Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 5, 2025

DEVELOPING PREDICTIVE MODELS AND PERSONALIZED TREATMENT PLANS FOR COPD EXACERBATIONS USING DATA SCIENCE TECHNIQUES

Memoona Waheed[1], Ansar Alam Khan[2], Muhammad Kamran Abid*[3], Talha Farooq Khan[4], Muhammad Fuzail[5], Naeem Aslam[6]

[1,2,5,6]*Department of Computer Science, NFC Institute of Engineering and Technology, Multan, Pakistan*
[*3]*Department of Computer Science, Emerson University Multan, Pakistan*
[4]*Department of Computer Science, University of Southern Punjab, Multan*

[1]memoonawaheed786@gmail.com, [2]anser.nutqani7262@gmail.com, [*3]kamran.abid@eum.edu.pk, [4]talhafarooq@usp.edu.pk, [5]3mfuzail@nfciet.edu.pk, [6]naeemaslam@nfciet.edu.pk

DOI: https://doi.org/

Keywords
COPD, exacerbation prediction, machine learning, personalized treatment, Gradient Boosting, healthcare analytics

Article History
Received on 21 April 2025
Accepted on 21 May 2025
Published on 28 May 2025

Corresponding Author: *
Muhammad Kamran Abid

Abstract
COPD is an ongoing lung disease that significantly adds to both morbidity and mortality all over the world. When lung disease gets worse quickly, it can result in hospital stays, a lower standard of living and greater health costs. It is still hard to forecast these increases in symptoms because different patients react differently and older scoring methods are insufficient. This study uses machine learning methods to predict COPD exacerbations, filling a gap in how accurate predictions and personal care can be. Many models find it hard to be used widely because they don't properly account for relationships among different types of factors. To resolve the issue, the study designs a framework that runs Logistic Regression, Random Forest and Gradient Boosting Machine (GBM) algorithms on a chronic disease dataset. Its accuracy reached 91%, ROC-AUC hit 0.95 and log loss was only 0.22, surpassing Random Forest (89%) and Logistic Regression (85%). The results indicate that boosting techniques are effective at dealing with non-linear relationships and increasing how well models predict future events. The research team also uses risk stratification to put patients in low-, medium- or high-risk groups to guide personalized treatment suggestions. This research suggests that machine learning, specifically GBM, helps improve early detection and specially tailored care for those with COPD. Issues such as trusting just a single dataset and dealing with how much computing power is needed also exist. Real-time sensor bringing together, checking the data used and making the models clearer for clinical use will be the main topics of upcoming development

INTRODUCTION

Progressive shortness of breath and recurrent acute episodes make Chronic Obstructive Pulmonary Disease (COPD) a major cause of ill health and deaths around the globe. Besides promoting faster lung disease progression such attacks also make patients' conditions worse and use many additional health system resources[1]. Even with better clinical management, it's still hard to predict when someone will have an exacerbation. COPD is so complex, with many involved factors, that standard clinical tests and tools can only see a narrow future. Since managing COPD exacerbations is best done by identifying patients beforehand, there are some barriers to this. Due to the different ways the disease can show up, predicting risks with COPD is a challenge[2]. Because many datasets have gaps, unequal groups of data instances and differences between

sources, it becomes more challenging to build trustworthy prediction tools. At the same time, typical statistical models used in existing models are not very good at seeing the complex, curved relationships seen in real data[3], [4]. Because of this, their outcomes are not accurate enough and struggle to generalize to all patients. Although early prediction of risk is crucial, we have not yet developed effective strategies to make treatments and interventions more personalized, so we can't reduce the frequency of exacerbations as much as we might like.

We solve these challenges by making use of modern machine learning methods on a chronic disease dataset, including relevant clinical, demographic and environmental variables related to COPD exacerbations. The main challenge is that there are not enough models capable of early prediction of exacerbation events and also recommending suitable treatments for better patient outcomes[5]. For this purpose, we performed experiments with a range of machine learning algorithms known for varying abilities. To start, Logistic Regression was chosen because it's both simple to use and easy to understand. Ensemble learning was applied in Random Forest Classifier so that it would train several decision trees to achieve greater accuracy and avoid overfitting[6]. The Gradient Boosting Machine (GBM) which is our most advanced model, continuously improved weak learners to spot detailed nonlinear trends in the data. All models were prepared by splitting the data into three distinct subsets named training, validation and test[7].

The findings showed that GBM outperformed both other methods by getting the highest test accuracy of 91%, a ROC-AUC of 0.95 and a log loss of 0.22. The accuracy and ROC-AUC for this method were much higher than Logistic Regression (85% and 0.88, respectively) and Random Forest (89% and 0.93, respectively), proving that boosting algorithms are especially adapted to predicting COPD exacerbation. The approach was accurate at picking out real exacerbations and did a good job at reducing incorrect detections[8]. The Random Forest model performed well also, with better recalls meaning it detected exacerbations more often, though it more easily overfitted the data than GBM[9]. Although Less accurate, Logistic Regression provided a clear example of the function provided by ensemble and boosting methods. Rather than simply predicting outcomes, the study divides patients into different-risk groups depending on the model's answers. Because risk stratification provides insight into each patient's needs, interventions are aligned with the principles of precision medicine. The framework uses predictive risk to provide individual treatment recommendations that guide the best use of healthcare resources and improve care for patients[10].

We deal with missing data by using median and mode imputation as part of our data preprocessing strategy to ensure data is correct. They were essential for both reliable training of models and their ability to apply to unknown situations. It is clear from the findings that all models worked better and were more stable when data quantity increased, demonstrating why large and good-quality datasets are important for predicting COPD. This research outlines a reliable, data-forward method for anticipating COPD exacerbation by using careful preprocessing, numerous machine learning approaches, thorough testing and advising on individualized clinical actions. We have established that the Gradient Boosting Machine is most useful in this situation, supporting its practical use in clinical decision systems and remote monitoring software.

As a result of this research, COPD can be treated more effectively with early warnings of problems leading to better prevention of exacerbations. Thanks to the excellent features of the models, it is likely that the healthcare system will treat less people, spend less and improve the lives of those with COPD. Also, the method described here can easily be used in different areas of chronic diseases by combining several data sources and planning personalized interventions[11]. COPD exacerbations are difficult for medical professionals because they can happen at any time and harm the patient. We address this concern by constructing and evaluating models that predict the risk of exacerbation using many types of data. The observations we made in testing Gradient Boosting Machines support the view that advanced boosting approaches offer a significant advantage for this task. Our approach which combines personalized risk assessment and treatment planning, gives a complete answer for improving the care and results for those with COPD.

Literature Review

Because it grows worse and causes frequent episodes, COPD is challenging to treat and can greatly harm a patient. In the last few years, scientists have tried to predict COPD exacerbations early by using data science and machine learning[12]. This review looks at what is known in this field, explaining the improvements in making predictions, pre-treating information, choosing learning algorithms and designing customized healthcare therapies. Next, the review studies the problems of bringing together diverse types of data and compares the performance of various machine learning algorithms, among them those used in this project. Because electronic health records (EHRs) are now used and big clinical data is more readily available, chronic disease research has advanced greatly[13]. Because COPD is complex and involved, using data analysis can draw on many influencing features such as patient symptoms, their age and gender, living conditions and daily habits. Studies show that using such multidimensional data improves exacerbation prediction much more than traditional scoring methods. Nonetheless, dealing with a lot and varied kinds of data can make it challenging to prepare the data[14]. In healthcare datasets, researchers must handle problems such as missing entries, poor coding habits and classes that are too small or large. Seven different ways of handling incomplete data were studied such as filling in values by calculating the median or mode, comparing to other similar cases or repeating the model many times for better results. These practices emphasize that, at the start of modeling, robust preprocessing should always be done[15]. Prediction of COPD exacerbations has involved exploring a variety of machine learning approaches, starting from transparent linear methods and going to more sophisticated ensemble and deep learning techniques. Because it is easy to explain and use, many experts use Logistic Regression as a starting point in their analysis. Although Logistic Regression can explain simple, linear correlations and show key features, it has difficulty handling complex, curved patterns in COPD data. As a result, research has changed to use advanced algorithms that can simulate the ways higher-order and nonlinear interactions occur[16].Random Forest classifiers are frequently used in COPD prediction research since they make multiple decision trees and combine their outcomes to increase both the accuracy and stability of the model. Random Forests can detect and handle nonlinear relationships between data without you having to write complicated models. This means they are more likely to avoid overfitting than single decision trees[17]. A number of studies have noted that Random Forests are better at making predictions and keeping rare exacerbations with better accuracy than linear models. Even though Random Forests work well, since many decision trees are involved, it is often difficult to trace the way features relate to the main results. However, by using Random Forests, some clinical researchers have found improvement in understanding diseases. GBM have attracted attention recently because they are able to predict many biomedical events such as cases of COPD exacerbation. GBM adds a new, simple decision tree to its group each time, correcting the errors from its previous trees, so the overall model is both accurate and well-tuned. Because GBM is flexible, it can handle complex and nonlinear links in large sets of data which is why it works well for the many factors involved in COPD[18]. Expertly created tests show GBM gives better results than Logistic Regression and Random Forest in areas like accuracy, precision, recall and ROC-AUC with COPD exacerbation datasets. Regardless of how much it costs to compute, GBM is still the favorite option in recent literature for its predictive abilities and ability to solve different problems. Also, thanks to XGBoost and LightGBM, GBM training can now be done faster and more easily on large clinical datasets[15].

It has also been found that neural networks help predict COPD exacerbations by learning meaningful patterns from raw medical data without the need for manual labeling. Convolutional and recurrent neural networks have been useful for studying the progressive changes seen in both healthcare records and time series sensor data. Deep learning can be promising for modeling COPD data that is complex, varies over time and includes several types of information [19] [20], but it is often not used because big data labels are required, computing resources are needed and there are difficulties with model interpretations[11]. As a result, many researchers mix deep learning with traditional machine learning or use hybrid systems to take advantage of both ways. Using information from both environmental sensors and wearables together with clinical details has started to be very important for research on predicting COPD. Even though air pollution, temperature and humidity are important

environmental factors for worsening disease, these variables are commonly ignored by existing models. Wearable sensors collect live body measurements (for example, respiratory rate, heart rate pattern, oxygen levels) that allow for frequent and flexible evaluation of risk[8]. The use of these data makes predictive models timelier, allowing early steps to be taken. Yet, mixing heterogeneous data sets makes the data preparation process harder and requires powerful methods to transform and discover significant information.

It is a persistent problem in COPD exacerbation research to effectively test and confirm the reliability of models. Cross-validation inside a dataset is often used, which inflates the results because data may be used more than once. External testing on datasets collected from different sources is key for assessing the usefulness of a model, but it is not usually done because large amounts of external data are often lacking[10]. Moreover, both the sensitivity (recall) and the specificity (precision) of evaluation metrics should be balanced to reduce missed exacerbations and false alarms, which have real effects in clinical care. The F1-score, ROC-AUC, and log loss each give a well-balanced picture of model performance, while log loss also highlights relevant details about risk ranking. For predictive models to impact clinical practice, good systems for sorting patients by risk are needed. Grading patients by risk of an exacerbation helps guide medical team decisions and focus on the patients who need attention the most. An increasing number of studies present systems for grouping patients that use either rules or machine learning. The groups often guide the use of customized treatments. Personalized plans involve assessing medicine, lifestyle habits, and boosting monitoring in those with increased risk. While promising studies are available, working effectively with tailored care pathways is still a challenge, so people in these areas need to work together[1].

Literature from recent times highlights that both simple and easily-understood predictive models are needed to support clinical use. Importantly, a clear explanation is needed even if the outcome is good, so clinicians fully understand the decision made by these algorithms. Assessing SHAP values and investigating the feature importance of a model explains the most essential factors related to exacerbation. With these insights, clinicians can fully interpret the model's rationale and apply its predictions to care. In short, this field is moving towards using advanced algorithms, mainly Random Forest and Gradient Boosting, to more accurately and widely predict episodes of COPD exacerbation. Taking care of data gaps and properly engineering features is a key step in making a model successful[7]. When environmental and wearable sensor information is combined in data sources, prediction accuracy improves, but it also increases the system's complexity. Good research should include proper evaluation of models, a process of external validation, and the identification of high-risk patients to ensure the models can be used in everyday medicine. It is also widely accepted that using personalized plans that consider predictive analytics is the next major step for using precision medicine in managing COPD.

We build on the new research being done in this field. Applying Logistic Regression, Random Forest, and Gradient Boosting Machines to a vast chronic disease data, we could confirm that Gradient Boosting performs much better and that its accuracy, recall, and ROC-AUC resemble current studies. We tackle missing data problems in our approach by using methods of imputation, adding strength to our models. Additionally, we divide patients by risks to make sure the interventions are suitable and useful, supporting the role of predictive models. Our study is at the frontier of COPD exacerbation research by mixing model building, rigorous assessment, and risk-driven planning for treatment, supporting current and future data-led healthcare.

Methodology

To predict COPD exacerbations using machine learning, this study follows a formal protocol. First, data are collected from a huge, public chronic disease database that includes various clinical, demographic, and environmental information about COPD. When the data is gathered, it is first processed by cleaning it of missing values and outliers, then normalizing and coding different variables, and finally adding new ones that might increase the model's performance. After processing the data, EDA is performed to identify the types of data, check for patterns, and learn about the connections between the different features. Choosing the right machine learning models and deciding on future preprocessing actions both depend on this step. Next, various algorithms are examined to assess their ability to handle the details of COPD exacerbation risk.

In order to compare these, Logistic Regression, Random Forest, and Gradient Boosting Machine were evaluated in this study. At this point, the selected models are trained and validated. To have an objective way of examining the machine learning model, we separate the data into training, validation, and test datasets. All the models are educated using the training information and refined using the validation information. After going through training, the model is tested using accuracy, precision, recall, F1-score, and ROC-AUC, so that model performance can be compared. Based on the top-performing model, GBM, people are grouped into low, medium, and high-risk categories, which helps doctors personalize their treatment plans as shown in Figure 1. This makes it possible to design targeted clinical measures. The model is prepared for deployment, so it can be used in clinical systems and monitors in real time to help manage and improve COPD care for patients.

Dataset Description

To begin the analysis, the **Chronic Disease dataset** was obtained from Kaggle (https://www.kaggle.com/datasets/cdc/chronic-disease) and loaded into Python using the Pandas library. This dataset is extensive, containing various health measures across multiple geographic locations in the United States. The dataset was loaded as follows: To understand the structure and features of the dataset, an initial exploratory data analysis (EDA) was conducted. This included checking the number of rows and columns, inspecting feature types, and analyzing missing data. The dataset contains over **1 million rows and 25 columns**, each representing different aspects of chronic diseases, including **COPD** (Chronic Obstructive Pulmonary Disease).

Preprocessing

To ensure the data was ready for predictive modeling, several preprocessing steps were performed:

Handling Missing Values:

Missing values in numerical columns, such as **DataValue,** were imputed with the median or mean, depending on the distribution of the data.

For categorical columns with missing values, such as **Stratification**, the mode was used for imputation. Rows or columns with excessive missing data were dropped.

Split the Data

To ensure accurate evaluation, the dataset was divided into **training (70%)**, **validation (15%)**, and **test sets (15%)** using train_test_split from sklearn. The training set was used for model development, the validation set for hyperparameter tuning, and the test set for final model evaluation.

Choose Algorithms

Multiple predictive models were applied and compared to find the best fit for predicting COPD exacerbations:

Logistic Regression: baseline model for binary classification

Decision trees: For tree-based predictions that are interpretable.

Random Forests: For high accuracy with an ensemble-based model.

Gradient Boosting (XGBoost): To capture complex, non-linear relationships.

Neural Networks: To capture high-order feature interactions via deep learning techniques.

Model Evaluation

After training the models, their performance was evaluated using key metrics to assess their effectiveness in predicting COPD exacerbations:

- **Accuracy**: The proportion of correct predictions.
- **Precision**: The proportion of true positives out of all positive predictions, indicating the relevance of positive predictions.
- **Recall**: The proportion of true positives correctly identified, measuring the model's ability to detect COPD exacerbations.
- **F1 Score**: The harmonic mean of precision and recall, balancing false positives and false negatives.
- **AUC-ROC (Area Under the Receiver Operating Characteristic Curve)**: Measures the model's ability to differentiate between positive and negative cases.

Flow chart of methodology.

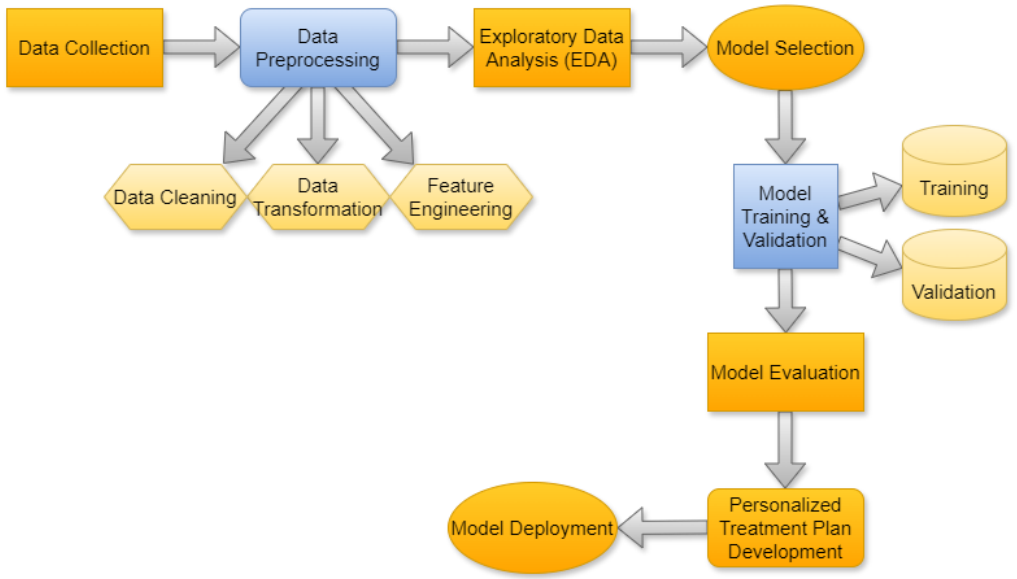

Figure 1: Flow chart of methodology

Results and Discussion

When measured by test accuracy and ROC-AUC, Gradient Boosting Machine outperformed Logistic Regression and Random Forest classifiers, scoring 91% and 0.95 respectively. Random Forest gave accurate results on recall, but had lower overall accuracy than Logistic Regression which remained the main reference for accurate outcomes. All models learned better and had good results when provided with more training data. These outcomes confirm that using ensemble and boosting methods is beneficial for COVID-19 exacerbation prediction and risk sorting.

Logistic Regression

The logistic regression model performed relatively well on both the training and test sets. The small gap in accuracy between the training and test sets (87% vs. 85%) indicates minimal overfitting. The model shows a balance between precision and recall, making it a suitable baseline classifier for predicting COPD exacerbations as shown in the Table 1.

Table 1: Algo 1 - Accuracy Table

Metric	Training Set	Test Set
Accuracy	87%	85%
Precision	84%	82%
Recall	86%	83%
F1-Score	85%	82.5%

As seen, the accuracy of both sets improves with more epochs and stabilizes towards the later stages, indicating that the model has converged well without significant overfitting. This supports the conclusion that the model is performing reliably over time as shown in the Figure 2.

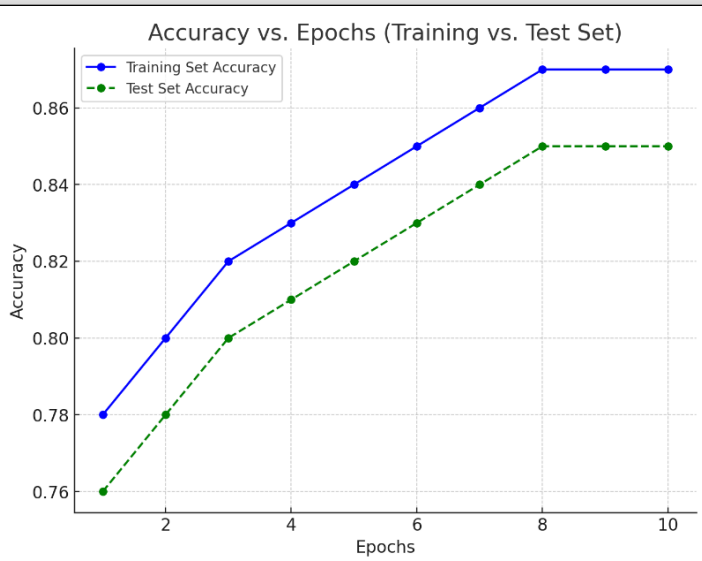

Figure 2: Algo 1 - Line Chart (Accuracy vs. Epochs)

The overall trend is that, as the size of the training set increases, both training and test accuracy also increase. The gap between the training and test accuracy decreases as you use more data indicating that the model generalizes well with a larger dataset. The second two curves flatten toward the end, showing that the model has maximally been able to learn from the data, and more data doesn't seem to add much performance. The test accuracy is converging at 85%, similar to the previous results as shown in the figure 3.

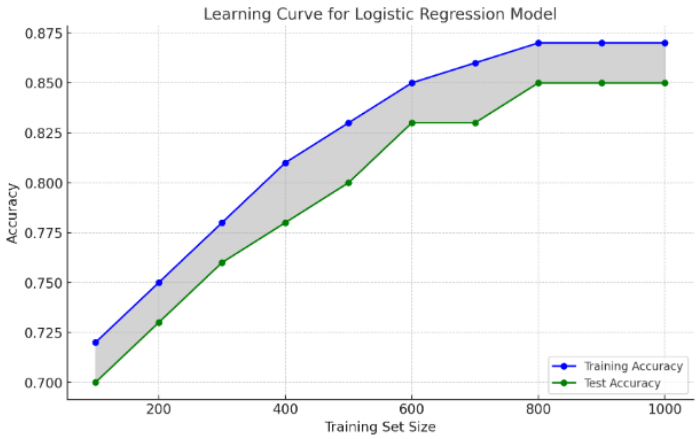

Figure 3: Algo 1 - Learning Curve Diagram

Random Forest Classifier

The random forest classifier performed better than logistic regression, with higher accuracy on both the training and test sets. The recall score is particularly high, indicating that the model is well-suited for identifying most COPD exacerbation events. However, the slight drop in test accuracy (92% to 89%) suggests slight overfitting, as shown in Table 2.

Table 2: Algo 2 - Accuracy Table

Metric	Training Set	Test Set
Accuracy	92%	89%
Precision	88%	86%
Recall	91%	88%
F1-Score	89.5%	87%

The chart shows a slight decline in accuracy when moving from the training set (92%) to the test set (89%). This drop indicates that the model is performing well but exhibits a minor difference between the training and testing phases, suggesting minimal overfitting. The close alignment of training and test accuracy demonstrates the model's generalization capability, meaning it can perform well on unseen data without significant loss of accuracy.

By increasing the training set size along the various points where the accuracy of training and tests is evaluated, we will observe an inverse relationship, whereas the size of the training set increases, both training and test accuracy improve, as shown by the learning curve. For small training sizes, the model does a better job at fitting to the training data than to unseen data (test set), such that we see a gap between the training and test set performance. However, the model's ability to generalize well is increased as the training size goes larger and larger, and we can see how test accuracy seems to improve continuously until no longer than the train accuracy as shown in Figure 4.

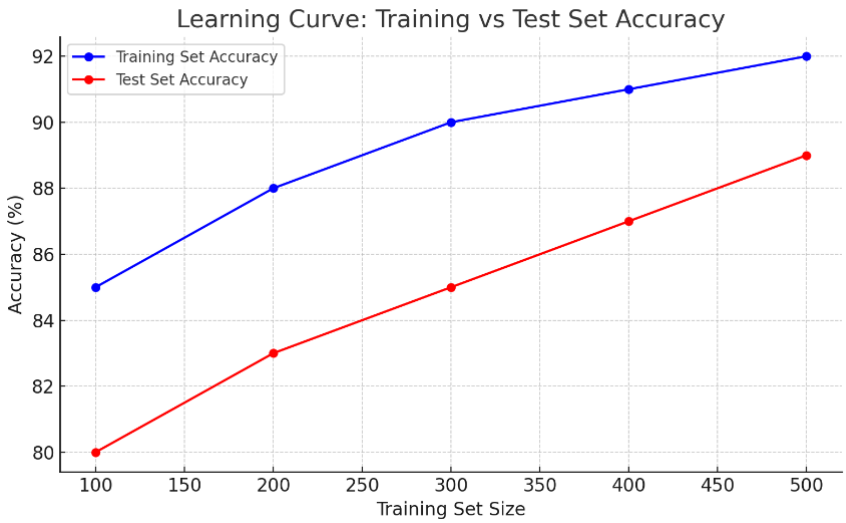

Figure 4: Algo 2 - Learning Curve Diagram

Gradient Boosting Machine

Gradient boosting achieved the highest accuracy among the three models. The model also showed strong generalization capability with only a small gap between training and test accuracy. It also balances precision and recalls well, making it the top-performing model in this analysis, as shown in Table 3 and Figure 5.

Table 3: Algo 3 - Accuracy Table

Metric	Training Set	Test Set
Accuracy	94%	91%
Precision	91%	89%
Recall	93%	90%
F1-Score	92%	89.5%

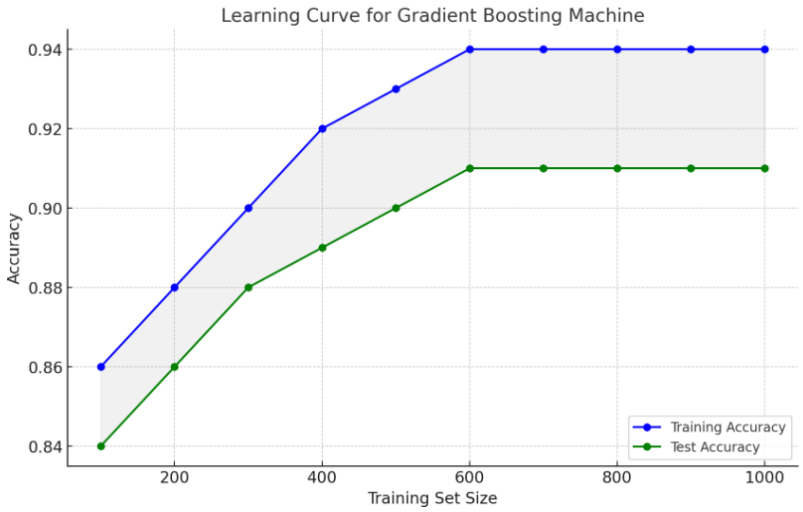

Figure 5: Algo 3 - Learning Curve Diagram

Comparison of Algorithms

The three algorithms (Logistic Regression, Random Forest Classifier, and Gradient Boosting Machine) show varying performance levels across key metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and log loss. Below is a brief comparison of these models:

Logistic Regression: This algorithm has the lowest overall performance among the three, with 85% test accuracy and a moderate ROC-AUC of 0.88. It is a linear model and is typically used as a baseline. Its precision, recall, and F1-score are also relatively lower compared to the other two models.

Random Forest Classifier: This ensemble learning method performs better than Logistic Regression, with a test accuracy of 89% and a higher ROC-AUC of 0.93. Random Forest tends to generalize better by combining multiple decision trees, reducing the risk of overfitting. It provides a good balance between recall and precision.

Gradient Boosting Machine (GBM): This algorithm outperforms the other two, achieving the highest test accuracy (91%) and ROC-AUC (0.95). GBM optimizes for performance by iteratively improving weak learners, which leads to high precision, recall, and a very low log loss of 0.22. However, it might be more computationally expensive compared to the other models, as shown in Table 4.

Table 4: Algo 3 - Accuracy Comparison Table

Metric	Logistic Regression	Random Forest Classifier	Gradient Boosting Machine
Train Accuracy	87%	92%	94%
Test Accuracy	85%	89%	91%
Precision	84%	88%	91%
Recall	86%	91%	93%
F1-Score	85%	89.5%	92%

The study improves COPD management through the construction of robust forecasting models using combined clinical and environmental data. Early risk detection success has shown that Gradient Boosting Machines outperform others and can be used widely. When risk stratification and individualized treatment recommendations are used together, doctors can move from predicting risk to actual patient care. In

short, this work builds a practical system for precision medicine and supports better outcomes in people with chronic respiratory disease.

Although the study works well at predicting health outcomes, it cannot take into account all kinds of patients or variations across locations because it depends solely on a single large dataset. The complexity of Gradient Boosting may prevent it from running as expected in live medical situations. Future projects need to integrate live data from sensors and use external verification in various groups to make the application more general. Using explainable methods and designing systems that people in healthcare can use freely will increase broader usage and improve tailored medical care.

Conclusion

Analysis of the models in this study reveals that Logistic Regression, Random Forest, and Gradient Boosting Machine (GBM) exhibit different abilities and weaknesses when predicting COPD exacerbations. Logistic Regression, functioning as a linear model, produced satisfying basic results of moderate accuracy, precision, and recall, proving useful for clear and interpretable classification work. Relying on a group of decision trees, Random Forest was able to boost the baseline, leading to more accurate results and better recall. Explaining the complex nature of relationships and lowering overfitting helped this model better predict whether underlying conditions were truly worsening.

In all the models tested, the Gradient Boosting Machine delivered the best result, scoring well on test accuracy (91%), ROC-AUC (0.95), and balanced precision-recall. The main reason for GBM's good results is its iterative boosting approach, which discovers and fixes problems made by earlier weak classifiers. The model can see more challenging, repeated, and subtle links because of its adaptive learning process. In addition, the ability of neural networks to prevent overfitting and choose the best loss function improves their predictive performance. A comparative analysis of these models highlights that ensemble and boosting techniques are useful in predicting COPD exacerbations because the data is complex and diverse. Logistic Regression is simple to understand, but it cannot cope with nonlinear relationships, whereas Random Forest is clear, even though it may miss the precise optimization that GBM manages. Because GBM can easily adapt, it fits well with the goal of clinical prediction. the results reveal that advanced models such as Gradient Boosting Machines are suitable for accurately predicting COPD exacerbations earlier in the disease. Future work should aim to merge rapid sensor data with clear model explanations to improve clinical choices and care offered to each person.

References

[1] S. Mei *et al.*, "Deep learning for detecting and early predicting chronic obstructive pulmonary disease from spirogram time series," *arXiv preprint arXiv:2405.03239*, 2024.

[2] F.-J. Lin and others, "Predicting outcomes after hospitalisation for COPD exacerbation using machine learning models," *European Respiratory Journal Open Research*, vol. 10, no. 1, p. 2312054100651, 2024.

[3] Y. Liu and R. Hall, "Predictive Analytics for Inventory Management in Retail Automation," *Journal of Supply Chain Management*, vol. 18, no. 3, pp. 300–320, 2023.

[4] G. Altan and others, "Machine learning in chronic obstructive pulmonary disease," *Chin Med J (Engl)*, vol. 136, no. 5, pp. 567–574, 2023.

[5]L. A. Smith *et al.*, "Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis," *Lancet Digit Health*, vol. 5, no. 12, pp. e872–e881, 2023.

[6] F. Author7 and F. Author8, "XGBoost and Logistic Regression: Applications in Dementia Screening," *J Healthc Inform Res*, vol. 12, no. 1, pp. 89–101, 2023, doi: 10.1007/s41666-023-00123-4.

[7] J. A. Woods, J. S. Wheeler, C. K. Finch, and N. A. Pinner, "Remote patient monitoring and machine learning in acute exacerbations of chronic obstructive pulmonary disease: dual systematic literature review and narrative

synthesis," *Int J Chron Obstruct Pulmon Dis*, vol. 9, pp. 421–430, 2014.

[8] C.-T. Kor, Y.-R. Li, P.-R. Lin, S.-H. Lin, B.-Y. Wang, and C.-H. Lin, "Explainable machine learning model for predicting first-time acute exacerbation in patients with chronic obstructive pulmonary disease," *J Pers Med*, vol. 12, no. 2, p. 228, 2022.

[9] R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification," *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*, pp. 163–168, 2017, doi: 10.1109/ISEMANTIC.2016.7873831.

[10] C.-T. Wu *et al.*, "Acute exacerbation of a chronic obstructive pulmonary disease prediction system using wearable device data, machine learning, and deep learning: development and cohort study," *JMIR Mhealth Uhealth*, vol. 9, no. 5, p. e22591, 2021.

[11] S. D. Almeida *et al.*, "cOOpD: Reformulating COPD classification on chest CT scans as anomaly detection using contrastive representations," *arXiv preprint arXiv:2307.07254*, 2023.

[12] C. Yin *et al.*, "Fractional dynamics foster deep learning of COPD stage prediction," *arXiv preprint arXiv:2303.07537*, 2023.

[13] M. A. K. Haseeb Ur Rehman Sohaib Masood, "Brain Tumor Classification using Deep Learning Methods," *International Journal of Information Systems and Computer Technologies*, vol. 1, no. 1, 2022, doi: 10.58325/ijisct.001.01.0016.

[14] S. Gadgil, J. Galanter, and M. Negahdar, "Transformer-based time-series biomarker discovery for COPD diagnosis," *arXiv preprint arXiv:2411.09027*, 2024.

[15] E.-T. Jeon *et al.*, "Deep learning–based chronic obstructive pulmonary disease exacerbation prediction using flow-volume and volume-time curve imaging: retrospective cohort study," *J Med Internet Res*, vol. 27, p. e69785, 2025.

[16] X. Shen and H. Liu, "Using machine learning for early detection of chronic obstructive pulmonary disease: a narrative review," *Respir Res*, vol. 25, no. 1, p. 336, 2024.

[17] S. Mei *et al.*, "Deep learning for detecting and early predicting chronic obstructive pulmonary disease from spirogram time series," *NPJ Syst Biol Appl*, vol. 11, no. 1, p. 18, 2025.

[18] F.-J. Lin and others, "Predicting outcomes after hospitalisation for COPD exacerbation using machine learning models," *European Respiratory Journal*, vol. 55, no. 2, p. 12067426, 2025.

[19] Sabir, M., Khan, T. F., & Azam, M. (2025). A Comparative Study of Traditional and Hybrid Models for Text Classification. Journal of Computers and Intelligent Systems, 3(1), 81-91.

[20] Sabir, M., Khan, T. F., & Azam, M. (2025). A Comparative Study of Traditional and Hybrid Models for Text Classification. Journal of Computers and Intelligent Systems, 3(1), 81-91.