Predicting treatment outcomes using explainable machine learning in children with asthma

Mario Lovric[1], Ivana Banic[2], Emanuel Lacic[1], Roman Kern[1], Kristina Pavlovic[1], and Mirjana Turkalj[3]

[1]Know-Center
[2]Children's Hospital Srebrnjak
[3]Srebrnjak Children's Hospital

July 28, 2020

Abstract

Background Asthma in children is a heterogeneous disease manifested by various phenotypes and endotypes. The level of disease control as well as the effectiveness of anti-inflammatory treatment is variable and inadequate in a significant portion of patients. Objectives By applying machine learning algorithms, we aimed to predict treatment success in a pediatric asthma cohort and to identify key variables for understanding underlying mechanisms. Methods We predicted treatment outcomes in children with mild to severe asthma (N=365), according to changes in asthma control, lung function (FEV1, MEF50) and FENO values after 6 months of controller medication use, using RandomForest and AdaBoost classifiers. Results The highest prediction power is achieved for control- and, to lower extend, for FENO-related treatment outcomes. The most predictive variables for asthma control are related to asthma severity and total IgE, which was also predictive for FENO-based outcomes. MEF50-related treatment outcomes were better predicted than FEV1-based response and one of the best predictive variables for this response was hsCRP. Conclusions Our results suggest that asthma control- and FENO-based outcomes can be more accurately predicted using machine learning than FEV1 and MEF50. This supports the symptom control-based asthma management approach and its complementary FENO-guided tool in children. T2-high asthma seemed to respond best to anti-inflammatory treatment. The prediction of MEF50-based treatment outcomes emphasizes the role of the distal airways in childhood asthma. The results of this study in predicting treatment success will help to enable treatment optimization and to implement the concept of precision medicine in pediatric asthma treatment.

Predicting treatment outcomes using explainable machine learning in children with asthma

Short title: Predicting asthma treatment outcomes in children

Mario Lovrić [a, b, 1], Ivana Banić [a, 1, *], Emanuel Lacić [b], Roman Kern [b], Kristina Pavlović [b], Mirjana Turkalj [a, c, d]

[a] Srebrnjak Children's Hospital, Srebrnjak 100, HR-10000 Zagreb

[b] Know-Center, Infeldgasse 13, AT-8010 Graz

[c] Faculty of Medicine, J.J. Strossmayer University of Osijek, Josipa Huttlera 4, HR-31000 Osijek

[d] Catholic University of Croatia, Ilica 242, HR-10000 Zagreb

[*] corresponding author:*ibanic@bolnica-srebrnjak.hr*, tel: +385 1 6391 191, fax: +385 1 6391 188

[1] these authors contributed equally

Word count: 3309

ML performed the ML analysis and wrote the manuscript. IB collected the data, performed the experiments and wrote the manuscript. EL and KP performed a part of the ML analysis. RK and MT designed the study and participated in writing the manuscript, MT additionally enrolled the participants into the study.

Abstract

Background Asthma in children is a heterogeneous disease manifested by various phenotypes and endotypes. The level of disease control as well as the effectiveness of anti-inflammatory treatment is variable and inadequate in a significant portion of patients.

Objectives By applying machine learning algorithms, we aimed to predict treatment success in a pediatric asthma cohort and to identify key variables for understanding underlying mechanisms.

Methods We predicted treatment outcomes in children with mild to severe asthma (N=365), according to changes in asthma control, lung function (FEV1, MEF50) and FENO values after 6 months of controller medication use, using RandomForest and AdaBoost classifiers.

Results The highest prediction power is achieved for control- and, to lower extend, for FENO-related treatment outcomes. The most predictive variables for asthma control are related to asthma severity and total IgE, which was also predictive for FENO-based outcomes. MEF50-related treatment outcomes were better predicted than FEV1-based response and one of the best predictive variables for this response was hsCRP.

Conclusions Our results suggest that asthma control- and FENO-based outcomes can be more accurately predicted using machine learning than FEV1 and MEF50. This supports the symptom control-based asthma management approach and its complementary FENO-guided tool in children. T2-high asthma seemed to respond best to anti-inflammatory treatment. The prediction of MEF50-based treatment outcomes emphasizes the role of the distal airways in childhood asthma. The results of this study in predicting treatment success will help to enable treatment optimization and to implement the concept of precision medicine in pediatric asthma treatment.

Keywords: asthma control, asthma controller medication, childhood asthma, machine learning, treatment outcome

Introduction

The aim of personalized medicine is to provide a target therapy for each individual or phenotype, based on the corresponding syndrome or disease[1]. Even though machine learning techniques have identified a number of structures and/or phenotypes in asthma, one has to be careful in clinical interpretation of these structures, as they may not represent true endotypes (underlying immunopathological mechanisms)[2]. Overlaps in endotypes as well as clinical presentation of the disease make the delivery of personalized asthma treatment quite elusive[3,4]. Furthermore, the same pattern of symptoms does not necessarily indicate the same underlying mechanism and moreover, different mechanisms are not mutually exclusive and may even act synergistically.

The emergence of advanced machine learning algorithms, abundance of clinically significant data and computing power can be attributed as key enablers in the development of personalized medicine. There is a substantial body of scientific work on data-driven methods in asthma phenotyping and the variables as well as the model chosen can largely affect the models and obtained results[5-11]. Careful selection of both the predictive model and the dataset are essential in such studies, with expert clinical interpretation being of utmost importance.

In childhood asthma, inhaled corticosteroids- ICS (in combination with long-acting beta-agonists, LABA and/or add-on leukotriene receptor antagonists, LTRA) remain controller medications of choice, although

evidence shows that a significant proportion of patients fail to respond adequately to such treatment.[5–8]The complexity of the disease, or better said, the "umbrella" diagnosis of asthma that encompasses a number of different phenotypes underpinned by different pathophysiological mechanisms or distinct inflammatory pathways (endotypes) seems to be the major obstacle in asthma management as well as in the development of personalized treatment approaches[9].

An important study on prediction was conducted by Belgrave et al.[10], focusing on preschool wheezers (N=150) with a large dataset ($N_{(variables)}$=636) using selected state-of-the-art techniques for data processing and machine learning and obtained 90%+ performance in Kappa statistics. The authors also reported robustness and performance quality when using Random Forest and that subjective variables are important in distinguishing ill patients from controls.

A key study focusing on asthma control-based response to controller medication was conducted in the Childhood Asthma Management Program (CAMP) cohort using novel machine learning algorithms[11]. They reported that asthma control, bronchodilator response and serum eosinophils were the most predictive features of asthma control, regardless of the medication used.

Our aim and contribution

In this work we present the data and results from an observational childhood asthma study in an ethnically homogeneous, age diverse cohort, reflecting real-life clinical situations with the majority of patients having mild to moderate disease. The primary aim was to test the predictive possibilities for treatment success in pediatric asthma patients and reveal key variables for understanding the mechanisms underlying such response. The primary endpoints were four different parameters of response to treatment, assessed by changes in lung function i) Forced expiratory volume in one second (FEV1) and ii) Maximal Expiratory Flow at 50% of Vital Flow Capacity (MEF50), iii) changes in airway inflammation- Fractional Exhaled nitric oxide (FENO) and iv) the level of asthma control assessed by a pediatric pulmonologist. Each of these targets was evaluated at baseline and after 6 months of treatment use alongside other parameters and biomarkers. The predictive possibilities were tested using the Random Forest (RF) and Adaptive Boosting (AdaBoost) machine learning algorithms. Since these algorithms are considered to be black-boxes, we introduced model explainability by the use of variable importance for evaluating the most important variables for differentiating between responders and non-responders. Furthermore, we discuss the use of different classification metrics for the validation of the results.

Methods

Population studied

365 pediatric patients (355 children aged 2-17 and ten adolescents aged 18-22) with atopic and non-atopic, mild to severe persistent asthma[12], were recruited in a prospective, non-interventional type of clinical study at the Srebrnjak Children's Hospital outpatient clinic. Informed consent was obtained from the children's parents/legal guardians. The study protocol was approved by the local Ethics Committee. All patients underwent physical examination, anthropometric measurements and standard diagnostic procedures to establish a diagnosis of asthma and guide its management (Table 1). The patients started treatment with ICS (alone or in combination with LABA) and/or LTRA, according to disease severity and previously assessed level of disease control. A follow-up visit with lung function and airway inflammation testing was made after 6 months of treatment use. Additionally, treatment outcomes and the level of asthma control (according to the Global Initiative for Asthma, GINA[12]) were assessed at the follow-up visit. In total 280 features (variables) were collected. The observational study is described in the supplementary file in detail.

Response variables

According to their response to treatment after 6 months of medication use, the patients were divided into "responders" and "non-responders" in accordance with the Minimal Clinically Important Difference (MCID) for lung function adjusted for children (% of predicted lung function) and data from other studies taking into account changes in the level of asthma control (LOAC) and changes in FENO[13–17]. The response variables are described in detail in Table 2.

Data preparation and balancing

We used Python scripts and methods previously described for data processing and modelling[18]. Variables with more than 10% missing values were removed. Those with fewer missing values were imputed by their respective median for continuous variables or mode for discrete variables. To avoid the "curse of dimensionality"[19], we aggregated individual variables describing allergic sensitization (skin prick test- SPT and allergen-specific immunoglobulin E- sIgE test results). These variables were binarized and summed into 4 categories: seasonal inhaled, perennial inhaled, insect venom and food allergens. Strong sensitization to house dust mite, cat dander and ragweed were treated separately due to their association with disease severity and more severe outcomes[20,21]. The dataset consisted of 365 patients and 73 variables. We dealt with an imbalanced classification problem (see Table 3), i.e. responders (1) or non-responders (0) could have been underrepresented. In imbalanced classification predictive models tend to recognize the major class better while struggling with the often scarce minor class, meaning that predictions may be biased towards the major class[18,22]. To avoid this, we employed synthetic data generation techniques, namely oversampling and under sampling (on the training set exclusively).

A powerful method for oversampling is Synthetic Minority Over-sampling Technique (SMOTE)[23], that has previously been reported for predicting lung disease outcomes[10]. Since our dataset was heterogeneous, we used the adapted algorithm, Synthetic Minority Over-sampling Technique for Nominal and Continuous (SMOTENC)[23]. We utilized Cluster Centroids (CC)[24,25] as the most promising under sampling approach.

Machine learning

Our aim was to estimate to which class a patient belongs (0/1) after treatment based on predictive variables and to predict the patients' future responses. The employed ensemble classification algorithms follow a paradigm where multiple "weak classifiers" are trained and averaged to improve the prediction abilities and lower the prediction error.

The basis for ensemble classifiers are decision trees (Figure 1, left)[26]. With the appearance of boosting[27] and bootstrapping strategies[26] combined predictors started to emerge. Boosting algorithms[28] are often utilized in industry[18] and personalized medicine[29]. RF (Figure 1, right) has shown good results in predicting pediatric asthma outcomes[10]. Except for their excellent performance, decision-tree-based classifiers do not require tedious data preparation and are convenient for working with heterogeneous data. We used two types of classifiers in our research: the AdaBoost and RF classifier. The data was split[22] into train (75%) and test (validation) sets (25%). The experimental matrix is described in Table 4.

For model explanation we used permutation importance (PI) which we used in our prior work[30]. It follows the rationale that a random permutation of a predictor variable values as well as the difference in the classification metrics before and after permuting a predictor variable are used as an importance measure[31]. This procedure is even more relevant when considering that bootstrapping (resampling with replacement) is used in ensemble classifiers, e.g. not all variables will appear in each tree. This adds up in revealing true predictors in the models and can be also used for feature selection in machine learning models[32]. Due to imbalance in the targets we stratified the minor class in a train and test set. The model quality metrics used in this work were Accuracy, Sensitivity, Specificity and the Matthews correlations coefficient (MCC).[33,34] A detailed description of these metrics is given in the supplementary data.

Results and discussion

Performance in prediction of treatment outcomes

Table 5 presents the best achieved classification results for each particular treatment outcome. The highest prediction accuracy is achieved for LOAC. This is due to a high number of correctly predicted outcomes for responders (indicated by specificity) and non-responders (indicated by sensitivity), as well as a high MCC. Although this trait is less objective than lung function and FENO as it encompasses symptom self-assessment, it reflects real-life treatment success best. This is in concordance with the control-based management approach, focusing on achieving adequate control of symptoms and minimizing future risks of exacerbations[12].

When predicting response according to FEV1, FENO and MEF50, an average accuracy between 65% and 70% was achieved. This suggests that lung function is not a preferred tool to be used to guide treatment in children with asthma, which is highlighted in current GINA guidelines. Lung function is a complex trait and reflects a number of structural and functional changes to the airways due to chronic inflammation. It does not correlate with symptom occurrence or severity well, especially in children, as certain patients with poor lung function may not exhibit severe symptoms and vice versa, certain patients with normal lung function may experience symptom aggravation[35,36]. Moreover, children with mild-to-moderate asthma using controller treatment exhibit a slower decline in lung function in comparison with deterioration of symptom control[37], which is probably why the model predicts these traits poorly, given the fact that most of the patients in our study had milder disease forms. Compared to lung function-based treatment outcomes predicting outcomes assessed by changes in FENO performed better as it showed a slightly higher accuracy and much better sensitivity (good prediction performance for responders), see Table 5. This suggests that FENO can be used as a predictor of steroid responsiveness even more consistently than other parameters, e.g. lung function[16,38]. FENO is a good biomarker of Th2- related allergic inflammatory response, as interleukin-13 promotes nitric oxide (NO)- synthase activity and NO production[39]. Moreover, the latest GINA guidelines[12] suggest that treatment guided by FENO in children and young adults is associated with a significant reduction in exacerbation rates and that it may be a good complementary approach compatible with control-based asthma management. Additionally, since FENO-based response was able to distinguish true responders quite well, it may be useful in identifying patients with ineffective or suboptimal treatment-those that require treatment adjustment[12]and those with poor adherence to treatment[40].

However, for treatment outcomes according to lung function and FENO, a much lower MCC (21%-26%) was achieved when compared to LOAC. This suggests that the model generates a significant proportion of false responders and non-responders for lung function- and FENO-based outcomes, which further supports the control-guided asthma management approach as a preferred option in guiding asthma treatment in children. Additional results by means of Receiver Operating Characteristic (ROC) curves[41] and confusion matrices are presented in the supplementary Figure s1.

Differences in the utilized classification algorithms and sampling method

We utilized two different classification methods and three sampling techniques. Figure 2 shows the distribution of MCC across sampling methods and classifiers. AdaBoost was the better performing classifier for LOAC, FEV1, FENO, except of MEF50 where RF outperformed marginally. This is of no surprise since boosting algorithms generally show good performance with imbalanced sets[42]. Overall, not sampling the data has in our case led to the best prediction results in combination with AdaBoost. Using oversampling resulted in a better MCC only for MEF50, which could indicate that when designing experiments like these one has to take care of non-uniform feature spaces for the rare or minor classes like responders here[42]. Even though the differences are marginal, in medicine even the slightest improvement may be important. These results can be explained by the advantage of AdaBoost which learns sequentially on misclassification from previous weak learners in the sequence and while over/under sampling improves the results for RF, it is not the case with AdaBoost. Additionally, since RF is trained in parallel, it is much faster in practice, hence for

training fast and large data sets RF with oversampling could be used.

Model interpretation

For each target modeled in the experimental matrix, we calculated the average PI for predictive variables (Table 6). The only predictive variable passing the 1% threshold for the classification of LOAC-based outcomes is "asthma severity", highlighting the importance of an expert assessment in asthma management from the start as well as the impact of asthma heterogeneity and its phenotypes on treatment success. Additionally, since both asthma control and severity encompass the patient's (caregiver's) self-estimation on symptom frequency and severity, these findings emphasize the importance of the patient's involvement in the management plan, an essential part of the "shared-care approach" in asthma management associated with improved outcomes[12].

Additionally, LOAC appears to be associated with the least complexity in regard to prediction, with only one important predictor, in spite of its great power of prediction (Table 6). For this reason, we submitted this subset and the target LOAC to a decision tree classifier (Figure 1, left) to follow the decisions of the complex classifiers which consist of many such weak classifiers, see Figure 3. This exemplary decision tree (only a subpart of an ensemble classifier) shows that children with milder forms of the disease respond well to treatment with ICS as well as that severe patients do not respond to treatment adequately, even though their treatment was adjusted according to disease severity, i.e. severe patients at baseline remain uncontrolled after 6 months of medication use. This suggests that the model is capable of identifying severe patients quite accurately, but the shortfall is that it does not inform about the potential mechanisms underlying treatment failure. Also, it seems that prominent markers of atopy (total IgE) are highly predictive of treatment success. The vast majority of childhood asthma patients have allergic asthma and a number of studies have shown that it is sensitive to treatment with ICS. More specifically, T helper 2 lymphocyte, T2-high endotypes respond best to ICS. It also seems that very high total IgE values are predictive of treatment failure (Figure 3), which is in consistency with previous findings that high serum IgE is observed in children with severe asthma.[43–45]

Overall, the PI for any of the targets did not include any available treatment variables, meaning the models did not use treatment variables in creating decisions on treatment outcomes. Even though treatment follows guidelines, these are not definite nor objective *sensu stricto* . Guidelines actually provide general choice recommendations and the physician is left to choose between several treatment options. Although this may represent a potential bias in identifying true responders vs. non-responders, it actually reflects the model's power of prediction and favors the current symptom control-guided asthma management approach.

Although FENO had a substantially lower accuracy and MCC than LOAC, treatment outcomes according to FENO changes was capable of identifying true responders quite well, indicating that FENO-guided treatment may be a complementary tool in guiding asthma management in children. PI revealed that predicting FENO is more complex in comparison to LOAC (Table 6). This may be due to the fact that FENO reflects the level and type of airway inflammation that drives the chronicity of the disease. Elevated IgE and sensitization to inhaled allergens are common markers of T2-high inflammation[43] which is known to respond better to anti-inflammatory treatment.

When comparing MEF50 and FEV1 to each other, MEF50-based outcomes were predicted by FEV1 to a lesser extent than FEV1-based outcomes by MEF50. Additionally, MEF50-based outcomes were also predicted by hsCRP (Table 6) which is a marker of subtly elevated systemic inflammation in asthma. Evidence shows that increased hsCRP is associated with more severe asthma outcomes[46]. This, in addition to the fact that the model predicting MEF50-related response performed better in almost all parameters (except specificity) compared to FEV1-related response (see Table 5) and the fact that oversampling further improved the models power in predicting true responders and non-responders (see Figure 2) for MEF50, highlights the importance of the distal airways in children with asthma[47]. The peripheral airways are the predominant site of airway inflammation[48] and may very well be a predominant site of airflow obstruction in asthmatic children, involved in the pathophysiology and resistance to treatment with ICS[49]. Moreover, distal airways

impairment may be present despite rare and mild asthma symptoms and normal FEV1 in pediatric patients.

Our results are in concordance with those of Ross et al.[11] who also identified asthma control as the strongest predictive variable for LOAC. These authors only focused on one type of ICS (budesonide) and chromones (nedocromil), while our study encompassed all commonly used classes of anti-inflammatory controller medication (ICS, LABA and LTRA). Ross et al.[11] only evaluated response to treatment according to symptom control, while our study involved lung function- and FENO-based treatment outcomes. Although the homogeneity of the population studied (a real-life situation with most of the children having allergic asthma and milder disease forms) could have been an advantage in identifying certain phenotypes and genetic traits associated with treatment outcomes, this was a disadvantage in identifying clear pathophysiological mechanisms involved. Moreover, the sample size in our study might have been small (N= 365 vs. N= 1019 in Ross et al.), possibly further hindering more detailed endotype characterization. Ross et al. identified serum eosinophils as one of the most predictive variables for asthma control, while we identified IgE, supporting previous findings that children with T2-high allergic asthma responds best to anti-inflammatory treatment.[50] Finally, even though GINA guidelines suggest treatment response review every 3-6 months, the assessment period in this study may have been too short to reflect biologically significant and measurable effects, especially on complex traits such as lung function changes in response to treatment.

Conclusion

Our goal was to evaluate the prediction of treatment outcomes in childhood asthma after six months of medication use based on initial assessment. Our results show that asthma control (LOAC) was well predicted, while the prediction quality of lung function- based treatment outcomes (FEV1, MEF50) was rather low. These results are in concordance with the GINA control-based management approach, while lung function may not be the tool of choice to be used in guiding treatment in children with asthma[21]. The prediction model for FENO-based treatment response performed better in almost all aspects than lung function-related outcomes, which suggests that treatment success guided by changes in FENO might be a complementary tool in childhood asthma management[21].

Our results also suggest that the current guidelines in asthma management and current expertise in clinical assessment (assessment of severity and disease control) are satisfactory in most cases. Additionally, since total IgE was one of the best predictive variables for FENO-based outcomes, this indicates that T2-high asthma subtypes in children respond best to common controller medication. Although machine learning has shown how treatment outcome prediction can be driven, it has revealed certain issues that need to be addressed in future studies:

- In respect with asthma chronicity the assessment period of 6 months may not be enough for valid predictions. Longitudinal and prospective studies are essential;
- Additional studies involving larger numbers of patients with even more clinically relevant parameters are required to increase the success of treatment outcome prediction and in further characterization of specific disease phenotypes and endotypes.

Recently, much focus has been given to the implementation of precision medicine in asthma, and experts emphasize that the time for action is now. The use of big data and machine learning in predicting treatment success such as the one in this study might enable treatment optimization and the development of new therapies for each defined endotype.

References

1. Hamburg MA, Collins FS. The Path to Personalized Medicine. *N Engl J Med* . 2010;363(4):301-304. doi:10.1056/NEJMp1006304

2. Saglani S, Custovic A. Childhood asthma: Advances using machine learning and mechanistic studies. *Am J Respir Crit Care Med* . 2019;199(4):414-422. doi:10.1164/rccm.201810-1956CI

3. Pavord ID, Beasley R, Agusti A, et al. After asthma: redefining airways diseases. *Lancet* . 2018;391(10118):350-400. doi:10.1016/S0140-6736(17)30879-6

4. Custovic A, Belgrave D, Lin L, et al. Cytokine responses to rhinovirus and development of asthma, allergic sensitization, and respiratory infections during childhood. *Am J Respir Crit Care Med* . 2018;197(10):1265-1274. doi:10.1164/rccm.201708-1762OC

5. Blais L, Kettani FZ, Lemire C, et al. Inhaled corticosteroids vs. leukotriene-receptor antagonists and asthma exacerbations in children.*Respir Med* . 2011;105(6):846-855. doi:10.1016/j.rmed.2010.12.007

6. Turkalj M, Erceg D. Terapijski pristup astmi u djece. *Medicus* . 2013;22(1):49-56.

7. Wu W, Bleecker E, Moore W, et al. Unsupervised phenotyping of Severe Asthma Research Program participants using expanded lung data. *J Allergy Clin Immunol* . 2014;133(5):1280-1288. doi:10.1016/j.jaci.2013.11.042

8. Szefler SJ, Phillips BR, Martinez FD, et al. Characterization of within-subject responses to fluticasone and montelukast in childhood asthma. *J Allergy Clin Immunol* . 2005;115(2):233-242. doi:10.1016/j.jaci.2004.11.014

9. Chung KF, Adcock IM. *Clinical Phenotypes of Asthma Should Link up with Disease Mechanisms* . Vol 15. Lippincott Williams and Wilkins; 2015:56-62. doi:10.1097/ACI.0000000000000134

10. Belgrave D, Cassidy R, Stamate D, et al. Predictive modelling strategies to understand heterogeneous manifestations of asthma in early life. In: *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017* . Vol 2017-December. Institute of Electrical and Electronics Engineers Inc.; 2017:68-75. doi:10.1109/ICMLA.2017.0-176

11. Ross MK, Yoon J, Van Der Schaar A, Van Der Schaar M. Discovering pediatric asthma phenotypes on the basis of response to controller medication using machine learning. *Ann Am Thorac Soc* . 2018;15(1):49-58. doi:10.1513/AnnalsATS.201702-101OC

12. Global Initiative for Asthma. *Global Strategy for Asthma Management and Prevention, 2020.* ; 2020.

13. Reddel HK, Taylor DR, Bateman ED, et al. An official American Thoracic Society/European Respiratory Society statement: Asthma control and exacerbations - Standardizing endpoints for clinical asthma trials and clinical practice. *Am J Respir Crit Care Med* . 2009;180(1):59-99. doi:10.1164/rccm.200801-060ST

14. Dweik RA, Boggs PB, Erzurum SC, et al. An official ATS clinical practice guideline. *Am J Respir Crit Care Med* . 2011;184(5):602-615. doi:10.1164/rccm.9120-11ST T4 - interpretation of exhaled nitric oxide levels (FENO) for clinical applications PM - 21885636 M4 - Citavi

15. de Jongste JC. Yes to NO: The first studies on exhaled nitric oxide-driven asthma treatment. *Eur Respir J* . 2005;26(3):379-381. doi:10.1183/09031936.05.00080705

16. Smith AD, Cowan JO, Brassett KP, Herbison GP, Taylor DR. Use of Exhaled Nitric Oxide Measurements to Guide Treatment in Chronic Asthma.*N Engl J Med* . 2005;352(21):2163-2173. doi:10.1056/NEJMoa043596

17. Pellegrino R, Viegi G, Brusasco V, et al. Interpretative strategies for lung function tests. *Eur Respir J* . 2005;26(5):948-968. doi:10.1183/09031936.05.00035205

18. Žuvela P, Lovric M, Yousefian-Jazi A, Liu JJ. Ensemble Learning Approaches to Data Imbalance and Competing Objectives in Design of an Industrial Machine Vision System. *Ind Eng Chem Res* . 2020;59(10):4636-4645. doi:10.1021/acs.iecr.9b05766

19. Friedman JH. On bias, variance, 0/1-loss, and the curse-of-dimensionality. *Data Min Knowl Discov* . 1997;1(1):55-77. doi:10.1023/A:1009778005914

20. Sheehan WJ, Phipatanakul W. Indoor allergen exposure and asthma outcomes. *Curr Opin Pediatr* . 2016;28(6):772-777. doi:10.1097/MOP.0000000000000421

21. Lombardi C, Savi E, Ridolo E, Passalacqua G, Canonica GW. Is allergic sensitization relevant in severe asthma? Which allergens may be culprit? *World Allergy Organ J* . 2017;10(1). doi:10.1186/s40413-016-0138-8

22. Chicco D. Ten quick tips for machine learning in computational biology. *BioData Min* . 2017;10(1). doi:10.1186/s13040-017-0155-3

23. Chawla N V., Bowyer KW, Hall LO, Kegelmeyer WP. *SMOTE: Synthetic Minority Over-Sampling Technique* . Vol 16.; 2002:321-357. doi:10.1613/jair.953

24. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng* . 2009;21(9):1263-1284. doi:10.1109/TKDE.2008.239

25. Zhang YP, Zhang LN, Wang YC. Cluster-based majority under-sampling approaches for class imbalance learning. In: *Proceedings - 2010 2nd IEEE International Conference on Information and Financial Engineering, ICIFE 2010* . ; 2010:400-404. doi:10.1109/ICIFE.2010.5609385

26. Breiman L. Random forests. *Mach Learn* . 2001;45(1):5-32. doi:10.1023/A:1010933404324

27. Freund Y, Schapire RE. Experiments with a New Boosting Algorithm. In: *Machine Learning : Proceedings of the Thirteenth International Conference* . ; 1996:148-156.

28. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In:*Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* . Vol 13-17-August-2016. Association for Computing Machinery; 2016:785-794. doi:10.1145/2939672.2939785

29. Malva A, Arpinelli F, Recchia G, Micheletto C, Alexander R. Artificial Intelligence applied to asthma biomedical research: a systematic review. In: *Medical Education, Web and Internet* . European Respiratory Society; 2019:PA1482. doi:10.1183/13993003.congress-2019.PA1482

30. Havaš Auguštin D, Šarac J, Lovrić M, et al. Adherence to Mediterranean Diet and Maternal Lifestyle during Pregnancy: Island–Mainland Differentiation in the CRIBS Birth Cohort.*Nutrients* . 2020;12(8):2179. doi:10.3390/nu12082179

31. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A. Conditional variable importance for random forests. *BMC Bioinformatics* . 2008;9:1-11. doi:10.1186/1471-2105-9-307

32. Šimić I, Lovrić M, Godec R, Kröll M, Bešlić I. Applying machine learning methods to better understand, model and estimate mass concentrations of traffic-related pollutants at a typical street canyon.*Environ Pollut* . 2020;263:114587. doi:10.1016/j.envpol.2020.114587

33. Lučić B, Batista J, Bojović V, et al. Estimation of Random Accuracy and its Use in Validation of Predictive Quality of Classification Models within Predictive Challenges. *Croat Chem Acta* . 2019;92(3). doi:10.5562/cca3551

34. Chicco D, Rovelli C. Computational prediction of diagnosis and feature selection on mesothelioma patient health records. *PLoS One* . 2019;14(1). doi:10.1371/journal.pone.0208737

35. Groningen EZ, Kerstjens HAM, Brand PLP, Koeter GH, Postma DS, De Jong PM. Influence of treatment on peak expiratory flow and its relation to airway hyperresponsiveness and symptoms. *Thorax* . 1994;49(11):1109-1115. doi:10.1136/thx.49.11.1109

36. Brand PLP, Duiverman EJ, Waalkens HJ, Van Essen-Zandvliet EEM, Kerrebijn KF. Peak flow variation in childhood asthma: Correlation with symptoms, airways obstruction, and hyperresponsiveness during long term treatment with inhaled corticosteroids. *Thorax* . 1999;54(2):103-107. doi:10.1136/thx.54.2.103

37. Tonascia J, Adkinson NF, Bender B, et al. Long-Term Effects of Budesonide or Nedocromil in Children with Asthma. *N Engl J Med* . 2000;343(15):1054-1063. doi:10.1056/NEJM200010123431501

38. Smith AD, Cowan JO, Brassett KP, et al. Exhaled nitric oxide: A predictor of steroid response. *Am J Respir Crit Care Med* . 2005;172(4):453-459. doi:10.1164/rccm.200411-1498OC

39. Bagnasco D, Ferrando M, Varricchi G, Passalacqua G, Canonica GW. A critical evaluation of Anti-IL-13 and Anti-IL-4 strategies in severe asthma. *Int Arch Allergy Immunol* . 2016;170(2):122-131. doi:10.1159/000447692

40. Price D, Ryan D, Burden A, et al. Using fractional exhaled nitric oxide (FeNO) to diagnose steroid-responsive disease and guide asthma management in routine care. *Clin Transl Allergy* . 2013;3(1):1-10. doi:10.1186/2045-7022-3-37

41. Tharwat A. Classification assessment methods. *Applied Computing and Informatics* . 2018.

42. G. Weiss. Mining with rarity: A unifying framework. *SIGKDD Explor* . 2004;6(1):7-19. doi:10.1145/1007730.1007734

43. Froidure A, Mouthuy J, Durham SR, Chanez P, Sibille Y, Pilette C. Asthma phenotypes and IgE responses. *Eur Respir J* . 2016;47(1):304-319. doi:10.1183/13993003.01824-2014

44. Licari A, Castagnoli R, Brambilla I, et al. Asthma endotyping and biomarkers in childhood asthma. *Pediatr Allergy, Immunol Pulmonol* . 2018;31(2):44-55. doi:10.1089/ped.2018.0886

45. Fitzpatrick AM, Jackson DJ, Mauger DT, et al. Individualized therapy for persistent asthma in young children. *J Allergy Clin Immunol* . 2016;138(6):1608-1618.e12. doi:10.1016/j.jaci.2016.09.028

46. Monadi, Mahmoud; Firouzjahi, Alireza; Hosseini, Amin; Javadian, Yahya; Sharbatdaran, Majid; Heidari B. Serum C-reactive protein in asthma and its ability in predicting asthma control, a case-control study. *Casp J Intern Med* . 2016;7(1):37-42.

47. Torrego Fernández A, Muñoz Cano RM. [Clinical relevance of distal airway involvement in asthma]. *Arch Bronconeumol* . 2011;47 Suppl 2:17-19. doi:10.1016/S0300-2896(11)70016-3

48. Corren J. Small airways disease in asthma. *Curr Allergy Asthma Rep* . 2008;8(6):533-539. doi:10.1007/s11882-008-0097-4

49. Ye ZH, Huang Y, Wang Y, Wang DJ. Association between body mass index and lung function in children with asthma after corticosteroids inhalation. *Chinese J Contemp Pediatr* . 2013;15(11):983-986. doi:10.7499/j.issn.1008-8830.2013.11.014

50. Walker C, Bode E, Boer L, Hansel TT, Blaser K, Johann-Christian Virchow J. Allergic and Nonallergic Asthmatics Have Distinct Patterns of T-Cell Activation and Cytokine Production in Peripheral Blood and Bronchoalveolar Lavage. *http://dx.doi.org/101164/ajrccm/1461109* . December 2012. doi:10.1164/AJRCCM/146.1.109

Table 1. The variables used in this study, described in more detail in the supplementary file. AR- allergic rhinitis, AD- atopic dermatitis, FVC- forced vital capacity, SPT- skin prick test, IgE- immunoglobulin E, ENT- ear/nose/throat, *GLCCI1* - glucocorticoid induced 1,*TBX21* - t-box 21, *CRHR1* - corticotropin releasing hormone receptor 1, *ADRB2* - beta-2 adrenergic receptor, *MMP9* - matrix metalloproteinase-9.

demographics	gender, age

subjective clinical data	at baseline- personal and family medical history- atopy status, allergic rhinitis (AR), atopic dermatitis (AD), food allergy and other comorbidities
objective clinical data	at baseline and after 6 months - symptom control, frequency and severity of exacerbations in the period since the last visit, lung function (FVC, FEV1, MEF50), airway inflammation (FENO) measurement and medication use; at baseline - skin prick (SPT) and total and specific IgE to inhaled allergens, blood eosinophils and neutrophils, anthropometric measures (height, weight, body mass index) and for certain patients with suggestive history for comorbidities -ENT examination, pH probing with impedance for diagnostics of laryngopharyngeal reflux and gastroesophageal reflux disease, polysomnography for diagnostics of obstructive sleep apnoea syndrome, SPT and specific IgE to food and insect venom allergens for diagnostics of food/insect venom allergy
genetic data	genotypes for rs37973 (*GLCCI1*), rs9910408 (*TBX21*), rs242941 (*CRHR1*), rs1876828 (*CRHR1*), rs1042713 (*ADRB2*) and rs17576 (*MMP9*)

Table 2. Patient stratification according to their response to treatment (target variables). Response to treatment is defined into more detail in the supplementary file. ppb- parts per billion.

	FEV_1	**MEF_{50}**	**FENO**
Responders	Increase[?]10% predicted	Increase[?]15% predicted	Decrease <20% for values >35 (50) ppb or <10
Non-responders	Change<10% predicted	Change<15% predicted	Decrease [?]20% FENO [?]20% for values over 35

Table 3. Distribution of responding and non-responding patients per measured outcome after 6 months of treatment. The 13 missing FENO response values were imputed.

Treatment outcome after 6 months of therapy	**Responders (1)**	**Non-responders (0)**
LOAC	230	135
FENO	248	104
FEV1	129	236
MEF50	126	239

Table 4. The experimental matrix consists of training two different classifiers x three sampling methods x four targets, meaning there 24 options (2x3x4), each resampled 100x train-test splitting (giving a total of 600 models trained per target).

Classification algorithm	**Sampling methods**	**Targets after six months of treatment use**
AdaBoost	No sampling (base)	MEF50
Random Forest	Under sampling (cluster centroids)	FEV1

	Oversampling	FENO
		LOAC

Table 5. Average classification results for the treatment targets FEV1, FENO, MEF50 and LOAC. The results are reported for the best performing model (classifier and sampling method) and are calculated by the mean of accuracy, specificity, sensitivity and the MCC.

Metric	**Treatment Outcome FEV1**	**Treatment Outcome FENO**	**Treatment Outcome MEF50**	**Treatment Outcome LOAC**
Accuracy	0.6503	0.7005	0.6753	0.9698
Specificity	0.8986	0.8531	0.8817	0.9661
Sensitivity	0.7854	0.9560	0.7855	0.9781
MCC	0.2190	0.2146	0.2608	0.9366

Table 6. Top important variables for each of the targets. The variables were aggregated by the median value of permutation importance per target (600 runs each). The permutation importance is divided by the respective MCC value from Table 5 and calculated as % of weight respective to the MCC i.e. contribution to MCC. For each target only several variables returned an aggregated median above 1%. hsCRP- high-sensitivity C-reactive protein.

Variable	**LOAC**	**FENO**	**FEV1**	**MEF50**
Seasonal allergens (SPT)		1.1%		
Asthma severity baseline	47.0%			
hsCRP				1.2%
IgE total		1.5%	3.2%	
FENO baseline		12.8%		
FEV1 baseline			14.8%	1.8%
MEF50 baseline			8.2%	30.3%

Figure 1. Left: Typical (weak) tree classifier. A group of patients with both responders and non-responders is to be separated based on given predictive variables. The first split happens with a variable which gives the best split for the two groups. The algorithms split the groups until it gets leaf nodes (tree bottom) as pure as possible for the aimed classes. Right: Simplified scheme of the random forest algorithm. The trees represent weak classifiers which are aggregated via voting to form a strong one. Every tree trains on a random part of the training data (bootstrapping). The AdaBoost classifier trains the trees sequentially instead of parallel.

Figure 2. Boxplot of the classification results by means of the MCC (x-axis). The comparison includes classification results for the four targets (FEV1, LOAC, FENO, MEF50), two classification algorithms (AB – Ada Boost) and two sampling methods (Oversampling- OS and Cluster Centroids- CC) compared to no-sampling (base). The best models assigned per target by a red square surrounding the box.

Figure 3. An exemplary decision tree classifier where the treatment outcome LOAC after six months was predicted by three predictive variables (LOAC baseline, Asthma severity baseline, IGE_total). The responders are assigned as R-LOAC, while the non-responders are assigned as NR-LOAC. Asthma severity baseline is the first split. Most of the responders will respond well to treatment if their Asthma severity was estimated to have a value of 1. In an ensemble classifier a few hundreds of these are trained on bootstrapped samples and averaged for prediction which is explained in Figure 1. Ass_asthma_sev_basline- asthma severity (according to GINA) grade assessed at baseline, ass_asthma_ctrl_basline- asthma control assessed at baseline, biom_ige_total- total serum IgE.