POLITECNICO DI TORINO
Repository ISTITUZIONALE

Explainable artificial intelligence for cough-related quality of life impairment prediction in asthmatic patients

Terms of use:

Publisher copyright

(Article begins on next page)

11 March 2025

RESEARCH ARTICLE

Explainable artificial intelligence for cough-related quality of life impairment prediction in asthmatic patients

Sara Narteni[1,2]*, Ilaria Baiardini[3], Fulvio Braido[3], Maurizio Mongelli[1]

1 CNR-IEIIT, Genoa, Italy, 2 DAUIN Department, Politecnico di Torino, Turin, Italy, 3 Respiratory Diseases and Allergy Department, IRCCS Polyclinic Hospital San Martino, Genoa, Italy

* sara.narteni@ieiit.cnr.it

Abstract

Explainable Artificial Intelligence (XAI) is becoming a disruptive trend in healthcare, allowing for transparency and interpretability of autonomous decision-making. In this study, we present an innovative application of a rule-based classification model to identify the main causes of chronic cough-related quality of life (QoL) impairment in a cohort of asthmatic patients. The proposed approach first involves the design of a suitable symptoms questionnaire and the subsequent analyses via XAI. Specifically, feature ranking, derived from statistically validated decision rules, helped in automatically identifying the main factors influencing an impaired QoL: pharynx/larynx and upper airways when asthma is under control, and asthma itself and digestive trait when asthma is not controlled. Moreover, the obtained *if-then* rules identified specific thresholds on the symptoms associated to the impaired QoL. These results, by finding priorities among symptoms, may prove helpful in supporting physicians in the choice of the most adequate diagnostic/therapeutic plan.

Introduction

Nowadays Artificial Intelligence (AI) is revolutionizing medicine by leveraging powerful technologies and advanced learning algorithms. This has the potential to support several clinical processes, from prognostics to diagnostics, from treatment management to drug discovery, and also can aid hospital administrative tasks. However, AI real application in healthcare needs to be approached very carefully, since failures may cause harm to human lives. For this reason, AI research is increasing its interests in *trustworthy AI* [1], a broad paradigm establishing how to properly design, develop and deploy real-world AI applications. Between its principles, *transparency* requires providing the user with an understanding of the autonomous decisions generated by the model: this topic is subject of *eXplainable AI (XAI)* research [2, 3]. XAI comprehends a wide range of methodologies, which can be broadly categorized as post-hoc explanations of black box models and transparent-by-design techniques [4]. In the latter category, rule-based models are characterized by understandable decision rules expressed in the *if-then* format. These kinds of models are particularly suitable in medicine, since their intrinsic interpretability allows clinicians to enter models' logic and increase trust in them. In

light of this, our work focuses on the usage of such techniques to characterize the quality of life of asthmatic patients with chronic cough.

Asthma is a frequent cause of cough in adults [5]. In addition to coughing, asthmatic patients may also wheeze or feel short of breath. However, some people have a condition known as cough variant asthma, in which cough is the only symptom of asthma. For these reasons, tools for the assessment of asthma, such as Asthma Control Test (ACT) [6], consider cough among the asthma features. While in patients with uncontrolled asthma the disease itself can be the cause of cough, the persistence of cough despite good asthma control can be related to concomitant disorders (i.e., postnasal drip, pharynx/larynx disorders, and acid reflux from the stomach [7]) or inability of asthma drugs to fully remove the symptoms.

In light of these considerations, it is very useful to design a method that allows to define the priority of choice among different diagnostic techniques, starting from patients' self-reported presence and entity of symptoms and their impact on the quality of life. Methods based on XAI, thanks to their transparent and interpretable methods, can offer a great opportunity in this direction.

Contribution

In this study, we propose the usage of a rule-based XAI model to support clinicians in the diagnostic procedure for determining the origins of chronic cough in asthmatic patients. More precisely, our main contributions are the following:

- We introduce a new block-based questionnaire, devoted to collect (respiratory) symptoms perceived by asthmatic patients with chronic cough.

- We train a rule-based model, the Logic Learning Machine (LLM), for predicting chronic cough-related quality of life based only on self-reported responses to the questionnaire of symptoms, by distinguishing patients with high or low asthma control level.

- By validating and analyzing the model, we discover which symptoms and corresponding values are mainly involved in a quality of life exacerbation.

The remaining part of the paper is organized as follows. In Section Related Work we report some recent examples of machine learning for chronic cough. Section Methodology describes the workflow, the dataset structure and the adopted methodologies. Section Results shows and discusses the obtained results. Finally, Section Conclusion concludes the paper and reports future research on the topic.

Related work

Different machine learning (ML) and AI-based studies on chronic cough and asthma have been carried out in recent years, by leveraging the newest medical technologies [8]. An AI-based cough count, CoughyTM [9], system was recently developed that quantifies cough sounds collected through a smartphone application. Study results showed that suggest that CoughyTM could be a novel solution for objectively monitoring cough in a clinical setting. A vocal biomarker-based machine learning approaches have shown promising results in the detection of various health conditions, including respiratory diseases, such as asthma [10]. Also, a deep learning model for identifying chronic cough patients with even higher sensitivity and specificity when structured and unstructured electronic health records EHR data are utilized has been proposed [11].

In [12], well established ML models like gradient boost and random forest were adopted in a retrospective study to predict the risk of persistent chronic cough (PCC) in patients with

chronic cough (CC). The work proposed in [13] used a statistical approach (Latent Class Analysis) on the Swedish Twin study On Prediction and Prevention of Asthma (STOPPA [14]) and the Child and Adolescent Twin Study in Sweden (CATSS) questionnaires responses to identify asthma and wheeze phenotypes in children. In [15], four adult chronic cough phenotypes were identified through a cluster analysis method applied to questionnaire data such as the COugh Assessment Test (COAT) [16] and the Korean version of the Leicester Cough Questionnaire [17].

However, all these literature examples do not provide their outcomes in an explainable way.

Methodology

Workflow

The overall methodology followed in the proposed analyses is depicted in Fig 1. The dataset was first split in a 70% training and 30% test sets, then an explainable Artificial Intelligence (XAI) model was considered for data classification. The adopted classifier is called Logic Learning Machine and provides its predictions through a set of rules. In order to verify the statistical significance of the resulting ruleset, this was validated through a statistical test. Rules that did not pass the test were then filtered out from the model, thus obtaining a final, validated, set of rules. Also, feature ranking was investigated to identify which of the inputs have the higher impact on the model outcome. Finally, the overall performance of the validated ruleset was measured on the test set, by considering some common metrics for machine learning models evaluation.

Next Sections provide the description of the dataset and some fundamentals about the adopted XAI, the rule validation test and the definition of the evaluation metrics.

Dataset description

We implemented a retrospective study on a cohort of asthmatic patients [from the NCT04796844 trial approved by local ethical committee (CER Liguria: 456/2020—DB id 10481)], who have been asked to answer to three different kinds of questionnaires (data were accessed on 2023/03/08; the authors had no access to information that could allow to identify individual participants during or after data collection).

The first questionnaire collects patients' feedback about a variety of symptoms. Specifically, it contains 19 items relating to four domains related to the more frequent causes of chronic cough, as shown in the diagram of Fig 2.
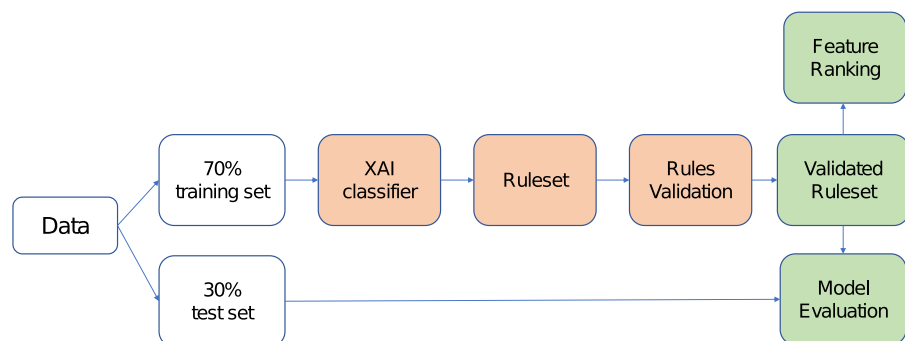

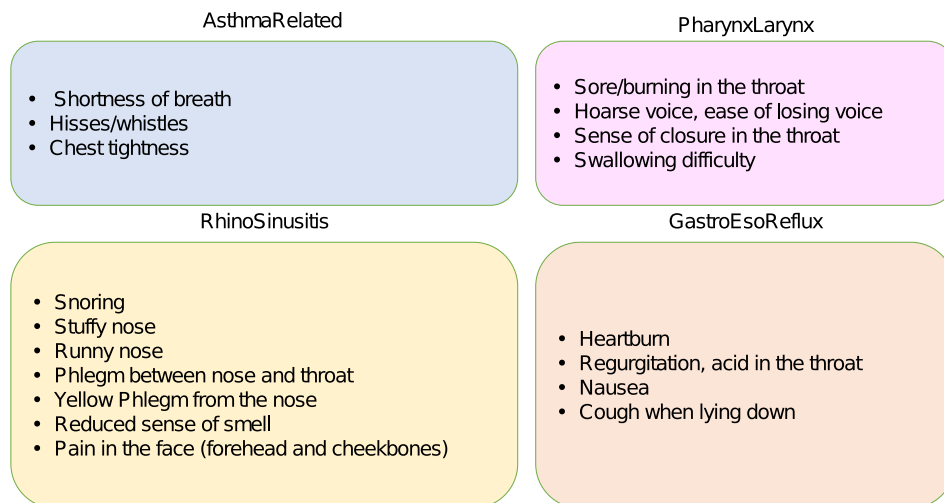**Fig 1. Workflow of the analyses carried out in the proposed XAI-based approach.**

https://doi.org/10.1371/journal.pone.0292980.g001

Fig 2. Symptoms questionnaire. Schematic representation of the four blocks (*AsthmaRelated, PharynxLarynx, RhinoSinusitis, GastroEsoReflux*) of the symptoms questionnaire and their related items.

For each item, the patients answered to the question *"How intense/annoying has the symptom been in the last month?"*, by self-reporting a level between *None* and *Very Much* expressing the perceived entity of the corresponding symptom. These levels were then proportionally converted to a score in the 0–100 scale. The average of the responses within each block was computed, thus individuating a set of four features that will be used as input to the ML model, each referred to a different body organ.

The second questionnaire involved in this study is the Chronic Cough Impact Questionnaire (CCIQ) [18]. It is useful to measure the impact of cough on health-related quality of life, namely *impact on daily life (CCIQ IDL)*, on *sleep/concentration (CCIQ SC)*, on *mood (CCIQ M)* and *relationship (CCIQ R)*. A score for each group is derived and contributes to compute a global score, called *CCIQ GLS*: based on this, we defined two classes of patients. Those scoring *CCIQ GLS*≥20 were labelled as *impaired Quality of Life (QoL)*, while those with *CCIQ GLS*<20 were associated to a *near normal QoL*. This threshold value choice and the subsequent labelling follows previous studies on the same questionnaire [19].

The last questionnaire considered is the Asthma Control Test (ACT) [6]. It is a 5-item questionnaire aimed at assessing at which extent the asthmatic patient has control of the pathology. We used the score obtained from this test to further distinguish patients between two populations: subjects with $ACT \geq 20$ were identified as the *controlled asthma* group, whereas those scoring $ACT < 20$ formed the *not controlled asthma* group.

The analyses carried out in this work thus considered three different cases: i) *all* patients were included; ii) only *controlled* asthma patients were included; iii) only *not controlled* asthma patients were included.

The adopted eXplainable AI classifier

For each patients group, we trained a XAI classifier that, fed with the 4 input features (referred to as *AsthmaRelated, PharynxLarynx, RhinoSinusitis* and *GastroEsoReflux*) representing the average scores on each block of the symptoms questionnaire (Fig 2), provided a prediction of the patient's cough-related QoL, which can be either *impaired* or *near normal*.

The analyses on the first group (i.e., *all* patients) did not explicitly use the knowledge acquired from the ACT questionnaire. Indeed, the classification model that is designed for this group represents a tool to individuate which areas and values of symptoms drive an impaired QoL in a generic asthmatic population, but without any previous knowledge on the asthma control level. Conversely, the analyses performed on the *controlled asthma* and *not controlled asthma* groups also exploited the information from the ACT, thus the results of the XAI predictive models provide indications that are specifically tailored to the different asthma control level.

Logic learning machine. In this Section, we provide some basic description of the adopted classifier, the Logic Learning Machine (LLM). It is a rule-based explainable AI model, designed and developed by Rulex [20] as the efficient implementation of Switching Neural Networks [21].

Given the input data, the LLM provides a classification model described by a set of rules $\mathcal{R} = \{r_k\}_{k=1,\dots,N_r}$, where each r_k is expressed with the form: **if** *premise* **then** *consequence*. The *premise* constitutes the antecedent of the rule and is a logical conjunction (**and**) of conditions on the input features. The *consequence* reports the outcome of the classification, i.e. the predicted class label.

The performance of any rule $r_k \in \mathcal{R}$ can be evaluated by covering $C(r_k)$ and error $E(r_k)$ metrics, defined as:

$$C(r_k) = \frac{TP(r_k)}{TP(r_k) + FN(r_k)} \tag{1}$$

$$E(r_k) = \frac{FP(r_k)}{TN(r_k) + FP(r_k)} \tag{2}$$

where $TP(r_k)$ and $FP(r_k)$ are the number of patients that, respectively, correctly and wrongly verify rule r_k; $TN(r_k)$ and $FN(r_k)$ are the number of patients that, respectively, correctly and wrongly do *not* verify the rule. The covering quantifies how many patients correctly satisfy the rule with respect to all the patients belonging to the same output label expressed by the rule consequence: therefore the larger it is, the higher is the probability that the rule is valid on new unseen patients. On the contrary, the error $E(r_k)$ measures how many patients wrongly satisfy the rule with respect to all patients not belonging to the same output label expressed by the rule consequence and its maximum value is usually fixed as a model hyperparameter (by default, it is of 5%).

Both covering and error are useful to define *feature ranking*. It allows to gain insights on which input attributes contribute the most to predict a given class; to this aim, values of relevance for each feature are computed and typically represented in bar plots in descending order.

Given a feature X_j and a rule r_k (predicting class label $\tilde{y}$) containing in its premise a condition c_j on variable X_j, covering and error are first combined to compute the relevance of c_j as $R(c_j) = (E(r'_k) - E(r_k))C(r_k)$, where r'_k is the rule obtained by removing condition c_j from r_k. The relevance $R_j^{\tilde{y}}$ for feature X_j is then derived by the following Eq 3:

$$R_j^{\tilde{y}} = 1 - \prod_k (1 - R(c_j)), \tag{3}$$

where the product is computed on the rules r_k that include a condition c_j on the feature of interest.

Given a patient characterized by measurements $\mathbf{x}$ (i.e., the collection his/her scores for AsthmaRelated, PharynxLarynx, RhinoSinusitis and GastroEsoReflux), the LLM makes a final decision about his/her QoL status, by assigning a class label $\hat{y}$, which can either be *near normal QoL* or *impaired QoL*. Such label assignment depends on the (possibly multiple) rules, generated by the LLM inner process, that are verified by the patient.

Consider the set $\mathcal{R}_{\mathbf{x}}^{near\ normal\ QoL}$ of all rules r_k satisfied by x predicting class label *near normal QoL*, and the set $\mathcal{R}^{near\ normal\ QoL}$ of all rules generated by the model and predicting *near normal QoL* class (i.e., not necessarily satisfied by the considered patient); then, the following quantity expresses a score for the near normal QoL label

$$w(\mathbf{x})_{near\ normal\ QoL} = \frac{\sum_{r_k \in \mathcal{R}_{\mathbf{x}}^{near\ normal\ QoL}} C(r_k)(1 - E(r_k))}{\sum_{r_k \in \mathcal{R}^{near\ normal\ QoL}} C(r_k)(1 - E(r_k))}$$

In the very same way, a score $w(\mathbf{x})_{impaired\ QoL}$ can be computed for the other class label. At the end, the label corresponding to the highest score is assigned as the final label $\hat{y}$ for the considered patient.

Rules statistical validation

In order to assess the statistical significance of the set of rules generated by the LLM, we decided to use the Pearson's χ^2 independence test [22].

The statistical validation of the rules aims at verifying which of the XAI-generated rules can be associated to the output class they predict for a real dependence, and not by chance. In particular, we considered two binary events involving the available data samples, namely their membership to an output class and their satisfaction of the rules in $\mathcal{R}$. Since these are categorical events, we considered Pearson's χ^2 independence test [22] a suitable statistical test for this purpose.

A 2×2 contingency table was built for each rule r_k belonging to ruleset $\mathcal{R}$, as shown in Table 1, reporting the counts of how many samples of the two classes are covered or not by the rule.

Let the input dataset be $\mathcal{T} = \{(\mathbf{x}_i, y_i)\}_{i=1,\dots,N}$, with binary output labels $y_i = 0$ (i.e., *near normal QoL* class in our case) or $y_i = 1$ (i.e., *impaired QoL* class).

The following quantities are defined for each point:

$$a_i = \begin{cases} 1 & \text{if } y_i = 1 \text{ and } \mathbf{x}_i \text{ satisfies } r_k \\ 0 & otherwise \end{cases} \quad b_i = \begin{cases} 1 & \text{if } y_i = 1 \text{ and } \mathbf{x}_i \text{ does not satisfy } r_k \\ 0 & otherwise \end{cases}$$

$$c_i = \begin{cases} 1 & \text{if } y_i = 0 \text{ and } \mathbf{x}_i \text{ satisfies } r_k \\ 0 & otherwise \end{cases} \quad d_i = \begin{cases} 1 & \text{if } y_i = 0 \text{ and } \mathbf{x}_i \text{ does not satisfy } r_k \\ 0 & otherwise \end{cases}$$

Table 1. 2×2 contingency matrix for rule r_k.

	r_k satisfied	r_k not satisfied
$y = 1$	a	b
$y = 0$	c	d

Finally, the elements of Table 1 can be computed as:

$$a = \sum_{i=1}^{N} a_i, \quad b = \sum_{i=1}^{N} b_i, \quad c = \sum_{i=1}^{N} c_i, d = \sum_{i=1}^{N} d_i$$

χ^2 statistic was then computed starting from the matrix. The test was carried out with a null hypothesis of independence between class label and rule membership, with a significance level of 0.05 for the p-value. Rules with a p-value <0.05 were then proved as statistically significant [23] and those that did not pass the test were removed from the ruleset $\mathcal{R}$, giving rise to a set of validated rules $\mathcal{R}_{val} \subseteq \mathcal{R}$.

Model performance evaluation

To evaluate the overall performance of the validated ruleset, the confusion matrix reporting the True Positives (TP, i.e., patients correctly predicted as *impaired QoL*), False Positives (FP, i.e., *near normal QoL* patients wrongly predicted as *impaired QoL*), True Negatives (TN, i.e., patients correctly predicted as *near normal QoL*) and False Negatives (FN, i.e., *impaired QoL* patients wrongly predicted as *near normal QoL*) obtained by applying such rules to a test set was first built. It is the basis to define the following measurements, particularly useful when evaluating the outcomes of a clinical ML model [24]:

$$ACC = \frac{TP+TN}{TP+FP+TN+FN} \qquad F_1 = \frac{2TP}{2TP + FP + FN}$$

$$PPV = \frac{TP}{TP+FP} \qquad NPV = \frac{TN}{TN + FN}$$

$$TPR = \frac{TP}{TP+FN} \qquad TNR = \frac{TN}{TN + FP}$$

While accuracy (ACC) and F_1-score (F_1) provide an evaluation of the model taking into account its performance on both the classes, the other ones assess the performance on single classes. In detail, Positive Predictive Value (or precision, PPV) and True Positive Rate (or sensitivity or recall, TPR) reflect the number of TPs over the total amount of positive predictions and the total amount of positive samples, respectively. Viceversa, Negative Predictive Value (NPV) and True Negative Rate (or specificity, TNR) represent the number of TNs over the total amount of negative predictions and the total quantity of negative samples, respectively.

Results

This study involved a population of 283 asthmatic patients (i.e., the *all* group), with age 11–79 years old and characterized by a Forced Expiratory Volume in the first second (FEV1) of 96.5% ± 19.09 (i.e, varying in the 27–143 range, with a median value of 94%), and an ACT score of 19.09±4.98. 146 patients belong to the *controlled asthma* group (i.e., the 52% of the whole population), while the remaining 137 patients form the *not controlled asthma* group.

The dataset is available at the following link: https://github.com/saranrt95/Cough_QoL_In_Asthma.

Data statistics at a first glance

Fig 3 provides a first glance on how the four blocks of symptoms are distributed between the two classes (impaired QoL and near normal QoL) both in the controlled and not controlled asthma patients. Each colored bar individuates a different group of patients and its length (the interquartile range, or IQR) varies between the 25th and 75th percentiles, while the vertical
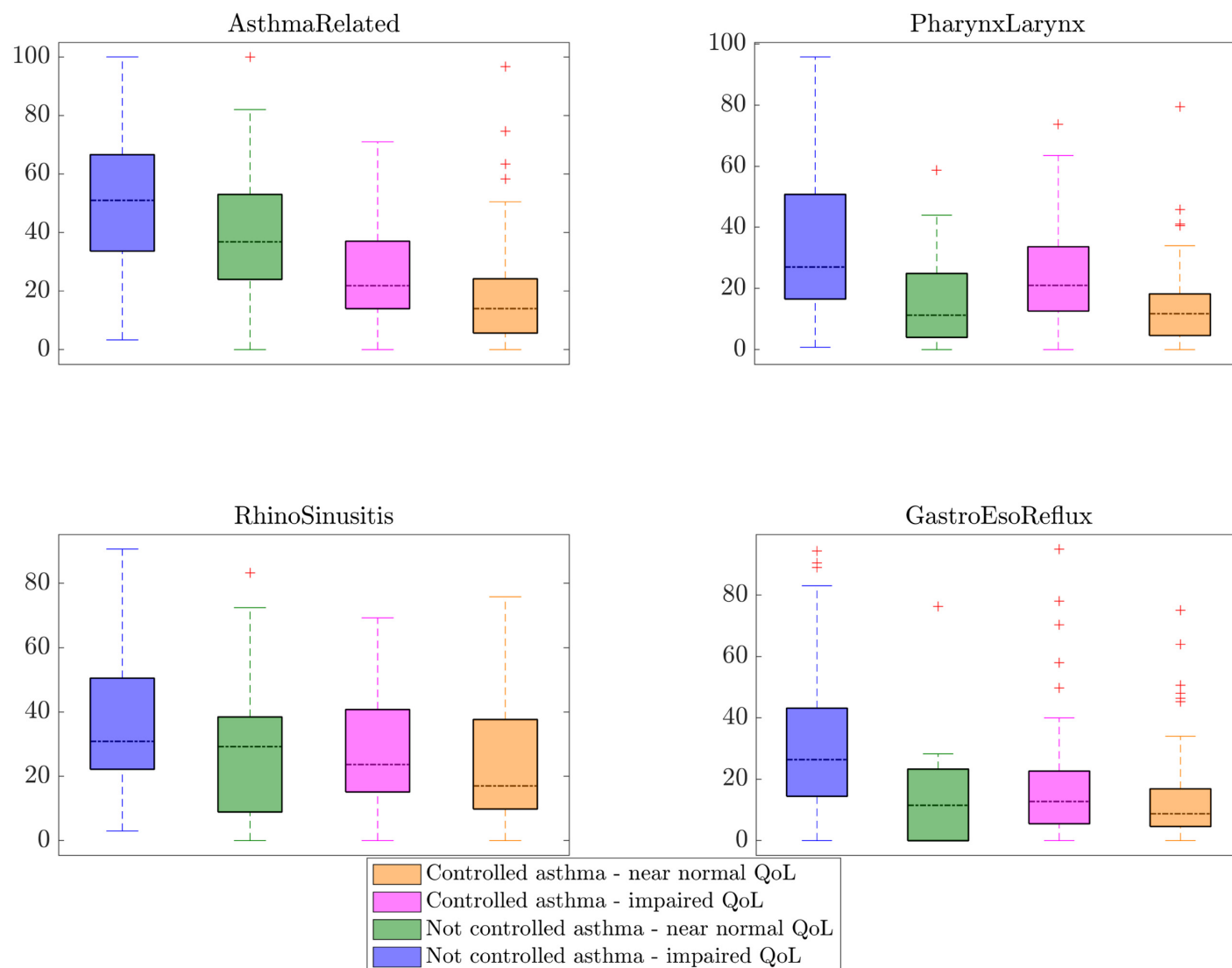
Fig 3. Box plots. Graphs showing the class distributions in the controlled versus not controlled patients groups, for each of the considered features.

dashed lines (i.e., the whiskers) range from the minimum to the maximum values and, finally, the horizontal dot-dashed black line points out the median value of the corresponding symptoms group. The red '+' markers represent outlier points. By observing the plots the following points arise. *AsthmaRelated* median values and the related IQRs are pretty different among all the groups. As expected, the median scores for the not controlled asthma (purple and green boxes) are higher than for the patients with controlled asthma (pink and orange boxes). Also, the values are larger for the impaired QoL than the near normal QoL class. *PharynxLarynx* can help distinguishing the two classes both in the controlled and not controlled asthma groups, since the median values for the impaired QoL class in the two groups are 21 and 27, respectively, against the 11.75 and 11.25 values for the near normal groups. Analyzing the box for *RhinoSinusitis* variable, we can observe that the medians of the not controlled asthma group are very close to each other (around 30) in both impaired and near normal QoL classes. In the controlled group, the median of the impaired QoL class (23.64) is slightly larger than for the

other class (17.00), suggesting some role of this feature in differentiating the classes in this group. Lastly, the *GastroEsoReflux* factor shows higher values for the impaired QoL class in the not controlled asthma group, while the other boxes are aligned over the same range of values.

However, this kind of evaluation is based on visual analytics and simple statistics, and the results do not provide any guarantee of validity on new, unseen, patients. Also, such kind of exploration only accounts for one variable at a time, and not for the relationships among them. This is why we decided to explore a machine learning-based approach.

Explainable AI-based analysis

For each of the considered cases, the LLM algorithm was trained on a 70% training set and generated a set of rules. In particular, for the *all* group, 19 rules were generated (8 predicting *impaired QoL* class and 11 the *near normal QoL*); from the *controlled asthma* case, we got 13 rules (4 for the *impaired QoL* and 9 for the *near normal QoL* class); lastly, 9 rules derived from the *not controlled asthma* group (5 referring to the *impaired QoL* and 4 to the *near normal QoL* class).

The Pearson's χ^2 validation test was then carried out to statistically proof the obtained rule-sets, as per the procedure detailed in Section. After the test, 2 rules out of 8 for the *impaired QoL* class and 4 out of 11 for the *near normal QoL* class were validated in the *all* case; 2 of the 4 rules predicting the *impaired QoL* class in the *controlled asthma* group resulted significant, while 3 out of 9 rules for the other class was validated in the same group; similarly, in the *not controlled asthma* patients, 2 rules out of 5 for the *impaired QoL* class passed the test, while 3 out of 4 rules related to the *near normal QoL* did.

Model performance metrics. After validating the rules, we thus have been able to define a final set of rules for each case, by leaving out from the original rulesets all those which tested not significant. The predictive performance of the validated rulesets was assessed on the test set, by computing the metrics described in Sec. Model Performance Evaluation; their values are depicted and compared in Fig 4 for the three groups.

The accuracy reached at least 70% in all cases, thus showing good performance of the validated rulesets. While also F_1 score value was high for the *all* and *not controlled asthma* groups (75% and 83%, respectively), it was lower (57%) for the *controlled asthma* group, denoting both poorer precision and recall. Indeed, PPV and TPR metrics, related to the positive class (i.e., impaired QoL), were found 66% and 50%, respectively, whereas NPV and TNR (reflecting the model's performance on the negative class, i.e., the near normal QoL) were sensitively larger (74% and 85%, respectively). In contrast, the *not controlled asthma* reached a high F_1 due to larger values of precision and recall, with a PPV of 77% and TPR of 89%; on the other hand, NPV and TNR resulted in lower values. A similar reasoning holds for the *all* group, even if the model performance on the two classes was more balanced, with less difference among the metrics for the positive and the negative class.

Most relevant symptoms questionnaire items. Further insights on the LLM results were obtained by visualizing the feature ranking. Bar plots, obtained for the three cases under analysis, are shown in Fig 5, representing the *impaired QoL* class feature ranking, that highlights which of the features influenced more the LLM decision towards that class. Concerning the *all* group, from Fig 5A *AsthmaRelated* and *PharynxLarynx* were individuated as the main factors leading to an impaired cough-related quality of life. In contrast, the main attributes for the *controlled asthma* group (Fig 5B) were *PharynxLarynx* and *RhinoSinusitis*. Finally, dominant features for the *not controlled asthma* resulted *AsthmaRelated* and *GastroEsoReflux* (Fig 5C).

The presence of *AsthmaRelated* as a relevant factor for the *not controlled asthma* group is in line with our expectation, since the deterioration of these patients' QoL reasonably depends on
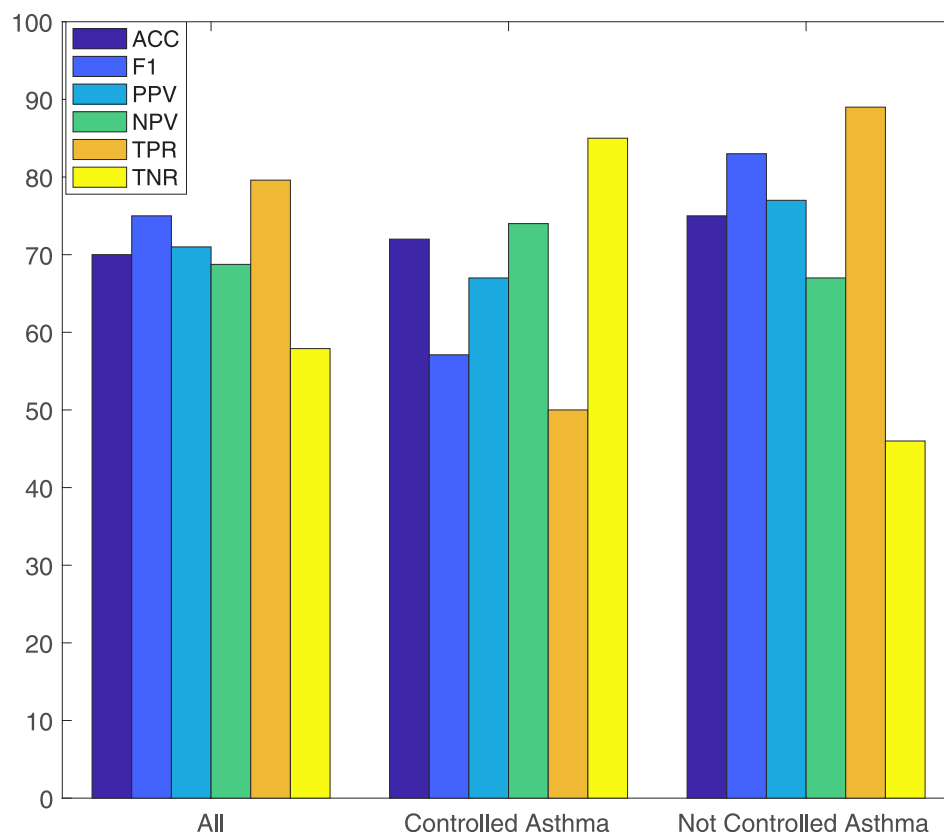
Fig 4. Validated rules performance. Percentage values of the accuracy (ACC), F_1-score (F1), Positive Predictive Value (PPV), Negative Predictive Value (NPV), True Positive Rate (TPR) and True Negative Rate (TNR) of the LLM in the three patients' groups.

https://doi.org/10.1371/journal.pone.0292980.g004

the asthma itself and the clinical investigation should be primarily addressed to it. Secondarily, the digestive tract should be considered. Conversely, the feature ranking for the *controlled asthma* patients provides the indication that further clinical assessments should focus first on the throat and, then, on the nose. By using the symptoms questionnaire, in absence of any information about the patient's asthma control level, results suggest to first consider the asthma and then the nose.

Symptoms questionnaire scores driving impaired QoL. While previous Section provided which are the main factors involved in the impaired QoL, in this Section our focus is posed on the information we can derive by inspecting the validated rules predicting the *impaired QoL* class, which are reported in Table 2. Their aim is to define useful criteria to support clinicians in the diagnostic process, by individuating, in the three cases, which values assumed by the symptoms questionnaire scores are more probably associated to an *impaired QoL* status. In this regard, we note that for a more practical use of these criteria, rules shown in Table 2 have been truncated to express integer thresholds (with no decimals); in the largest part of the cases, this approximation did not heavily impact performance. The only exception concerns the second rule of the non controlled asthma group, whose original threshold of 28.375 achieved the 0% error, while cutting it to 28 raises the error to 13%. This is probably due to the presence of near normal QoL points (i.e., of the opposite class) around these values, and the slight variation introduced by truncation also causes these points to satisfy the rule,
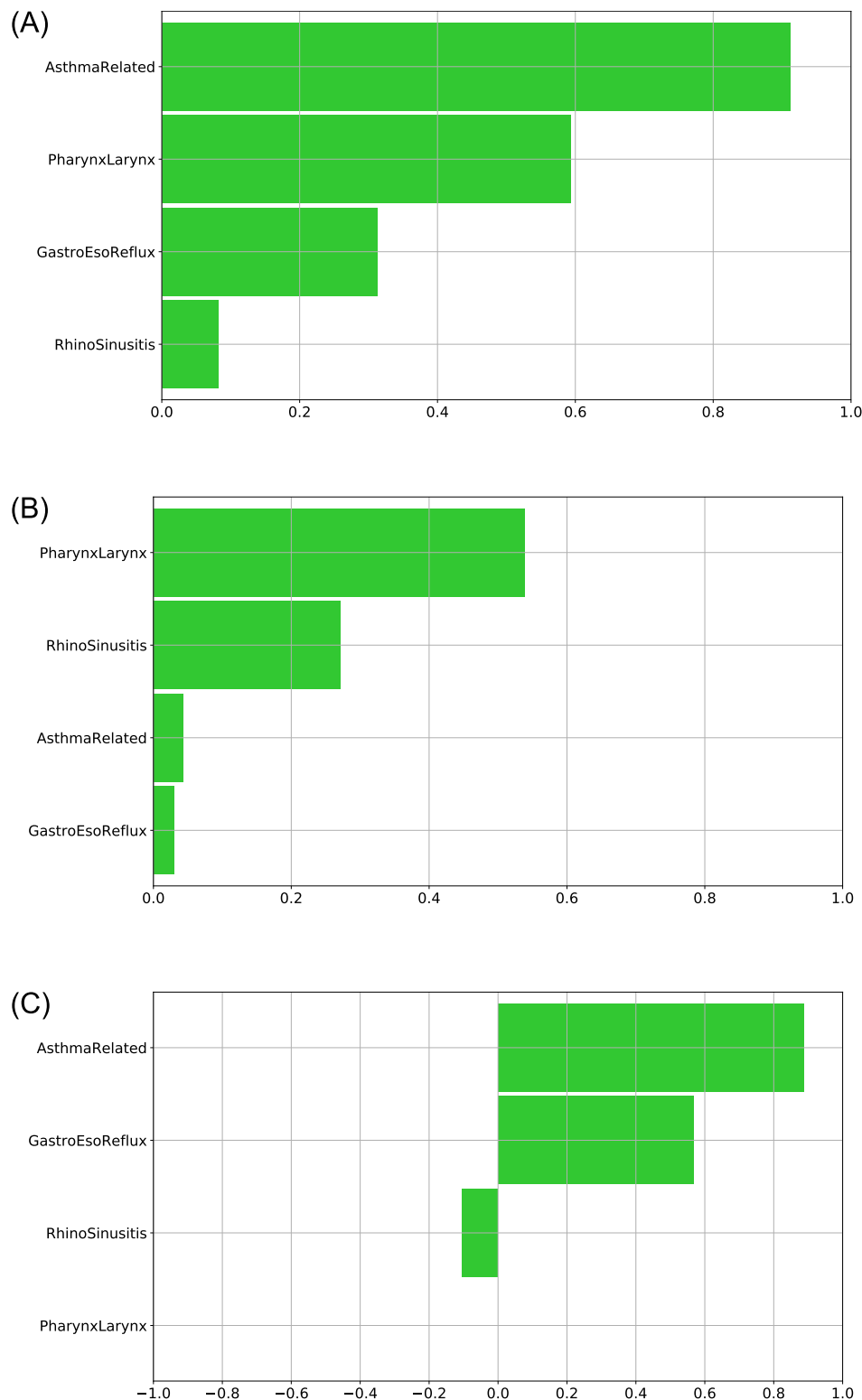
Fig 5. Feature ranking. LLM feature ranking for the impaired QoL class in the three cases. (A): all group; (B): controlled asthma group; (C): Not controlled asthma group.

https://doi.org/10.1371/journal.pone.0292980.g005

Table 2. Criteria for *impaired QoL* prediction through symptoms questionnaire, as emerged from LLM rules validated through the χ^2 independence test, for each considered patient group. Pink-colored cells highlight the rules that were proved the most performing even on previously unseen patients.

Case	Significant Rules	Covering (%)	Error (%)
All	1. **if** AsthmaRelated > 8 **and** PharynxLarynx > 15 **and** RhinoSinusitis ≤ 70 **and** GastroEsoReflux > 8 **then** *impaired QoL*	57.7	5.6
	2. **if** AsthmaRelated > 28 **and** 8 < RhinoSinusitis ≤ 70 **and** GastroEsoReflux > 4 **then** *impaired QoL*	56.7	5.6
Controlled Asthma	1. **if** AsthmaRelated > 11 **and** PharynxLarynx > 23 **and** RhinoSinusitis ≤ 70 **then** *impaired QoL*	44	5.7
	2. **if** PharynxLarynx ≤ 13 **and** 11 < RhinoSinusitis ≤ 27 **and** GastroEsoReflux > 1 **then** *impaired QoL*	20	5.7
Not controlled Asthma	1. **if** AsthmaRelated > 51 **and** RhinoSinusitis ≤ 70 **then** *impaired QoL*	46	0
	2. **if** GastroEsoReflux > 28 **then** *impaired QoL*	41	13

https://doi.org/10.1371/journal.pone.0292980.t002

thus increasing the error value (that expresses how many points satisfy the rule, but do not belong to the output predicted by the rule). By looking at the threshold values of a same indicator in the two rules for a given group, it can be noticed that they can be pretty different or even conflicting. For example, the 8 and the 28 in the *AsthmaRelated* score for the *all* group have a difference of 20 percentage points, which cannot be disregarded; also, the condition on *PharynxLarynx* in the *controlled asthma* group is discordant in the two related rules, the first stating that values larger than 23 lead to QoL deterioration, while the second states the same for values lower than 13. Regarding the *not controlled asthma* case, the two rules seem to individuate two clusters of patients, one depending on increasing (> 51) *AsthmaRelated* score and decreasing (≤ 70) *RhinoSinusitis* score, and the other depending on *GastroEsoReflux* score only. Therefore, rule generation alone is able to individuate several clusters of patients, each described by a pretty different set of conditions on the questionnaire scores. Nevertheless, our final goal is to provide, through the ML system, more general information to be used in clinical practice, especially valid in the case of new, never seen before, patients.

Further evaluations of the models are then carried out for a better knowledge extraction suitable to our objective. Covering and error percentages reported in the Table have been derived during model training on the training data portion. Hence, their values, even when considerably high (as in the cases of > 50% covering), do not guarantee the same performance on test (previously unseen) data. Thus, percentages of impaired QoL test points satisfying either one, both or even none of the two rules were computed to understand how the original covering changes on new data; the obtained values are outlined in Table 3. When points satisfy both rules, the most important one can still be individuated as the highest-covering one (from Table 2). Thus, the 44.90% rate of satisfaction of both rules in the *all* group contributes to the rate of rule 1 (of the same group), which then reaches a total value of 65.31% of satisfaction. Thus, this rule should be taken as a reference for individuating the factors with higher impact on the impaired QoL. The same reasoning holds for the *not controlled asthma* group, where rule 1 reaches the about the 52%. Regarding the *controlled asthma* case, rule 1 proves as the most frequently validated by the unseen patients. Moreover, it is worth noting that the sum of the percentages shown in Table for Rule 1, Rule 2 and Both rules columns corresponds to the

Table 3. Satisfaction percentages of validated rules for the *impaired QoL* class on unseen data. For each group, Rule 1 refers to rule number 1 of Table 2, and, similarly, Rule 2 here refers to rule number 2.

	Rule 1	Rule 2	Both rules	No rules
All	20.41%	14.28%	44.90%	20.41%
Controlled asthma	41.67%	8.33%	0%	50%
Not controlled asthma	37.04%	37.04%	14.81%	11.11%

https://doi.org/10.1371/journal.pone.0292980.t003

TPR computed in Fig 4. Hence, in this analysis we can see the specific contribution of the two rules in determining its value.

In summary, for each of the three groups, a rule has emerged as the one with the best predictive ability for an *impaired QoL* status and it can be considered as a helpful decision-making support for clinicians, especially at the beginning of the clinical evaluation process. Indeed, by using the information from the feature ranking (Fig 5), we discovered the main blocks of symptoms associated to an impaired QoL status due to chronic cough and the individuated decision rules define which ranges of values should be considered alarming on those variables.

Conclusion

In this work, we proposed the evaluation of the quality of life of asthmatic patients, with lower or higher degree of asthma control, experiencing chronic cough. To this end, we first developed a questionnaire to collect patients' symptoms in relation to the most frequent causes of chronic cough (i.e., upper airways, pharynx/larynx, digestive tract, lower airways). The LLM-based analysis of patients' responses to the questionnaire items, through feature ranking, helped in automatically identifying priorities among these causes: pharynx/larynx and upper airways when asthma is sufficiently controlled, and asthma itself and digestive trait when asthma is not controlled. Moreover, the adopted rule-based model, with proper statistical validation, identified which specific values of the symptoms are associated to an impairment of cough-related quality of life. The obtained results could support the physician in choosing the right diagnostic/therapeutic plan. However, sensitivity and specificity of the developed model need to be verified in further prospective studies. Furthermore, future research in this direction may investigate the adoption of other rule-based models than the LLM, as well as the usage of black-box algorithms with subsequent rule extraction.

Supporting information

S1 Table. Respiratory symptoms questionnaire. Structure of the questionnaire used to collect patients' respiratory symptoms. The groups of items forming the features used in our analysis are defined as follows: items 1–3 (*AsthmaRelated*); 4–7 (*PharynxLarynx*); 8–14 (*RhinoSinusitis*); 15–18 (*GastroEsoReflux*).
(PDF)

Author Contributions

Conceptualization: Ilaria Baiardini, Fulvio Braido, Maurizio Mongelli.

Data curation: Sara Narteni, Ilaria Baiardini, Fulvio Braido.

Formal analysis: Ilaria Baiardini, Maurizio Mongelli.

Investigation: Ilaria Baiardini, Maurizio Mongelli.

Methodology: Sara Narteni, Ilaria Baiardini, Fulvio Braido, Maurizio Mongelli.

Software: Sara Narteni.

Supervision: Ilaria Baiardini, Fulvio Braido, Maurizio Mongelli.

Validation: Fulvio Braido, Maurizio Mongelli.

Writing – original draft: Sara Narteni.

Writing – review & editing: Sara Narteni, Ilaria Baiardini, Fulvio Braido.

References

1. European Commission, Directorate-General for Communications Networks, Content and Technology, (2019). Ethics guidelines for trustworthy AI—, Publications Office. https://data.europa.eu/doi/10.2759/346720

2. Bharati S., Mondal M. R. H. and Podder P., "A Review on Explainable Artificial Intelligence for Healthcare: Why, How, and When?," in IEEE Transactions on Artificial Intelligence https://doi.org/10.1109/TAI.2023.3266418

3. Saraswat D. et al., "Explainable AI for Healthcare 5.0: Opportunities and Challenges," in IEEE Access, vol. 10, pp. 84486–84517, 2022. https://doi.org/10.1109/ACCESS.2022.3197671

4. Hulsen T. Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare. AI. 2023; 4 (3):652–666. https://doi.org/10.3390/ai4030034

5. Morice AH, Millqvist E, Bieksiene K, et al. ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. Eur Respir J 2020; 55:1901136. https://doi.org/10.1183/13993003.01136-2019 PMID: 31515408

6. Nathan R.A.; Sorkness C.A.; Kosinski M.; Schatz M.; Li J.T.; Marcus P.; et al. Development of the asthma control test: a survey for assessing asthma control. Journal of Allergy and Clinical Immunology 2004, 113, 59–65. https://doi.org/10.1016/j.jaci.2003.09.008 PMID: 14713908

7. Irwin RS, French CL,Chang AB, et al. Classification of Cough as a Symptom in Adults and Management Algorithms: CHEST Guideline and Expert Panel Report. Chest 2018; 153:196–209. https://doi.org/10.1016/j.chest.2017.10.016 PMID: 29080708

8. Tsang KCH, Pinnock H, Wilson AM, Shah SA. Application of Machine Learning Algorithms for Asthma Management with mHealth: A Clinical Review. J Asthma Allergy. 2022 Jun 29; 15:855–873. https://doi.org/10.2147/JAA.S285742 PMID: 35791395

9. Shim JS, Kim BK, Kim SH, Kwon JW, Ahn KM, Kang SY, et al. A smartphone-based application for cough counting in patients with acute asthma exacerbation. J Thorac Dis. 2023 Jul 31; 15(7):4053–4065. https://doi.org/10.21037/jtd-22-1492 PMID: 37559656

10. Kaur S, Larsen E, Harper J, Purandare B, Uluer A, Hasdianda MA, et al. Development and Validation of a Respiratory-Responsive Vocal Biomarker-Based Tool for Generalizable Detection of Respiratory Impairment: Independent Case-Control Studies in Multiple Respiratory Conditions Including Asthma, Chronic Obstructive Pulmonary Disease, and COVID-19. J Med Internet Res. 2023 Apr 14; 25:e44410. https://doi.org/10.2196/44410 PMID: 36881540

11. Luo X, Gandhi P, Zhang Z, Shao W, Han Z, Chandrasekaran V, et al. Applying interpretable deep learning models to identify chronic cough patients using EHR data. Comput Methods Programs Biomed. 2021 Oct; 210:106395. https://doi.org/10.1016/j.cmpb.2021.106395 PMID: 34525412

12. Chen W, Schatz M, Zhou Y, Xie F, Bali V, Das A, et al. Prediction of persistent chronic cough in patients with chronic cough using machine learning. ERJ Open Res. 2023 Mar 27; 9(2):00471-2022. https://doi.org/10.1183/23120541.00471-2022 PMID: 37009024

13. Brew BK, Chiesa F, Lundholm C, Ortqvist A, Almqvist C (2019) A modern approach to identifying and characterizing child asthma and wheeze phenotypes based on clinical data. PLoS ONE 14(12): e0227091. https://doi.org/10.1371/journal.pone.0227091 PMID: 31887128

14. Almqvist C, Örtqvist AK, Ullemar V, Lundholm C, Lichtenstein P, Magnusson PK. Cohort Profile: Swedish Twin Study on Prediction and Prevention of Asthma (STOPPA). Twin Res Hum Genet. 2015 Jun; 18 (3):273–80. https://doi.org/10.1017/thg.2015.17 PMID: 25900604

15. Kang J, Seo WJ, Kang J, Park SH, Kang HK, Park HK, et al. (2023) Clinical phenotypes of chronic cough categorised by cluster analysis. PLoS ONE 18(3): e0283352. https://doi.org/10.1371/journal.pone.0283352 PMID: 36930618

16. Koo H-K, Jeong I, Kim J-H, et al. Development and validation of the COugh Assessment Test (COAT). Respirology. 2019; 24: 551–557. https://doi.org/10.1111/resp.13462 PMID: 30681246

17. Kwon JW, Moon JY, Kim SH, Song WJ, Kim MH, Kang MG, et al. Reliability and validity of a korean version of the leicester cough questionnaire. Allergy Asthma Immunol Res. 2015 May; 7(3):230–3. https://doi.org/10.4168/aair.2015.7.3.230 PMID: 25749761

18. Baiardini I., Braido F., Fassio O., Tarantini F., Pasquali M., Tarchino F., et al. (2005), A new tool to assess and monitor the burden of chronic cough on quality of life: Chronic Cough Impact Questionnaire. Allergy, 60: 482–488. https://doi.org/10.1111/j.1398-9995.2005.00833.x PMID: 15727580

19. Braido F, Baiardini I, Menoni S, Gani F, Senna GE, Ridolo E, et al. (2012) Patients with Asthma and Comorbid Allergic Rhinitis: Is Optimal Quality of Life Achievable in Real Life? PLoS ONE 7(2): e31178. https://doi.org/10.1371/journal.pone.0031178 PMID: 22363573

20. https://www.rulex.ai

21. Muselli, M. Switching Neural Networks: A New Connectionist Model for Classification. In: Apolloni, B., Marinaro, M., Nicosia, G., Tagliaferri, R. (eds) Neural Nets. WIRN NAIS 2005 2005. Lecture Notes in Computer Science, vol 3931. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11731177_4

22. Whatley M. One-Way ANOVA and the Chi-Square Test of Independence. In Introduction to Quantitative Analysis for International Educators; Springer, 2022; pp. 57–74. https://doi.org/10.1007/978-3-030-93831-4_5

23. Vaccari I.; Orani V.; Paglialonga A.; Cambiaso E.; Mongelli M. A Generative Adversarial Network (GAN) Technique for Internet of Medical Things Data. Sensors 2021, 21, 3726. https://doi.org/10.3390/s21113726 PMID: 34071944

24. Hicks S.A., Strümke I., Thambawita V. et al. On evaluation metrics for medical applications of artificial intelligence. Sci Rep 12, 5979 (2022). https://doi.org/10.1038/s41598-022-09954-8 PMID: 35395867