Detecting asthma exacerbations using daily home monitoring and machine learning

Zhang, Olivier; Minku, Leandro L; Gonem, Sherif

DOI:
10.1080/02770903.2020.1802746

Document Version
Peer reviewed version

Citation for published version (Harvard):
Zhang, O, Minku, LL & Gonem, S 2020, 'Detecting asthma exacerbations using daily home monitoring and machine learning', *Journal of Asthma*. https://doi.org/10.1080/02770903.2020.1802746

Detecting asthma exacerbations using daily home monitoring and machine learning

[1]Olivier Zhang

[2]Leandro L. Minku*

[3,4]Sherif Gonem*

[1]INSA Rennes, Rennes, France

[2]School of Computer Science, University of Birmingham, UK

[3]Department of Respiratory Medicine, Nottingham City Hospital, Nottingham, UK

[4]Department of Respiratory Science, University of Leicester, Leicester, UK

*Joint senior authors

Address for correspondence:

Dr. Sherif Gonem

Department of Respiratory Medicine

Nottingham City Hospital

Hucknall Road

Nottingham

NG5 1PB, UK

Tel: +44 115 969 1169

E-mail: sg330@le.ac.uk

Funding: This paper presents independent research funded by the National Institute for Health Research (NIHR). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. SG was funded by an NIHR Clinical Lectureship.

Acknowledgement: The authors are grateful to AstraZeneca for providing the SAKURA dataset used in the study. AstraZeneca did not contribute to the data analysis or the decision to publish.

Abstract

Objective

Acute exacerbations contribute significantly to the morbidity of asthma. Recent studies have shown that early detection and treatment of asthma exacerbations leads to improved outcomes. We aimed to develop a machine learning algorithm to detect severe asthma exacerbations using easily available daily monitoring data.

Methods

We analysed daily peak expiratory flow and symptom scores recorded by participants in the SAKURA study (NCT00839800), an international multicentre randomised controlled trial comparing budesonide/formoterol as maintenance and reliever therapy versus budesonide/formoterol maintenance plus terbutaline as reliever, in adults with persistent asthma. The dataset consisted of 728,535 records of daily monitoring data in 2010 patients, with 576 severe exacerbation events. Data post-processing techniques included normalisation, standardisation, calculation of differences or slopes over time and the use of smoothing filters. Principal components analysis was used to reduce the large number of derived variables to a smaller number of linearly independent components. Logistic regression, decision tree, naïve Bayes, and perceptron algorithms were evaluated. Model accuracy was assessed using stratified cross-validation. The primary outcome was the detection of exacerbations on the same day or up to three days in the future.

Results

The best model used logistic regression with input variables derived from post-processed data using principal components analysis. This had an area under the receiver operating

characteristic curve of 0.85, with a sensitivity of 90% and specificity of 83% for severe asthma exacerbations.

Conclusion

Asthma exacerbations may be detected using machine learning algorithms applied to daily self-monitoring of peak expiratory flow and asthma symptoms.

Key words: asthma; exacerbation; peak expiratory flow; home monitoring; machine learning

Running title: Asthma exacerbations and home monitoring

Introduction

Acute exacerbations of asthma are episodes of deteriorating symptoms, often with concomitant reductions in lung function, requiring a change in treatment such as a short course of oral corticosteroids[1]. Acute exacerbations are an important cause of morbidity in patients with asthma, and can result in days off work or school, hospital admission, or even death. Preventing exacerbations is a key priority in the management of asthma[2]. Regular use of inhaled corticosteroids at an appropriate dose and with correct technique is the mainstay of preventative asthma treatment, but does not completely eliminate exacerbations[3].

The concept of detecting exacerbations at an early stage of development in order to intervene and avert them has recently gained ground. McKeever *et al* showed that a self-management plan which involved quadrupling the dose of inhaled corticosteroids at the first signs of an asthma exacerbation (increased symptoms and/or reduced peak expiratory flow [PEF]) reduced exacerbation rates compared to standard treatment[4]. The typical changes in peak expiratory flow and asthma symptom scores leading up to asthma exacerbations were initially described by Tattersfield *et al*[5]. These authors showed that PEF began to gradually fall approximately 10 days prior to an exacerbation, followed by a much steeper fall from 3 days prior to an exacerbation, culminating in a 15-20% fall from baseline on the day of exacerbation. Asthma symptom scores followed a very similar pattern, with a gradual rise starting from 10 days prior to an exacerbation, followed by a steeper rise from 3 days prior to an exacerbation. These results suggest that detecting asthma exacerbations up to three days in advance using daily monitoring of PEF and symptoms is potentially feasible. Since then, a number of researchers have investigated the sensitivity and specificity of algorithms based upon daily electronic monitoring of symptoms and PEF to detect impending asthma exacerbations[6,7]. These studies used fairly simple statistical cut-offs for PEF and symptom

scores to detect exacerbation events, and moreover the datasets used were relatively small, thus precluding more complex analyses such as examining temporal trends.

Machine learning is a branch of artificial intelligence in which statistical models are used to learn patterns from data in order to accomplish a specific task. Applications of machine learning within respiratory and other branches of medicine have grown significantly during the past five years[8]. The most common applications are those in which cases are classified into a small number of categories such as 'low-risk' and 'high-risk'. Although machine learning models have the potential to be more accurate than simpler predictive tools, their complexity means that they require large training datasets of labelled cases for their development.

The use of machine learning techniques to predict asthma exacerbations based on daily PEF and symptom monitoring has been investigated in one previous study by Finkelstein *et al*[9]. These authors utilised a moderately sized dataset of 7001 records submitted by adults with asthma using home telemonitoring software. They investigated the predictive value of three machine learning algorithms, namely naïve Bayesian classifier, adaptive Bayesian network, and support vector machine. However, it should be noted that exacerbations in this study were not defined as clinician-diagnosed events requiring treatment, but were instead based on 'alert levels' defined using the home telemonitoring data itself.

We hypothesised that a predictive algorithm derived using machine learning techniques in conjunction with a large training dataset of daily monitoring data would provide superior accuracy for detecting asthma exacerbations compared to previously published models.

Methods

Study dataset

We utilised a large dataset of daily PEF and symptom scores which were recorded by participants in the SAKURA study (NCT00839800), an international multicentre randomised controlled trial comparing budesonide/formoterol as maintenance and reliever therapy versus budesonide/formoterol maintenance plus terbutaline as reliever, in patients age ≥ 16 years with persistent asthma[10]. Eligibility criteria included a documented history of persistent asthma for at least 6 months, reversible airway obstruction (increase in forced expiratory volume in one second [FEV_1] of at least 12% relative to baseline with administration of a bronchodilator), use of maintenance inhaled corticosteroids (ICS) for at least 3 months before study entry, and having at least one asthma exacerbation in the 12 months prior to study entry. Current or previous smokers with a smoking history of ≥ 10 pack years were excluded. The study population had a mean age of 46 years with 68% being female. The mean beclometasone dipropionate equivalent ICS dose at study entry was 1023 µg/day, and 62% of patients were using long-acting β_2 agonists at study entry. The mean baseline FEV_1 was 70% predicted, with mean reversibility following administration of a bronchodilator of 23%.

Participants in this study kept a paper diary in which they recorded on a daily basis:

i) PEF twice daily (best of three blows each time)

ii) Morning symptoms on an integer scale from 0 (no symptoms) to 3 (severe symptoms)

iii) Evening symptoms on an integer scale from 0 (no symptoms) to 3 (severe symptoms)

iv) Number of puffs of reliever inhaler taken overnight

v) Number of puffs of reliever inhaler taken during the day

vi) Whether or not they had woken up due to asthma during the previous night

These data were entered into an electronic database together with a record of days in which a severe asthma exacerbation occurred. Severe exacerbations were defined as deterioration in asthma leading to oral corticosteroid treatment for at least 3 days, or hospitalisation or emergency room treatment due to asthma. Access to the dataset was provided to the investigators by AstraZeneca using a secure online data repository and analysis platform. Participants in the study gave informed consent for the secondary use of anonymised study data for research.

Data analysis

The dataset consisted of 728,535 records of daily monitoring data in 2010 patients, with a total of 576 severe exacerbation events. The mean length of follow-up for each patient was 362 days. The primary goal of the analysis was to derive and validate a predictive model which could detect exacerbation events occurring on the same day or up to three days in the future. The analysis consisted of a number of steps as described in the text below and summarised in Figure 1. At each stage of the analysis a number of options were available, each of which was systematically investigated. Once the most favourable option had been selected this was then used for the remainder of the analysis until the final model was reached. This process is described in the results section. Further details of the analysis techniques are given in the Online Supplement.

Processing of daily monitoring variables

The nine basic daily monitoring variables entered into the predictive models were:

i) Morning, evening and mean peak expiratory flow rate.

ii) Morning and evening symptom scores

iii) Number of puffs of reliever inhaler used during the overnight and daytime periods

iv) Total of morning and evening symptom scores, and overnight and daytime reliever inhaler usage

v) Waking during the previous night (yes/no)

We utilised a number of variable post-processing techniques, alone or in combination, resulting in a total of 432 basic and derived variables:

i) Normalisation of variables as a percentage of the mean value for that patient (normalisation), or as the number of standard deviations above or below the mean for that patient (standardisation)[13]. The rationale for this is that some parameters (particularly PEF) are heavily dependent on demographic characteristics such as age, sex and height. Therefore it is logical to standardise values according to the mean value for each individual, thus accentuating within-person rather than between-person variability.

ii) Calculating the difference or the slope between the current value and the value observed 1, 2, 3, 4 or 5 days ago, as an indication of the short-term trend. We chose to explore this method since previous evidence has shown that exacerbations are often preceded by short-term reductions in PEF and increases in symptom scores[5].

iii) Applying filters in order to smooth short-term variability[14-16]. These were used since a number of home monitoring measurements (particularly PEF) exhibit a degree of random variability which may mask the underlying trend. Figure 2 shows an example of PEF data before and after application of a smoothing filter.

Variable selection and reduction

As the total number of basic and derived variables (432) is very large and it is unclear which of them are most predictive of exacerbations, both recursive feature elimination and principal component analysis (PCA)[17] were investigated as variable selection and reduction techniques. Recursive feature elimination is a variable selection method which is used in combination with a particular machine learning model and with cross-validation. Starting with the full list of 432 variables, the weakest (least predictive) variables are eliminated from the model one by one until the optimal sensitivity is reached. PCA is a data reduction method that is used to reduce a large number of variables into a smaller number of linearly independent (uncorrelated) components, each of which is a weighted linear combination of one or more of the original variables. The purpose of PCA is to capture the variance or information content of a dataset with many variables using a smaller number of components, which can then be entered into predictive models. Since the components derived using PCA are linearly independent (uncorrelated) they are more likely to have independent value when entered into a predictive model. The numbers of components can be specified *a priori*. We investigated PCA using 3, 5, 9, 20, 40, 60, 80 and 100 components. It should be noted that PCA is a standalone procedure which occurs prior to entering data into a machine learning model, whereas recursive feature elimination is integrated into the process of tuning and testing a machine learning model.

Application of class imbalance learning techniques

Predicting asthma exacerbations from this dataset was a class imbalanced learning problem[11], in which there were much fewer examples of exacerbation cases (approximately 0.08%) than non-exacerbation cases (approximately 99.92%) in the dataset. Therefore, we investigated class imbalance learning techniques that operate by resampling the training data. These

techniques increase the proportion of the training set that represents the minority (exacerbation) class, aiming at producing models that are able to better recognise cases of the minority class. Importantly, it should be noted that these techniques were only applied to the training data, not the validation data from which the final model accuracy was determined. The following three techniques were investigated:

i) Random under-sampling: Randomly discarding training data from the majority (non-exacerbation) class.

ii) Random over-sampling: Randomly duplicating training data from the minority (exacerbation) class.

iii) Synthetic minority over-sampling technique (SMOTE): Adding synthetic training data that have been generated from the minority (exacerbation) class[11,12].

For each of these techniques we investigated different ratios of exacerbation to non-exacerbation training data to determine which produced the best balance between sensitivity and specificity.

Development and validation of machine learning models

We investigated a number of machine learning models:

i) Logistic regression: Statistical model in which the log odds of an event are assumed to be linearly related to one or more predictor variables.

ii) Naïve Bayes: Conditional probability model in which the probability of an event is assumed to be related independently to one or more predictor variables.

iii) Decision tree: Classification algorithm which assigns a category in a hierarchical manner based upon decision points with respect to the predictor variables.

iv) Perceptron: Classification algorithm which assigns a category based upon whether a weighted combination of the predictor variables exceeds a particular threshold.

The ability of the machine learning models to recognise exacerbation and non-exacerbation cases was evaluated using sensitivity, specificity, and area under the receiver operating characteristic curve (AUC). Sensitivity was defined as the true positive rate. We considered a prediction to be a true positive if an exacerbation occurred on the same day or up to 3 days after the prediction. Specificity was defined as the true negative rate. AUC was the area under the curve formed by true positive and false positive rates obtained by varying the decision thresholds within the machine learning models. We used stratified cross-validation[18] to evaluate each of the machine learning models. This procedure was chosen due to the small number of exacerbation examples in the data set. It separates the data into *k* folds. *k*-1 folds are used to train a predictive model, and the remaining fold is used for evaluation purposes. In this study we used k=5 folds and for most analyses repeated the procedure 10 times. The average sensitivity, specificity and AUC was calculated across the 10 repetitions (if applicable).

Results

Qualitative examination of the dataset revealed that when all 576 exacerbation events were taken in aggregate, each of the raw daily monitoring variables displayed a distinct pattern in the run-up to exacerbation events, as shown in Figure 3. However, there was a great deal of individual variability, meaning that none of these variables alone was sufficient to predict asthma exacerbations with high sensitivity or specificity.

Developing a predictive model of asthma exacerbations presented a number of options at each step such as the choice of variable processing techniques, variable selection method,

class imbalance learning technique and machine learning model. These choices were made sequentially until the final model was reached, as described below.

i) Exploratory analysis of variable processing methods

We initially investigated the effect of different variable processing techniques on the predictive ability of home monitoring variables. For this part of the study, three basic variables were used (mean PEF, total symptom score and night-time waking), entered into a logistic regression model with the use of SMOTE to address class imbalance. Results were assessed using 5-fold cross-validation repeated 10 times. Table S1 in the Online Supplement shows the predictive performance of the basic variables compared to when smoothing filters were applied. When applied alone the smoothing filters did not confer an advantage compared to the basic variables. The best performing filter was Savitzky-Golay with window width s=3 and polynomial order d=2, so this was retained for subsequent analyses.

Table S2 shows the performance of standardised and normalised variables, with and without the additional use of the Savitzky-Golay smoothing filter. Standardised variables are expressed as the number of standard deviations above or below the mean value for that patient, while normalised variables are expressed as the percentage of the mean value for that patient. Standardisation improved the predictions compared to the basic variables whereas normalisation worsened them. Therefore, only standardisation was retained as a variable processing method for subsequent analyses.

Table S3 shows the performance of differenced measurements and slope (over 1 to 5 days), with and without the additional use of standardisation and Savitzky-Golay filter. For each of these tests, a total of 15 processed variables were entered into the model, since the difference

or slope was calculated over a period of 1, 2, 3, 4 or 5 days for each of the three basic variables. It was observed that differenced values were moderately sensitive and specific, whereas slopes were more sensitive but rather non-specific. Both variable processing methods were retained for future analyses.

ii) Comparison of machine learning algorithms

In light of the exploratory analysis described in the previous section, a list of predictor variables was chosen in order to test the four machine learning models (logistic regression, naïve Bayes, decision tree and perceptron). These were the three basic variables used in the previous section (mean PEF, total symptom score and night-time waking) smoothed using the Savitzky-Golay filter, with standardisation, or with differencing (over 1, 2, 3, 4 or 5 days), or with calculation of the slope (over 1, 2, 3, 4 or 5 days). These analyses were performed using SMOTE to address class imbalance, and a grid search to tune parameters based on one run of 5-fold stratified cross-validation for each combination of parameter values investigated. Table S4 shows the parameter values investigated and the performance obtained by each of the machine learning models using these input data. Logistic regression gave the best balance between sensitivity and specificity and was therefore used in subsequent analyses.

iii) Comparison of class imbalance learning techniques

We found that using over-sampling, under-sampling or SMOTE was essential to overcome the class imbalance problem and enable the logistic regression algorithm to recognise exacerbation cases. Table S5 shows the results obtained by logistic regression using no resampling and using different class imbalance learning techniques. These analyses were performed with 5-fold cross-validation repeated 10 times. The most balanced results in terms of sensitivity and specificity were provided with a 1:1 ratio of exacerbation and non-

exacerbation cases. The three class imbalance learning techniques performed equally with respect to sensitivity and specificity when using a 1:1 ratio of exacerbation and non-exacerbation cases. For subsequent analyses under-sampling was used since this was the simplest and least computationally intensive option.

iv) Comparison of variable selection and data reduction techniques

In order to develop and validate the final predictive model, the variable processing techniques detailed in section (i) above were applied alone or in combination to the full list of nine raw monitoring variables to produce a total of 432 raw and processed variables. The final model used logistic regression as the machine learning model with under-sampling as the class imbalance technique. Recursive feature elimination and PCA were applied as described in the Methods section, with the results shown in Table S6. These analyses were performed with 5-fold cross-validation repeated 10 times. PCA with the number of components (c) = 80 achieved the best overall results, with sensitivity of 90% and specificity of 83% for asthma exacerbations, and an AUC of 85%, as shown in Figure 4.

Discussion

We have shown that machine learning techniques in combination with simple daily monitoring data such as PEF and patient-reported symptom scores can predict asthma exacerbations with good sensitivity and specificity. In particular our best algorithm, using logistic regression in combination with PCA for feature extraction, achieved sensitivity of 90% and specificity of 83% for asthma exacerbations, with an AUC of 85%. This was achieved using class imbalance techniques to better balance the positive and negative training data, enabling the resulting models to better recognise minority cases. This was necessary due

to the severe class imbalance in the original dataset (0.08% exacerbation cases, 99.92% non-exacerbation cases). Without using class imbalance learning techniques, most statistical and machine learning models would simply predict all cases as being in the majority class (ie. non-exacerbation cases) since this would yield an accuracy of 99.92% - however, such a model would clearly not have clinical utility. Therefore, class imbalance learning techniques are essential to develop predictive models that give meaningful results. It should be noted that class imbalance techniques were only used to balance cases in the training data, not the validation data. Therefore the sensitivity, specificity and AUC values we have reported are applicable to prospectively collected home monitoring data. Using predictive models in clinical practice requires consideration of additional factors such as the relative 'cost' of false negative and false positive results. For instance it may be decided that a given number of false alarms will be tolerated to correctly diagnose one exacerbation event. This will determine the threshold of the model output that is chosen to initiate further action such as contact with a health professional.

The primary outcome of this study was the accuracy of predicting asthma exacerbations occurring on the same day or up to 3 days in the future. This was chosen based on previous work by Tattersfield *et al* showing that significant changes in PEF and asthma symptoms start to occur 3 days prior to exacerbations[5]. Given that the anti-inflammatory actions of inhaled and oral corticosteroids commence within 2-3 days and 3-8 hours of administration respectively[19,20], intervention within this timeframe would be expected to have a favourable effect, potentially averting incipient exacerbations before they become severe. McKeever *et al* showed that a strategy of quadrupling inhaled corticosteroid dose in response to a drop in PEF or increase in asthma symptoms had the effect of reducing asthma exacerbations[4]. It is

likely that improving the algorithm for early detection of exacerbations would further enhance the efficacy of this strategy.

A number of smartphone apps already exist for daily monitoring and self-management of asthma[21]. The algorithms we have developed could be readily incorporated into a smartphone app, providing patients and clinicians with an early warning of impending exacerbations. Although development of machine learning models is often computationally intensive due to the need to tune the model to a large training dataset, applying the final model to new data is usually much less so. The final model we generated uses relatively simple manipulations of data which would be well within the capacity of modern smartphones.

Further treatment or management studies are needed to determine the best response to algorithm-generated early warnings, with the goal of reducing severe exacerbations, use of oral corticosteroids and hospital admissions. Potential options include contact with a healthcare professional or a patient-initiated increase in therapy such as a quadrupling of inhaled corticosteroid dose[4]. Moreover, the health economic benefits of such an approach require evaluation, given the potential for false alarms and unnecessary healthcare contacts.

Strengths of our study include the use of a large international dataset, incorporating 728,535 patient-days of data in 2010 patients, with a total of 576 severe exacerbation events. We investigated a wide variety of machine learning techniques in order to optimise the potential of this dataset. However we acknowledge a number of potential limitations of the study. This was a post hoc analysis of daily diary data that were collected as part of a randomised controlled trial, and were not originally intended to be used for exacerbation prediction. The data were collected using paper diaries which may have been prone to inaccurate transcribing

or fabrication. Moreover, there was no way of verifying correct technique with home peak expiratory flow measurements. Electronic real-time data collection using a smartphone app wirelessly linked to a digital spirometer with in-built quality control would have provided more reliable data. It is also possible that the simple three-point symptom scores utilised in this study were not maximally predictive. Validated daily outcome measures such as the Asthma Control Diary[22] and the Asthma Daily Symptom Diary[23] are available, but these instruments are subject to licencing restrictions which prevent their free use on electronic platforms. Reliever inhaler usage was self-reported in our study whereas there is now the potential to objectively monitor this using digital inhaler attachments[24-27]. There is emerging evidence that monitoring reliever inhaler usage in real time may provide important predictive information. Objectively monitored reliever inhaler use has been shown to increase in the days leading up to asthma exacerbations[26] and hospital admissions[27]. It is possible that daily monitoring of additional variables such as exhaled nitric oxide would also improve the predictive power of home monitoring, albeit with the drawback of increasing cost and complexity. Exhaled nitric oxide is a biomarker of steroid-responsive airway inflammation which can be measured in a variety of settings[28]. A recent systematic review and meta-analysis has shown that tailoring asthma treatment based on exhaled nitric oxide measurements can reduce exacerbations in both adults and children with asthma[29]. Home monitoring of exhaled nitric oxide using portable devices has been shown to be feasible by a number of investigators[30-34], and Van der Walk *et al* observed increases in exhaled nitric oxide in the days leading up to moderate exacerbations in children with asthma[32].

In conclusion, we have shown that machine learning algorithms have the potential to improve the early detection of asthma exacerbations when compared to traditional paper-based action plans. We anticipate that electronic data collection using smartphone apps linked to digital

spirometers and inhalers will further improve the predictive ability of these algorithms. Further studies are needed to assess whether this can translate into improved clinical outcomes, and whether asthma self-management using predictive algorithms is cost-effective.

References

1) Reddel HK, Taylor DR, Bateman ED, Boulet LP, Boushey HA, Busse WW, Casale TB, Chanez P, Enright PL, Gibson PG, de Jongste JC, Kerstjens HA, Lazarus SC, Levy ML, O'Byrne PM, Partridge MR, Pavord ID, Sears MR, Sterk PJ, Stoloff SW, Sullivan SD, Szefler SJ, Thomas MD, Wenzel SE; American Thoracic Society/European Respiratory Society Task Force on Asthma Control and Exacerbations. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. *Am J Respir Crit Care Med*. 2009; 180(1): 59-99.

2) Global Initiative for Asthma, 2020. Global strategy for asthma management and prevention. Available from: https://ginasthma.org. Accessed 1st June 2020.

3) Desai D, Siddiqui S, Brightling C. Can inhaled corticosteroids prevent asthma exacerbations? *Curr Opin Pulm Med*. 2011; 17(1): 16-22.

4) McKeever T, Mortimer K, Wilson A, Walker S, Brightling C, Skeggs A, Pavord I, Price D, Duley L, Thomas M, Bradshaw L, Higgins B, Haydock R, Mitchell E, Devereux G, Harrison T. Quadrupling Inhaled Glucocorticoid Dose to Abort Asthma Exacerbations. *N Engl J Med*. 2018; 378(10): 902-910.

5) Tattersfield AE, Postma DS, Barnes PJ, Svensson K, Bauer CA, O'Byrne PM, Löfdahl CG, Pauwels RA, Ullman A. Exacerbations of asthma: a descriptive study of 425 severe exacerbations. The FACET International Study Group. *Am J Respir Crit Care Med*. 1999; 160(2): 594-9.

6) Kupczyk M, Haque S, Sterk PJ, Niżankowska-Mogilnicka E, Papi A, Bel EH, Chanez P, Dahlén B, Gaga M, Gjomarkaj M, Howarth PH, Johnston SL, Joos GF, Kanniess F, Tzortzaki E, James A, Middelveld RJ, Dahlén SE; BIOAIR investigators. Detection of exacerbations in asthma based on electronic diary data: results from the 1-year prospective BIOAIR study. *Thorax*. 2013; 68(7): 611-8.

7) Honkoop PJ, Taylor DR, Smith AD, Snoeck-Stroband JB, Sont JK. Early detection of asthma exacerbations by using action points in self-management plans. *Eur Respir J*. 2013; 41(1): 53-9.

8) Gonem S, Janssens W, Das N, Topalovic M. Applications of artificial intelligence and machine learning in respiratory medicine. *Thorax*. 2020 May 14. [Epub ahead of print]

9) Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann NY Acad Sci*. 2017; 1387(1): 153-65.

10) Atienza T, Aquino T, Fernández M, Boonsawat W, Kawai M, Kudo T, Ekelund J, Ivanov S, Carlsson LG. Budesonide/formoterol maintenance and reliever therapy via Turbuhaler versus fixed-dose budesonide/formoterol plus terbutaline in patients with asthma: phase III study results. *Respirology*. 2013; 18(2): 354-63.

11) He H, Garcia EA. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*. 2009; 21(9): 1263-84.

12) Wang S, Minku LL, Yao X. Resampling-Based Ensemble Methods for Online Class Imbalance Learning. *IEEE Transactions on Knowledge and Data Engineering*. 2015; 27(5): 1356-68.

13) http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.scale.html

14) https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.signal.medfilt.html

15) https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.savgol_filter.html

16) https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.wiener.html#scipy.signal.wiener

17) http://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

18) Bishop, C. Pattern Recognition and Machine Learning. *Springer*: New York, 2006.

19) Anderson WJ, Short PM, Williamson PA, Lipworth BJ. Inhaled Corticosteroid Dose Response Using Domiciliary Exhaled Nitric Oxide in Persistent Asthma: The FENOtype Trial. *Chest*. 2012; 142(6): 1553-1561.

20) Williams DM. Clinical Pharmacology of Corticosteroids. *Resp Care.* 2018; 63(6): 655-670

21) Kagen S, Garland A. Asthma and Allergy Mobile Apps in 2018. *Curr Allergy Asthma Rep*. 2019; 19(1): 6.

22) Juniper EF, O'Byrne PM, Ferrie PJ, King DR, Roberts JN. Measuring asthma control. Clinic questionnaire or daily diary? *Am J Respir Crit Care Med*. 2000; 162(4 Pt 1): 1330-4.

23) Gater A, Nelsen L, Fleming S, Lundy JJ, Bonner N, Hall R, Marshall C, Staunton H, Krishnan JA, Stoloff S, Schatz M, Haughney J; Patient-Reported Outcome Consortium's Asthma Working Group. Assessing Asthma Symptoms in Adolescents and Adults: Qualitative Research Supporting Development of the Asthma Daily Symptom Diary. *Value Health*. 2016; 19(4): 440-50.

24) Patel M, Pilcher J, Munro C, Hosking A, Pritchard A, Shaw D, Black P, Weatherall M, Beasley R; SMART Study Group. Short-acting β-agonist use as a marker of current asthma control. *J Allergy Clin Immunol Pract*. 2013; 1(4): 370-7.

25) Patel M, Pilcher J, Reddel HK, Pritchard A, Corin A, Helm C, Tofield C, Shaw D, Black P, Weatherall M, Beasley R; SMART Study Group. Metrics of salbutamol use as predictors of future adverse outcomes in asthma. *Clin Exp Allergy*. 2013; 43(10): 1144-51.

26) Pilcher J, Patel M, Pritchard A, Thayabaran D, Ebmeier S, Shaw D, Black P, Braithwaite I, Weatherall M, Beasley R. Beta-agonist overuse and delay in obtaining

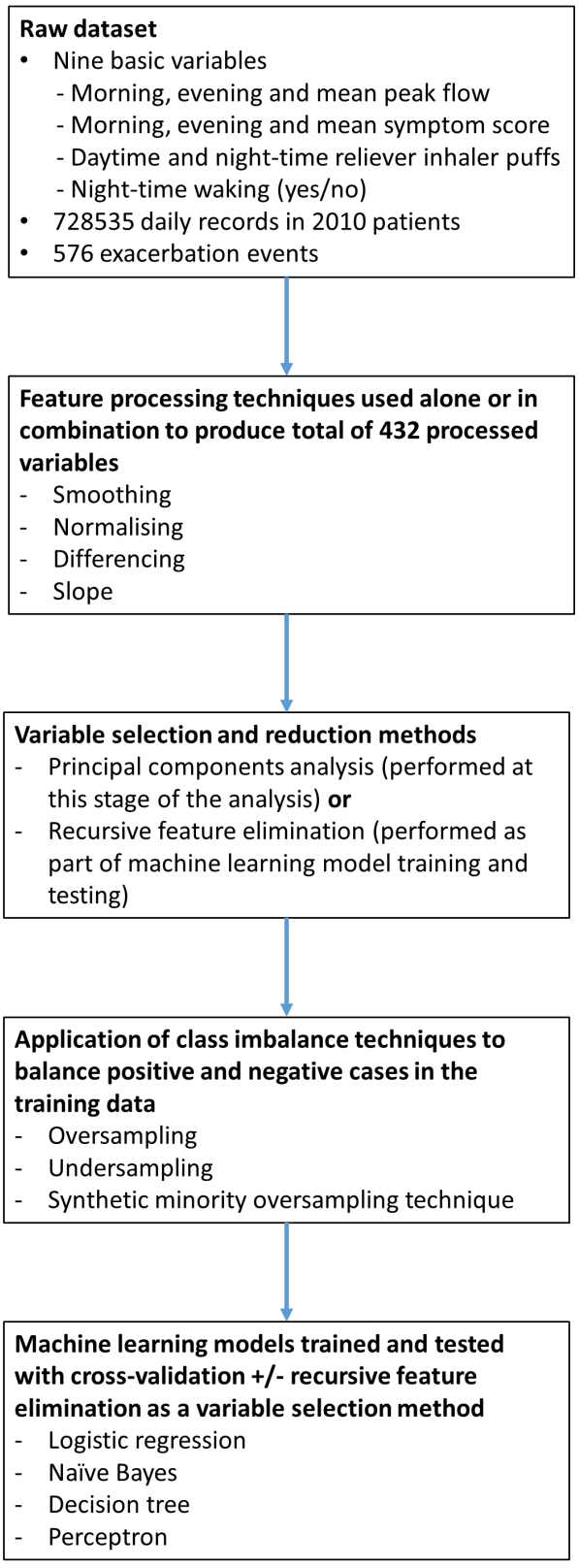
573 medical review in high risk asthma: a secondary analysis of data from a randomised
574 controlled trial. *NPJ Prim Care Respir Med*. 2017; 27: 33.
575
576 27) Patel M, Pilcher J, Hancox RJ, Sheahan D, Pritchard A, Braithwaite I, Shaw D, Black
577 P, Weatherall M, Beasley R, SMART Study Group. The Use of β2-agonist Therapy
578 Before Hospital Attendance for Severe Asthma Exacerbations: A Post-Hoc Analysis.
579 *NPJ Prim Care Respir Med*. 2015; 25: 14099.
580
581 28) Menzies-Gow A, Mansur AH, Brightling CE. Clinical utility of fractional exhaled
582 nitric oxide in severe asthma management. *Eur Respir J*. 2020; 55(3): 1901633.
583
584 29) Petsky HL, Cates CJ, Kew KM, Chang AB. Tailoring Asthma Treatment on
585 Eosinophilic Markers (Exhaled Nitric Oxide or Sputum Eosinophils): A Systematic
586 Review and Meta-Analysis. *Thorax*. 2018; 73(12): 1110-1119.
587
588 30) Pijnenburg MW, Floor SE, Hop WC, De Jongste JC. Daily ambulatory exhaled nitric
589 oxide measurements in asthma. *Pediatr Allergy Immunol*. 2006; 17(3): 189-93.
590
591 31) Hashimoto S, Ten Brinke A, Roldaan AC, van Veen IH, Möller GM, Sont JK,
592 Weersink EJM, van der Zee JS, Braunstahl GJ, Zwinderman AH, Sterk PJ, Bel EH.
593 Internet-based Tapering of Oral Corticosteroids in Severe Asthma: A Pragmatic
594 Randomised Controlled Trial. *Thorax*. 2011; 66(6): 514-20.
595

32) van der Valk RJP, Baraldi E, Stern G, Frey U, de Jongste JC. Daily exhaled nitric oxide measurements and asthma exacerbations in children. *Allergy*. 2012; 67(2): 265-71.

33) Saito J, Gibeon D, Macedo P, Menzies-Gow A, Bhavsar PK, Chung KF. Domiciliary diurnal variation of exhaled nitric oxide fraction for asthma control. *Eur Respir J*. 2014; 43(2): 474-84.

34) Nanda CR, Singapuri A, Soares M, Monteiro W, Siddiqui S, Gonem S. Domiciliary exhaled nitric oxide and eosinophilic airway inflammation in adults with asthma. *Eur Respir J*. 2016; 48(1): 242-4.

621 **Figure 1: Summary of data analysis steps**



622

Figure 2: Application of a data smoothing filter

Daily peak expiratory flow data (L/min) is shown before and after application of a Savitzky-Golay filter.

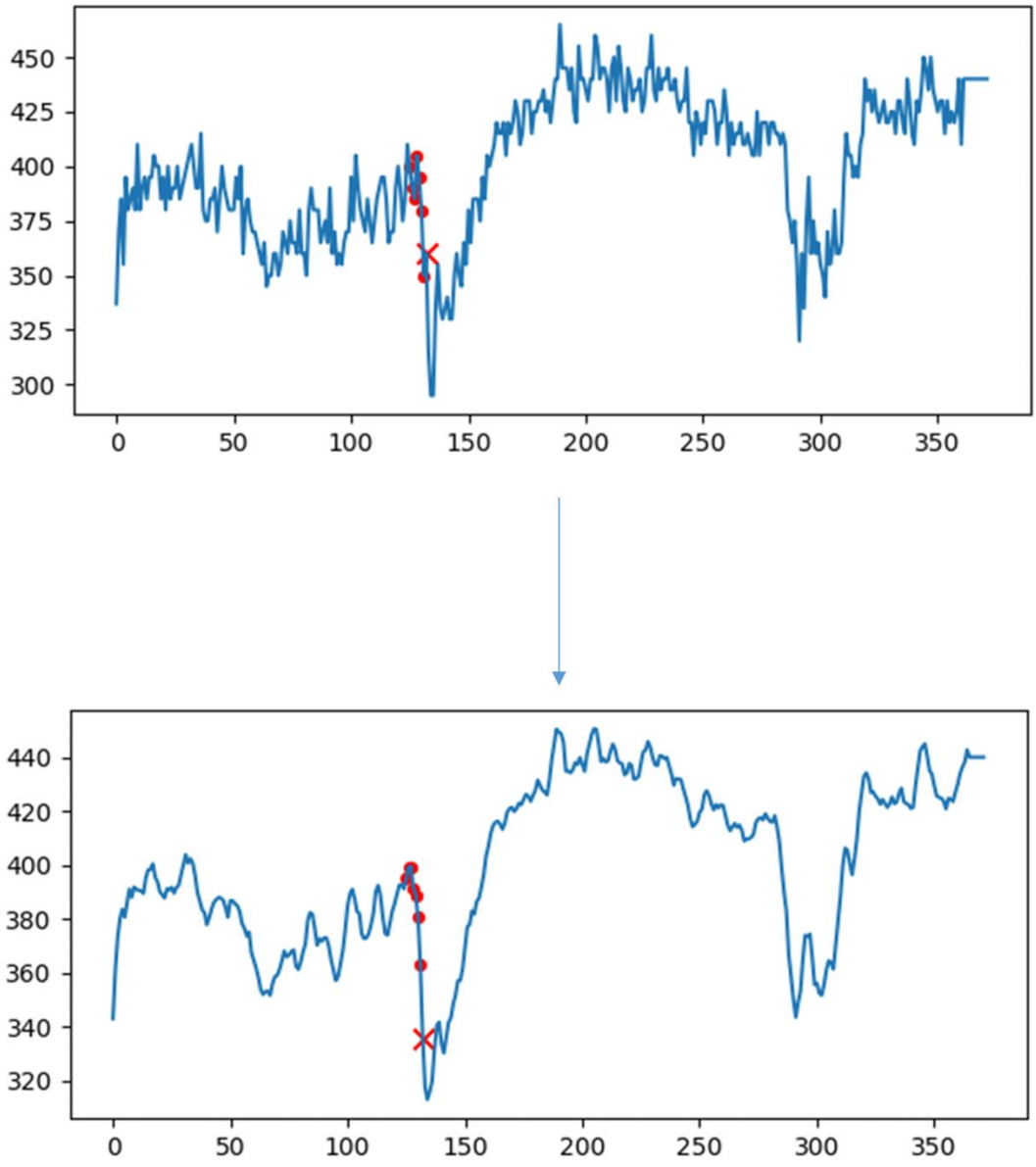


Figure 3: Changes in daily monitoring variables in the period preceding and following exacerbations

Panels show the average value of daily monitoring variables immediately preceding and following exacerbation events occurring on Day 0. PEF = peak expiratory flow (L/min).

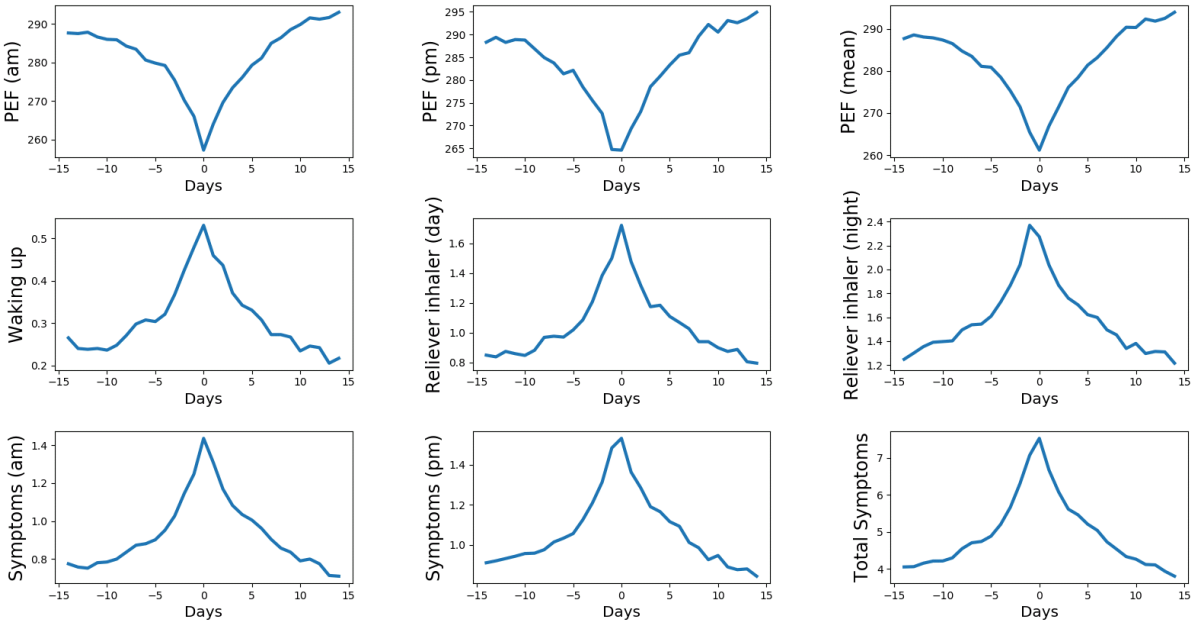
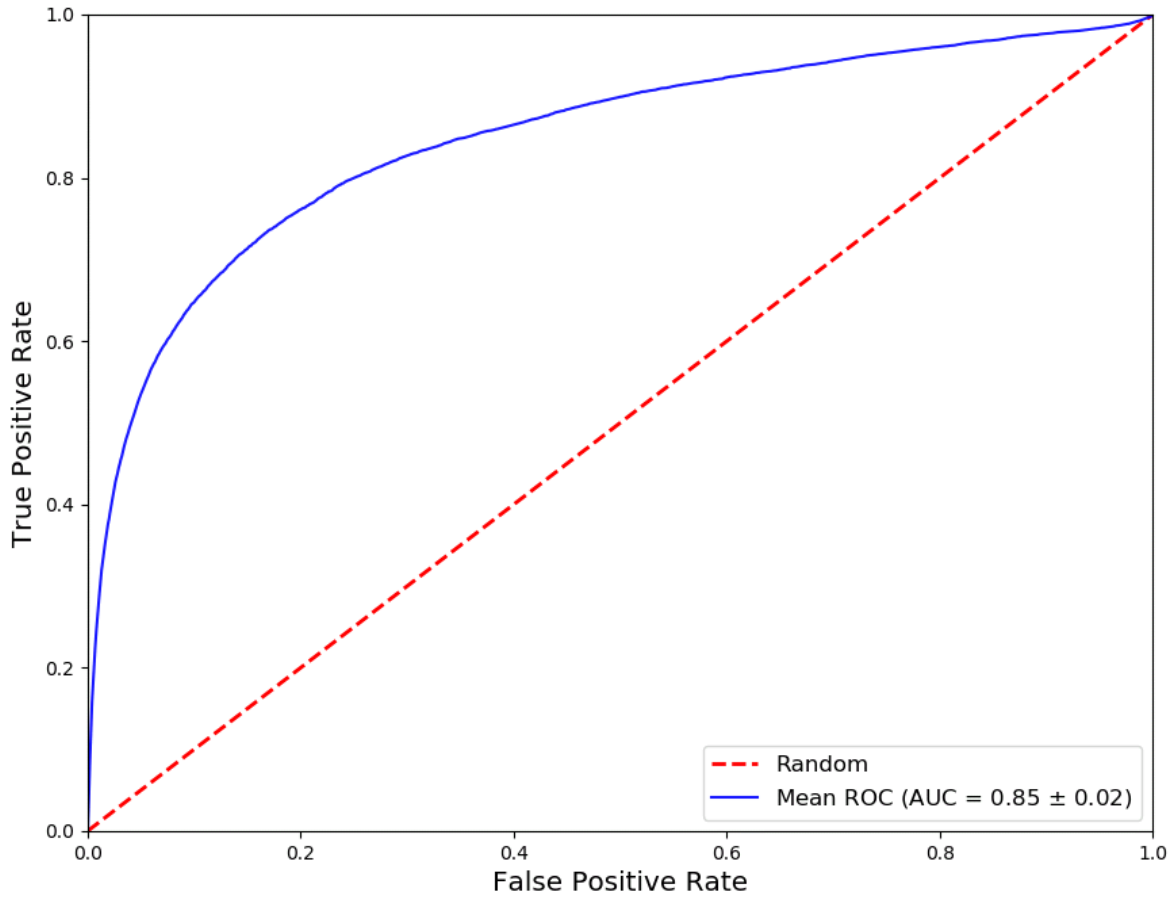


Figure 4: Receiver operating characteristic curve for the detection of asthma exacerbation using the final logistic regression model



Detecting asthma exacerbations using daily home monitoring and machine learning

Supplementary methods and data

Further details of post-processing techniques

Filter techniques:

- Median Filter – this filter keeps a sliding window over the data produced over time and uses the median value in this sliding window as a variable.
- Savitzky-Golay Filter – this filter fits a low order polynomial to the examples in a sliding window of the data based on linear least squares.
- Wiener Filter – this filter uses Wiener deconvolution to smooth signals based on a sliding window of the data.
- Median Filter, Savitzky-Golay Filter and Wiener Filter's parameter *s* refers to the sliding window used for smoothing the data.
- Savitzky-Golay Filter's parameter *d* is the polynomial order used by the filter.

Difference and slope techniques:

- Difference refers to the difference between the current value of a monitored variable and its value *d* days ago.
- Slope refers to the slope of the regression line that fits a sequence of values of the variables. The sequence includes the current value and all values up to and including *d* days ago.

Principal Component Analysis (PCA):

- PCA is a variable transformation technique that converts a set of values from possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. PCA's parameter *c* refers to the number of principal components to be used.

Table S1: Predictive performance obtained using different smoothing filters

Post-processing technique	Sensitivity	Specificity	AUC
Basic variables	80	78	82
Median Filter $s = 2$	79	77	81
Median Filter $s = 5$	77	76	80
Median Filter $s = 7$	75	75	79
Median Filter $s = 9$	73	74	77
Savitzky-Golay Filter $s = 3$, $d = 2$	80	78	82
Savitzky-Golay Filter $s = 5$, $d = 2$	79	78	82
Savitzky-Golay Filter $s = 5$, $d = 3$	79	78	82
Savitzky-Golay Filter $s = 7$, $d = 2$	78	78	82
Savitzky-Golay Filter $s = 7$, $d = 3$	78	78	82
Savitzky-Golay Filter $s = 9$, $d = 2$	78	77	82
Savitzky-Golay Filter $s = 9$, $d = 3$	78	77	82
Wiener Filter $s = 2$	79	78	82
Wiener Filter $s = 5$	78	78	82
Wiener Filter $s = 7$	78	78	82
Wiener Filter $s = 9$	78	78	82

AUC = area under the receiver operating characteristic curve.

Table S2: Predictive performance obtained using standardisation and normalisation +/- smoothing filter

Post-processing technique	Sensitivity	Specificity	AUC
Basic variables	80	78	82
Standardisation applied to basic variables	87	84	83
Standardisation applied to variables with Savitzky-Golay Filter $s = 3$, $d = 2$	87	84	83
Normalisation applied to basic variables	75	89	83
Normalisation applied to variables with Savitzky-Golay Filter $s = 3$, $d = 2$	75	89	83

AUC = area under the receiver operating characteristic curve.

Table S3: Predictive performance obtained using difference and slope +/- standardisation +/- smoothing filter

Post-processing technique	Sensitivity	Specificity	AUC
Basic variables	80	78	82
Difference d = {1,2,3,4,5} applied to basic variables	84	84	72
Difference d = {1,2,3,4,5} applied to variables with Savitzky-Golay Filter s = 3, d = 2	84	84	72
Difference d = {1,2,3,4,5} applied to variables with Standardisation	84	84	72
Difference d = {1,2,3,4,5} applied to variables with Savitzky-Golay Filter s = 3, d = 2 and Standardisation	84	84	72
Slope, d = {1,2,3,4,5} applied to basic variables	91	72	54
Slope d = {1,2,3,4,5} applied to variables with Savitzky-Golay Filter s = 3, d = 2	91	72	54
Slope d = {1,2,3,4,5} applied to variables with Standardisation	92	68	54
Slope d = {1,2,3,4,5} applied to variables with Savitzky-Golay Filter s = 3, d = 2 and Standardisation	92	68	54
Slope d = {1,2,3,4,5} applied to variables with Savitzky-Golay Filter s = 3, d = 2, Standardisation and Difference d = {1,2,3,4,5}	92	75	63

AUC = area under the receiver operating characteristic curve.

Table S4: Comparison of machine learning model performance

Machine learning model	Sensitivity	Specificity	AUC
Decision tree	8	100	52
Naïve Bayes	80	84	82
Perceptron	96	69	-
Logistic regression	86	86	84

AUC = area under the receiver operating characteristic curve.

These values have been obtained after a grid search to tune the parameter values below, based on one run of 5-fold stratified cross validation for each combination of parameter values. The values in bold obtained the best results.

- Decision tree:
    - Split criterion {**gini index**, entropy}
    - Split strategy {best, **random**}
- Naïve Bayesian:
    - Prior probabilities of the classes {**none**, $(1 - 10^{-2}, 10^{-2})$, $(1 - 10^{-3}, 10^{-3})$, $(1 - 10^{-4}, 10^{-4})$}
- Perceptron:
    - Regularisation method {l1, **l2**}
    - Tolerance for stopping criterion {**none**, 10^{-3}, 10^{-4}}
- Logistic regression:
    - Regularisation method {l1, **l2**}
    - Tolerance for stopping criterion {10^{-3}, $\mathbf{10^{-4}}$}

Table S5: Comparison of class imbalance learning techniques

Technique	Sensitivity	Specificity	AUC
No resampling	0	100	82
Under-sampling *r*=25%	55	97	83
Under-sampling *r*=50%	72	93	83
Under-sampling *r*=75%	81	89	83
Under-sampling *r*=100%	87	84	83
Over-sampling *r*=25%	55	97	83
Over-sampling *r*=50%	72	93	83
Over-sampling *r*=75%	81	89	83
Over-sampling *r*=100%	87	84	83
SMOTE *r*=25%	54	97	83
SMOTE *r*=50%	72	93	83
SMOTE *r*=75%	81	89	83
SMOTE *r*=100%	87	84	83

AUC = area under the receiver operating characteristic curve; SMOTE = synthetic minority over-sampling technique; *r* = ratio of exacerbation and non-exacerbation training examples obtained by resampling.

Table S6: Comparison of variable selection and data reduction techniques

Post-processing technique	Sensitivity	Specificity	AUC
Recursive feature elimination	88	83	86
PCA *c* = 3	83	69	73
PCA *c* = 5	84	68	73
PCA *c* = 9	85	69	73
PCA *c* = 20	87	79	79
PCA *c* = 40	88	82	85
PCA *c* = 60	89	83	86
PCA *c* = 80	90	83	85
PCA *c* = 100	90	82	84

AUC = area under the receiver operating characteristic curve; PCA = principal components analysis.

c = number of components