

Modeling and Classifying Conditional Flare-Ups in Patients with Respiratory Diseases

Aakash Bhattacharya¹; Siddhartha Singh²

^{1,2} Amazon Web Services Inc. (*)
Seattle, Washington, U.S.A.

Publication Date: 2025/09/26

Abstract: Respiratory illnesses, including Acute Respiratory Distress Syndrome (ARDS), asthma, cystic fibrosis, and Chronic Obstructive Pulmonary Disease (COPD), represent an escalating global health concern, impacting a substantial segment of the population and ranking among the foremost causes of mortality worldwide. A critical aspect of managing these conditions is the occurrence of flare-ups, which mark a sudden and severe aggravation of symptoms like shortness of breath, palpitation, and persistent cough, demanding urgent medical care. These acute episodes are often triggered by airway inflammation (bronchitis), alveolar damage (emphysema), and exposure to various environmental irritants such as dust, smoke, chemicals, and fumes. To address this clinical challenge, a model has been presented that aims to uncover the most influential causative factors behind flare-ups and accurately classify a patient's risk of experiencing them based on distinct prognostic markers.

Keywords: Acute Respiratory Distress Syndrome (ARDS), Chronic Pulmonary Obstructive Disease (COPD), Ensembling, Genetic Algorithm, Neural Networks, Random Forest.

How to Cite: Aakash Bhattacharya; Siddhartha Singh (2025) Modeling and Classifying Conditional Flare-Ups in Patients with Respiratory Diseases. *International Journal of Innovative Science and Research Technology*, 10(9), 1563-1570.
<https://doi.org/10.38124/ijisrt/25sep783>

I. INTRODUCTION

Respiratory diseases are an important cause of disease and death worldwide. Over the years, the number of hospital admissions has grown due to respiratory diseases. In 2021, Chronic Obstructive Pulmonary Disease (COPD) was responsible for approximately 3.5 million deaths globally, representing 5% of all worldwide mortality and positioning it as the fourth leading cause of death (World Health Organization, 2023). It is estimated that 262 million people suffer from asthma (World Health Organization, 2024), more than 200 million people have COPD, 65 million endure moderate-to-severe COPD (World Health Organization - Burden of COPD, 2014), 1-6% of the adult population (more than 100 million people) experience sleep disordered breathing (World Health Organization, 2014), 8.7 million people develop Tuberculosis (TB) annually (World Health Organization, 2018), millions live with pulmonary hypertension (World Health Organization, 2014) and more than 50 million people struggle with occupational lung diseases, totaling more than 1 billion people suffering from chronic respiratory conditions (World Health Organization, 2018). Globally, at least two billion individuals face exposure to the toxic effects of biomass fuel consumption, with an additional one billion exposed to outdoor air pollution and another one billion to tobacco smoke.

Each year, 4 million people die prematurely from chronic respiratory disease (World Health Organization, 2018). Flare-ups result in sudden worsening of symptoms due to exposure to causative factors, often leading to hospital admissions and, in some cases, to death.

This manuscript reports a novel analysis to identify the prognostic markers leading to flare-ups. A model has been trained to classify a patient's probability of having flare-ups based on the reduced factors. A case study has been performed to deduce the dominant factors causing flare-ups using various machine learning and statistical techniques.

This paper is organized into seven sections. Section I introduces the research presented. Section II then provides a concise overview of the paper's main objective. Next, Section III reviews existing literature on respiratory diseases, focusing on flare-ups and their contributing factors. Section IV meticulously details the methodology employed to identify prognostic markers for flare-ups and to classify a patient's risk of experiencing them. A case study illustrating the analysis is presented in Section V. The key findings and results are then discussed in Section VI, and Section VII concludes the paper, also outlining future research directions.

II. OBJECTIVE

This paper presents a case study of patients with respiratory diseases, identifying the dominant factors contributing to respiratory flare-ups. A classification model has been proposed that predicts a patient's likelihood of experiencing flare-ups based on a reduced set of key features. This streamlined feature set not only enhances model efficiency but also serves as crucial prognostic markers, directly highlighting the factors most impactful to flare-up occurrences. Ultimately, this model offers a valuable tool for clinicians, enabling them to proactively suggest personalized precautionary measures to high-risk patients. From a pulmonologist's perspective, this accelerates the diagnostic process and significantly advances patient care.

III. RELATED WORK

In the USA, COPD affects approximately 24 million Americans, results in about 155,000 deaths a year, and is now the fourth leading cause of death (Xu et al., 2018). COPD poses a substantial threat to the healthcare industry worldwide; it is one of the major causes of morbidity, mortality, and poor health status (Donaldson et al., 2002). Exacerbation contributes to dangerous consequences like premature death (PasseportSante, 2019), abasement of quality of life (Naim et al., 2007), regression of respiratory functions (Burt and Corbridge, 2013). To date, there is no cure for exacerbation resulting from COPD. The existing medication distends the bronchi, allowing for greater oxygen absorption by alveoli (Soci' e Canadienne de Thoracologie, 2010). As we continue to reap the consequences of the 20th century tobacco smoking, COPD remains one of the prevalent diseases (Chapman et al., 2006). Cigarette smoking is the leading cause of COPD worldwide (Agarwal et al., 2023).

Under-diagnosis of COPD due to negligence in African-Americans is commonly observed (Chapman, 2018). Azithromycin is sometimes used in COPD patients to prevent acute exacerbations. Recent studies have demonstrated that azithromycin helps prevent COPD exacerbations (Uzun et al., 2014) (Albert et al., 2011). While azithromycin helps with COPD exacerbations, we need a study to analyze the causative factors causing exacerbations so that they can be independently targeted. A recent projection from the World Health Organization shows that COPD caused 3.5 million deaths in 2021, which was approximately 5% of the total global death count, making it the 4th leading cause of fatality worldwide. Also, COPD is the 8th leading cause for poor health globally (World Health Organization, 2024).

Context Relevant Prediction models using Bayesian Belief Network have been developed to predict the exacerbation in COPD with an accuracy of 81.5% using 17 relevant attributes (Hamid et al., 2017). An Artificial Neural Network (ANN) model has been developed which predicted an exacerbation within a 13-day window frame with an Area Under Curve (AUC) 0.89 (95% Confidence Interval 0.89–0.90) and identified an exacerbation median (interquartile range) 7 (5 to 9) days before clinical diagnosis (Yousuf et al., 2024).

Coronavirus Disease (COVID) has also impacted COPD adversely. The incidence of hospitalization and disease severity in patients with prior respiratory diseases, such as COPD, is much higher in patients with COVID-19 than with seasonal influenza (Beltramo et al., 2021).

IV. METHODS

The analysis utilized a dataset comprising 1985 observations across 62 distinct attributes. This dataset is partitioned into a training set (1687 entries) and a testing set (298 entries), with the latter intentionally excluding the response variable. The dependent variable, designated as 'Flare_Up', is binary categorical, where a value of '1' denotes the presence of a flare-up and '0' indicates its absence. Figure 1 outlines the comprehensive workflow for classifying the 'Flare_Up' condition. The methodological approach, consistent with previous work (Bhattacharya et al., 2019) [Unpublished Work], broadly encompasses data pre-processing, feature selection, classifier training, and subsequent prediction of the response.

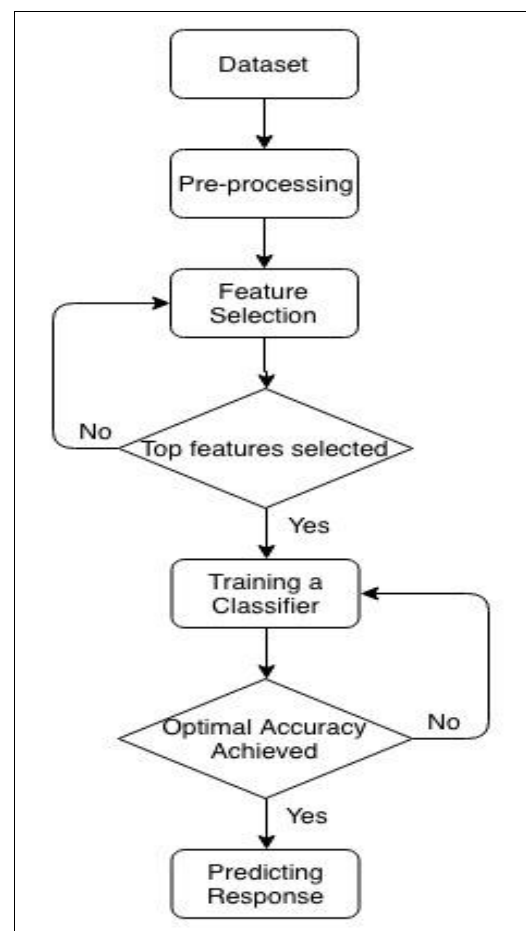


Fig 1. Framework for Classifying 'Flare Up' Condition

As a prerequisite, data pre-processing has been conducted to suitably condition the dataset for analysis. No missing values have been identified across any attributes, and continuous data have already been normalized to a standardized range of [0, 1]. Pearson's product-moment correlation coefficient has been computed for all continuous variables to mitigate redundancy and reduce model complexity. Variables exhibiting a correlation magnitude exceeding 0.85 have been subsequently

eliminated (Figure 2). The resulting filtered data has then been supplied as input to the feature selection phase, where the most salient variables for classifier training have been identified. Model development has incorporated repeated K-fold cross-validation ($K = 10$) combined with resampling techniques, and classification accuracy has been derived from the contingency table. The methodological approach employed in this manuscript has aligned with that previously reported in (Bhattacharya et al., 2017).

```
> corr_removed
[1] "LungFun9" "LungFun13" "LungFun14" "LungFun3" "RespQues1" "LungFun11" "LungFun5" "LungFun20" "LungFun18" "Demo4" "LungFun15"
```

Fig 2. Variables removed after correlation analysis

A. Feature Selection

Feature selection algorithms are crucial for enhancing model performance by eliminating irrelevant or redundant attributes that can negatively impact accuracy and efficiency. A reduced feature set not only enables more efficient model training but also significantly mitigates overfitting and improves the overall generalizability of the model. For this analysis, two distinct feature selection algorithms have been employed: Genetic Algorithm and Simulated Annealing. These methods are further explicated in the subsequent subsection.

➤ Simulated Annealing

Simulated Annealing (SA) is a meta-heuristic global optimization technique that employs a probabilistic approach to efficiently explore large search spaces. As a subset-selection-based algorithm, it evaluates a different subset of features in each iteration.

For our analysis, Simulated Annealing has been integrated as a wrapper around a Random Forest model. Compared to a Genetic Algorithm, SA's less aggressive search strategy aids in reducing overfitting. However, a potential drawback of Simulated Annealing is its probabilistic nature, which means it may not always converge to the global optimum when optimizing the objective function. Figure 3 displays the top 23 features identified by Simulated Annealing.

• Algorithm:

- ✓ An initial subset of features is chosen randomly.
- ✓ Evaluate the performance of the subset based on an objective function.
- ✓ Small perturbations are added to the initial candidate solution, and the performance of the new candidate set is evaluated.
- ✓ If the performance of the new candidate is greater than the previous candidate, it is selected for the next iteration.
- ✓ Else, the probability of accepting a move is calculated based on the two performance values and the current iteration of the algorithm. This probability decreases exponentially with the badness of the move.
- ✓ Steps 3-5 are cyclically executed until the predetermined iteration count is reached. The optimal solution is attained eventually from these suboptimal solutions, with the fact that the solution improves in the subsequent iterations.

```
> rf_sa$optVariables
[1] "Demo6" "DisHis3Times" "LungFun2" "LungFun10"
[5] "LungFun12" "LungFun16" "Dis3Times" "ResQues1a"
[9] "ResQues1b" "ResQues1c" "SmokHis1" "SmokHis3"
[13] "SmokHis4" "Demo1" "Demo5" "DisHis1"
[17] "DisHis3" "Dis1Treat" "Dis2" "Dis4"
[21] "Dis4Treat" "Dis5Treat" "Dis6"
```

Fig 3. Features Selected by Stimulated Annealing

➤ Genetic Algorithm

The Genetic Algorithm (GA) is a powerful subset-selection based feature extraction algorithm, drawing its principles from biological evolution and natural genetics. Its operation is fundamentally governed by a cyclical process involving population selection, fitness allocation, crossover, and mutation. By accurately approximating important features, the Genetic Algorithm consistently generates precise model estimates. Figure 4 illustrates the basic flowchart of the genetic algorithm.

• Algorithm:

- ✓ A random initial population is selected.
- ✓ A rank-based method is used to evaluate the fitness of each individual.
- ✓ An individual is selected having the highest rank, i.e., an individual possessing the least selection error.
- ✓ New offspring are produced using the pre-defined mutation and crossover probabilities.
- ✓ Fitness of the new offspring is evaluated.
- ✓ The unhealthiest individual in the population is replaced by the higher-ranking individual.
- ✓ Steps 3 to 6 are repeated until the specified number of generations is reached.

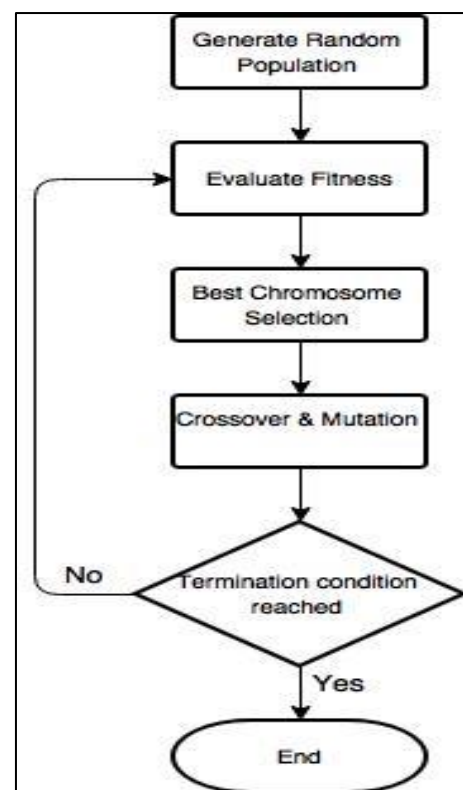


Fig 4. Genetic Algorithm Flowchart

For our analysis, a Genetic Algorithm (GA) has been employed as a wrapper around a Random Forest model. Optimal feature selection has been achieved using a crossover probability of 0.8 and a mutation probability of 0.1. The GA has been configured with a population size of 50 per generation and allowed to run for 250 generations. This process identified 29 top features, as depicted in Figure 5, which yielded a training accuracy of 87.55%.

```
> ga_model$optVariables
[1] "Demo2" "Demo6" "LungFun1" "LungFun6" "LungFun8"
[6] "LungFun12" "LungFun16" "LungFun17" "Dis2Times" "Dis3Times"
[11] "ResQues1a" "ResQues1b" "ResQues1c" "ResQues2a" "SmokHis1"
[16] "SmokHis2" "SmokHis4" "Demo1" "Demo5" "DisHis1"
[21] "DisHis3" "DisHis7" "DisStage1" "DisStage2" "LungFun19"
[26] "Dis1" "Dis1Treat" "Dis4" "Dis7"
```

Fig 5. Features Selected by Genetic Algorithm

The genetic algorithm typically overfits small datasets. However, our substantial training set is adequate for deriving optimal, well-generalized results without overfitting. Although fewer generations could further reduce overfitting, this trade-off would render the results impractical.

B. Classification

A primary objective of this analysis is constructing a model capable of classifying a patient's likelihood of experiencing flare-ups. Since 'Flare_Up' is a binary variable, various classification techniques have been employed to achieve optimal accuracy. Each of these methods is discussed subsequently.

➤ Naïve Bayes Classifier

The Naive Bayes Classifier operates on the principles of Bayes' theorem of probability to predict dataset classes. It stands as one of the fastest classification algorithms available today. Naive Bayes assumes the independence among the predictors. Essentially, Bayes' theorem offers a method for determining the posterior probability from the probability of the class, the probability of the predictor, and the likelihood of the predictor given the class.

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)}$$

Here,

$P(c|x)$ is the posterior probability of class given predictor $P(c)$ is the prior probability of the class
 $P(x|c)$ is the probability of predictor given class
 $P(x)$ is the prior probability of the predictor

➤ Decision Tree

A decision tree is a classification and prediction technique that uses tree-like structures to categorize or forecast values based on a scoring measure. In these structures, internal nodes represent tests on an attribute, branches extending from nodes depict specific outputs, and leaf nodes signify an outcome. A splitting criterion is responsible for dividing a node, which can be discrete, continuous, or binary. This splitting occurs recursively across multiple levels, hence the term recursive partitioning of trees (rpart). For our analysis,

we've utilized both rpart and classification trees (ctree) to classify data into discrete binary values.

➤ Random Forest

Random Forest is a versatile classification and regression approach that functions as an ensemble method built upon Classification and Regression Trees (CART). It operates by constructing multiple decision trees, each trained on diverse data samples and distinct subsets of initial variables. The final classification or regression is determined by aggregating the results from these numerous trees, a process repeated for a specified number of iterations.

• Algorithm:

- ✓ A decision tree is constructed randomly by selecting x features and y samples from the dataset with replacement.
- ✓ Net error is minimized by tree pruning.
- ✓ Steps 1 and 2 are repeated for a defined number of iterations.
- ✓ The final classification is based on the average of all the probability scores in the intermediate steps.

➤ Neural Network

Artificial neural networks (ANNs) mimic the biological nervous system's approach to information processing. Essentially, an ANN is an interconnected system of "neurons" that work simultaneously to solve a predefined problem. Neural networks are renowned for their ability to extract insights from imprecise data, often surpassing what humans or other computational methods can achieve manually.

Each neuron within the network has an input and an output. An activation function determines the neuron's output by processing its weighted input. In a multi-layer network, the output from one neuron then serves as an input for subsequent neurons, with the weights associated with each connection being updated through a learning process. Figure 6 illustrates a typical neural network featuring a single hidden layer. For this analysis, we have employed a back-propagation neural network model.

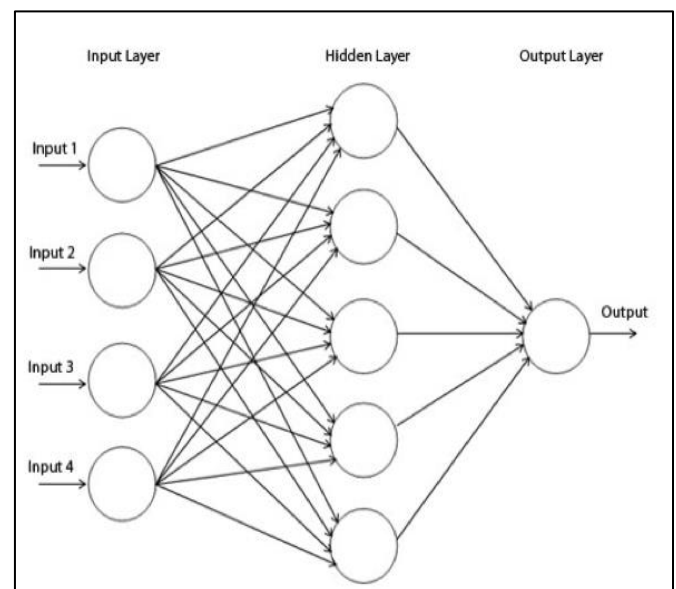


Fig 6. A Neural Network Model

➤ *Ada Boost/Bag*

Ensemble learning is a powerful technique that combines multiple algorithms, known as base learners, to create a more robust model. This approach typically results in higher accuracy and a reduced generalization error. The most popular types of ensemble techniques are bagging, boosting, and stacking, each differing in their implementation, and these are explained subsequently.

Bagging (Bootstrap Aggregating) essentially involves averaging the results from different bootstrapped samples fed into an algorithm. While averaging is common, weighted averages or majority voting can also be used. Bagging primarily serves to reduce the variance of a model. A classic example of bagging is the Random Forest algorithm, which utilizes variations of trees to make predictions.

Boosting is a sequential learning technique where the first algorithm trains on the entire dataset, and subsequent algorithms build upon the performance of the preceding model. It specifically assigns higher weights to observations that were not accurately predicted by the previous model, effectively building a new model using the residuals from the prior one.

AdaBoost (Adaptive Boosting) is one of the pioneering boosting algorithms developed for classification problems. Its core focus is on converting a set of weak classifiers into a single strong classifier. The generalized equation representing this classification algorithm is:

$$F(x) = \sum_{n=1}^S \theta_n f_n(x)$$

Here, f_n denotes the n^{th} weak classifier and θ_n is the weight associated with it, making $F(x)$ the weighted sum of all the weak classifiers.

• *Algorithm:*

- ✓ Initialize the weights for each data point in the dataset to $w_i = 1/n$, where n is the total number of data points.
- ✓ The dataset is then fitted to weak classifiers, and the classifier resulting in the least weighted classification error is selected.
- ✓ The weight of the n^{th} weak classifier is calculated as:
- ✓ $\theta_n = \frac{1}{2} \ln\left(\frac{1-\epsilon_n}{\epsilon_n}\right)$, where ϵ_n is the error associated with the n^{th} classifier.
- ✓ The weight for each data point is updated using an equation that also uses a normalization factor so that the sum of all the weights is equal to 1.
- ✓ Steps 2 to 4 are repeated for all n weak classifiers.

For the purpose of this analysis, both AdaBoost and its modified variation, AdaBag has been used. While AdaBoost employs a boosting approach, AdaBag incorporates bagging for its ensemble learning.

V. CASE STUDY

The dataset comprises 1985 observations with 62 distinct attributes. A detailed description of these attribute types is provided in Table 1. The response variable, 'Flare_Up', is categorical and binary, taking values of '1' to signify a patient experiencing a flare-up and '0' for its absence. Model accuracy has been evaluated using a contingency table.

As a preprocessing step, a few continuous features have been removed from the dataset due to their high correlation with other variables. Subsequently, feature selection algorithms have been employed to extract the most relevant features for classification. These selected features function as the prognostic markers for the 'Flare_Up' condition. Utilizing these top extracted features, various classification models have been built, resulting in increased accuracy.

Table 1. Attribute Description

Attribute	Description
Flare_Up	Patient experiences flare up or not
Demographics	Age, Gender, Height, Weight, Race and Nationality
Disease Stage	Severity of the disease (higher implies more severe)
Lung Function	Lung function measurements derived from spirometry
Disease History	Specific history of lung disease symptoms
Other Lung Diseases	History of cough, asthma and other lung ailments
Respiratory Questionnaire	Results of questionnaires to measure overall well-being in patients
Smoking History	Patients smoking habits

The dataset comprises 1985 observations with 62 distinct attributes. A detailed description of these attribute types is provided in Table 1. The response variable, 'Flare_Up', is categorical and binary, taking values of '1' to signify a patient experiencing a flare-up and '0' for its absence. Model accuracy has been evaluated using a contingency table.

As a preprocessing step, a few continuous features have been removed from the dataset due to their high correlation with other variables. Subsequently, feature selection algorithms have been employed to extract the most relevant features for classification. These selected features function as the prognostic markers for the 'Flare_Up' condition. Utilizing these top extracted features, various classification models have been built, resulting in increased accuracy.

VI. RESULTS

Feature selection has been performed using both simulated annealing and a genetic algorithm. Of these, the genetic algorithm demonstrated greater efficacy, with its derived feature set proving to be more indicative of flare-up prognostics. The prominent features identified by the genetic algorithm are presented in Figure 5.

In our investigation, various classification models were developed to predict a patient's propensity for flare-ups. The Naive Bayes classifier yielded the lowest accuracy at 81.39%. While the Decision Tree model offered a respectable accuracy of 86.54%, the Random Forest algorithm demonstrated superior performance. This ensemble method enhances predictive power by constructing multiple decision trees from diverse data subsets and aggregating their outputs, which ultimately led to an improved accuracy of 87.73%. Furthermore, the Random Forest's inherent ability to reduce variance and overcome the problem of overfitting, commonly associated with single decision trees, made it a more robust solution for this classification task.

Table 2. Classification Results

Model	Accuracy
Naive Bayes Classifier	81.39%
Recursive Partitioning of Trees	86.54%
Classification Tree	86.19%
Random Forest	87.73%
Neural Networks	87.01%
Model Average Neural Networks	87.25%
AdaBoost	88.20%
AdaBag	87.07%
Boosted Classification Trees	87.66%

An Artificial Neural Network (ANN) model has then been used for the classification problem. A back-propagation neural network with five hidden layers has been trained, observing that deeper architectures led to overfitting. We also explored a model-averaged neural network, which enhances generalization by averaging predictions from models trained with different random initializations; this model achieved 87.25% accuracy.

Beyond individual ANNs, ensemble methods substantially improved predictive accuracy. Specifically, AdaBoost, an algorithm based on boosting, achieved an accuracy of 88.20%. AdaBag, which employs bagging, yielded 87.5% accuracy. Additionally, boosted classification trees demonstrated strong performance with an accuracy of 87.66%. The performance metrics for all models are detailed in Table 2.

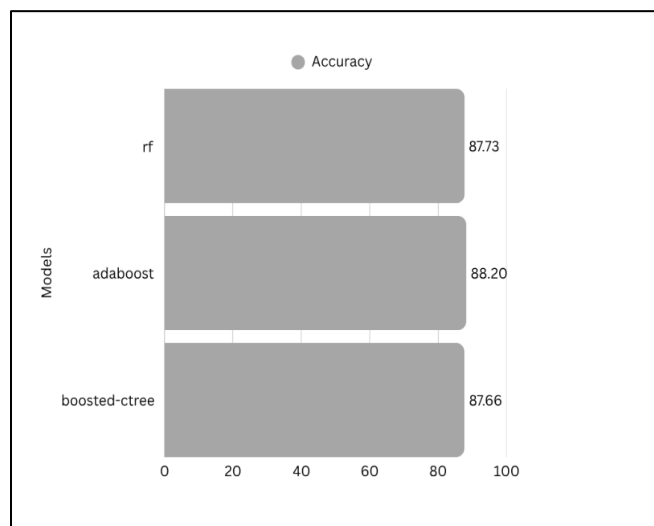


Fig 7. Top 3 Models

Among the models evaluated, AdaBoost, Random Forest, and Boosted Classification Trees emerged as the top three classifiers. Figure 7 visually represents the accuracy of these superior models for predicting a patient's propensity for flare-ups.

VII. CONCLUSION AND FUTURE WORK

This research effectively employed various machine learning techniques and statistical methods to identify the dominant factors contributing to respiratory flare-ups and develop a model capable of accurately classifying a patient's tendency for these events. The Genetic Algorithm proved instrumental in selecting key features that serve as crucial prognostic markers for flare-up conditions. Among the numerous classification models explored, AdaBoost consistently yielded the highest accuracy. This predictive model offers a valuable tool for medical practitioners, enabling them to forecast a patient's likelihood of experiencing flare-ups based on readily available symptoms and spirometry data. This capability allows for the proactive implementation of precautionary measures, ultimately improving patient care and potentially reducing acute exacerbations.

To further advance this research and enhance the model's clinical utility, several key areas warrant future investigation:

- **Dataset Expansion and Enrichment:** Expanding to a substantially larger and more diverse dataset is crucial for improving the model's robustness and generalizability. Future efforts should focus on integrating additional features, particularly those related to genetic predispositions or imperfections, which could offer deeper insights into individual susceptibility to flare-ups.
- **In-depth Feature Characterization:** A comprehensive descriptive study of each identified feature is essential. This would involve a detailed evaluation of their individual and synergistic impact on flare-up conditions, leading to a more profound understanding of the underlying pathophysiological mechanisms.
- **Exploration of Deep Learning Architectures:** While traditional machine learning models showed promising results, applying deep learning techniques to this dataset should be explored. These advanced architectures may uncover more complex and nuanced patterns, further enhancing the model's predictive power and generalization capabilities across diverse patient populations.
- **Prospective Clinical Validation:** To affirm the real-world applicability and predictive accuracy of the model, future work must include rigorous prospective validation using new, unseen patient cohorts over an extended period. This crucial step will establish its clinical utility and reliability.
- **Integration with Electronic Health Records (EHR):** Investigating the seamless integration of this predictive model directly into existing Electronic Health Record (EHR) systems could revolutionize clinical workflow. Such integration would enable real-time risk assessment, making the model more accessible and actionable for healthcare providers at the point of care.
- **Development of Personalized Intervention Strategies:** Building upon the model's predictive capabilities, future research could focus on designing and evaluating

personalized intervention strategies tailored to individual patient risk profiles. This would transition the work from mere prediction to actionable clinical guidelines for precision medicine.

- Health Economic Analysis: A thorough cost-benefit analysis of implementing this predictive model in routine clinical practice would be highly beneficial. This analysis would quantify the potential healthcare savings through reduced hospitalizations and emergency visits versus the investment in technology and clinician training.

REFERENCES

- [1]. Agarwal, A. K., A. Raja, and B. D. Brown (2023, August). Chronic obstructive pulmonary disease. In StatPearls. Treasure Island, FL: StatPearls Publishing. Updated 2023 Aug 7. <https://www.ncbi.nlm.nih.gov/books/NBK559281/>
- [2]. Albert, R. K., J. Connett, W. C. Bailey, R. Casaburi, J. A. Cooper, G. J. Criner, J. L. Curtis, M. T. Dransfield, M. K. Han, S. C. Lazarus, B. Make, N. Marchetti, F. J. Martinez, N. E. Madinger, C. McEvoy, D. E. Niewoehner, J. Porsasz, C. S. Price, J. Reilly, P. D. Scanlon, F. C. Sciurba, S. M. Scharf, G. R. Washko, P. G. Woodruff, and N. R. Anthonisen (2011, August). Azithromycin for prevention of exacerbations of COPD. *New England Journal of Medicine* 365 (8), 689–698. COPD Clinical Research Network. <https://doi.org/10.1056/NEJMoa1104623>
- [3]. Beltramo, G., J. Cottenet, A.-S. Mariet, M. Georges, L. Piroth, P. Tubert-Bitter, P. Bonniaud, and C. Quantin (2021). Chronic respiratory diseases are predictors of severe outcome in COVID-19 hospitalised patients: A nationwide study. *European Respiratory Journal* 58, 2004474. <https://doi.org/10.1183/13993003.04474-2020>
- [4]. Bhattacharya, A., R. Khatri, and T. Pradhan (2017, September). A proactive intelligent model for identification of reduced factors affecting severity of coronary artery disease and its prediction. In *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udipi, India, pp. 2032–2038. IEEE. <https://doi.org/10.1109/ICACCI.2017.8126056>
- [5]. Bhattacharya, A., S. Singh, and S. Sinha (2019, 04). Factors affecting conditional flare-ups in patients with respiratory diseases and its classification (unpublished work).
- [6]. Burt, L. and S. Corbridge (2013). Copd exacerbations. *American Journal of Nursing* 113, 34–43. <https://doi.org/10.1097/01.naj.0000426688.96330.60>
- [7]. Chapman, K. R. (2018). Increasing awareness of copd: two steps forward, one step back. *Chronic Obstructive Pulmonary Disease* 5 (4), 228–230. editorial. <https://doi.org/10.15326/jcopdf.5.4.2018.0154>
- [8]. Chapman, K. R., D. M. Mannino, J. B. Soriano, P. A. Vermeire, A. S. Buist, M. J. Thun, C. Connell, A. Jemal, T. A. Lee, M. Miravittles, S. Aldington, and R. Beasley (2006). Epidemiology and costs of chronic obstructive pulmonary disease. *European Respiratory Journal* 27(1), 188–207. <https://doi.org/10.1183/09031936.06.00024505>
- [9]. Donaldson, G. C., T. A. R. Seemungal, A. Bhowmik, and J. A. Wedzicha (2002). Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. *Thorax* 57(10), 847–852. <https://doi.org/10.1136/thorax.57.10.847>
- [10]. Hamid, M., S. Lokman, A. Hicham, and M. Hafedh (2017). Context relevant prediction model for copd domain using bayesian belief network. *Sensors* 17(7), 1486. <https://doi.org/10.3390/s17071486>
- [11]. Naim, P., P. H. Wuillemin, P. Leray, O. Pourret, and A. Becker (2007). *Reseaux Bayesiens*. Paris, France: Eyrolles. [Publication in French]. https://www.researchgate.net/publication/234882987_Reseaux_Bayesiens
- [12]. PasseportSante (2019). Bronchite et emphyseme. Last accessed 15 Jan 2019. [Source in French].
- [13]. Soci  t   Canadienne de Thoracologie (2010). *Le fardeau humain et financier de la mpoc*. Technical report, Soci  t   Canadienne de Thoracologie, Ottawa, ON, Canada. [Publication in French]. PUBNUM: 221681. OCLC: (OCoLC)758215903. https://librarysearch.mtroyal.ca/permalink/01MTROYAL_INST/o62j6r/alma9923218202704656
- [14]. Uzun, S., R. S. Djamin, J. A. J. W. Kluytmans, P. G. H. Mulder, N. E. van't Veer, A. A. M. Ermens, A. J. Pelle, H. C. Hoogsteden, J. G. J. V. Aerts, and M. M. van der Eerden (2014, May). Azithromycin maintenance treatment in patients with frequent exacerbations of chronic obstructive pulmonary disease (COLUMBUS): a randomised, double-blind, placebo-controlled trial. *Lancet Respiratory Medicine* 2(5), 361–368. [https://doi.org/10.1016/s2213-2600\(14\)70019-0](https://doi.org/10.1016/s2213-2600(14)70019-0)
- [15]. World Health Organization (2014). *Chronic Respiratory Diseases*. Accessed: 05-Oct-2018. www.who.int/gard/publications/chronic_respiratory_diseases.pdf
- [16]. World Health Organization (2014). *Chronic Respiratory Diseases - Burden of COPD*. Accessed: 05-Oct-2018. www.who.int/respiratory/copd/burden/en/index.html
- [17]. World Health Organization (2018). *Global Tuberculosis Report*. Accessed: 11-Jun-2025. <https://www.who.int/publications/i/item/9789241565646>
- [18]. World Health Organization (2024). *Chronic Obstructive Pulmonary Disease (COPD). Fact sheet*. Updated 2023. Accessed 20-May-2025. [https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd))
- [19]. Centers for Disease Control and Prevention (2018). *Asthma: Key Facts*. Accessed: 20-May-2025. Xu, J., S. L. Murphy, K. D. Kochanek, B. Bastian, and E. Arias (2018, July). Deaths: Final data for 2016. *National Vital Statistics Reports* 67 (5), 1–76. https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_05.pdf

- [20]. Yousuf, A. J., G. Parekh, M. Farrow, G. Ball, S. Graziadio, K. Wilson, C. Lendrem, L. Carr, L. Watson, S. Parker, J. Finch, S. Glover, V. Mistry, K. Porter, A. Duvoix, L. O'Brien, S. Rees, K. E. Lewis, P. Davis, and C. E. Brightling (2024). Artificial neural network risk prediction of chronic obstructive pulmonary disease (COPD) exacerbations using urine biomarkers. *ERJ Open Research Early Access*, 00797–2024. Published Online: January 31, 2024. <https://doi.org/10.1183/23120541.00797-2024>