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

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 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[3]. 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.
asthmatic bronchial wall was studied. There was an increased expres-
sition of interleukin-1 beta (IL-1beta) and IL-1ra, in the normal and
mucosa, in which mast prison cell, eosinophils and activated T cells
airway inflammation. Many cytokines are involved in the devel-
process for T cells in patient with asthma is mainly attributable to IL-
constriction, and bronchial hyperresponsiveness. The potential role
and airway secretions of affected role with asthma, and the effect of
particular inhibitors such as sense organ antagonists or specific anti-
Cytokine therapy has tremendous potential for treating a vari-
ity 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-
flammatory cytokines.

Asthma attack is a chronic inflammatory disease of bronchial
mucosa, in which mast cell prison, 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[16]. 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[16]. IL-8, IL-17, FGF, Eotaxin, GM-CSF, MIP-1a were
eliminated as there were no recordings for the same. The healthy sub-
jects were characterized as those who had no history of allergic dis-
case, 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 Informa-
tion 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 weight-
ing 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

Predicting severe exacerbation caused by uncontrolled asthma is
effort for patients, as avoiding critical symptoms which will need
special treatment or perhaps hospitalization will defend patients from
bronchoconstriction and its effects. Four Machine learning algorithms
were evaluated for their accuracy in predicting severe asthma at-
tack exacerbations by exploiting completely different parameters, of
which SVM and Naïve Bayesian approaches showed higher sensitiv-
ity, while Random Forests showed higher negative prognostic power
(very high specificity)[9].

An important problem in realizing personalized medicine is the de-
velopment of methods for identifying disease subtypes using quanti-
tative 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[9].

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) bronchohyperper-
sponsiveness (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 devel-
opment of the atopic state and of the chronic inflammatory process
of asthma, ultimately contributing to airway remodelling, broncho-
constriction, 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 var-
iety 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-
flammatory cytokines.

Asthma attack is a chronic inflammatory disease of bronchial
mucosa, in which mast cell prison, 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[16]. 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[16]. In patients with
asthma the strong correlation between the levels of IL-8 and the
percentage neutrophils was reported in[16][17]. 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[9], 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 expres-
sion of both IL-1beta and IL-1ra in the asthmatic bronchial epithe-
lium IL-1 receptor antagonist protein, an anti-inflammatory cytokine
that plays an important role in maintaining the balance between in-
flammatory and anti-inflammatory cytokines.

Machine Learning techniques play a significant role in the area
of predictive analytics as they aim to effectively predict disease out-
comes 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 pre-
dominant cytokines that contribute to the process of distinguishing
controlled asthmatics from uncontrolled asthmatics by employing
feature scoring technique, followed by a process of weighted averag-
ing, which tries to extract the best of the features. Further, we apply
binary classifiers to distinguish controlled and uncontrolled asthmat-
ic and deduce the best classifier that performs optimally in the pro-
cess.

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 inter-
national 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-1a were
eliminated as there were no recordings for the same. The healthy sub-
jects were characterized as those who had no history of allergic dis-
case, 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 Informa-
tion 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 weight-
ing 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 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

Pooja MR et al. Machine learning model to differentiate controlled and uncontrolled asthmatics


