European Journal of Internal Medicine 104 (2022) 66–72

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim

Original article

A machine learning approach to characterize patients with asthma exacerbation attending an acute care setting

Maria D'Amato [a,*,#], Pasquale Ambrosino [b,#], Francesca Simioli [a], Sarah Adamo [c], Anna Agnese Stanziola [a], Giovanni D'Addio [d], Antonio Molino [a], Mauro Maniscalco [a,e,*]

[a] Department of Respiratory Medicine, Federico II University, Naples, Italy
[b] Istituti Clinici Scientifici Maugeri IRCCS, Cardiac Rehabilitation Unit of Telese Terme Institute, Telese Terme, Italy
[c] Department of Information Technology and Electrical Engineering, University of Naples "Federico II", Napoli, Italy
[d] Istituti Clinici Scientifici Maugeri IRCCS, Bioengineering Unit of Telese Terme Institute, Telese Terme, Italy
[e] Istituti Clinici Scientifici Maugeri IRCCS, Pulmonary Rehabilitation Unit of Telese Terme Institute, Telese Terme, Italy

ARTICLE INFO

Keywords:
Asthma
Biomarker
Rehabilitation
Disability
Outcome
Occupational medicine
Exercise capacity
Chronic obstructive pulmonary disease
Chronic disease

ABSTRACT

Background: One of the main problems in poorly controlled asthma is the access to the Emergency Department (ED). Using a machine learning (ML) approach, the aim of our study was to identify the main predictors of severe asthma exacerbations requiring hospital admission.

Methods: Consecutive patients with asthma exacerbation were screened for inclusion within 48 hours of ED discharge. A k-means clustering algorithm was implemented to evaluate a potential distinction of different phenotypes. K-Nearest Neighbor (KNN) as instance-based algorithm and Random Forest (RF) as tree-based algorithm were implemented in order to classify patients, based on the presence of at least one additional access to the ED in the previous 12 months.

Results: To train our model, we included 260 patients (31.5% males, mean age 47.6 years). Unsupervised ML identified two groups, based on eosinophil count. A total of 86 patients with eosinophiles ≥370 cells/μL were significantly older, had a longer disease duration, more restrictions to daily activities, and lower rate of treatment compared to 174 patients with eosinophiles <370 cells/μL. In addition, they reported lower values of predicted FEV_1 (64.8±12.3% *vs.* 83.9±17.3%) and FEV_1/FVC (71.3±9.3 *vs.* 78.5±6.8), with a higher amount of exacerbations/year. In supervised ML, KNN achieved the best performance in identifying frequent exacerbators (AUROC: 96.7%), confirming the importance of spirometry parameters and eosinophil count, along with the number of prior exacerbations and other clinical and demographic variables.

Conclusions: This study confirms the key prognostic value of eosinophiles in asthma, suggesting the usefulness of ML in defining biological pathways that can help plan personalized pharmacological and rehabilitation strategies.

1. Introduction

Asthma is a respiratory disease with a wide spectrum of clinical presentations. The prevalence in adults is 8.2% and 9.4% in children in the European Union [1]. Severe asthma is a clinical condition affecting patients with different characteristics but similar medical needs. One of the main problems with poorly controlled asthma is the need for access to an emergency department (ED), which places a heavy economic burden on health care systems, given that each emergency visit costs on average 5-fold more than an outpatient visit for asthma [2]. Despite this heavy social and economic burden, a relatively low number of studies has been specifically designed to identify the factors associated with hospital admission in severe asthma patients [3,4]. If psychological dysfunction has been occasionally proposed as the strongest factor predicting frequent exacerbations [5], most literature evidence suggests an association between elevated blood or sputum eosinophil counts and severe exacerbations in patients with uncontrolled asthma [6], particularly in the presence of elevated baseline levels of fractional exhaled

Abbreviations: RF, random forest.

* Corresponding authors.
E-mail addresses: marielladam@hotmail.it (M. D'Amato), mauro.maniscalco@icsmaugeri.it (M. Maniscalco).
# These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.ejim.2022.07.019
Received 24 May 2022; Received in revised form 14 July 2022; Accepted 26 July 2022
Available online 31 July 2022

nitric oxide (FeNO) [7]. However, while these biomarkers of type 2 inflammation have been reported as the strongest predictors also in other studies [8], one of the largest population-based articles on this topic suggests that, regardless of disease severity and therapy adherence, prior exacerbations represent the best predictor of future events [9]. Overall, despite prediction scores and current guidelines [10], severe exacerbations still remain a major issue in asthma management, highly impacting prognosis and quality of life.

Recently, the usefulness of artificial intelligence in medical research has been widely demonstrated and reported in the literature [11–14]. Unlike conventional statistical models which draw population inferences from a sample, machine learning (ML) is able to find generalizable predictive patterns [15]. Thus, ML has the potential to manage large and heterogeneous sources of data, identifying new patterns and predicting outcomes [15]. Given the above, both supervised and unsupervised ML have been effectively applied to explore asthma, with the aim of identifying phenotypes and comparing clinical characteristics and comorbidities [16,17]. However, to the best of our knowledge, ML has never been applied to exploring clinical patterns related to the severity of asthma exacerbations.

Thus, using a ML approach, the aim of our study was to identify the main predictors of severe asthma exacerbations requiring hospital admission.

2. Methods

We designed a prospective cohort study of asthma patients with recent exacerbation in the Campania Region, Italy.

2.1. Study population

From 2018 to 2019, consecutive asthmatic patients referring to the Respiratory Unit of the A.O.R.N. dei Colli, Naples, Italy, within the first 48 hours after discharge from an acute care setting, they went to for a recent asthma exacerbation, were screened for study entry according to inclusion and exclusion criteria. Inclusion criteria were: age $\geq$18 years; history of asthma; recent visit to the ED for an acute asthma attack. Patients were excluded in the presence of one of the following criteria: history of chronic obstructive pulmonary disease (COPD) or interstitial lung disease; previous lung surgery; history of drug or alcohol abuse; inability to understand the informed consent or poor compliance with the study procedures. Patients with missing data for the outcome of interest were excluded from the study.

The study was conducted following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines [18]. All patients signed written informed consent for voluntary participation to use their de-identified data, and the study was approved by the Institutional Review Board of SUN-A.O. dei Colli, Naples, Italy, with reference number 2018-002266-45.

2.2. Study procedures

After signing the informed consent, the main demographic and clinical data were collected for all patient, including age of asthma onset, disease duration, frequency of exacerbations, therapies, smoking habit, allergy, family history, and comorbidities (*e.g.*, obesity, gastro-oesophageal reflux disease, bronchiectasis, nasal polyposis, chronic sinusitis). Moreover, each patient underwent functional assessment and spirometry with a Master Screen Body Jaeger-Carefusion spirometer (22745 Savi Ranch Parkway, Yorba Linda, CA, USA) following the American Thoracic Society (ATS) and the European Respiratory Society (ERS) Task Force Standardisation of Lung Function Testing [19]. All patients were instructed to avoid the use of inhaled short- and long-acting bronchodilators for $\geq$12 hours before testing. Forced expiratory manoeuvres were judged to be acceptable if they met or exceeded the ATS criteria [19]. The best forced expiratory volume in 1 second (FEV_1) and forced vital capacity (FVC) were selected for data analysis. The FEV_1/FVC ratio was calculated. Data on eosinophil counts (cells/μL) at the time of ED access, prior to initiating any asthma exacerbation therapy, were extracted from regional electronic health records.

2.3. Unsupervised ML analysis

Preliminary assessments using MATLAB software (v. 2021b; MathWorks, Natick, MA, USA) and unsupervised ML using the KNIME Analytics Platform (v. 4.5.0; KNIME, Zurich, Switzerland) were performed. Previous biomedical studies in the literature have demonstrated the effectiveness of KNIME, which allows workflows to be developed to implement ML analyses with good results in cardiology [20] and other clinical settings [21,22].

A *k*-means clustering algorithm was implemented to evaluate whether a potential distinction could be identified between different phenotypes depending on the clinical and respiratory parameters of patients with asthma. *k*-means is an unsupervised iterative algorithm that identifies k clusters assigning n similar data to them through minimization of the in-cluster sum of squares, and each cluster is represented by a centroid [23]. The optimal number of clusters was determined using the silhouette criterion, which assigns a score ranging from −1 to 1 to the appropriateness of clustering depending on how well separated and clearly distinguished the groups are [24]. Then, multidimensional data were mapped and visualized in a two-dimensional set using a parallel coordinates plot. In particular, the names of clinical variables were reported on the x-axis and their values on the y-axis, with each line of the graph plotting the value of each variable of a clustered subject. This may allow to compare the features of several individuals on a set of numeric variables and to understand whether or not there is a discrimination among them.

2.4. Supervised machine learning analysis

Binary supervised classification learning was performed in KNIME using a leave-one-out cross validation. *k*-nearest neighbour (KNN, an instance-based algorithm) and random forest (RF, a tree-based algorithm) were implemented to classify the patients based on the presence of at least one additional visit to the ED in the previous 12 months. Tree-based learning algorithms use decision tree classifiers and are particularly suitable for target problems with discrete values, in addition to being robust to errors and imbalance. Instance-based learning algorithms perform classification depending on the similarity of instances and associating similar neighbours in terms of attributes, thus allowing the identification of phenotypes with analogue characteristics.

KNN is one of the simplest but most effective classification methods that defines groups of k similar samples according to a query point in the features space, measuring the similarity by the distance in the neighbourhood [25]. RF is an ensemble learning algorithm that combines the predictions of a high number of decision trees according to the bagging technique, performing a randomization.

Finally, the importance of the features was computed to identify the most relevant features in the classification through information gain (IG). IG is an entropy-based feature evaluation method, which considers how much information a feature can provide and how much this feature can be used in the classification process. The IG of all the features was normalized and expressed as a percentage to calculate the contribution of each feature to the prediction.

Performances were evaluated in terms of accuracy, sensitivity, specificity and area under the receiver operating characteristics curve (AUROC).

2.5. Statistical analysis

The SPSS software (v. 27.0, IBM, Armonk, NY, USA) was used for the statistical analyses. A t test for independent samples was performed to

compare demographic and clinical features with normal distribution. Otherwise, a Mann-Whitney *U* test was used. Normality of the data was assessed with the Kolmogorov-Smirnov test and then Levene's test was performed for normally distributed data to assess the homoscedasticity of variances between groups.

3. Results

A total of 260 patients with an asthma exacerbation (31.5% male; mean age, 47.6 years) were enrolled. The main characteristics of the study population are reported in Table 1. All patients had been treated in the ED with a single intramuscular or intravenous injection of betamethasone 4 mg or methylprednisolone 40 mg, respectively, while, at the time of study enrolment, an ongoing therapy with inhaled corticosteroids (ICS) was documented in 161 patients, using a daily dose of budesonide or equivalent of 200-800 μg/day in combination with LABA (formoterol, 6–24 μg/day, or salmeterol, 12–24 μg/day) or LABA *plus* a leukotriene receptor antagonist (montelukast, 10 mg/day) (Supplementary Table 1).

Overall, the dataset was composed of 260 instances and 9 numerical features for unsupervised ML (Supplementary Table 1). The optimal number of clusters was obtained for $k = 2$, as demonstrated by the silhouette coefficients related to the number of clusters from 2 to 4; the silhouette score for $k = 2$ was 0.78, and 0.73 and 0.67 for $k = 3$ and $k = 4$, respectively. The parallel coordinates plot highlighted that the eosinophil counts well discriminated and separated two clusters (Fig. 1). Thus, two groups were identified, depending on whether their eosinophil count was greater or less than 370 cells/μL. As shown in Table 2, 86 patients with an eosinophil count ≥370 cells/μL (mean, 530.81 ± 127.29 cells/μL) were significantly older, had a longer disease duration, more restrictions to daily activities, and lower rate of treatment and ICS use compared with 174 patients with an eosinophil count <370 cells/μL (mean, 196.56 ± 84.33 cells/μL). In addition, patients with a higher eosinophil count reported lower values of FEV_1 (64.8 ± 12.3 % predicted *vs.* 83.9 ± 17.3 % predicted, $p < 0.001$) and FEV_1/FVC (71.3 ± 9.3 *vs.* 78.5 ± 6.8, $p < 0.001$), with a significantly higher number of exacerbations per year (5.3 ± 2.2 *vs.* 3.18 ± 2.1, $p < 0.001$) (Fig. 2).

Table 1
Characteristics of patients with asthma exacerbation attending the emergency department.

	Patients (n=260)
Age, years	47.6 ±13.8
Age at onset, years	31.7 ± 17.3
Disease duartion, years	16.3 ± 11.7
FEV_1, % predicted	77.6 ±18.2
FVC, % predicted	86.5 ± 16.7
FEV_1/FVC	76.2 ± 8.4
Exacerbations, n/year	3.9 ± 2.3
Males, n (%)	82 (31.5)
Smokers, n (%)	73 (28.1)
Former smokers, n (%)	50 (19.2)
Family history, n (%)	78 (30)
Restriction to daily activities, n (%)	164 (63.1)
GERD, n (%)	48 (18.5)
Bronchiectasis, n (%)	9 (3.5)
Obesity, n (%)	12 (4.6)
CS without NP, n (%)	77 (29.6)
CS with NP, n (%)	40 (15.4)
ASA-induced asthma, n (%)	7 (2.7)
Allergy, n (%)	122 (46.9)
Eosinophil count, cells/μL	307.1 ± 186.0
ICS, n (%)	161 (61.9)
No ICS, n (%)	48 (18.5)
No treatment, n (%)	51 (19.6)

n: number; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; GERD, gastro-esophageal reflux disease; ICS, inhaled corticosteroids; CS, chronic sinusitis; NP, nasal polyposis; ASA, acetylsalicylic acid. Data are expressed as mean ± standard deviation unless otherwise indicated.

Supervised ML was then used to classify patients based on the presence of at least one additional visit to the ED in the previous 12 months. In this case, the dataset was composed of 260 instances and 19 features, of which 9 were numerical variables and 10 were nominal attributes transformed in binary variables (Supplementary Table 2). As shown in Table 3 and Supplementary Figure 1, KNN achieved the best performance with an accuracy of 98.5% (95%CI: 96.1–99.4), a sensitivity of 98.7% (95% CI: 96.2–99.6), a specificity of 96.7% (95% CI: 83.3–99.4), and an AUROC of 96.7% (95% CI: 83.3–99.4). On the other hand, RF exhibited low specificity (40.0%) and high accuracy (92.3%), due to an imbalance in the data, thus confirming overfitting (Supplementary Figure 2).

The 10 most important features for predicting an additional visit to the ED with the corresponding importance rankings are reported in Table 4. When stratifying these features based on the presence/absence of the target characteristic, we documented significant differences between the two groups. In particular, frequent exacerbators were older (51.43 ± 8.3 years *vs.* 47.04 ± 14.3 years, $p = 0.017$), with lower FEV_1/FVC values (69.3 ± 13.5 *vs.* 77.0 ± 7.1, $p = 0.004$), higher frequency of nasal polyposis (46.7% *vs.* 15.7%, $p < 0.001$) and lower frequency of allergy (26.7% *vs.* 49.6%, $p = 0.018$). A trend towards lower FEV_1 values ($p = 0.056$) was also documented (Table 5).

4. Discussion

In this study, a ML approach was used to identify the main characteristics related to severe asthma exacerbations requiring ED admittance. In our study population, unsupervised ML revealed the prominent role of eosinophil counts in discriminating two phenotypic clusters among asthma patients who had access to the ED, with a more severe pattern for those who exhibited eosinophil counts greater than 370 cells/μL. Moreover, based on the presence of at least one additional access to the ED in the previous 12 months, supervised learning identified a model with a good performance, in which eosinophil count was confirmed as one of the most important features of frequent exacerbators, along with the number of prior exacerbations, pulmonary function parameters and other clinical and demographic variables.

The identification of the biological pathways that are predominant in asthma patients with frequent exacerbations is still a matter of debate [26]. Several cross-sectional and longitudinal studies identified circulating and sputum eosinophil counts as the main risk factors for exacerbations, also suggesting a role for FeNO as an additional biomarker of type 2 inflammation [5,27,28]. Overall, the prognostic role of eosinophil count in asthma has been widely documented. In one of the largest observational studies on over 130,000 asthmatics, patients with blood eosinophiles greater than 400 cells/μL experienced more severe exacerbations and had poorer asthma control [29]. Other observational studies using the same cut-off value reported similar findings [30–32], thus confirming the key role of monitoring eosinophil count to maintain a good asthma control. On the other hand, when analysing the Severe Asthma Research Program-3 (SARP-3) population, Peters et *al* [33] failed in demonstrating a strong association between eosinophil counts and the frequency of exacerbations, while suggesting that prior exacerbations may represent the strongest predictor of subsequent additional events. Moreover, they identified a number of clinical and demographic variables (*e.g.*, age, female gender, body mass index, diabetes, spirometry parameters) that somehow impacted the risk of exacerbations. The fact that neither blood or sputum eosinophil count nor FeNO differed when stratifying patients according to exacerbation rate may be related to differences in asthma severity or treatment regimens in their study population but, nevertheless, it raises the question of the need for specifically designed population-based studies and alternative methods for identifying hidden patterns of exacerbation-prone asthma [26].

In recent times, the growing popularity and applicability of ML in medical research has been observed, as ML is able to handle large and heterogeneous data sources, identifying new patterns and predicting

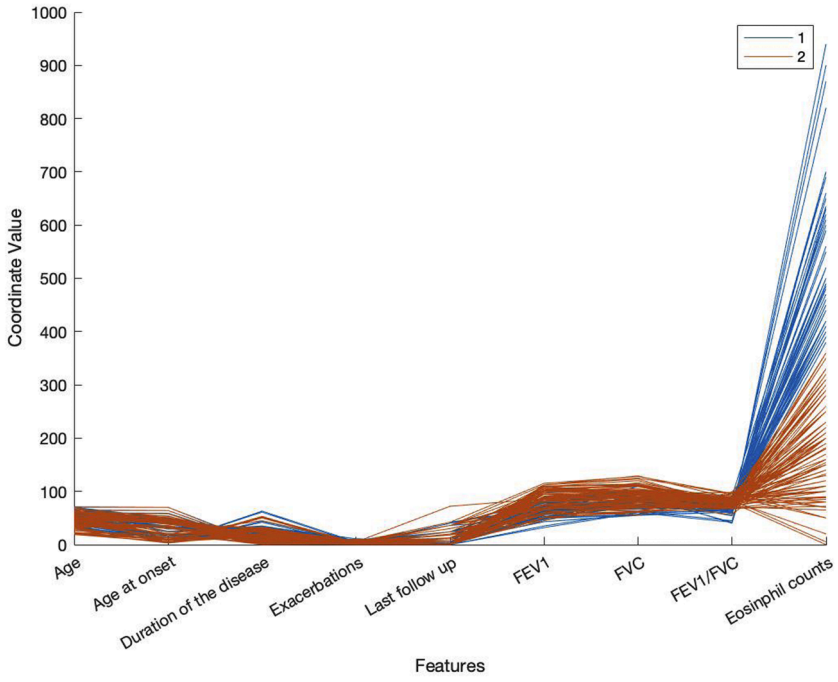

Fig. 1. Parallel coordinates: each line represents a clustered subject, with clinical features reported on the x-axis and their values on the y-axis. The clusters are presented with distinct colours. If a variable can discriminate clustered subjects, all the lines are distinguishable. Otherwise, all the lines are overlapping. FEV_1, forced expiratory volume in 1 second; FVC: forced vital capacity.

Table 2
Main demographic and clinical characteristics of 260 patients with asthma exacerbation based on blood eosinophil count.

	Eosinophil ≥ 370 cells/μL (n=86)	Eosinophil < 370 cells/μL (n=174)	P-value
Age, years	52.9 ± 10.7	44.9 ± 14.4	**<0.001**
Males, n (%)	34 (39.5)	48 (27.6)	0.051
Age at onset, years	34.5 ± 15.4	30.3 ± 18.1	0.054
Disease duration, days	19.0 ± 11.8	15.0 ± 11.5	**0.010**
FEV_1, % pred	64.8 ± 12.3	83.9 ± 17.3	**<0.001**
FVC, % pred	77.8 ± 12.5	90.9 ± 16.8	**<0.001**
FEV_1/FVC	71.3 ± 9.3	78.5 ± 6.8	**<0.001**
Exacerbations, n/year	5.3 ± 2.2	3.18 ± 2.1	**<0.001**
Smoker, n (%)	26 (30.2)	48 (27.6)	0.802
Family history, n (%)	8 (9.3)	70 (40.2)	**<0.001**
Restriction to daily activities, n (%)	75 (87.0)	89 (51.1)	**<0.001**
GERD, n (%)	19 (22.1)	29 (16.7)	0.289
Bronchiectasis, n (%)	3 (3.5)	6 (3.4)	0.987
Obesity, n (%)	7 (8.1)	5 (2.9)	0.057
CS without NP, n (%)	13 (15.1)	64 (36.8)	**<0.001**
CS with NP, n (%)	37 (43.0)	3 (1.7)	**<0.001**
ASA-induced asthma, n (%)	4 (4.7)	3 (1.7)	0.170
Allergy, n (%)	25 (29.1)	97 (55.7)	**<0.001**
ICS, n (%)	74 (86.0)	87 (50.0)	**<0.001**
Non ICS, n (%)	9 (10.4)	39 (22.4)	
No treatment, n (%)	3 (3.5)	48 (27.6)	

n: number; FEV_1, forced expiratory volume in 1 second; FVC, forced vital capacity; GERD, gastro-esophageal reflux disease; ICS, inhaled corticosteroids; CS, chronic sinusitis; NP, nasal polyposis; ASA, acetylsalicylic acid. Data are expressed as mean ± standard deviation unless otherwise indicated.

outcomes [15]. Therefore, the use of ML in addition to conventional statistics has been highly recommended in asthma [26], given the urgent need to recognize the biological pathways that are more prevalent among frequent exacerbators. Previous evidence on the use of supervised and unsupervised ML in asthma patients have been reported, substantially confirming the key prognostic role of eosinophiles [16,17].

It is interesting to highlight that our unsupervised model was able to identify a cut-off value of 370 cells/μL, which was not predetermined but automatically identified by the computational process, being very similar to the value associated with poorer asthma control in other large population-based studies [29]. Moreover, in line with previous evidence [34], we documented that the participants with higher eosinophiles experienced more exacerbations, with worse spirometry parameters and pulmonary function, thus confirming the prognostic value of this variable. Interestingly, while the presence of atopic conditions and nasal polyps were expected to be linked to eosinophilia [35], only nasal polyposis was found to be more frequent among patients with higher eosinophil counts. Allergy, on the other hand, was less frequent. This apparently contrasting result may have different interpretations. First, patients with an eosinophil count greater than 370 cells/μL had significantly higher rates of steroid treatment in our study population, potentially reflecting the higher need for therapy due to the more severe disease pattern, and it is important to highlight that blood eosinophiles may not reflect airway eosinophilia in patients with severe asthma treated with high doses of steroids [36]. Moreover, although blood eosinophils have been identified as a surrogate marker of airway eosinophilic inflammation [37], they may not unconditionally reflect an allergic aetiology, as eosinophilic inflammation is not a prerogative of atopy [38]. Accordingly, more than half of the participants in our study were nonallergic, suggesting that severe eosinophilia may be a characteristic related to severity rather than aetiology [39] and thus participating in the debate on what molecular mechanisms may be involved in airway eosinophilic inflammation in case of nonallergic asthma [38]. While allergen-specific T helper 2 (Th2) cells drive the pathogenesis of allergic asthma, it has been demonstrated that high levels of type 2 innate lymphoid cells (ILC2s) can be found in both nasal polyps and in nonallergic asthma [40]. Therefore, two different pathways activated by allergen-specific Th2 lymphocytes or non-specific ILC2 may both result in an increased interleukin-5 (IL-5) synthesis, which in turn is responsible for eosinophilic inflammatory response of the airways [38]. Overall, it may be argued that our unsupervised ML model confirmed the key prognostic role of blood eosinophil count in asthma exacerbations,

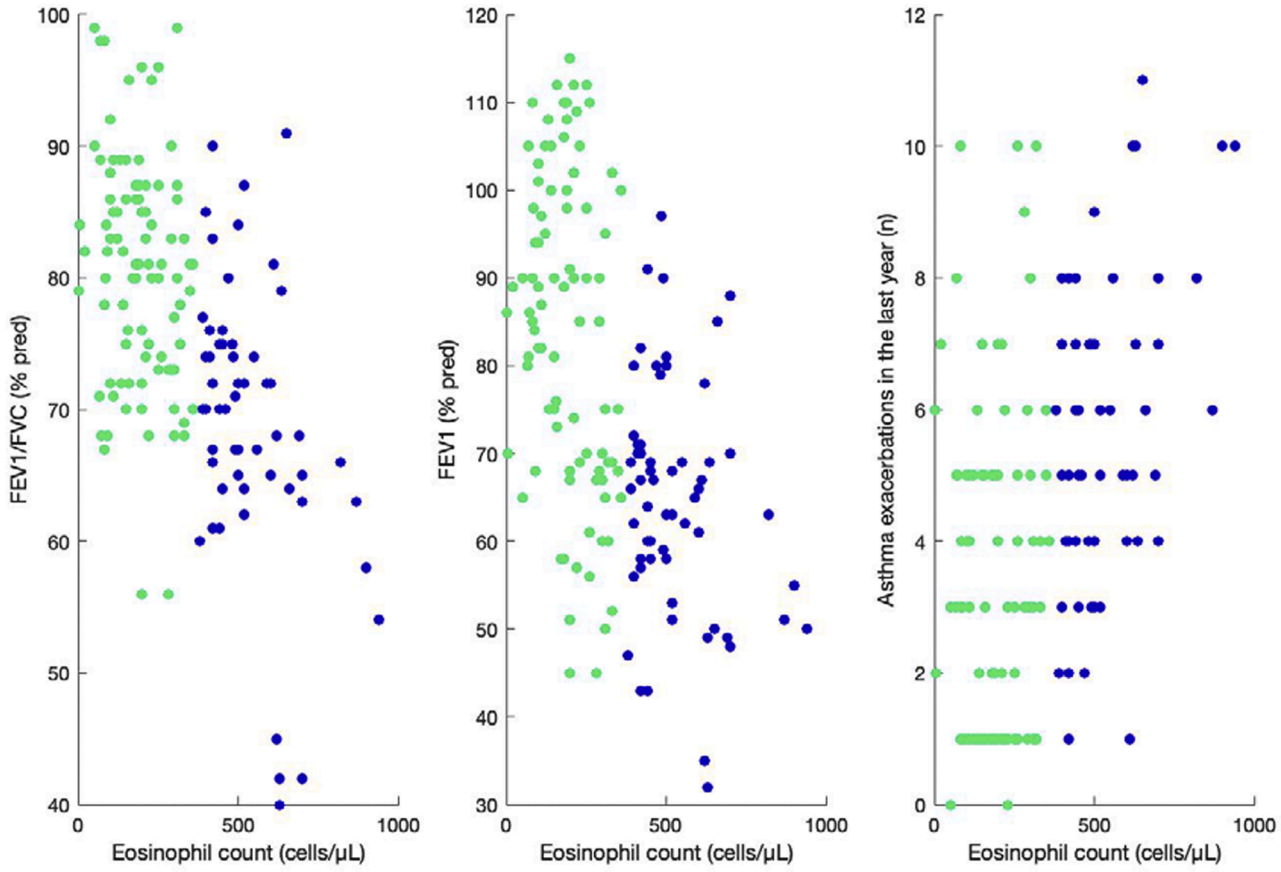

Fig. 2. The figure shows a scatter plot of the obtained clusters, corresponding to the identified phenotypes, and highlights the relationship of FEV_1/FVC, FEV_1, and the number of asthma exacerbations in the last year with eosinophil count (cells/μL). Clusters are represented with distinct colours. In particular, blue dots represent participants (n=86) with eosinophil count ≥ 370 cells/μL, green dots refer to those (n=174) with eosinophil count < 370 cells/μL.

Table 3
Evaluation metrics for each algorithm.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUROC (%)
KNN	98.5 (95%CI: 96.1-99.4)	98.7 (95%CI: 96.2-99.6)	96.7 (95%CI: 83.3-99.4)	96.3
RF	92.3 (95%CI: 88.4-95.0)	99.1 (95%CI: 96.9-99.8)	40.0 (95%CI: 24.6-57.7)	90.1

AUROC: area under the receiver operating characteristic curve; RF: random forest; KNN: *k*-nearest neighbors.

Table 4
Supervised machine learning analysis: features Information Gain (IG) normalized and transformed into percentage for the 10 most important parameters predicting the presence of at least one additional access to the emergency department (ED) in the previous 12 months.

Feature	IG
FEV_1/FVC	11.9%
Exacerbations	10.4%
FEV_1	9.2%
Nasal polyposis	8.1%
Age	7.8%
Eosinophil count	6.9%
Disease duration	6.6%
Follow-up	6.4%
Age at onset	6.1%
Allergy	6.0%

FEV_1, forced expiratory volume in 1 second; FVC: forced vital capacity.

being able to identify a more severe pattern among those who had access to the ED, regardless of the allergic or nonallergic aetiology. On the other hand, the fact that allergy was less frequent among patients with high blood eosinophiles may support – at least in part – the evidence of a more severe phenotype and lower response rate to standard therapy for patients with nonallergic asthma [41].

Interestingly, when a supervised approach was used, eosinophil count was confirmed as one of the features related to frequent exacerbations, along with age, spirometry parameters, prior exacerbations and other clinical variables. However, when stratifying participants according to the presence of at least one additional access to the ED in the previous 12 months, no significant difference was found for eosinophil count, possibly due to the unbalanced distribution of patients between the two groups. An atopic status was also one of the most important features, with allergy being less represented in frequent exacerbators. In line with the results of the unsupervised approach, this may be consistent with the hypothesis of a worse outcome for patients with nonallergic asthma [41], further suggesting that eosinophilia is related to severity of asthma rather than its aetiology [40].

Overall, these results should be considered in light of the current scenario, where severe asthma has become a major issue in terms of public health and social costs [42]. About 5% of asthma patients experience a severe form, which is resistant to standard therapy and is characterized by frequent exacerbations, often requiring ED admittance [43]. Therefore, severe asthma can be considered a disabling condition, having a significant impact on quality of life and even ability to work [44]. Accordingly, similar to COPD and other respiratory disorders [45,

Table 5
Distribution of the most important features used for supervised machine learning among patients with and without at least one additional access to the emergency department (ED) in the previous 12 months.

	Further access to ED (n=30)	No further access to ED (n=230)	P-value
FEV_1/FVC	69.3 ± 13.5	77.0 ± 7.1	**0.004**
Exacerbations, n/year	5.5 ± 2.1	3.7 ± 2.3	**<0.001**
FEV_1, % predicted	71.6 ± 19.3	78.3 ± 17.9	0.056
Nasal polyposis, n (%)	14 (46.7)	36 (15.7)	**<0.001**
Age, years	51.43 ± 8.3	47.04 ± 14.3	**0.017**
Eosinophil count, cells/μL	352.7 ± 223.0	301.2 ± 181.3	0.233
Disease duration, years	18.8 ± 8.8	16.0 ± 12.0	0.221
Follow-up duration, months	15.2 ± 16.3	9.2 ± 12.2	**0.016**
Age at onset, years	33.1 ± 14.2	31.5 ± 17.7	0.631
Allergy, n (%)	114 (49.6)	8 (26.7)	**0.018**

n: number; FEV_1, forced expiratory volume in 1 second. Data are expressed as mean ± standard deviation unless otherwise indicated.

46], it has been reported that severe asthma patients may benefit from multidisciplinary pulmonary rehabilitation [47–49], which has been shown to be effective in improving symptoms and functional exercise capacity [50]. Overall, given this health and social burden, the identification of hidden patterns and prognostic models is mandatory in order to predict long-term outcomes, plan adequate pharmacological and rehabilitation strategies, and monitor response to therapies in such a disabling condition.

Some relevant limitations of our protocol should be considered. First, patients included in this study were all local residents from the Campania Region in Italy. Therefore, we cannot exclude that this predictive model may be generalized to other populations/ethnic groups. In addition, the relatively low number of included patients and their unbalanced distribution between groups when considering a specific variable is another major limitation of this study. Therefore, considering that ML performs better with large datasets, further studies on a larger sample are needed, aimed at confirming the role of the features chosen for modelling in determining disease severity and frequency of exacerbations in asthma.

5. Conclusions

In conclusion, our results confirm the prognostic role of blood eosinophil count in asthma exacerbations, being able to identify a more severe pattern among those requiring hospital admission, regardless of the allergic or nonallergic aetiology. Moreover, eosinophil count was confirmed as one of the most important features of frequent exacerbators, along with the number of prior exacerbations, pulmonary function parameters and other clinical and demographic variables. Finally, our study reveals the potential of ML in identifying specific phenotypic patterns for asthma, thus suggesting the usefulness of artificial intelligence in defining models with good performance that can help predict long-term outcomes and plan personalized pharmacological and rehabilitation strategies while monitoring response to therapies.

Funding

This work was supported by the "Ricerca Corrente" funding scheme of the Ministry of Health, Italy. The editorial support was funded by AstraZeneca.

The funder was not involved in the collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Ethical statement

The protocol was approved by the Institutional Review Board of SUN-AO Dei Colli, Naples, Italy with reference number 2018-002266-45. All patients provided written informed consent to use their de-identified data.

Data availability statement

The data are available upon request to the corresponding author (M. D.A.).

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgments

Sarah Adamo wishes to thank the "Gruppo per l'Armonizzazione delle Reti della Ricerca (GARR)" for her research grant. Editorial and writing assistance was provided by Edra S.p.A., Milan, Italy.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejim.2022.07.019.

References

[1] Selroos O, Kupczyk M, Kuna P, Lacwik P, Bousquet J, Brennan D, et al. National and regional asthma programmes in Europe. Eur Respir Rev. 2015;24:474–83.
[2] Accordini S, Corsico AG, Braggion M, Gerbase MW, Gislason D, Gulsvik A, et al. The cost of persistent asthma in Europe: an international population-based study in adults. Int Arch Allergy Immunol. 2013;160:93–101.
[3] Emerman CL, Woodruff PG, Cydulka RK, Gibbs MA, Pollack Jr CV, Camargo Jr CA. Prospective multicenter study of relapse following treatment for acute asthma among adults presenting to the emergency department. MARC investigators. Multicenter Asthma Research Collaboration. Chest 1999;115:919–27.
[4] Weber EJ, Silverman RA, Callaham ML, Pollack CV, Woodruff PG, Clark S, et al. A prospective multicenter study of factors associated with hospital admission among adults with acute asthma. Am J Med. 2002;113:371–8.
[5] ten Brinke A, Sterk PJ, Masclee AA, Spinhoven P, Schmidt JT, Zwinderman AH, et al. Risk factors of frequent exacerbations in difficult-to-treat asthma. Eur Respir J. 2005;26:812–8.
[6] Vedel-Krogh S, Fallgaard Nielsen S, Lange P, Vestbo J, Nordestgaard BG. Association of blood eosinophil and blood neutrophil counts with asthma exacerbations in the copenhagen general population study. Clin Chem. 2017;63:823–32.
[7] Busse WW, Wenzel SE, Casale TB, FitzGerald JM, Rice MS, Daizadeh N, et al. Baseline FeNO as a prognostic biomarker for subsequent severe asthma exacerbations in patients with uncontrolled, moderate-to-severe asthma receiving placebo in the LIBERTY ASTHMA QUEST study: a post-hoc analysis. Lancet Respir Med 2021;9:1165–73.
[8] Semprini R, Williams M, Semprini A, McDouall A, Fingleton J, Holweg C, et al. Type 2 biomarkers and prediction of future exacerbations and lung function decline in adult asthma. J Allergy Clin Immunol Pract 2018;6:1982–8. e1.
[9] Bloom CI, Palmer T, Feary J, Quint JK, Cullinan P. Exacerbation patterns in adults with asthma in England. A population-based study. Am J Respir Crit Care Med 2019;199:446–53.
[10] Global Initiative on Asthma. www.ginasthma.org, last accessed June 24, 2022.
[11] Ricciardi C, Jonsson Jr H, Jacob D, Improta G, Recenti M, Gislason MK, et al. Improving prosthetic selection and predicting BMD from biometric measurements in patients receiving total hip arthroplasty. Diagnostics (Basel) 2020:10.
[12] Ricciardi C, Edmunds KJ, Recenti M, Sigurdsson S, Gudnason V, Carraro U, et al. Assessing cardiovascular risks from a mid-thigh CT image: a tree-based machine learning approach using radiodensitometric distributions. Sci Rep 2020;10:2863.
[13] Mannarino T, Assante R, Ricciardi C, Zampella E, Nappi C, Gaudieri V, et al. Head-to-head comparison of diagnostic accuracy of stress-only myocardial perfusion imaging with conventional and cadmium-zinc telluride single-photon emission computed tomography in women with suspected coronary artery disease. J Nucl Cardiol 2021;28:888–97.
[14] Adamo S, Ambrosino P, Ricciardi C, Accardo M, Mosella M, Cesarelli M, et al. A machine learning approach to predict the rehabilitation outcome in convalescent COVID-19 patients. J Pers Med 2022:12.
[15] Weissler EH, Naumann T, Andersson T, Ranganath R, Elemento O, Luo Y, et al. The role of machine learning in clinical research: transforming the future of evidence generation. Trials. 2021;22:537.

[16] Zhan J, Chen W, Cheng L, Wang Q, Han F, Cui Y. Diagnosis of asthma based on routine blood biomarkers using machine learning. Comput Intell Neurosci. 2020; 2020:8841002.
[17] Su MW, Lin WC, Tsai CH, Chiang BL, Yang YH, Lin YT, et al. Childhood asthma clusters reveal neutrophil-predominant phenotype with distinct gene expression. Allergy 2018;73:2024–32.
[18] von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007; 370:1453–7.
[19] Graham BL, Steenbruggen I, Miller MR, Barjaktarevic IZ, Cooper BG, Hall GL, et al. Standardization of spirometry 2019 update. An Official American Thoracic Society and European Respiratory Society Technical Statement. Am J Respir Crit Care Med. 2019;200:e70–88.
[20] Donisi L, Ricciardi C, Cesarelli G, Coccia A, Amitrano F, Adamo S, et al. Bidimensional and tridimensional poincaré maps in cardiology: a multiclass machine learning study. Electronics 2022;11:448.
[21] Ricciardi C, Improta G, Amato F, Cesarelli G, Romano M. Classifying the type of delivery from cardiotocographic signals: a machine learning approach. Comput Methods Programs Biomed. 2020;196:105712.
[22] Scrutinio D, Ricciardi C, Donisi L, Losavio E, Battista P, Guida P, et al. Machine learning to predict mortality after rehabilitation among patients with severe stroke. Sci Rep. 2020;10:20127.
[23] Likas A VN, Verbeek V J. The global k-means clustering algorithm. Pattern Recognit 2003;36:451–61.
[24] PJ R. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–65.
[25] Hu Q YD, Xie Z. Neighborhood classifiers. Expert Syst Appl 2008;34:866–76.
[26] Sterk PJ, Sinha A. Emerging complexity in the biomarkers of exacerbation-prone asthma. Am J Respir Crit Care Med 2020;202:915–7.
[27] Pavord ID, Holliday M, Reddel HK, Braithwaite I, Ebmeier S, Hancox RJ, et al. Predictive value of blood eosinophils and exhaled nitric oxide in adults with mild asthma: a prespecified subgroup analysis of an open-label, parallel-group, randomised controlled trial. Lancet Respir Med 2020;8:671–80.
[28] Denlinger LC, Phillips BR, Ramratnam S, Ross K, Bhakta NR, Cardet JC, et al. Inflammatory and Comorbid Features of Patients with Severe Asthma and Frequent Exacerbations. Am J Respir Crit Care Med 2017;195:302–13.
[29] Price DB, Rigazio A, Campbell JD, Bleecker ER, Corrigan CJ, Thomas M, et al. Blood eosinophil count and prospective annual asthma disease burden: a UK cohort study. Lancet Respir Med 2015;3:849–58.
[30] Schleich FN, Chevremont A, Paulus V, Henket M, Manise M, Seidel L, et al. Importance of concomitant local and systemic eosinophilia in uncontrolled asthma. Eur Respir J 2014;44:97–108.
[31] Zeiger RS, Schatz M, Li Q, Chen W, Khatry DB, Gossage D, et al. High blood eosinophil count is a risk factor for future asthma exacerbations in adult persistent asthma. J Allergy Clin Immunol Pract 2014;2:741–50.
[32] Tran TN, Khatry DB, Ke X, Ward CK, Gossage D. High blood eosinophil count is associated with more frequent asthma attacks in asthma patients. Ann Allergy Asthma Immunol 2014;113:19–24.
[33] Peters MC, Mauger D, Ross KR, Phillips B, Gaston B, Cardet JC, et al. Evidence for exacerbation-prone asthma and predictive biomarkers of exacerbation frequency. Am J Respir Crit Care Med 2020;202:973–82.
[34] Casciano J, Krishnan JA, Small MB, Buck PO, Gopalan G, Li C, et al. Value of peripheral blood eosinophil markers to predict severity of asthma. BMC Pulm Med 2016;16:109.
[35] Wardlaw AJ, Brightling C, Green R, Woltmann G, Pavord I. Eosinophils in asthma and other allergic diseases. Br Med Bull 2000;56:985–1003.
[36] Ullmann N, Bossley CJ, Fleming L, Silvestri M, Bush A, Saglani S. Blood eosinophil counts rarely reflect airway eosinophilia in children with severe asthma. Allergy 2013;68:402–6.
[37] Nadif R, Siroux V, Oryszczyn MP, Ravault C, Pison C, Pin I, et al. Heterogeneity of asthma according to blood inflammatory patterns. Thorax 2009;64:374–80.
[38] Brusselle GG, Maes T, Bracke KR. Eosinophils in the spotlight: Eosinophilic airway inflammation in nonallergic asthma. Nat Med 2013;19:977–9.
[39] Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON, et al. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. Lancet. 2012;380:651–9.
[40] Mjosberg JM, Trifari S, Crellin NK, Peters CP, van Drunen CM, Piet B, et al. Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are defined by expression of CRTH2 and CD161. Nat Immunol 2011;12:1055–62.
[41] Peters SP. Asthma phenotypes: nonallergic (intrinsic) asthma. J Allergy Clin Immunol Pract 2014;2:650–2.
[42] Nunes C, Pereira AM, Morais-Almeida M. Asthma costs and social impact. Asthma Res Pract 2017;3:1.
[43] Chung KF, Wenzel SE, Brozek JL, Bush A, Castro M, Sterk PJ, et al. International ERS/ATS guidelines on definition, evaluation and treatment of severe asthma. Eur Respir J 2014;43:343–73.
[44] Eisner MD, Yelin EH, Katz PP, Lactao G, Iribarren C, Blanc PD. Risk factors for work disability in severe adult asthma. Am J Med. 2006;119:884–91.
[45] Mosher CL, Nanna MG, Jawitz OK, Raman V, Farrow NE, Aleem S, et al. Cost-effectiveness of pulmonary rehabilitation among US adults with chronic obstructive pulmonary disease. JAMA Netw Open. 2022;5:e2218189.
[46] Ambrosino P, Papa A, Maniscalco M, Di Minno MND. COVID-19 and functional disability: current insights and rehabilitation strategies. Postgrad Med J 2021;97: 469–70.
[47] Zampogna E, Zappa M, Spanevello A, Visca D. Pulmonary rehabilitation and asthma. Front Pharmacol 2020;11:542.
[48] Turk Y, van Huisstede A, Franssen FME, Hiemstra PS, Rudolphus A, Taube C, et al. Effect of an outpatient pulmonary rehabilitation program on exercise tolerance and asthma control in obese asthma patients. J Cardiopulm Rehabil Prev 2017;37: 214–22.
[49] Salandi J, Icks A, Gholami J, Hummel S, Schultz K, Apfelbacher C, et al. Impact of pulmonary rehabilitation on patients' health care needs and asthma control: a quasi-experimental study. BMC Pulm Med 2020;20:267.
[50] ATS statement–snowbird workshop on standardization of spirometry. Am Rev Respir Dis 1979;119:831–8.