

RESEARCH ARTICLE

A machine learning approach to triaging patients with chronic obstructive pulmonary disease

Sumanth Swaminathan^{1,2*}, Klajdi Qirko^{1,2}, Ted Smith¹, Ethan Corcoran³, Nicholas G. Wysham^{4,5}, Gaurav Bazaz¹, George Kappel¹, Anthony N. Gerber⁶

1 Revon Systems Inc, Louisville, KY, United States of America, 40014, **2** Department of Mathematics, University of Delaware, Newark, DE, United States of America, 19716, **3** Department of Pulmonology, Kaiser Permanente, Clackamas, OR, United States of America, 97015, **4** Vancouver Clinic Division of Pulmonology & Critical Care, Vancouver, WA, United States of America, 98664, **5** Washington State University School of Medicine, Spokane, WA, United States of America, 99210, **6** Department of Medicine, National Jewish Health, Denver, CO, United States of America, 80206

* sumanth@revonsystems.net



OPEN ACCESS

Citation: Swaminathan S, Qirko K, Smith T, Corcoran E, Wysham NG, Bazaz G, et al. (2017) A machine learning approach to triaging patients with chronic obstructive pulmonary disease. PLoS ONE 12(11): e0188532. <https://doi.org/10.1371/journal.pone.0188532>

Editor: Manlio Milanese, Ospedale S. Corona, ITALY

Received: August 20, 2017

Accepted: November 8, 2017

Published: November 22, 2017

Copyright: © 2017 Swaminathan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data is available as a Supporting Information file.

Funding: The authors received no specific funding for this work.

Competing interests: We would like to disclose that Anthony N. Gerber is a consultant for Revon Systems Inc and holds stock options. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Abstract

COPD patients are burdened with a daily risk of acute exacerbation and loss of control, which could be mitigated by effective, on-demand decision support tools. In this study, we present a machine learning-based strategy for early detection of exacerbations and subsequent triage. Our application uses physician opinion in a statistically and clinically comprehensive set of patient cases to train a supervised prediction algorithm. The accuracy of the model is assessed against a panel of physicians each triaging identical cases in a representative patient validation set. Our results show that algorithm accuracy and safety indicators surpass all individual pulmonologists in both identifying exacerbations and predicting the consensus triage in a 101 case validation set. The algorithm is also the top performer in sensitivity, specificity, and ppv when predicting a patient's need for emergency care.

Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a serious long-term lung condition that progressively restricts airflow from the lungs and imposes a significant burden on patients' daily lives. COPD includes a spectrum of pulmonary phenotypes with emphysema and chronic bronchitis being the two most prominent members. Flare-ups (or exacerbations) are a frequent trigger of physician and hospital visits that are both costly and distressing to patients. Moreover, exacerbations are associated with long term declines in lung function and health status [1, 2]. A World Health Organization report anticipates that by 2030, COPD will become the third leading cause of mortality and the seventh leading cause of morbidity worldwide [3].

Despite the recognized impact of exacerbations on morbidity, mortality, and health status, there is no standardized clinical approach to improve self-identification of COPD exacerbations by patients at home. Perhaps the most widely used system is a physician provided paper checklist or "action plan". In these instances, patients are instructed to refer to a document

when they are feeling concerned about their breathing [4, 5]. The document generally has green, yellow, and red zones, which guide patients to continue usual treatment, call a physician, or go to the emergency room if their symptoms match those designated in a particular zone [6, 7]. While the type of medical guidance offered in these checklists has demonstrated some utility in patient education [8, 9], the method of delivering that guidance through a hard-coded list lacks rigor, validation, and robustness at the level of the individual patient [10]. It is not surprising then, that urgent calls or visits to the emergency room may provide the fastest path to feedback especially during hours when a doctor's office is closed. In fact, COPD is one of the leading chronic conditions driving potentially avoidable hospital admissions [11]. The need for novel solutions that limit the impact of exacerbations on patient health is abundantly apparent.

Both in COPD and many other chronic diseases, telemonitoring and mobile application-based tools have generated a great deal of excitement as novel, nonpharmacologic strategies to improve home-based disease management [12, 13]. In many cases, however, clinical examinations of such approaches have struggled to show statistically significant efficacy [14–16]. In COPD, one difficulty in enabling early diagnosis and potential treatment of exacerbation is the lack of a specific predictive diagnostic criteria. For example, the American Thoracic Society (ATS) defines a COPD exacerbation as, “an event in the natural course of the disease characterized by a change in the patient’s baseline dyspnea, cough, and/or sputum and beyond normal day-to-day variations, that is acute in onset and may warrant a change in regular medication in a patient with underlying COPD” [17]. This definition is highly ambiguous given the range, duration, and severity of possible COPD symptoms, which makes a definitive diagnosis of a COPD exacerbation challenging.

Compounding the issue of diagnosis is the inherent complexity in the interdependence of clinical features. For example, a rule-based system that dictates recommendations based on oxygen saturation or pulse may struggle to deliver appropriate guidance to a Gold Stage 1 COPD patient, who likely has normal baseline pulse and oxygen saturation in comparison to a Gold Stage 4 patient, who is likely to have abnormal baseline values [18]. Moreover, the myriad of patient physiologic profiles within individual Gold stages confounds efforts to create effective nested rules systems. Thus, app-based solutions that simply mimic paper-based home action flowcharts are unlikely to result in important improvements in patient outcomes. Machine-learning methods have gained considerable attention as novel strategies for capturing the interdependence of health variables when making predictions of complex health events [19–21]. In this study, we developed one such approach to provide both at-home decision support and an assessment of the possibility of a disease flare-up to COPD patients. We started by performing a detailed literature search and conducting an expert opinion review to define key patient characteristics including demographics, comorbid conditions, history, symptoms and vitals signs that are sufficiently and robustly predictive of exacerbation risk [22–31]. We used these variables to generate clinically diverse, simulated patient cases, and we asked physicians to provide their opinion on 1) the severity of the patient’s baseline health, vital signs, and current symptoms, 2) whether or not the patient was experiencing an exacerbation, and 3) the appropriate triage category for the patient. Physician labeled data sets were used to train a supervised machine-learning algorithm that predicts the likelihood that a patient is having a COPD flare-up and provides guidance on the appropriate responsive action. The algorithm feature set included a diverse mix of current and baseline health data. The model’s performance was validated by comparing its predictions to the consensus decision of a panel of physicians in an out-of-sample representative patient set. Analysis of the algorithm performance and the physician provided data showed 1) the algorithm showed exceptional performance when compared to individual pulmonologists in assessing the likelihood that a patient is

experiencing an exacerbation and identifying the appropriate consensus triage, 2) the algorithm triaged in favor of the safety of the patient, when disagreeing with consensus, more often than individual physicians, and 3) the algorithm decision making was transparent and consistent when compared to participating pulmonologists.

Methods

Physician selection

Physician input was used to facilitate three major aspects of the algorithm development process:

1. Algorithm feature (clinical variable) selection,
2. Algorithm training data,
3. Algorithm validation data.

All participating physicians were board certified pulmonologists and/or critical care specialists from both private and academic institutions. Refer to [S1 Table](#) for the profiles of the physicians and their respective roles in this study.

Algorithm feature selection & patient case generation

The most relevant patient symptoms, vital signs, and baseline characteristics in relationship to COPD triage were identified through a multi-tier process. First, a comprehensive literature review of common institutional practices, published guidelines, and COPD assessment tests [32–35], clinical predictors and prediction models of exacerbations [25, 29, 36, 37], and current COPD management applications [38–41] was carried out. Once selected, the features were put under consideration by a panel of three board certified pulmonologists and one critical care specialist. This panel scrutinized and modified the variable list based on consensus practice methodologies and clinical experience. Finally, the questions, responses, and measures of each variable were generated and reviewed for content, conciseness, and patient appropriate language.

The question and response list from the aforementioned process defines a space of possible patient cases. To create the optimum set of data for training and validation, a statistical experimental design using the R optFederov package from the AlgDesign library was used. Each feature was modeled linearly. This method was applied to the profile variables and baseline vital signs to generate a diverse test set of 100 patient types. Once generated, the remaining symptom, current vital sign, and comorbidity features for each patient case were randomly selected in a Monte Carlo simulation based on known distributions and correlations in the literature [42–45] to create realistic patient scenarios.

The test set was shuffled and sent to a group of 6 pulmonologists to separately triage and assess the likelihood of an exacerbation. This set gave the physicians an opportunity to better assess the suite of patient health variables and provide feedback on the appropriateness of question language, completeness of clinical features, and realism of cases while actively triaging cases. The feedback from physicians was used to update the algorithm feature list and redesign a larger set of 2501 patient scenarios replete with baseline, vitals, and symptom data. In total, 101 cases were randomly selected for validation and 2400 cases were used for training.

Each of the 6 pulmonologists provided exacerbation and triage data in the training and validation datasets. An additional 3 pulmonologists contributed labels to the validation set. In particular, they provided,

1. A 1-5 rating of the severity of baseline, symptom, and vital sign variables
2. An assessment of whether or not the patients current health indicated a COPD exacerbation with 0-100% confidence
3. A recommendation for the appropriate triage action to take with 0-100% confidence.

The triage categories from which the physicians could choose were,

1. **Ok:** No additional medical attention needed,
2. **Plan:** Continue normal treatment and check back in 1-2 days,
3. **Doc:** Call the doctor,
4. **ER:** Go to the emergency room.

Data was sent to physicians in 100-case batches. Triage and exacerbation assessments were recorded in spreadsheets akin to the sample shown in [S1 Spreadsheet](#). Cases that were used in the training were individually labeled by physicians, while cases used in the validation set included the opinion of all 9 previously mentioned physicians. The process is depicted in [Fig 1](#).

Algorithm training and validation

The strategy used to find the optimal prediction model is shown in [Fig 2](#). This process was identical both for predicting the presence of an exacerbation and predicting the appropriate

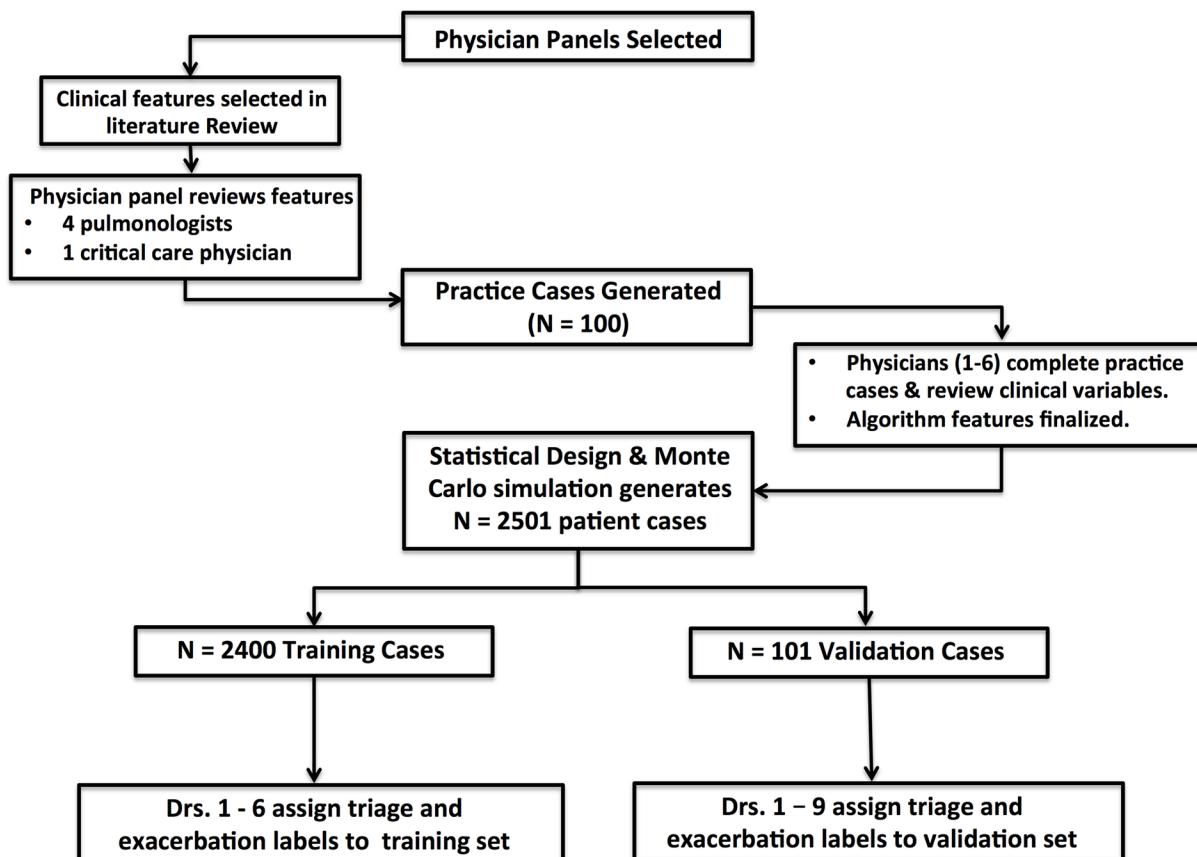


Fig 1. Process for generating patient case scenarios and collecting algorithm training and validation data.

<https://doi.org/10.1371/journal.pone.0188532.g001>

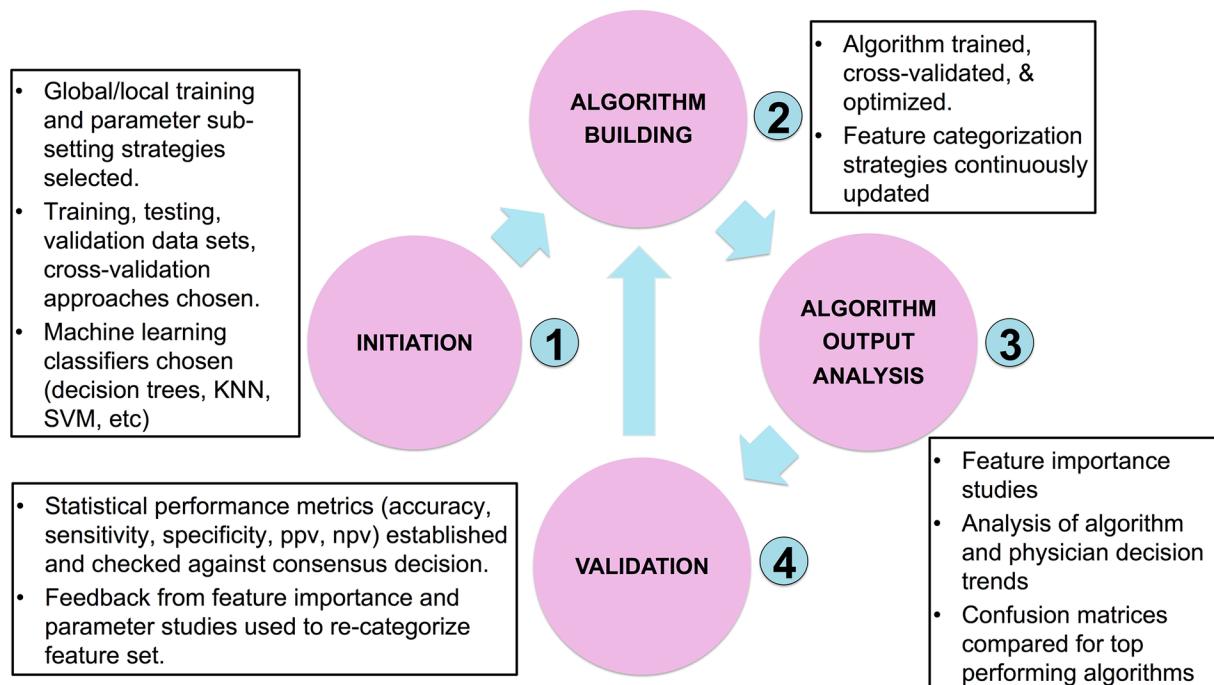


Fig 2. Training, validation, and optimization procedure for building COPD exacerbation and triage prediction algorithms.

<https://doi.org/10.1371/journal.pone.0188532.g002>

triage recommendation. Initially, several candidate supervised learning classifiers were selected including support vector machines, logistic regression, Naive Bayes, KNN and a variety of gradient boosted and ensemble decision tree methods. For each classifier type, thousands of algorithms were trained on each combination of physician training data using Python's Scikit-Learn suite. All algorithms went through a hyper-parameter optimization process including a grid search with 5-folds cross-validation. The top performing algorithms of each class were selected based on how they performed when making predictions on the out-of-sample validation test.

Algorithm predictions were validated by comparing the algorithm's triage and exacerbation (y/n) classifications to the consensus decision of a panel of physicians on 101 hypothetical patient cases. Each individual physician and the algorithm were tested for how often their particular recommendation for a patient case matched the majority opinion. In cases of ties, the more conservative medical decision (higher triage/exacerbation category) was accepted as the correct one. The performance of the algorithm was compared to the other physicians in three scenarios: 1) The algorithm voted in the majority opinion, 2) The algorithm did not vote in the majority opinion, and 3) Neither the algorithm nor any individual physician voted in the majority opinion. The 101 validation cases were removed from the 2501 case set prior to training, which made them statistically diverse, clinically relevant, and truly out-of-sample. Statistical measures of performance used in this study included:

$$\text{Classification Accuracy}(\%) = \text{ACC} = \frac{TC}{TC + FC}, \quad (1)$$

$$\text{Sensitivity} = \text{TPR} = \frac{TP_{ER}}{TP_{ER} + FN_{ER}}, \quad (2)$$

$$\text{Specificity} = \text{TNR} = \frac{TN_{ER}}{TN_{ER} + FP_{ER}}, \quad (3)$$

$$\text{Positive Predictive Value} = \text{PPV} = \frac{TP_{ER}}{TP_{ER} + FP_{ER}}, \quad (4)$$

$$\text{Negative Predictive Value} = \text{NPV} = \frac{TN_{ER}}{TN_{ER} + FN_{ER}}, \quad (5)$$

$$\text{Confusion Matrix Proximity to Upper Triangular} = \text{UTP} = 1 - \frac{LT}{101}, \quad (6)$$

$$\text{Misclassifications Greater Than One Category} = E_{G1} = \frac{C_{G1}}{101}, \quad (7)$$

where,

1. TC = Total classifications matching consensus
2. FC = Total classifications not matching consensus
3. TP_{ER} = Emergency classifications matching consensus
4. TN_{ER} = Non-emergency classifications matching
5. FP_{ER} = Emergency classifications not matching consensus
6. FN_{ER} = Non-emergency classifications not matching consensus
7. LT = Number of lower triangle entries in confusion matrix
8. C_{G1} = Number of triage missclassifications greater than 1 category

Additional analysis was done to assess the algorithm's performance on identifying situations in which "medical attention" was required. In particular, medical attention is defined as triage circumstances which call for physician assistance (triage category = 3) or emergency care (triage category = 4). The remaining two triage categories define instances in which no medical attention is needed. The statistical measures of performance for this study included:

$$\text{ACC}_M = \frac{TC_M}{TC_M + FC_M}, \quad (8)$$

$$\text{TPR}_M = \frac{TP_M}{TP_M + FN_M}, \quad (9)$$

$$\text{TNR}_M = \frac{TN_M}{TN_M + FP_M}, \quad (10)$$

$$\text{PPV}_M = \frac{TP_M}{TP_M + FP_M}, \quad (11)$$

$$\mathbf{NPV}_M = \frac{TN_M}{TN_M + FN_M}, \quad (12)$$

where,

1. TC_M = Total classifications matching the consensus view on the need or lack of need for medical attention
2. FC_M = Total classifications not matching the consensus view on the need or lack of need for medical attention
3. TP_M = Sum of category 3 and 4 classifications made on cases with a consensus triage of at least category 3
4. TN_M = Sum of category 1 and 2 classifications made on cases with a consensus triage of at most category 2
5. FP_M = Sum of category 3 and 4 classifications made on cases with a consensus triage of at most category 2
6. FN_M = Sum of category 1 and 2 classifications made on cases with a consensus triage of at least category 3

Confusion matrices were used in this study to visualize the extent of algorithm and physician agreement with consensus for each triage class. Perfect confusion matrices are diagonal, indicating complete agreement between the triager and the consensus. Off-diagonal entries below the diagonal indicate under-triage with comparison to consensus while entries above the diagonal indicate over-triage. Subsequent result sections show such results.

Algorithm feature importance

The importance of clinical variables in machine-learning predictions is calculated based on the methodology used by the particular model. In this study, the feature ranking of the Gradient-Boosted Decision Trees (GB) classifier was determined by the expected fraction of samples (case outcomes) to which a particular feature contributed across all trees in the ensemble. Higher fractions indicate higher feature importance. Ultimately, the fractional contribution was determined as an average across the entire forest.

In the case of Logistic Regression, the feature importance was determined by the size of the coefficient effect. When predicting triage, the Logistic Regression included a separate prediction model for each class relative to the other classes, conforming to one-vs-rest methodology. Hence, the feature importance was determined as the average rank of each feature effect over the four prediction models.

Robustness of validation set consensus

The size and scope of the physician panel used in the validation set was a topic of great importance in this study. In order to assess the robustness of the consensus on the validation set, we started by selecting a minimum number of doctors (five) from the complete validation panel of 9 doctors + algorithm. After finding the majority triage opinion of the 5 physician panel on each case, we added a 6th doctor and calculated the percent of 101 total cases that changed triage labels. This process was repeated for all other physicians not in the 5-member panel and the results were averaged. Finally, the outcome of this procedure was averaged for every possible initial combination of 5 physicians to yield the average, max, and min percentage of cases

where the majority decision changed after adding a 6th physician panel member. Using this method for all initial panel sizes generates a quantitative assessment of how many physicians are needed to establish a robust consensus in the validation set.

Results

Algorithm feature set

The clinical variables selected for algorithm training were found through the multi-tier process described in the method's section, *Algorithm Feature Selection & Patient Case Generation*. The final variable list is shown in [Table 1](#), and includes 1) patient background characteristics that are associated with COPD exacerbation risk and severity, 2) current clinical symptoms that encompass widely accepted features of exacerbations, and 3) physiologic measurements that are predicted to influence physician perception of exacerbation severity.

Table 1. List of patient profile, comorbidity, vital sign, and symptom factors, with respective measures, used in the COPD triage and exacerbation algorithms.

Variable		Units—Type
Patient Profile	Age	years—continuous
	Weight	lb—continuous
	Height	feet + inches—continuous
	Gender	Male/Female—categorical
	COPD GOLD STAGE	1,2,3,4—categorical
	Baseline MMRC Dyspnea	1,2,3,4,5—categorical
	Recent Exacerbations & Hospitalizations	Yes/No—categorical
	Lives Alone?	Yes/No—categorical
	Smoker	Yes/No—categorical
	Long-Term Oxygen User	Yes/No—categorical
Comorbidities	Assisted Daily Activity	Yes/No—categorical
	Congestive Heart Failure	Yes/No—categorical
	High Blood Pressure	Yes/No—categorical
	Coronary Artery Disease	Yes/No—categorical
	Diabetes	Yes/No—categorical
	Anemia	Yes/No—categorical
	Pulmonary Hypertension	Yes/No—categorical
Symptoms	Acid Reflux	Yes/No—categorical
	Shortness of Breath	1,2,3—categorical
	Cough	1,2,3—categorical
	Wheezing	1,2,3—categorical
	Change in Sputum Color	Yes/No—categorical
	Increased Sputum Volume	Yes/No—categorical
	Cold/URI	Yes?NO—categorical
	Medication Compliance	1,2,3—categorical
Vital Signs	Sleeplessness	Yes/No—categorical
	Current MMRC Dyspnea	1,2,3,4,5—categorical
	Oxygen Saturation	%—continuous
	FEV1	Vol/sec—continuous
	Heart Rate	BPM—continuous
	Temperature	°F—continuous

<https://doi.org/10.1371/journal.pone.0188532.t001>

As the feature list included both continuous and categorical variables with different units and responses, the detailed questions and responses are included in [S1 Document](#). The numerical levels of each categorical variable correspond to patient level responses. For example, in the case of *cough*, the levels 1,2,3 correspond to *less than usual*, *same as usual*, and *more than usual* respectively. All features have an additional response of *unknown* except for age, weight, height, gender, baseline dyspnea, and symptom questions. This was done to train the algorithm on cases in which patient data could be missing.

Top performing algorithms

As detailed in the methods, patient cases generated using the variables shown in [Table 1](#) were labeled by physicians, and the resultant data were used to train algorithms using a variety of strategies. [Fig 3](#) includes a comparison of the top performing algorithms of each classifier type for out-of-sample classification accuracy. Among the different machine-learning classifiers tested, The top 2 performers were the Gradient-Boosted Decision Tree and the Logistic Regression. All classifier algorithm types were trained in a comparable way inclusive of hyper-parameter optimization and cross-validation.

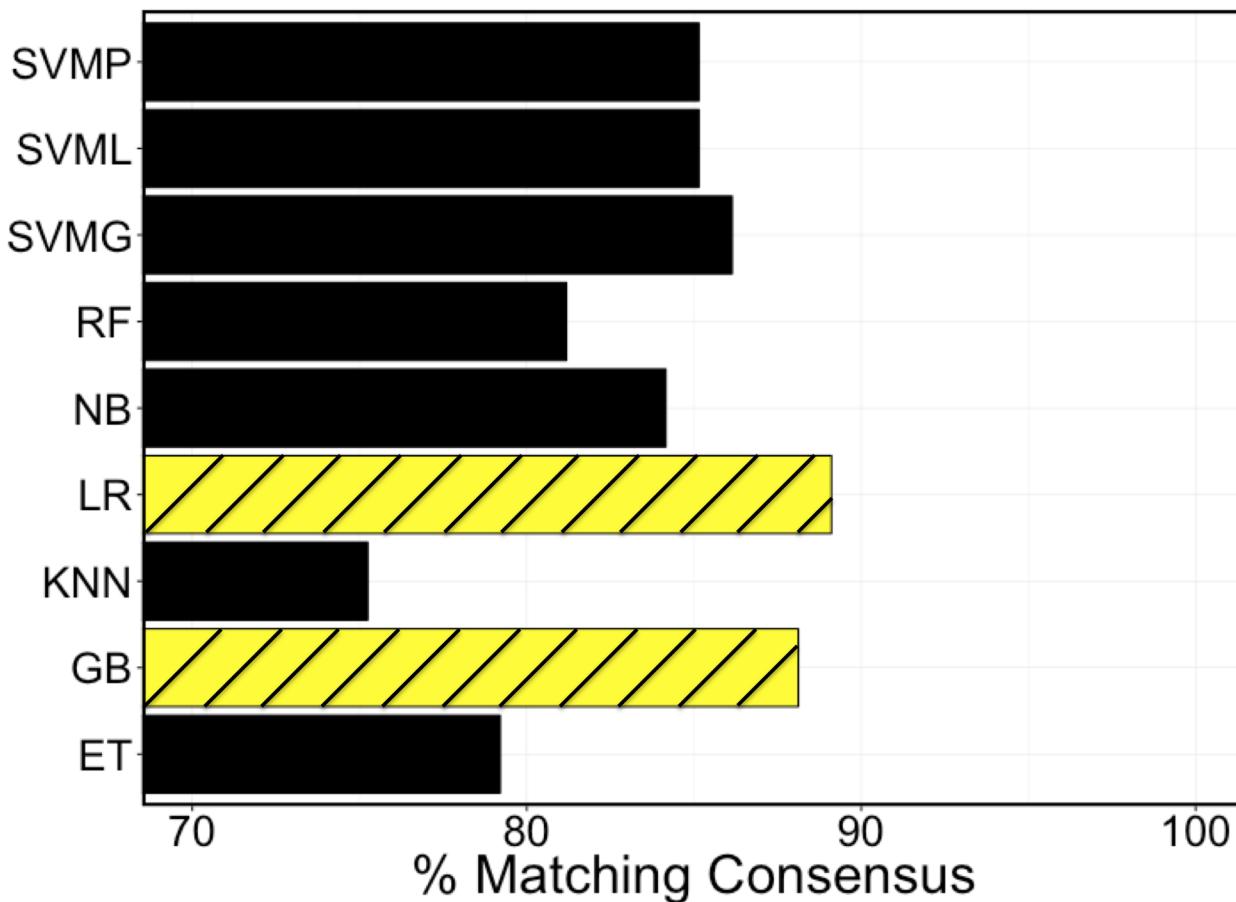


Fig 3. Comparison between ML classifiers at matching consensus decision in the validation set. SVMP, SVML, and SVMG are all support vector machine algorithms with polynomial, linear, and Gaussian kernels respectively. RF = Random Forest, NB = Naïve Bayes, LR = Logistic Regression, KNN = K-Nearest Neighbors, GB = Gradient Boosted Random Forest, and ET = Extra Decision Tree Classifier.

<https://doi.org/10.1371/journal.pone.0188532.g003>

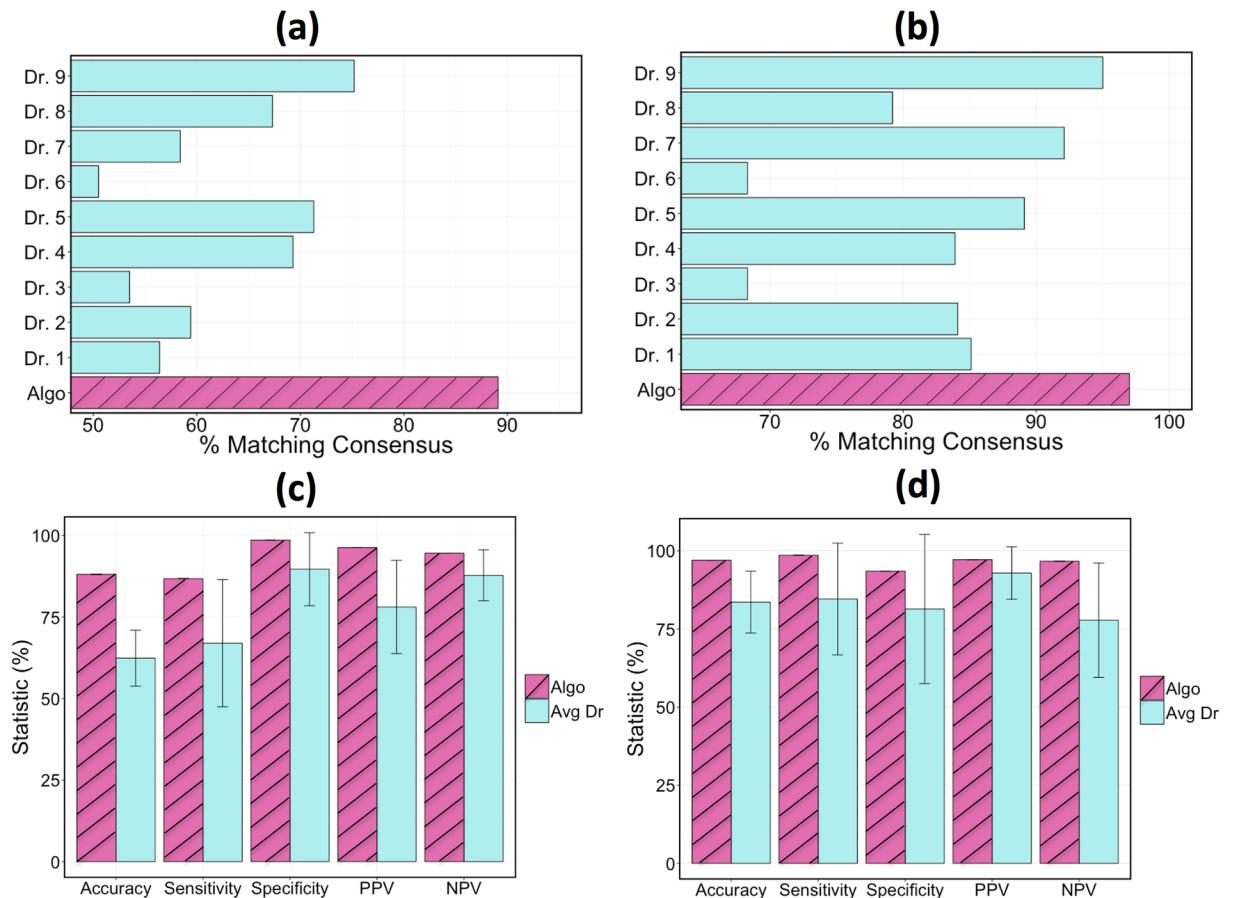


Fig 4. Performance comparison when the algorithm and all of the physicians got a vote in the consensus opinion.
Comparison of the algorithm and individual physicians at predicting the consensus triage and exacerbation (y/n) in the validation set: (a) triage identification, (b) exacerbation identification. A comparison of the algorithm with the average physician in accuracy, sensitivity, specificity, ppv, and npv for: (c) triage identification, (d) exacerbation identification. Triage statistics were computed as defined in Eqs 1–7.

<https://doi.org/10.1371/journal.pone.0188532.g004>

Algorithm performance

Model accuracy was measured as the percentage of classifications that matched the consensus triage and exacerbation labels in the validation set. The accuracy results of the GB classifier when the classifier voted in the consensus are depicted in Fig 4. The algorithm agreed with the consensus opinion in 88% of triage cases, whereas an individual physician agreed with the consensus 74% of the time at best. When determining if an exacerbation had occurred, the algorithm assessment again agreed with the consensus determination more than any individual doctor with a success rate of 97% as compared to 95% from the top performing physician. A comparison of the algorithm to the average physician performance is also shown in Fig 4. In the case of triage accuracy, sensitivity, and ppv, the algorithm performed more than 1 standard deviation better than the average physician (more than 2 standard deviations in the case of accuracy). The exhaustive set of statistical performance metrics for the top algorithms and the top physician are shown in Table 2.

It is noteworthy that the algorithm maintained its classification performance relative to the other physicians even in the assessments where it did not vote in the consensus. Results of these tests are shown in Fig 5. When the consensus opinion did not include the algorithm but

Table 2. Statistical measures (Eqs 1–7) of triage and exacerbation identification ability for the top 2 performing algorithms and top physician.

Metric	Triage			Exacerbation		
	GB	LR	Top Dr.	GB	LR	Top Dr.
ACC	88.1	89.1	74.3	97.0	97.0	95.0
TPR	86.7	90.3	70.0	100	98.6	97.1
TNR	98.6	98.6	100	90.6	93.5	90.3
PPV	96.3	96.6	100	95.8	97.2	95.8
NPV	94.6	95.8	88.7	100	96.7	93.3
UTP	96.0	95.0	77.0	0.0	99.0	98.0
E _{G1}	0.0	0.0	1.0	NA	NA	NA

<https://doi.org/10.1371/journal.pone.0188532.t002>

included all individual physicians (a test that inherently favors the physicians), the algorithm had a triage/exacerbation classification accuracy of 82%/96% compared to the top performing physician at 77%/94%. In the case where no member's vote was included in the consensus when calculating that member's accuracy, the top physician dropped considerably in performance with triage and exacerbation accuracies of 62% and 93% respectively.

Confusion matrix analysis

The confusion matrices shown in Fig 6 give a comprehensive performance summary of both the algorithm and the top-performing physician (the physician with the highest classification accuracy on the validation set) in triage and exacerbation identification. On exacerbation identification, the top performing algorithm and top performing physician showed comparable performance when compared to consensus. On triage category identification, however, a number of performance and safety differences were observed

1. Performance observations

- a. The algorithm showed a total classification accuracy of 88% while the top performing physician's accuracy was 74%
- b. The algorithm showed a sensitivity of 95% in triaging patients to the doctor and 87% in triaging patients to the emergency room. The top physician, by contrast, had a sensitivity of 78% and 70% respectively.

2. Safety Analysis

- a. The algorithm never under or over triaged a patient by more than one category while the top physician under triaged one patient with *No additional medical attention needed* when the consensus triage was *Call your doctor*.
- b. The algorithm never under triaged a patient who should be sent to the doctor while the top physician under triaged 22% of such patients. In ten out of eleven of those cases, the physician suggested that the patient consult their normal treatment plan and check back in 1-2 days and in one case that No additional medical attention is needed.
- c. Out of 101 total cases, the algorithm under triaged the consensus less than 4% of the time. The top physician, by comparison, under triaged approximately 23% of the time.
- d. For patients who have a consensus triage of *Go to the ER*, the physician under triaged 30% of them to the doctor while the algorithm under triaged less than 14% of them to the doctor.

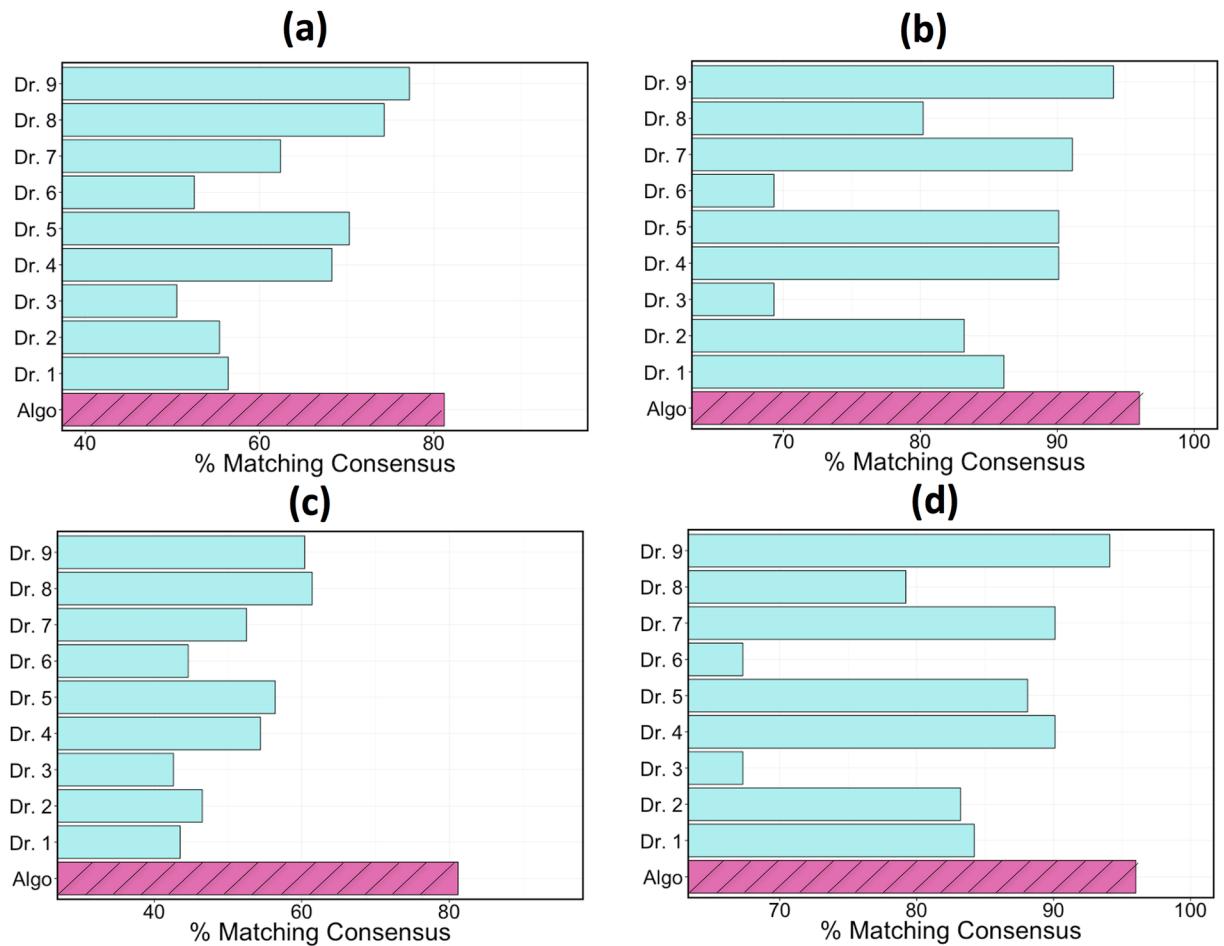


Fig 5. Performance comparison of algorithm and individual physicians at predicting the consensus of the validation sets.

(a) triage performance, algorithm was not included in consensus, (b) exacerbation performance, algorithm was not included in consensus, (c) triage performance, no member votes when assessing their accuracy, (d) exacerbation performance, no member votes when assessing their accuracy.

<https://doi.org/10.1371/journal.pone.0188532.g005>

- e. When the algorithm didn't agree with the consensus category 2 triage, it always triaged the patient to the doctor.
- f. In comparison to the perfect confusion matrix the algorithm had 12 off-diagonal entries (88% accuracy) with 4 below the diagonal indicating that the algorithm under-triaged with respect to consensus 4% of the time and under-triaged when misclassifying 33% of the time. The top performing pulmonologist, by contrast, had 26 off-diagonal entries (74% accuracy) with 23 of those below the diagonal indicating that the top physician under-triaged with respect to consensus 23% of the time and under-triaged when misclassifying 88% of the time.

In the final study of algorithm performance, the algorithm and physicians were examined for their ability to discern the presence of general medical need (i.e: triage category of “Call your doctor” or “Go to the ER”). Statistical performance metrics of this study are given in [Table 3](#) with confusion matrices shown in [Fig 7](#). Similar observations can be made about the algorithms effectiveness though the top physician did exhibit superior performance in

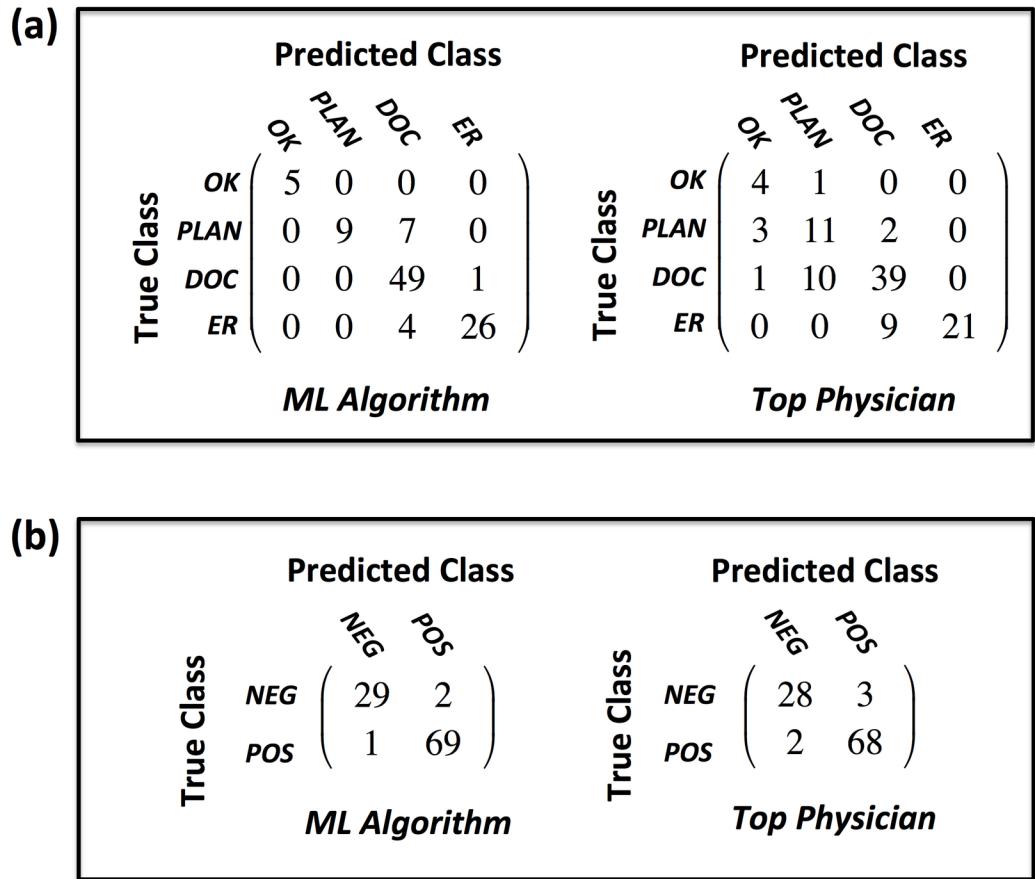


Fig 6. Confusion matrices comparing assessment performance of the GB algorithm to the top physician.
 (a) triage, (b) exacerbation. Note: top physician = the physician with highest classification accuracy.

<https://doi.org/10.1371/journal.pone.0188532.g006>

specificity and PPV over both algorithms. This is likely explained by the fact that the algorithm tended to over-triage in cases when it disagreed with consensus. It is further noteworthy that the algorithm never failed to identify the need for medical attention in the 101 validation cases, while the top performing physician misclassified 11 out of 80 such consensus instances (13.75% of the time).

Physician decision making trends

Physician-labeled data on exacerbation and triage categories were compared in the validation sets to better understand physician decision making. Fig 8 below shows the distribution of

Table 3. Statistical measures of performance of the top 2 algorithms (highest classification accuracy) and top performing physician when classifying the need for medical attention.

Metric	GB	LR	Top Dr.
ACC_M	93.1	93.1	87.1
TPR_M	100	97.5	86.3
TNR_M	66.7	77.3	90.5
PPV_M	92.0	93.9	97.2
NPV_M	95.6	95.8	88.8

<https://doi.org/10.1371/journal.pone.0188532.t003>

		Predicted Class		Predicted Class	
		NEG	POS	NEG	POS
True Class	NEG	14	7	29	2
	POS	0	80	11	69
		ML Algorithm		Top Physician	

Fig 7. Comparing the performance of the GB algorithm to the top physician in assessing the need for medical attention.
Note: top physician = physician with highest classification accuracy.

<https://doi.org/10.1371/journal.pone.0188532.g007>

triage and exacerbation labels in the validation set per doctor. Plots of the average triage and exacerbation classes are also shown with error bars indicating 1 standard deviation intervals.

A variety of observations can be made about outlier opinions, inter-physician consistency, physician treatment of risk, and correlation between triage and exacerbation categories. Doctor 2, for example, triaged 63% of patients to the emergency room and Doctor 3 triaged 32% of patients as needing no additional medical attention, which are both over 2 standard deviations outside of the respective means. Physician triage assessment was also often highly independent of exacerbation assessment. Drs. 3 and 6, for example, had nearly identical exacerbation class distributions, but Dr. 3 triaged 32% of patients as needing no additional medical attention as opposed to 11% for that of Dr. 6. This could be partially due to a belief that an alternate diagnosis was driving symptoms. Moreover, the shape of the consensus triage distribution was matched closely only by the algorithm plus Drs. 5, 7, and 9. This suggests that the remaining physicians used a qualitatively different logic when choosing triage categories.

Robustness of validation set consensus

The study of how many physicians constituted a robust validation panel (detailed in the [Methods](#) section) resulted in the convergence plot shown in [Fig 9](#). Looking at the graph one can notice that each case of the validation set converged to an unchanging correct answer as more doctors were added. 7 physicians marked the region where the set showed good convergence with only 8% of cases changing on average when adding another doctor.

Machine-learning feature importance

The patient variables that most influence triage and exacerbation assessment are shown in [Table 4](#). [Table 4](#) shows the hierarchical importance of the top 15 features in predicting the correct triage category based on the process described in the methods section *Algorithm Feature Importance*. Interestingly, the GB algorithm favored patient profile characteristics like age, BMI, and height as the most influential factors for triage while the same variables failed to

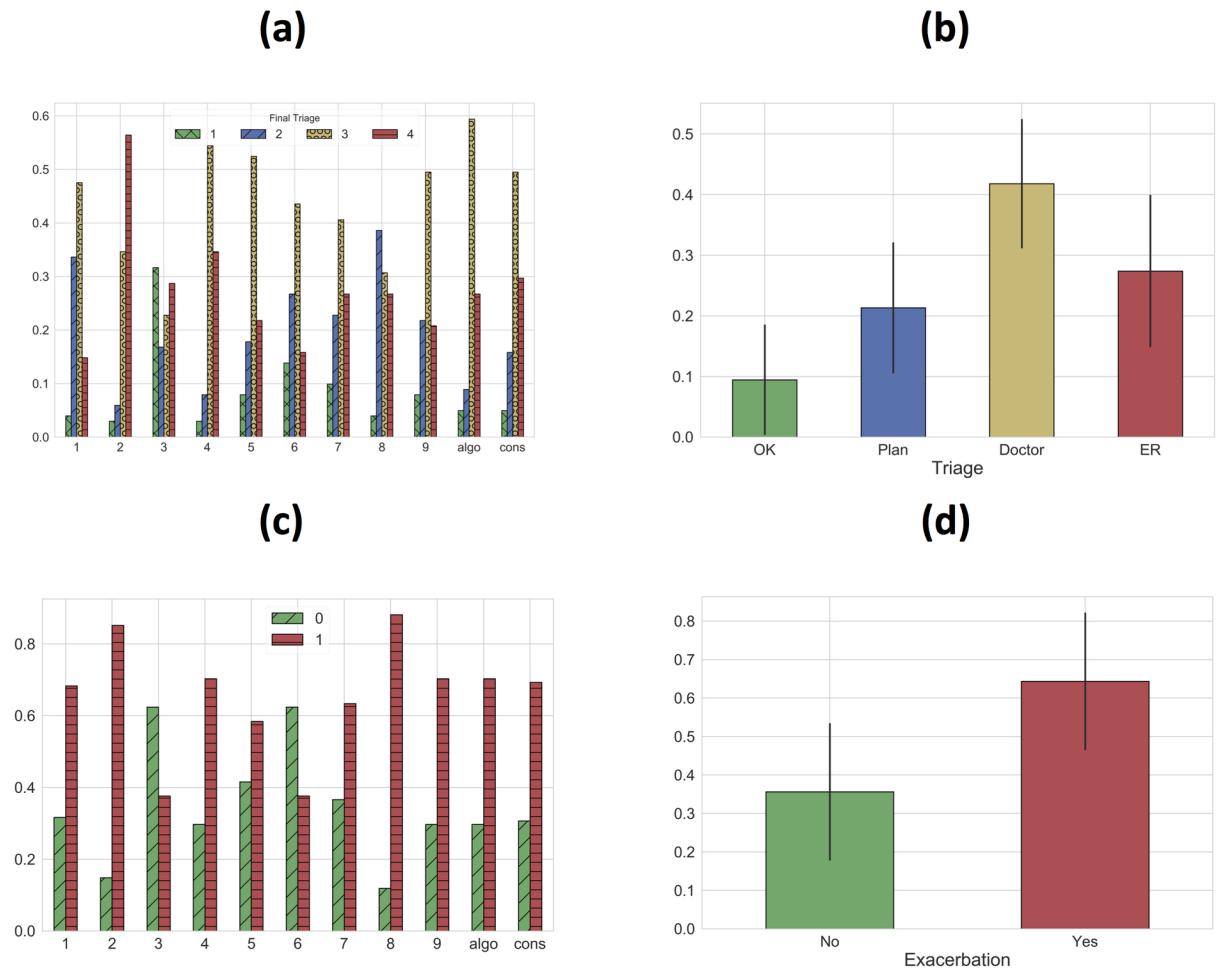


Fig 8. Distributions for each physician in the validation set (left) and the averaged distributions (right). (a) triage distribution, (b) averaged triage distribution, (c) exacerbation distribution, (d) averaged exacerbation distribution. Note: error bars indicate 1 standard deviation about the mean.

<https://doi.org/10.1371/journal.pone.0188532.g008>

reach the top 15 list for LR triage. LR also tended to weight vital sign levels more heavily when predicting on both triage and exacerbation classes. Despite these differences, both algorithms maintained comparably high statistical measures of performance.

Discussion

The machine-learned triage approach in this study performed favorably when compared to individual physicians in a broad range of statistical performance measures both in triage and in predicting the presence of a COPD exacerbation. Unlike existing paper checklist type tools, the models incorporated the baseline medical health of the patient in a way that robustly accounted for the complex interactions of patient health variables. Gradient-Boosted Decision Trees and Logistic Regression showed the highest performance when making out-of-sample predictions on the validation set. The performance metrics used to evaluate the algorithms demonstrated accuracy, safety, consistency, and edge case prediction performance comparable to or better than the top performing physicians in all studies with three different assessments of consensus.

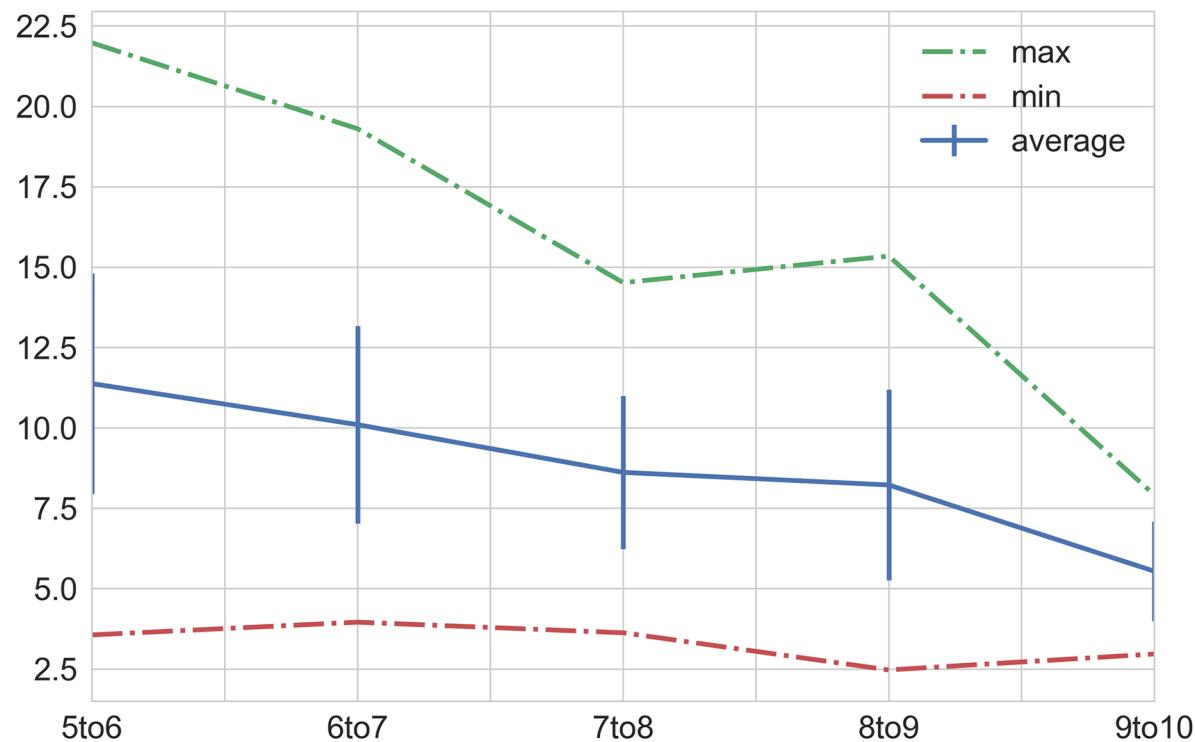


Fig 9. Plot of % change in consensus triage answers as additional doctors are added to the validation panel (plus algo). The average change when the panel reaches 10 members (from 9) is 5.5%.

<https://doi.org/10.1371/journal.pone.0188532.g009>

Strengths

The use of machine-learning predictions on a clinical feature set to identify disease flare-ups and provide subsequent patient decision support is a unique contribution of this study. To date, we are not aware of any other work that has produced a comparable result. The use of

Table 4. Feature importance for the top two performing algorithms for both triage and exacerbation models.

Triage		Exacerbation	
GB	LR	GB	LR
Age	O ₂ Sat_(87, 89]	cough_3	cough_3
BMI	cough_3	sputum(col+vol)_0	shortbreath_3
Height	sputum(col+vol)_0	shortbreath_3	sputum(col+vol)_0
O ₂ Sat_(87, 89]	shortbreath_3	wheeze_3	wheeze_3
shortbreath_3	wheeze_3	sputum(col+vol)_1	sputum(col+vol)_1
Infection	Sleeplessness	Δfev1_(-15, -5]	Δfev_(0, 10]
Sleeplessness	Infection	Δfev1_(0, 10]	med_comp_3
cough_3	hr_(100, 110]	O ₂ Sat_(91, 93]	cough_1
wheeze_3	fev1_(0, 20]	sputum_vol_1	Δfev1_(-15, -5]
fev1_(0, 20]	O ₂ Sat_(85, 87]	fev1_(90, 100]	O ₂ Sat_(87, 89]
sputum(col+vol)_0	O ₂ Sat_(0, 85]	med_comp_3	Sleeplessness
hr_(100, 110]	Δfev1_(-100, -15]	Age	ΔO ₂ Sat_-[4,-2]_base < 93
O ₂ Sat_(85, 87]	hr_(110, 120]	shortbreath_1	O ₂ Sat_(85, 87]
heartrate_(110, 120]	sputum(vol+col)_1	Δfev_(-100, -15]	sputum_col_1
cough_1	hr>120	Sleeplessness	Δfev_(-5, 0]

<https://doi.org/10.1371/journal.pone.0188532.t004>

consensus physician opinion as a validation standard and the analysis of individual physician performance on that standard is also a unique contribution of this study.

The current study has demonstrated that the top two performing algorithms, GB and LR, both yield a suite of statistical performance metrics that compare favorably with individual physicians, and yet, the clinical variables that most influence each model's output maintain a different rank order of importance. Logistic regression generally weighted vital sign data with more importance. This result suggests that a good recommendation based on the validation standard in this study can be achieved through different logic (modeling type) with diverse health data (algorithm features).

Limitations

While the algorithm exhibits very strong performance when predicting on the out-of-sample validation set, ultimately the algorithm training is done on cases with individual physician labels. Training on cases with more opinions could facilitate a more robust in-sample cross-validation process. Given the data collection methodology used in this study, it would become increasingly expensive and intractable to collect orders of magnitude more cases for training, but with increased access to electronic medical records, one could consider using a much larger patient dataset with historic patient outcomes guiding the training process. This approach would require considerable thought and further investigation given the lack of gold standard on what constitutes a correct triage assessment.

Thus far, the algorithm has been both developed and tested on hypothetical patient cases. Unlike clinical datasets for post-hoc analysis of hospitalized patients, data for outpatient triage and evaluation is not readily available, necessitating the use of simulated data. Although the feature set within the training data is certainly comprehensive and large when compared to the information that would generally be available to a pulmonologist, internist, or nurse in the clinical setting, an additional level of validation would be to compare the prediction of the algorithm in a real patient setting with a set of physicians actively triaging the same set of patients. This type of clinical data would provide additional insight based on current medical practices.

It is further recognized that the black-box nature of ensemble decision tree methods makes the decision making logic in triage recommendations difficult to interpret. The feature importance studies previously discussed shed light on which patient variables most influence the final outcome, but ultimately, the inherent complexity and interactions of the feature set make it difficult to give a simple, linear causal explanation of the algorithm output based on the inputted features.

Future work

Mobile applications geared toward improved at-home patient care and self-management of chronic illnesses have substantially grown in use due largely to the availability of technology and the rising costs of health care [46]. While the growing popularity of mhealth (health care and public health practice supported by mobile devices [47]) is evident, its impact and efficacy is not [46]. This study has shown that machine-learning based applications offer the exciting prospect of accurate and personalized triage of COPD patients. Early detection of disease flare-ups and accurate council to patients has the potential to both reduce the severity of exacerbations and prevent unnecessary hospitalizations for otherwise healthy, anxious patients. This may assist the drive towards personalized medicine by better guiding decision support for individual patients.

The current algorithm is deployed in a mobile app that is primarily meant for at-home patient use, though it could also be used by nurses and internists less familiar with a patient's baseline health as a tool to confirm their assessments during patient calls. The app should be further explored for its effectiveness in real patient populations with respect to various clinically relevant endpoints both for improving patient decision making and for engendering clinical reduction in severity and frequency of COPD exacerbations. Future investigation of the feasibility of machine-learning applications in clinical trials will be needed. Moreover, a robust clinical study on the influence of these applications on patient anxiety, stress, and overall health would elucidate.

With modern computational capability and continuously better access to health data, the opportunities to train machine-learning algorithms on large patient outcome datasets will improve. This may have particular relevance in COPD, where emerging data from large phenotyping and genotyping efforts, such as COPDGene and Spiromics, are delineating novel variables that impact exacerbation and disease risk [48]. Blood-based biomarkers and cellular content such as eosinophils, for example, are known to be correlated with increased risk of COPD exacerbations [49]. Although physician opinion is currently the gold standard for many clinical decisions, including diagnosis and triage of COPD exacerbations, active cloud-based training that integrates patient data in electronic medical records with available scientific knowledge may eventually provide specific predictions and recommendations that support medical-decision making. Such cloud-based information could be returned to a patient at home or to a provider in a clinical setting as APIs for computers and mobile devices.

Conclusion

This study has shown that a machine-learning approach to triaging patients with COPD is a viable and robust method when compared to individual pulmonologists at facilitating at-home triage and exacerbation self-identification. The ML algorithm exhibited higher accuracy than all individual, board certified physicians in predicting the consensus opinion on both the presence of an exacerbation and the appropriate triage category in a representative set of patient cases. Furthermore, the algorithm erred in favor of patient safety more often than any individual pulmonologist and exhibited greater consistency in its recommendations. While the app is not meant to be a substitute for physician examinations or physician guided patient care, it does provide simple, easily accessible, safe, and highly accurate at-home decision support which can direct patients to the right care. Furthermore, it is generalizable to other chronic illnesses in which relevant symptom, signs, and patient profile data are available.

Supporting information

S1 Table. List of physicians, their affiliations and their contribution to the smart COPD development process. Pulmonologists who participated in this study provided expert opinion in clinical selection/review, algorithm training, and validation. The physician profiles are indicated in the supporting table.

(XLSX)

S1 Document. Detailed questions and responses.

(DOCX)

S1 Spreadsheet. Training data sample. Simulated patient cases are issued to physicians to provide exacerbation and triage labels for the purpose of algorithm training and validation. This file includes a sample batch of 100 cases with the corresponding physician entered data.

(XLSX)

S1 Dataset. Training and validation data. This file includes all the data used for training and validation.
(ZIP)

Acknowledgments

The authors thank Dr. Lucas Abraham and Dr. Ervin Anaya for useful consultation through algorithm development and validation.

Author Contributions

Conceptualization: Sumanth Swaminathan, Anthony N. Gerber.

Data curation: Sumanth Swaminathan, Klajdi Qirko.

Formal analysis: Sumanth Swaminathan, Klajdi Qirko.

Funding acquisition: Sumanth Swaminathan.

Investigation: Sumanth Swaminathan, Klajdi Qirko, Ted Smith, Ethan Corcoran, Nicholas G. Wysham, Anthony N. Gerber.

Methodology: Sumanth Swaminathan, Klajdi Qirko, Ethan Corcoran, Nicholas G. Wysham, Anthony N. Gerber.

Project administration: Sumanth Swaminathan.

Resources: Sumanth Swaminathan, Ted Smith.

Software: Sumanth Swaminathan, Gaurav Bazaz, George Kappel.

Supervision: Sumanth Swaminathan, Ted Smith, Anthony N. Gerber.

Validation: Sumanth Swaminathan, Klajdi Qirko, Ethan Corcoran, Nicholas G. Wysham, Anthony N. Gerber.

Visualization: Sumanth Swaminathan, Klajdi Qirko.

Writing – original draft: Sumanth Swaminathan.

Writing – review & editing: Sumanth Swaminathan, Klajdi Qirko, Anthony N. Gerber.

References

1. Dransfield MT, Kunisaki KM, Strand MJ, Anzueto A, Bhatt SP, Bowler RP, et al. Acute Exacerbations and Lung Function Loss in Smokers with and without Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med.* 2017; 195(3):324–330. <https://doi.org/10.1164/rccm.201605-1014OC> PMID: 27556408
2. Rennard SI, Farmer SG. Exacerbations and Progression of Disease in Asthma and Chronic Obstructive Pulmonary Disease. *Proc Am Thorac Soc.* 2004; 1(2):88–92. <https://doi.org/10.1513/pats.2306026> PMID: 16113418
3. World Health Organization. Global Surveillance, prevention and control of chronic respiratory diseases: A comprehensive approach; 2007.
4. Watson PB, Town GI, Holbrook N, Dwan C, Toop LJ, Drennan CJ. Evaluation of a Self-Management Plan for Chronic Obstructive Pulmonary Disease. *European Respiratory Review.* 1997; 10:1267–1271. <https://doi.org/10.1183/09031936.97.10061267>
5. Kaplan A. The COPD Action Plan. *Canadian Family Physician.* 2009; 55(1):58–59. PMID: 19155371
6. American Lung Association. My COPD Action Plan; 2012. “<http://www.lung.org/assets/documents/copd/copd-action-plan.pdf>”.
7. Australian Lung Foundation. COPD Action Plan; 2015. “<http://lungfoundation.com.au/health-professionals/clinical-resources/copd/copd-action-plan/>”.

8. Turnock AC, Walters EH, Walters JA, Wood-Baker R. Action Plans for Chronic Obstructive Pulmonary Disease. *Cochrane Database System Rev.* 2005; 19(4). <https://doi.org/10.1002/14651858.CD005074.pub2>
9. Walters JA, Turnock AC, Walters EH, Wood-Baker R. Action plans with limited patient education only for exacerbations of Chronic Obstructive Pulmonary Disease. *Cochrane Database System Rev.* 2010; 12(5). <https://doi.org/10.1002/14651858.CD005074.pub3>
10. Nici L, Bontly TD, Zuwallack R, Gross N. Self-Management in Chronic Obstructive Pulmonary Disease. Time for a Paradigm Shift? *Ann Am Thorac Soc.* 2014; 11(1):101–107. <https://doi.org/10.1513/AnnalsATS.201306-150FR> PMID: 24460443
11. Porter J, Herring J, Lacroix J, Levinton C. CIHI Survey: Avoidable Admissions and Repeat Admissions: What Do They Tell Us? *Healthcare Quarterly.* 2007; 10(1):26–29.
12. Ambrosino N, Vagheggi G, Mazzoleni S, Vitacca M. Telemedicine in Chronic Obstructive Pulmonary Disease. *Medical Clinics of North America.* 2016; 12(4):351–356.
13. Sohrabi K, Mursina L, Seifert O, Scholtes M, Hoehle L, Hildebrandt O, et al. Telemonitoring and medical care supporting of patients with Chronic Respiratory Diseases. *ehealth2015.* 2015; 212:141–145.
14. Porter J, Herring J, Lacroix J, Levinton C. Systematic review of Telemonitoring in COPD: an Update. *Pneumonologia Alergologia Polska.* 2015; 83(6):476–484. <https://doi.org/10.5603/PiAP.2015.0077>
15. Vianello A, Fusello M, Gubian L, Rinaldo C, Dario C, Concas A, et al. Home telemonitoring for patients with acute exacerbation of Chronic Obstructive Pulmonary Disease: a randomized controlled trial. *BMC Pulmonary Medicine.* 2016; 16(157):1–12.
16. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. *BMJ.* 2013; 347:1–16. <https://doi.org/10.1136/bmj.f6070>
17. Makris S, Bouros D. COPD Exacerbation: Lost in Translation. *BMC Pulmonary Medicine.* 2009; 9(6):1–3.
18. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. *Am J Respir Crit Care Med.* 2001; 163:1256–1276. <https://doi.org/10.1164/ajrccm.163.5.2101039> PMID: 11316667
19. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS ONE.* 2017; 12(4). <https://doi.org/10.1371/journal.pone.0174944>
20. Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. *Ann N Y Acad Sci.* 2017; 1387(1):153–165. <https://doi.org/10.1111/nyas.13218> PMID: 27627195
21. Song X, Mitnitski A, Cox J, Rockwood K. Comparison of machine learning techniques with classical statistical models in predicting health outcomes. *Stud Health Technol Inform.* 2004; 107(1):736–740. PMID: 15360910
22. Balkissoon R, Lommatsch S, Carolan B, Make B. Chronic Obstructive Pulmonary Disease: a concise review. *Medical Clinics of North America.* 2011; 95(6):1125–1141. <https://doi.org/10.1016/j.mcna.2011.08.009> PMID: 22032431
23. Bahadori K, Fitzgerald JM. Risk factors of hospitalization and re-admission of patients with COPD exacerbation a systematic review. *Medical Clinics of North America.* 2007; 2(3):241–251.
24. Stiell IG, Clement CM, Aaron SD, Rowe BH, Perry JJ, Brison RJ, et al. Clinical Characteristics Associated with Adverse Events in Patients with Exacerbation of Chronic Obstructive Pulmonary Disease: a Prospective Cohort Study. *CMAJ.* 2014; 186(6):193–204. <https://doi.org/10.1503/cmaj.130968>
25. Roche N, Zureik M, Soussan D, Neukirch F, Perrotin D, the Urgence BPCO (COPD Emergency) Scientific Committee and investigators. Predictors of outcomes in COPD exacerbation cases presenting to the emergency department. *European Respiratory Journal.* 2008; 32(4):953–961. <https://doi.org/10.1183/09031936.00129507> PMID: 18508819
26. Singanayagam A, Schembri S, Chalmers JD. Predictors of mortality in hospitalized adults with acute exacerbation of Chronic Obstructive Pulmonary Disease: A systematic review and meta-analysis. *Ann Am Thorac Soc.* 2013; 10(2):81–89. <https://doi.org/10.1513/AnnalsATS.201208-043OC>
27. Velickovski F, Ceccaroni L, Roca J, Burgos F, Galdiz JB, Marina N, et al. Clinical decision support systems (CDSS) for preventive management of COPD patients. *J Translational Medicine.* 2014; 12(2):81–89.
28. Henny S, Dieterle T, Brack T, Brutsche M, Clarenbach C, Irani S, et al. Clinical characteristics of patients with frequent COPD exacerbations in a Swiss multicenter COPD cohort study (TOPDOCS). *European Respiratory Journal.* 2016; 48.

29. Urwyler P, Hussein NA, Brudevaux PO, Chhajed PN, Dieterle T, Fitting JW, et al. Predictive factors for exacerbation and re-exacerbation in COPD: Results from the Swiss COPD cohort. *European Respiratory Journal*. 2016; 48.
30. Lange P, Marott JL, Vestbo J, Nordestgaard BG. Prevalence of Night-Time Dyspnoea in COPD and its Implications for Prognosis. *European Respiratory Journal*. 2014; 43:1590–1598. <https://doi.org/10.1183/09031936.00196713> PMID: 24488571
31. Arnaud CavaillÃšs A, Brinchault-Rabin G, Dixmier A, Goupl G, Gut-Gobert C, Marchand-Adam S, et al. Comorbidities of COPD. *European Respiratory Review*. 2013; 22:454–475. <https://doi.org/10.1183/09059180.00008612>
32. Vestbo J, Hurd SS, AgustÃ AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *CMAJ*. 2013; 187(4):347–365.
33. Launois C, Barbe C, Bertin E, Nardi J, Perotin JM, Dury S, et al. The Modified Medical Research Council Scale for the Assessment of Dyspnea in Daily Living in besity: a Pilot Study. *BMC Pulmonary Medicine*. 2012; 12(61):1–7.
34. Sobnath DD, Philip N, Kayyali R, Gebara SN, Pierscionek B, Vaes AW, et al. Clinical Characteristics Associated with Adverse Events in Patients with Exacerbation of Chronic Obstructive Pulmonary Disease: a Prospective Cohort Study. *JMIR Mhealth Uhealth*. 2017; 5(2):1–11.
35. American Thoracic Society. COPD Assessment Test (CAT); 2016.
36. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. *European Respiratory Review*. 2017; 26:1–13. <https://doi.org/10.1183/16000617.0061-2016>
37. Kerkhof M, Freeman D, Jones R, Chisholm A, Price DB. Predicting Frequent COPD Exacerbations Using Primary Care Data. *Int J COPD*. 2015; 10:2439–2450.
38. Alwashmi M, Hawboldt J, Davis E, Marra C, Gamble JM, Ashour WA. The Effect of Smartphone Interventions on Patients With Chronic Obstructive Pulmonary Disease Exacerbations: A Systematic Review and Meta-Analysis. *JMIR Mhealth and Uhealth*. 2016; 4(3). <https://doi.org/10.2196/mhealth.5921> PMID: 27589898
39. Canadian Lung Association. Mobile App to Manage COPD. *BCMJ*. 2016; 58(1):39.
40. Mt Sinai & National Jewish Health Respiratory Institute. COPD Navigator; 2016. “<http://www.lifemap-solutions.com/products/cupd-navigator/>”.
41. DocTot and Global Initiative for Chronic Obstructive Lung Disease. GOLD COPD Strategy; 2016. “<http://www.doctot.com/clinical-diagnosis-apps-paid>”.
42. Gupta SS, Gothi D, Narula G, Sircar J. Correlation of BMI and Oxygen Saturation in Stable COPD in Northern India. *Lung India*. 2017; 31(1):29–34. <https://doi.org/10.4103/0970-2113.125891>
43. Ardestani ME, Abbaszadeh M. Telemedicine in Chronic Obstructive Pulmonary Disease. *J Res Med Sci*. 2014; 19(3):257–261.
44. Falaschetti E, Laiho J, Primatesta P, Purdon S. Prediction equations for normal and low lung function from the Health Survey for England. *European Respiratory Journal*. 2004; 23:456–463. <https://doi.org/10.1183/09031936.04.00055204> PMID: 15065839
45. McSharry DG, Ryan S, Calverly P, Eewards JC, McNicholas WT. Sleep quality in chronic obstructive pulmonary disease. *Respirology*. 2012; 17:1119–1124. <https://doi.org/10.1111/j.1440-1843.2012.02217.x> PMID: 22758620
46. Hamine S, Guyette EG, Faulk D, Green BB, Ginsburg AS. Impact of mHealth Chronic Disease Management on Treatment Adherence and Patient Outcomes: A Systematic Review. *J Med Internet Res*. 2015; 17(2):e52. <https://doi.org/10.2196/jmir.3951> PMID: 25803266
47. World Health Organization. Second Global Survey on eHealth (Global Observatory for eHealth) Geneva: World Health Organization; 2011.
48. Keene JD, Jacobson S, Kechris K, Kinney GL, Foreman MG, Doerschuk CM, et al. Biomarkers predictive of exacerbations in the SPIROMICS and COPDGene cohorts. *AJRCCM*. 2017; 195(4).
49. Watz H, Tetzlaff K, Wouters E, Kirsten A, Magnussen Hea. Blood eosinophil count and exacerbations in severe chronic obstructive pulmonary disease after withdrawal of inhaled corticosteroids: a post-hoc analysis of the WISDOM trial. *Lancet Respir Med*. 2016; 4(5). [https://doi.org/10.1016/S2213-2600\(16\)00100-4](https://doi.org/10.1016/S2213-2600(16)00100-4) PMID: 27066739