

ORIGINAL ARTICLE

Analysis of A Panel Of Cytokines in BAL Fluids to Differentiate Controlled and Uncontrolled Asthmatics Using Machine Learning Model

Pooja MR¹, Pushpalatha MP²

¹ Department of Computer Science & Engg, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India; Research Scholar at Department of Computer Science & Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka, India

² Department of Computer Science & Engg, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka, India

Conflict-of-interest statement: The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Correspondence to: Pooja MR, Department of Computer Science & Engg, Vidyavardhaka College of Engineering, Mysuru, Karnataka, India

Email: pooja.mr@vvce.ac.in

Telephone: +91-9886819448

Received: February 23, 2019

Revised: April 7, 2019

Accepted: April 10, 2019

Published online: June 8, 2019

ABSTRACT

Cytokines play an integral part in the coordination and persistence of the inflammatory process in the chronic inflaming of the airways in bronchial asthma since they are capable of inducing many of the pro-inflammatory effects characteristic of this disease. Here, we adopt a machine learning model to identify the important cytokines that cause the inflammation of the lungs and use them exclusively to differentiate between controlled and uncontrolled asthma subjects. The predominant cytokines are primarily identified by a combined

feature scoring technique. The asthma controls are classified by deploying the binary classifiers. The best results were obtained with Logistic regression classification technique when applied on the optimal cytokines combination.

Key words: Cytokines; Controlled, BAL; Feature scoring; Classification

© 2019 The Author(s). Published by ACT Publishing Group Ltd. All rights reserved.

Pooja MR, Pushpalatha MP. Analysis Of A Panel Of Cytokines In BAL Fluids To Differentiate Controlled And Uncontrolled Asthmatics Using Machine Learning Model. *Journal of Respiratory Research* 2019; **5(1)**: 142-145 Available from: URL: <http://www.ghrnet.org/index.php/jrr/article/view/2611>

INTRODUCTION

Asthma, a chronic inflammatory disease of the airways is currently regarded as the most common chronic respiratory disease in both rural and industrialized regions. The disease inflames and narrows the airways in the lung. Symptoms include coughing, shortness of breath and pectus tightness. Asthma control may be defined in a variety of ways. In general, the term “control” may indicate disease prevention or even cure. However, in asthma, where neither of these are realistic options at present, it refers to control of the manifestations of the disease^[1]. Patients who experience prolonged periods of uncontrolled asthma have a higher incidence of exacerbations and increased morbidity and mortality rates. The ability to determine and to predict levels of asthma control and the occurrence of exacerbations is crucial in asthma management^[2]. Asthma is heterogeneous with reference to clinical options, cellular sources of inflammation, and response to common therapies^[3]. Asthma control assessment is of particular importance as current asthma guidelines suggest that treatment should be maintained, stepped-up, or stepped-down based on the level of control^[5]. Accordingly, the necessity for tools which will dependably determine and monitor the management parameter has become a vital issue. The application of valid prognostic models for outcome prediction can considerably impact the clinical management of asthma

attack^[11].

Predicting severe exacerbation caused by uncontrolled asthma is very vital for patients, as avoiding critical symptoms which will need special treatment or perhaps hospitalization will defend patients from bronchostilation and its effects. Four Machine learning algorithms were evaluated for their accuracy in predicting severe asthma attack exacerbations by exploiting completely different parameters, of which SVM and Naïve Bayesian approaches showed higher sensitivity, while Random Forests showed higher negative prognostic power (very high specificity)^[3].

An important problem in realizing personalized medicine is the development of methods for identifying disease subtypes using quantitative proteomics. Recently it was found that bronchoalveolar lavage (BAL) cytokine patterns contain information about dynamic lung responsiveness^[10]. In the meta-analysis of patients with persistent asthma, the use of single maintenance and reliever therapy compared with inhaled corticosteroids as the controller therapy (with or without a long-acting β -agonist) and short-acting β -agonists as the relief therapy was associated with a lower risk of asthma exacerbations^[4].

Asthma is a very complex and difficult terminal figure to define in simple manner, currently it is considered to be a group of diverse disorders described by three major characteristics: (1) intermittent and reversible airway obstruction leading to recurrent episodes of wheezing, SOB, chest tightness and cough; (2) bronchohyperresponsiveness (BHR) which is defined as an increased sensitivity to bronchoconstrictors such as histamine and cholinergic agonist and (3) airway inflammation. Many cytokines are involved in the development of the atopic state and of the chronic inflammatory process of asthma, ultimately contributing to airway remodelling, bronchoconstriction, and bronchial hyperresponsiveness. The potential role of each cytokine in these processes can be evaluated by studying their properties, their presence and localization in the airway wall and airway secretions of affected role with asthma, and the effect of particular inhibitors such as sense organ antagonists or specific anti bodies. Cytokine therapy has tremendous potential for treating a variety of disease. Anti-inflammatory drug for asthma may be developed by targeting inhibition of cytokine production and effects (such as cytokine antibodies, cytokine sense organ antagonists, or blockers of specific signaling transduction effects) or by using or modifying anti-inflammatory cytokines.

Asthma attack is a chronic inflammatory disease of bronchial mucosa, in which mast prison cell, eosinophils and activated T cells are of considerable importance. The increased chemotactic bodily process for T cells in patient with asthma is mainly attributable to IL-16^[7]. A strong association between asthma and allergic rhinitis exists from a clinical and epidemiologic standpoint of view. The increased levels of IL-8 in the airway secretions from both patients with asthma and COPD may be markers of an ongoing inflammatory process, which is more pronounced in patients with COPD^[6]. In patients with asthma the strong correlation between the levels of IL-8 and the percentage neutrophils was reported in^[9,13]. Accumulating evidence suggests that the cytokine network is central to the immunopathology of bronchial asthma and recent findings have suggested that naturally occurring cytokine antagonists may also be involved. In the study carried out in^[8], the expression of interleukin-1 beta (IL-1beta) and its naturally occurring receptor antagonist, IL-1ra, in the normal and asthmatic bronchial wall was studied. There was an increased expression of both IL-1beta and IL-1ra in the asthmatic bronchial epithelium IL-1 receptor antagonist protein, an anti-inflammatory cytokine that plays an important role in maintaining the balance between inflammatory and anti-inflammatory cytokines.

Machine Learning techniques play a significant role in the area of predictive analytics as they aim to effectively predict disease outcomes by identifying risk factors contributing to the disease. Here, we develop a machine learning model that is adequately tuned to get the best performance possible. Initially, we try to analyze the predominant cytokines that contribute to the process of distinguishing controlled asthmatics from uncontrolled asthmatics by employing feature scoring technique, followed by a process of weighted averaging, which tries to extract the best of the features. Further, we apply binary classifiers to distinguish controlled and uncontrolled asthmatics and deduce the best classifier that performs optimally in the process.

MATERIAL AND METHODS

Dataset

The dataset contains information of 36 subjects including 11 healthy, 15 controlled and 10 uncontrolled subjects acquired from the study data of Department of Asthma, Allergy and Lung Biology, King's College London School of Medicine, U.K. which was available on the Dryad repository. A panel of 48 cytokines and chemokines in BAL fluids from healthy control subjects and subjects with controlled and uncontrolled asthma was used for the study. The asthma severity was defined based on FEV1 while on treatment, according to international ERS/ATS guidelines. The asthma attribute was coded as 0 for healthy subjects, 1 for controlled subjects and 2 for uncontrolled subjects. IL-2, IL-4, IL-17, FGF, Eotaxin, GM-CSF, MIP-1 α were eliminated as there were no recordings for the same. The healthy subjects were characterized as those who had no history of allergic disease, had PC20 of more than 32 mg per milliliter and normal FEV1 whereas asthmatic subjects were included on the basis of history and a demonstrated airflow limitation which is reversible and increased airway responsiveness to meth choline or both. However no history of other respiratory diseases was reported.

Identification of inflammatory cytokines

Features constituting the feature subset along with their feature scores obtained using different feature scoring techniques namely Information gain, Gain ratio, Gini, ReliefF and FCBF. About 12 features constituting approximately 25% of the total features were selected by combined feature scoring method which performed average weighting of features chosen by the individual feature ranking techniques.

The following feature scoring techniques have been used in our approach.

Information gain: the amount of information expected resulting in reduction of entropy Gain ratio: the proportion of gained information and the information present intrinsically within the attribute Gini: the index that discriminates the frequency distance values ANOVA (Analysis of one way variance): the difference between the average values of the feature in different classes Chi2: the dependence as measured using chi-square statistics between the class and the feature ReliefF: the capacity of an attribute to distinguish between classes with respect to the data instances those of which are similar FCBF (Fast Correlation Based Filter): an entropy-based measurement which identifies redundancy that arises because of pairwise correlations between features.

Of the 12 features in the reduced feature subset, we used 4 features that have comparatively higher scores namely, IL-8, IL-16, IL1-RA and percentage of neutrophils. We excluded features having higher scores namely ICS dosage, percentage of predicted FEV1, LABA, ICS usage if any and age, as our study aimed at analyzing only the

association between the asthma control levels and the inflammatory cytokines in BAL fluids.

Though initially we perform feature selection using filtering approach to arrive at the reduced feature set containing 12 attributes, we further apply wrapper method on the 4 features mentioned above, to tune the model to accommodate the best of the features that offer an optimal performance. With wrapper methods, we use classification performance of a classifier such as Classification accuracy to evaluate the optimal features that help in better prediction. Experimentally we infer that the combination of IL-8, IL-16, IL1-RA and percentage of neutrophils is an optimal feature subset that yields a very good performance and thus we conclude that wrapper methods can always provide the best subset of features.

RESULTS AND DISCUSSION

Table 1 presents 20 percent of the total features/attributes that are extracted into the feature subset by applying feature scoring technique. The score with respect to the individual feature ranking techniques is also displayed. However, we neglect the other features in the feature subset as our work aims at identifying only the cytokines and chemokines constituting the BAL fluid that influence the decision of identifying the asthma control level (controlled/uncontrolled). Further we explore the variable combinations of cytokines to investigate into the optimal combination that offers the best performance by applying the binary classifier presented in the Tables 2, 3, 4 and 5. It can be concluded that the cytokines IL-8, IL-16, IL1-RA and neutrophil percentage play an important role in distinguishing the two categories of asthma control levels namely controlled and uncontrolled by observing the performance metrics namely AUC, CA, F1, Precision and Recall. Table 5 reveals that Logistic regression technique yields a value of 1.0 for all the above mentioned metrics indicating its optimal performance over other binary classifiers.

Our results indicate that IL-8, IL-16 and IL1-RA along with the neutrophil percentage predict the correct category of asthma controls with a very high degree of precision. Logistic regression technique predicted the outcome with a precision and recall of 1.0 and 1.0 respectively indicating its optimal performance. Further, Classification accuracy and AUC also yielded a value of 1.0 in this case. As per the work carried out by Hosoki et al, IL-8 was the only cytokine that was seen to distinguish controlled from uncontrolled asthma subjects^[14]. However when IL-8 was alone used to predict the control category, prediction metrics yielded a precision of 0.883 and recall of 0.750 with both Naïve Bayes and Support Vector Machine classifiers. Further, we evaluated the classification performance using IL-8, IL-16 and IL1-RA as the cytokines to distinguish the two classes and obtained a precision of 0.883 and recall of 1.000 with SGD, and SVM techniques. Thus using the inference from Hosoki et.al, we tried to deduce the most important cytokines in order to profile asthma categories. The results obtained via different classification techniques in varying scenarios are presented. Such inferences are at most important in deriving therapeutic implications to treat the disease at right time.

CONCLUSION

Our deeper analysis investigating the relationship between asthma control level and the components of BAL fluids via machine learning model proposed, has revealed that few cytokines and chemokines namely IL-8, IL-16 and IL1-RA as well as neutrophil percentage play a significant role in the process of distinguishing controlled

Table 1 Feature subset obtained via scoring technique.

Features	Information gain	Gain Ratio	Gini	Relieff	FCBF
ICS (doseµg/day)	0.971	0.551	0.48	0.462	0.708
%predict FEV1	0.728	0.364	0.36	0.226	5.10E-5
LABA	0.717	0.767	0.388	0.72	0.721
ICS (use)	0.610	0.610	0.32	0.564	6.62E-5
Age	0.590	0.297	0.3	0.314	4.05E-5
IL-1RA	0.367	0.184	0.2	-0.034	2.63E-5
IL-8	0.367	0.184	0.2	0.070	2.63E-5
%neutrophil	0.272	0.137	0.147	0.045	1.99E-5
IL-16	0.243	0.121	0.12	0.007	1.79E-5
M-CSF	0.243	0.121	0.12	-0.013	1.79E-5
total cells (x 10 ⁶)	0.243	0.121	0.12	-0.017	1.79E-5
% mo0/Mac	0.190	0.096	0.122	-0.032	1.32E-5

Table 2 Classification results obtained using IL-8,IL-16 AND IL1-RA.

Method	AUC	CA	F1	Precision	Recall
SVM Learner	1	0.75	0.733	0.833	0.75
SGD	0.75	0.75	0.733	0.833	0.75
Random Forest Learner	0.375	0.5	0.333	0.25	0.5
Naive Bayes	0.25	0.5	0.333	0.25	0.5
Logistic Regression	0.75	0.75	0.733	0.833	0.75

Table 3 Classification results obtained using Only IL-8.

Method	AUC	CA	F1	Precision	Recall
SVM Learner	0.5	0.75	0.733	0.833	0.75
SGD	0.5	0.5	0.333	0.25	0.5
Random Forest Learner	0.75	0.5	0.333	0.25	0.5
Naive Bayes	0.75	0.75	0.733	0.833	0.75
Logistic Regression	1	0.5	0.333	0.25	0.5

Table 4 Classification results obtained using IL-8 and IL-16.

Method	AUC	CA	F1	Precision	Recall
SVM Learner	0.75	0.75	0.733	0.833	0.75
SGD	0.75	0.75	0.733	0.833	0.75
Random Forest Learner	0.5	0.5	0.333	0.25	0.5
Naive Bayes	0.625	0.75	0.733	0.833	0.75
Logistic Regression	0.75	0.5	0.5	0.5	0.5

Table 5 Classification results obtained using IL-8,IL-16, IL1-RA and neutrophil percentage.

Method	AUC	CA	F1	Precision	Recall
SVM Learner	1	0.75	0.733	0.833	0.75
SGD	0.75	0.75	0.733	0.833	0.75
Random Forest Learner	1	0.75	0.733	0.833	0.75
Naive Bayes	0.75	0.75	0.733	0.833	0.75
Logistic Regression	1	1	1	1	1

and uncontrolled asthmatics. This has an important implication as cytokine or anti-cytokine therapy presents an important alternative or adjunct to current therapies offered today.

REFERENCES

1. Koshak EA. Classification of asthma according to revised 2006 GINA: evolution from severity to control. *Annals of thoracic medicine*. 2007 Apr; **2**(2): 45, [DOI: 10.4103/1817-1737.32228]

2. Honkoop PJ, Simpson A, Bonini M, Snoeck-Stroband JB, Meah S, Chung KF, Usmani OS, Fowler S, Sont JK. MyAirCoach: the use of home-monitoring and mHealth systems to predict deterioration in asthma control and the occurrence of asthma exacerbations; study protocol of an observational study. *BMJ open*. 2017 Jan 1; **7**(1): e013935. [DOI: 10.1136/bmjopen-2016-013935]
3. Anastasiou A, Kocsis O, Moustakas K. Exploring machine learning for monitoring and predicting severe asthma exacerbations. *In Proceedings of the 10th Hellenic Conference on Artificial Intelligence* 2018 Jul 9 (p. 45). ACM. [DOI: 10.1145/3200947.3201036]
4. Sobieraj DM, Weeda ER, Nguyen E, Coleman CI, White CM, Lazarus SC, Blake KV, Lang JE, Baker WL. Association of inhaled corticosteroids and long-acting β -agonists as controller and quick relief therapy with exacerbations and symptom control in persistent asthma: a systematic review and meta-analysis. *Jama*. 2018 Apr 10; **319**(14): 1485-96. [DOI: 10.1001/jama.2018.2769]
5. Braido, Fulvio. "Failure in asthma control: reasons and consequences." *Scientifica* **2013** (2013). [DOI: 10.1155/2013/549252]
6. Nocker RE, Schoonbrood DF, van de Graaf EA, Hack E, Lutter R, Jansen HM, Out TA. Interleukin-8 in airway inflammation in patients with asthma and chronic obstructive pulmonary disease. *International archives of allergy and immunology*. 1996; **109**(2): 183-91. [DOI: 10.1159/000237218]
7. Afifi SS, ElArab AE, Mostafa SY. Interleukin 16 (IL-16) in asthma and allergic rhinitis. A comparison between upper and lower airways. *The Egyptian journal of immunology*. 2004; **11**(2): 31-6. [PMID: 16734115]
8. Gohlke H, Illig T, Bahnweg M, Klopp N, André E, Altmüller J, Herbon N, Werner M, Knapp M, Pescolliderung L, Boner A. Association of the interleukin-1 receptor antagonist gene with asthma. *American journal of respiratory and critical care medicine*. 2004 Jun 1; **169**(11): 1217-23. [DOI: 10.1164/rccm.200302-281OC]; [PMID: 15020290]
9. Ennis M. Neutrophils in asthma pathophysiology. *Current allergy and asthma reports*. 2003 Mar 1; **3**(2): 159-65. [DOI: 10.1007/s11882-003-0029-2]
10. Brasier AR, Victor S, Ju H, Busse WW, Curran-Everett D, Bleecker E, Castro M, Chung KF, Gaston B, Israel E, Wenzel SE. Predicting intermediate phenotypes in asthma using bronchoalveolar lavage-derived cytokines. *Clinical and translational science*. 2010 Aug 1; **3**(4): 147-57. [DOI: 10.1111/j.1752-8062.2010.00204.x]
11. Brasier AR, Ju H. Analysis and predictive modeling of asthma phenotypes. *In Heterogeneity in Asthma* 2014 (pp. 273-288). Humana Press, Boston, MA. [DOI: 10.1007/978-1-4614-8603-9_17]
12. Bhakta NR, Woodruff PG. Human asthma phenotypes: from the clinic, to cytokines, and back again. *Immunological reviews*. 2011 Jul 1; **242**(1): 220-32. [DOI: 10.1111/j.1600-065X.2011.01032.x]
13. Zhang J, Bai C. Elevated Serum Interleukin-8 Level as a Preferable Biomarker for Identifying Uncontrolled Asthma and Glucocorticosteroid Responsiveness. *Tanaffos*. 2017 Jun; **16**(4): 260. [PMID: 29849682]
14. Hosoki K, Ying S, Corrigan C, Qi H, Kurosky A, Jennings K, Sun Q, Boldogh I, Sur S. Analysis of a panel of 48 cytokines in BAL fluids specifically identifies IL-8 levels as the only cytokine that distinguishes controlled asthma from uncontrolled asthma, and correlates inversely with FEV1. *PloS one*. 2015 May 26; **10**(5): e0126035. [DOI: 10.1371/journal.pone.0126035]