SUPPLEMENT ARTICLE

WILEY

Machine learning: A modern approach to pediatric asthma

Giovanna Cilluffo[1] | **Salvatore Fasola[1]** | **Giuliana Ferrante[2]** | **Amelia Licari[3,4]** | **Giuseppe Roberto Marseglia[5]** | **Andrea Albarelli[5]** | **Gian Luigi Marseglia[3,4]** | **Stefania La Grutta[1]**

[1]Institute for Biomedical Research and Innovation, National Research Council, Palermo, Italy

[2]Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties, University of Palermo, Palermo, Italy

[3]Pediatric Clinic, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

[4]Department of Clinical-Surgical, Diagnostic and Pediatric Sciences, University of Pavia, Pavia, Italy

[5]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy

Correspondence
Amelia Licari, Pediatric Clinic, Fondazione IRCCS Policlinico San Matteo, University of Pavia, viale C. Golgi 19, 27100 Pavia, Italy.
Email: amelia.licari@unipv.it

Editor: Elena Chiappini

Abstract

Among modern methods of statistical and computational analysis, the application of machine learning (ML) to healthcare data has been gaining recognition in helping us understand the heterogeneity of asthma and predicting its progression. In pediatric research, ML approaches may provide rapid advances in uncovering asthma phenotypes with potential translational impact in clinical practice. Also, several accurate models to predict asthma and its progression have been developed using ML. Here, we provide a brief overview of ML approaches recently proposed to characterize pediatric asthma.

KEYWORDS
asthma, children, machine learning, phenotypes

1 | INTRODUCTION

Modern methods of statistical and computational analysis offer a powerful tool to mine knowledge from various sources of healthcare data. In this context, machine learning (ML) has increasingly provided a new perspective for characterizing the heterogeneity of asthma among children and predicting its outcomes (Figure 1). Here, we provide a brief overview of ML approaches recently proposed to characterize pediatric asthma.

2 | MACHINE LEARNING

ML can be defined as the study of computer algorithms that improve automatically through experience. In this sense, ML is an umbrella term encompassing all computational methods designed for learning from experience (available data) to improve performance and make accurate predictions. Unsupervised ML is concerned with the identification of data patterns in the absence of any pre-defined outcome. Supervised ML involves learning a rule for predicting an outcome based on input-output examples.

Regarding unsupervised ML, data-driven approaches using clustering methods could help characterize heterogeneous features of diseases among distinct patients. By unveiling the underlying structure of the data, cluster analysis can gather a set of samples into different clusters. For example, the k-means algorithm is one of the most popular iterative descent clustering methods, aiming to minimize the sum of variance within clusters and maximize separation between clusters, thereby identifying distinct groups within the

---

Giovanna Cilluffo and Salvatore Fasola: co-first authors.

population. Conversely, agglomerative hierarchical clustering algorithms follow a tree structure where the elementary nodes represent the samples to be clustered, and the root node represents a supercluster containing all the samples.

Regarding supervised ML, different classifiers have been implemented using regression or classification methods. The most used approaches remain linear regression for quantitative outcomes and logistic regression for categorical outcomes. However, in the era of big data, the mining potential of ML has increased substantially, and more advanced models are expanding; some of them are briefly introduced below.

Regression trees are obtained by splitting the variables, constituting the tree's root node, into subsets that constitute the child nodes. Random forests are an ensemble learning method based on a family of decision trees. Support vector machines build a model that predicts new observations using a non-probabilistic binary linear classifier. Deep learning has developed as a modern sub-field of ML based on deep neural networks (NN), which can reproduce the mechanisms of the human brain in processing complex and high-dimensional data such as images, video, or text. Deep learning is also linked to natural language processing, which is concerned with the analysis of large amounts of human language data to perform speech recognition and text classification.

FIGURE 1 Word cloud analysis using the title of articles published in the last five years. The list of publications was obtained using the following search strategy (PUBMED): machine learning AND asthma AND children. The pre-processing procedures applied were as follows: (1) removing words in the search strategy, non-English words, or common words that do not provide information; (2) changing words into lower case, and (3) removing punctuation and white spaces. The size of the words is proportional to the observed frequency

> **Key Message**
>
> Machine Learning (ML) is a modern comprehensive approach to characterize pediatric asthma phenotypes effectively and may be a promising tool to predict asthma and its progression.

3 | PHENOTYPING OF PEDIATRIC ASTHMA

There is increasing recognition that pediatric asthma is a heterogeneous disease with multiple subtypes, which may have overlapping observable characteristics (phenotypes), but different underlying pathophysiological causes. In this context, ML has been recently used to uncover phenotypes of pediatric asthma and derive clusters based on a series of characteristics derived from high-dimensional clinical data sets (Table 1). The most used techniques for phenotyping asthma in children include clustering methods (k-means and hierarchical) and latent class analysis.

Different features of asthma have been identified to distinguish the clusters in several studies. Integrating the data derived from 613 children with asthma with clinical expert domain knowledge, Deliu et al. identified four distinctive features (age of onset, allergic sensitization, severity, and exacerbations in the previous year) informative of five different phenotypes of pediatric asthma.[1] In the Taiwanese Consortium of Childhood Asthma Study, Su et al. combined clinical and functional features with gene expression profiles of 351 asthmatic children. They obtained five distinct phenotypes of childhood asthma, differing in lung function, symptom frequency, healthcare utilization, percentages of eosinophils and neutrophils in peripheral blood, and serum IgE.[2] Also, five latent classes differing for demographic features, asthma control, sensitization, type 2 inflammatory markers, and lung function were identified from Fitzpatrick et al. among a large heterogeneous cohort of more than 2500 children with mild to moderate persistent asthma. Notably, multiple allergic sensitizations and partially reversible airflow limitation emerged as clinically valuable features of asthma in identifying children at the greatest risk for future exacerbation.[3] Similarly, clusters with prominent type 2 inflammation and features of greater asthma severity have also been noted in previous cluster analyses of children with asthma.[4,5] Although valuable information has been gained, the results of these data-driven approaches require further validation to expand our knowledge of asthma phenotypes in children.

4 | PREDICTION OF ASTHMA AND ITS PROGRESSION

Advanced predictive analytic techniques such as ML have been very recently used to achieve the prediction of asthma and its progression. Recently, a segmented logit regression model was used to

TABLE 1 ML approaches used in phenotyping asthma in children

ML approach	Study design and participants	Distinctive features of asthma clusters	Clusters identified	Ref.
Hierarchical clustering	Cross-sectional, 613 asthmatic children	Age of onset, allergic sensitization, severity, and exacerbations in the previous year	– Early-onset mild atopic asthma – Early-onset mild non-atopic asthma – Late-onset asthma – Difficult asthma – Exacerbation-prone asthma	Deliu et al.[1]
k-means clustering	Cross-sectional, 351 asthmatic children from the Taiwanese Consortium of Childhood Asthma Study	Lung function, symptom frequency, healthcare utilization, percentages of eosinophils and neutrophils in peripheral blood, and serum IgE	– Asthma with elevated RBC and wheeze episodes – Neutrophil-predominant asthma – Allergic asthma with preserved pulmonary function – Eosinophil-predominant asthma with poor pulmonary function – Asthma with low wheeze episodes	Su et al.[2]
LCA	Cross-sectional, 2593 children with mild to moderate persistent asthma	Demographic features, asthma control, sensitization, type 2 inflammatory markers, and lung function	– Multiple sensitization with partially reversible airflow limitation – Multiple sensitization with reversible airflow limitation – Lesser sensitization with reversible airflow limitation – Multiple sensitization with normal lung function – Lesser sensitization with normal lung function	Fitzpatrick et al.[3]

Abbreviations: LCA, latent class analysis; ML, machine learning; RBC, red blood cells.

compute the probability of having asthma in children In particular, 7.9% and 14.7% resulted in the estimated bronchodilator response cutoff values at which the probability of predicting asthma was found to change significantly.[6]

ML models have also been trained on electronic health record data to distinguish between persistent and transient asthma cases in childhood. All the five models tested (naïve Bayes, logistic regression, k-nearest neighbors, random forest, and gradient boosted trees) in the study were found to be able to predict childhood asthma persistence.[7]

ML has the potential to overcome the poor sensitivity and specificity of current prediction models for asthma exacerbations, including several associated factors such as epidemiologic, environmental, and physiologic factors.[8] By means of an automatic ML model, Luo et al developed a predictive model for severe asthma exacerbation. The model achieved an area under the curve of 0.86 (95% CI, 0.846–0.871) and showed a negative predictive value of 97.8%, demonstrating an ability to identify children who are not at risk for exacerbation.[9] In line with these findings, Sills et al. proved that advanced automatic ML techniques are superior to conventional ML approaches, such as random forest and logistic regression, in terms of the predictive power of the need for hospitalization of pediatric asthma patients in emergency departments (ED), suggesting that ML models could be successfully implemented to improve the ED workflow as well as to spare resources.[10]

Notwithstanding the aforementioned studies, further research is warranted to test and validate the model's generalizability on external datasets; moreover, ML models would benefit from achieving better predictive accuracy.

5 | CONCLUSION

ML has recently become an increasingly used analytic tool to phenotyping for many disease processes. In asthma research, ML approaches have provided rapid advances in addressing pediatric asthma heterogeneity and uncover different disease phenotypes with potential translational impact in clinical practice. The ability of ML to include several data from multiple sources yields the advantage of improving predictive accuracy over conventional statistical methods. However, the number of variables required for ML models could also be a limitation since it may be challenging to choose variables that contribute most to the model. In addition, some variables may not always be available, which can limit replication of the models in external populations and their potential for application in clinical practice. Therefore, we can conclude that, in the field of pediatric asthma, ML is promising but still a work in progress.

CONFLICT OF INTEREST

Authors declared they have no conflict of interests.

AUTHOR CONTRIBUTIONS

Giovanna Cilluffo: Methodology (equal); Writing-review & editing (equal). **Salvatore Fasola**: Methodology (equal); Writing-review & editing (equal). **Giuliana Ferrante**: Conceptualization (equal); Writing-review & editing (equal). **Amelia Licari**: Conceptualization (equal); Writing-original draft (equal). **Giuseppe Roberto Marseglia**: Methodology (equal); Writing-review & editing (equal). **Andrea Albarelli**: Methodology (equal); Writing-review & editing (equal). **Gian Luigi Marseglia**: Supervision (lead); Writing-review & editing (equal). **Stefania La Grutta**: Conceptualization (equal); Supervision (lead); Writing-review & editing (equal).

ORCID

Giovanna Cilluffo https://orcid.org/0000-0003-4305-3981
Salvatore Fasola https://orcid.org/0000-0002-5378-5505
Giuliana Ferrante https://orcid.org/0000-0001-9917-2387
Amelia Licari https://orcid.org/0000-0002-1773-6482
Giuseppe Roberto Marseglia https://orcid.org/0000-0002-3901-8334
Andrea Albarelli https://orcid.org/0000-0002-3659-5099
Gian Luigi Marseglia https://orcid.org/0000-0003-3662-0159
Stefania La Grutta https://orcid.org/0000-0001-8026-0715

REFERENCES

1. Deliu M, Yavuz TS, Sperrin M, et al. Features of asthma which provide meaningful insights for understanding the disease heterogeneity. *Clin Exp Allergy*. 2018;48(1):39-47.
2. Su MW, Lin WC, Tsai CH, et al. Childhood asthma clusters reveal neutrophil-predominant phenotype with distinct gene expression. *Allergy*. 2018;73(10):2024-2032.
3. Fitzpatrick AM, Bacharier LB, Jackson DJ, et al. Heterogeneity of mild to moderate persistent asthma in children: confirmation by latent class analysis and association with 1-year outcomes. *J Allergy Clin Immunol Pract*. 2020;8(8):2617-2627.e4.
4. Fitzpatrick AM, Teague WG, Meyers DA, et al. Heterogeneity of severe asthma in childhood: confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. *J Allergy Clin Immunol*. 2011;127(2):382-389.e1-13.
5. Just J, Gouvis-Echraghi R, Rouve S, Wanin S, Moreau D, Annesi-Maesano I. Two novel, severe asthma phenotypes identified during childhood using a clustering approach. *Eur Respir J*. 2012;40(1):55-60.
6. Sottile G, Ferrante G, Cilluffo G, et al. A model-based approach for assessing bronchodilator responsiveness in children: the conventional cutoff revisited. *J Allergy Clin Immunol*. 2021;147(2):769-772.e10.
7. Bose S, Kenyon CC, Masino AJ. Personalized prediction of early childhood asthma persistence: a machine learning approach. *PLoS One*. 2021;16(3):e0247784.
8. Navanandan N, Hatoun J, Celedón JC, Liu AH. Predicting severe asthma exacerbations in children: blueprint for today and tomorrow. *J Allergy Clin Immunol Pract*. 2021;9(7):2619-2626.
9. Luo G, He S, Stone BL, et al. Developing a model to predict hospital encounters for asthma in asthmatic patients: secondary analysis. *JMIR Med Inform*. 2020;8:e16080.
10. Sills MR, Ozkaynak M, Jang H. Predicting hospitalization of pediatric asthma patients in emergency departments using machine learning. *Int J Med Inform*. 2021;151:104468.