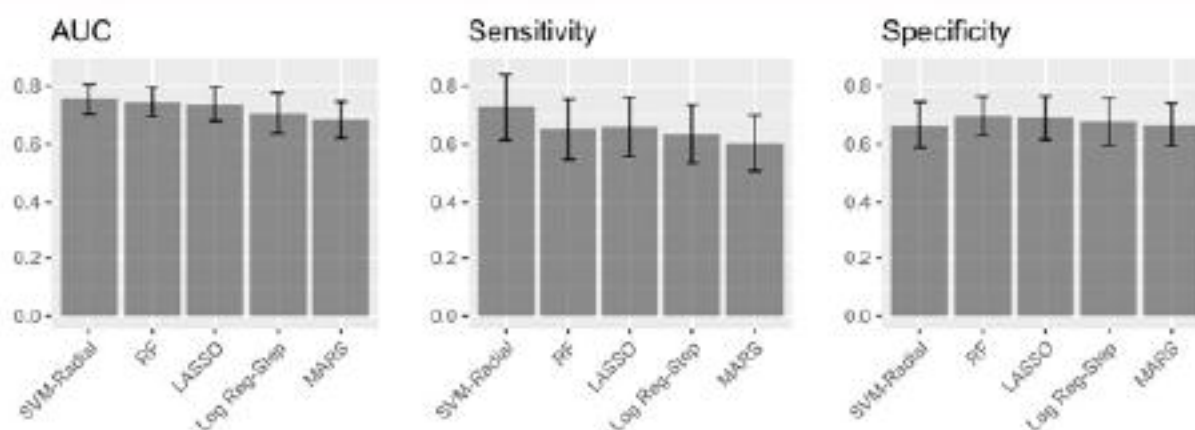
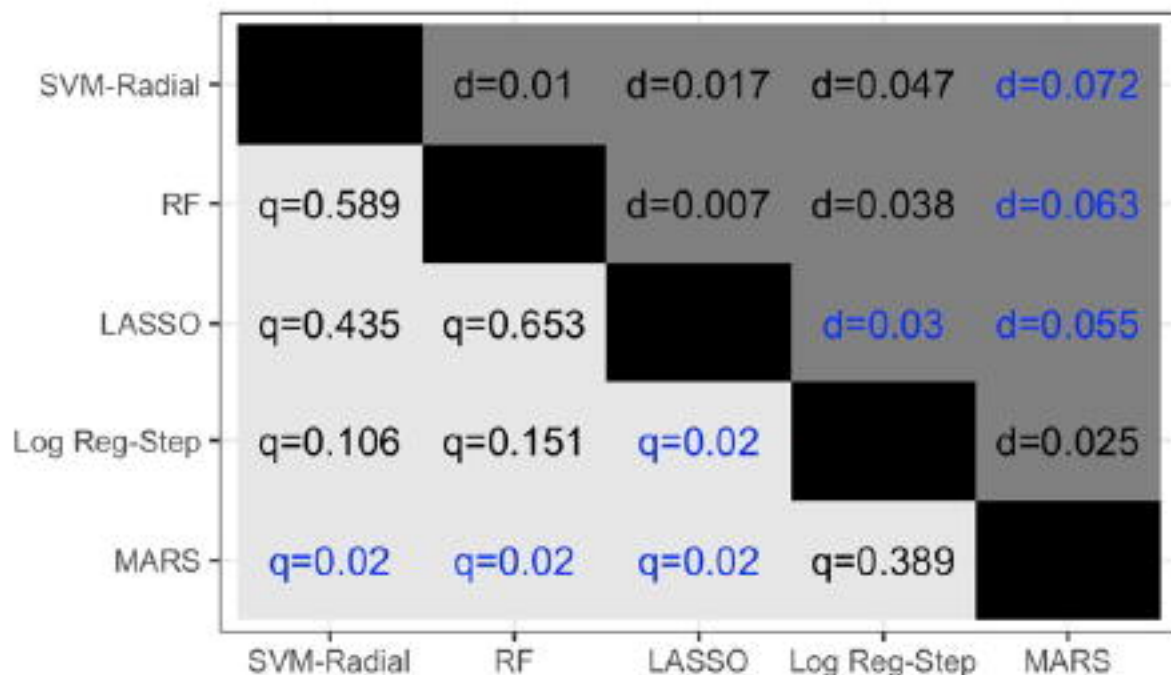
513
514
515 **Figure 2. Variable importance display for most accurate classification model (Support**
516 **Vector Machine with a radial basis kernel).** Note the inclusion of many predictor variables –
517 32 are included with > 10% variable importance, suggesting potential interaction between
518 predictors. *SVM-Radial = support vector machine with a radial basis kernel.*

519
520
521 **Figure 3. Top 10 Predictors and Variable importance for Random Forest and LASSO**
522 **models.** *AERD = aspirin exacerbated respiratory disease; CF = cystic fibrosis.*

523
524
525 **Supplementary Figure 1. Complete list of predictors for Random Forest and LASSO**
526 **models.** Notably, the LASSO model incorporates significantly more features of interest than the
527 Random Forest algorithm (26 features with >10% importance in the prediction algorithm).
528

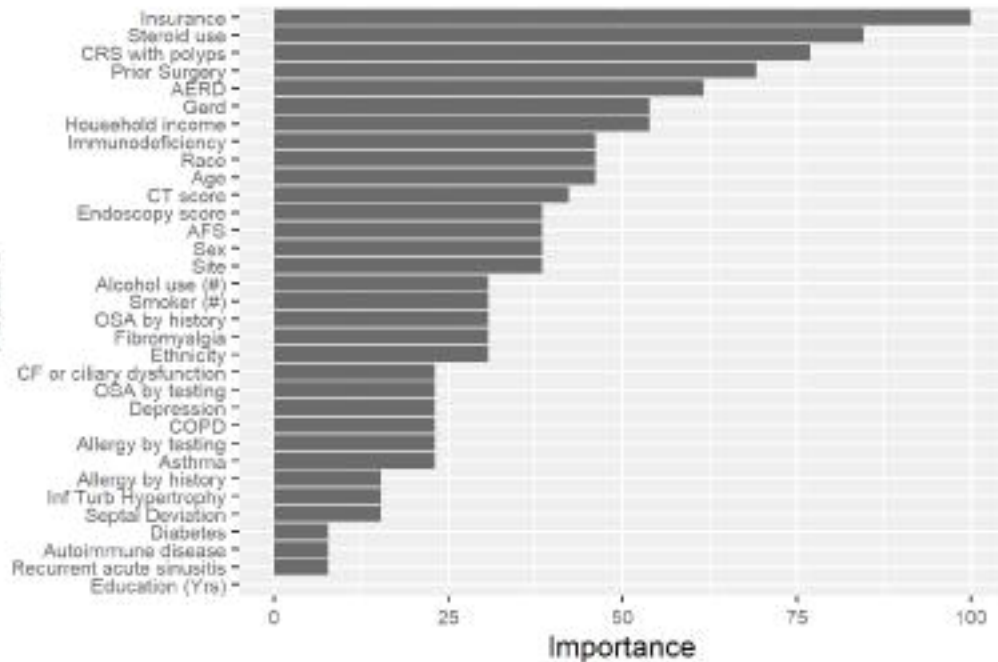
A

Model	AUC (SD)	Sensitivity (SD)	Specificity (SD)
SVM-Radial	0.754 (0.051)	0.726 (0.114)	0.665 (0.079)
Random Forest	0.744 (0.049)	0.649 (0.103)	0.696 (0.069)
LASSO	0.737 (0.059)	0.659 (0.102)	0.689 (0.076)
Log Reg-Step	0.707 (0.070)	0.635 (0.100)	0.676 (0.080)
MARS	0.682 (0.063)	0.604 (0.096)	0.667 (0.074)

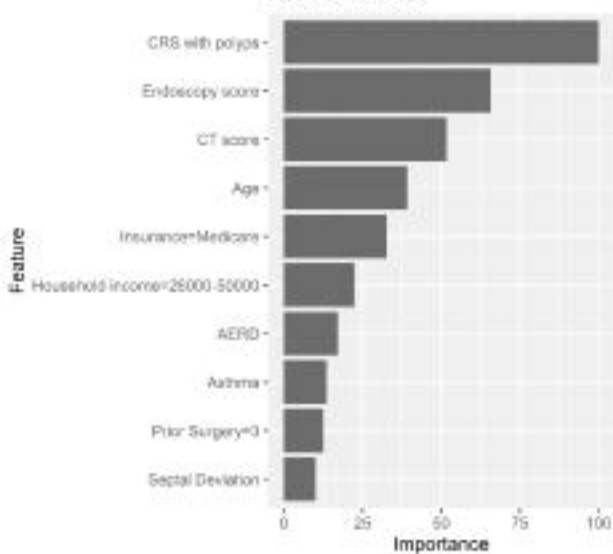
B**C**

SVM-Radial

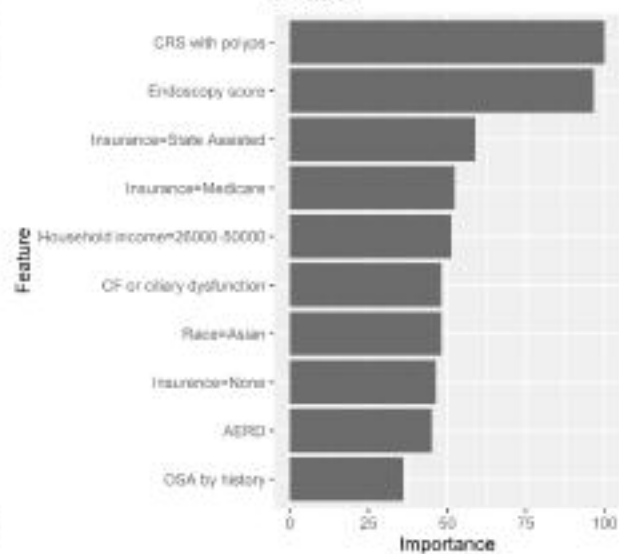
Feature



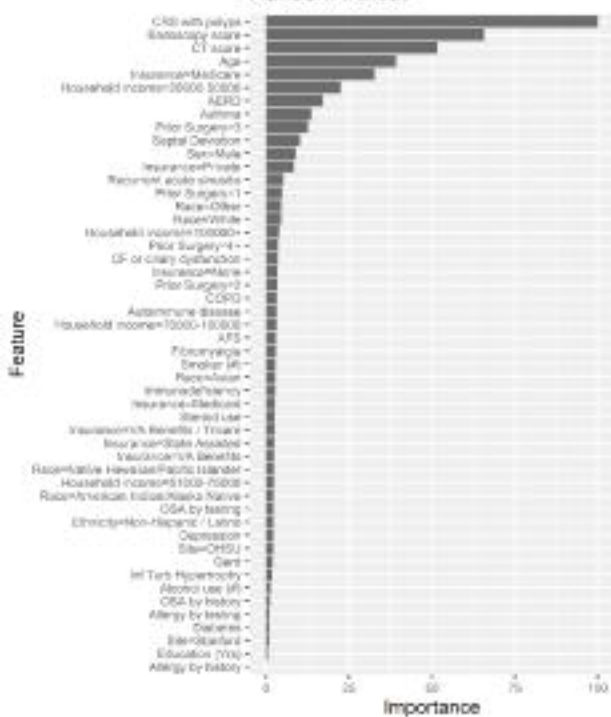
Random Forest



LASSO



Random Forest



LASSO

