Predicting the Causal Effect Relationship Between COPD and Cardio Vascular Diseases

Debjani Panda, Satya Ranjan Dash and Ratula Ray
Kalinga Institute of Industrial Technology, Bhubaneswar, India
E-mail: PANDAD@indianoil.in, sdashfca@kiit.ac.in, and ratularay@gmail.com

Shantipriya Parida (corresponding author)
Idiap Research Institute, Martigny, Switzerland
E-mail: shantipriya.parida@idiap.ch

Coronary Obstructive Pulmonary Disease (COPD) is one of the critical factors that are affecting the health of the population worldwide and in most cases affects the patient with cardiovascular diseases and their mortality. The onset of COPD in a patient in most of the cases affects him/her with cardiovascular disease and the management of the disease becomes more complex for medical practitioners to handle. The factors affecting COPD and cardiovascular disease in patients are most of the time, concurrent, and are responsible for their mortality. The list of factors and their underlying causes have been identified by experts and are treated with utmost importance before the patient suffers from an emergency condition and its management becomes even more difficult.

This paper discusses the need to study COPD and the factors affecting it to avoid cardiovascular deaths. The dataset used for the study is a novel one and has been collected from a Government Medical College, for study and experimentation. Classification methods like Decision Trees, Random Forest (RF), Logistic Regression (LR), SVM (Support Vector Machine), KNN (K-Nearest Neighbours), and Naïve Bayes have been used and Random Forests have given the best results with 87.5% accuracy. The importance of the paper is in the attempt to infer important links between the associated features to predict COPD. To the best of our knowledge, such an attempt to infer the interrelation between cardiac disease and COPD using Machine Learning classifiers has not been made yet. The paper focuses on determining the important correlation between the associated features of COPD and compare different supervised classifiers to check their prediction performance. Coronary Pulmonate, Age, and Smoking have shown a strong correlation with the presence of COPD and the performance analyses of the classifiers have been shown using the ROC (Receiver Operating Characteristic) curve.

Povzetek: Več metod strojnega učenja je bilo uporabljenih na povezovanju učinkov pljučnih in srčno-žilnih bolezni.

1 Introduction

The concern for deaths due to COPD and Cardiovascular diseases is increasing worldwide in an exponential manner. Experts have identified a causal effect relationship between these two diseases where the presence of one determines the onset of the other or vice versa. COPD occurs in people complaining about severe difficulty in breathing or arrhythmia (irregular heartbeats). The presence of COPD in patients mostly makes them vulnerable to cardiovascular diseases and their mortality. It has also been observed that a patient suffering from cardiovascular disease also complains about COPD. This paper reveals the factors to be considered for patients with COPD for determining whether he is suffering from cardiovascular disease or not. COPD is characterized by obstruction in the air passages, which persist over a significant amount of time. It typically refers to bronchial asthma, bronchitis, and emphysema. Bronchial asthma is an allergic reaction that affects the respiratory tract, caused due to significant amounts of histamine in the blood. Bronchitis is defined as the inflammation of the bronchi caused due to infection. Emphysema is the inflation of bronchioles or alveolar sacs in the lungs. The common factors affecting the patients with COPD and Heart Disease have been shown in our below mentioned in Figure 1.

1.1 Symptoms of COPD

The common symptoms include coughing, wheezing, difficulty in breathing mainly during exhalation, excess mucous discharge, fatigue, pressure in the chest, anxiety, and loss
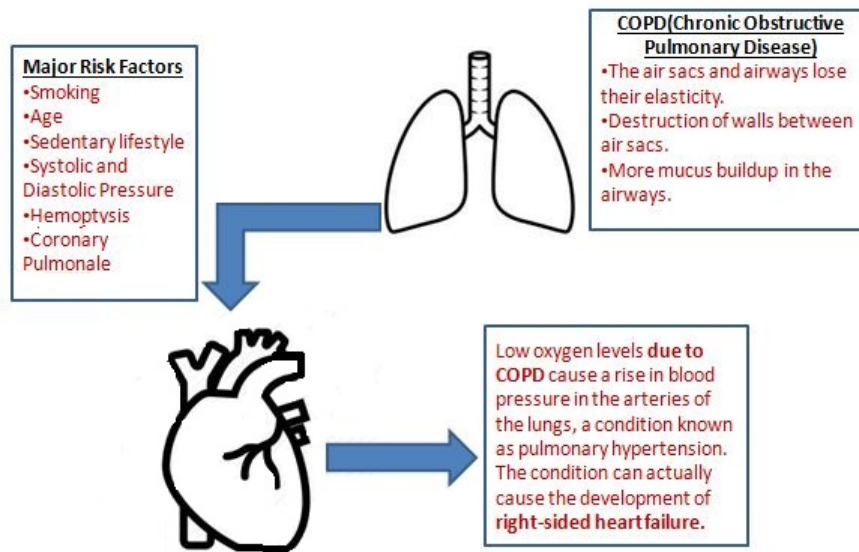
Figure 1: Causal Effect Relationship of heart disease and COPD.

in muscle tissue and weight.

1.2   Complications of COPD

People with COPD are more prone to fatal diseases such as pneumonia, pneumothorax (lung collapse), osteoporosis, edema, enlargement of liver and cor pulmonale (right side of the heart), diabetes, sleep apnea (repeated starting and stopping of breathing during sleep), stroke, high blood pressure, arrhythmia (irregular heartbeat), and heart failure. There has been evidence that patients with COPD [24] have a higher risk of Myocardial Infarction (MI) and it becomes worse when it is not properly managed in hospitals and adequate treatment is not given. Studies have revealed that factors like smoking contribute to around 3.8 % to 16 % of patients suffering from COPD and Congestive Cardiac Failure [7, 25]. Cardiovascular and COPD have co-morbidities [17] like diabetes, smoking, hypertension, atria fibrillation, Congestive heart failure, and several other factors.

There is a necessity to study the factors that correlate COPD and Cardiovascular diseases [13, 9] so that their underlying factors can be identified and help in treating the patient on time to avoid premature deaths. COPD in most patients causes cardiovascular deaths, and it remains a challenge to identify its occurrence and treat the patient on time.

In this paper, we have used the real patient data from the Srirama Chandra Bhanja Medical College and Hospital [1], Cuttack, Odisha, India. The data was collected for 200 Patients by the Regional Medical Research Center (RMRC), Bhubaneswar [2]. RMRC is a medical Research Institute

[1] http://scbmch.nic.in/
[2] http://www.rmrcbbsr.gov.in/

that collects patient's data for research and study of ailments. RMRC is the authorized body of Government to carry out research activities on available medical data of patients from different organizations. The data set used in our study was contributed by RMRC for research purposes. The data was collected by RMRC through questionnaires and from test reports of the patients with the consent of the patients. The raw data was pre-processed to drop identification labels of patients and missing values were imputed. The data was then split to train the classifiers. The experiment uses the heat map to identify the most important factors which affected the patients with COPD. Our paper reveals important factors like age, Coronary Pulmonate, and smoking as major contributing factors that are highly responsive to cause COPD in patients. The other factors include the systolic and diastolic pressure of the patients. Figure 2 shows the pipeline of the work that has been carried out in this paper.

2   Literature survey

Various supervised ML classifiers have been used for the prediction of health conditions for a long time. Works have been done previously to study the correlation between COPD and related risk factors. We have attempted to consolidate the most relevant works in this field that have been carried out in the past. Rabe et al. [21] establishes the strong connection between COPD and Cardio-Vascular diseases. Whenever a patient is suffering from either or both the diseases there are pathophysiological changes in the body which includes inflamed lungs and heart. The focus has been on suggesting various treatments to be administered to the patients suffering from COPD who have
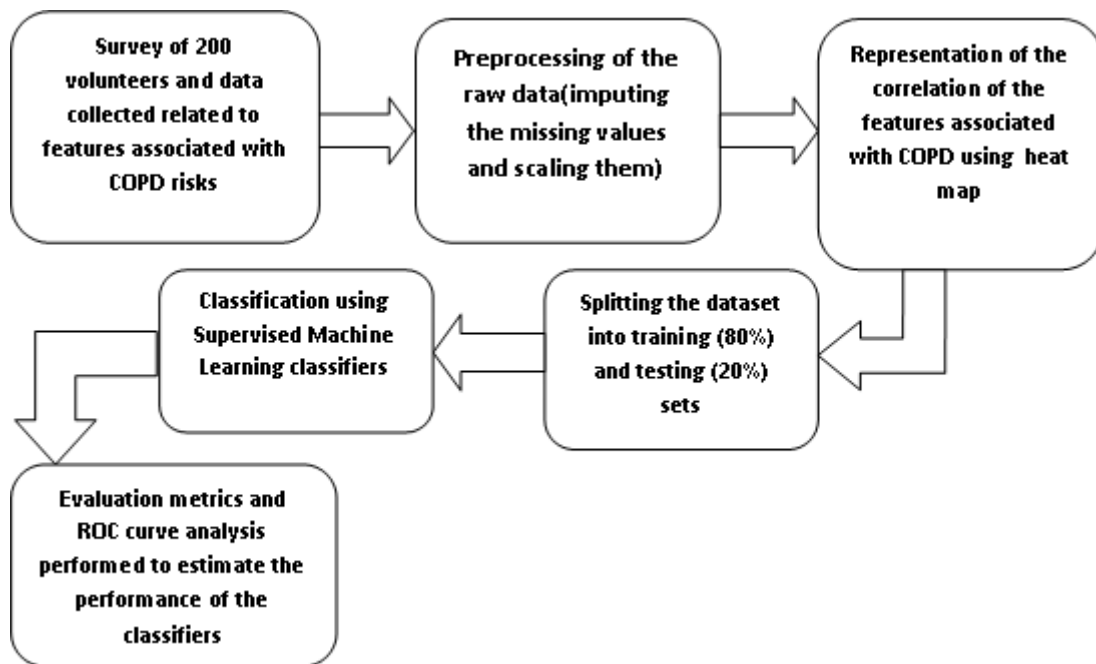
Figure 2: The figure highlights the workflow of the paper where the features associated with COPD are classified after pre-processing of the crude dataset to detect the presence of the disease. Also, the supervised classification and performance analysis of the different classifiers, using ROC analysis and other evaluation metrics are important prediction measures of the processed dataset.

these underlying heart diseases. The critical factors that have been identified to coexist with COPD and CVD are smoking, age, diabetes, and a sedentary lifestyle. The paper suggests the use of Beta-blockers for treating patients with Cardio Vascular ailments that do not have much interference with COPD drugs. Cazzola et al. [2] has suggested effective ways to manage the two interrelated deadly diseases COPD and CVD which coexist in the patients. The prime objective is to reduce the COPD causing symptoms by first treating the swollen lungs and breathlessness of the patients. Suggested drugs to reduce the urge to smoking, bronchodilators, and inhaled corticosteroids are used in the majority of the cases to treat COPD. The use of Angiotensin-converting enzyme inhibitors, angiotensin II type 1 receptor blockers, statins, antiplatelet drugs, or β-adrenoceptor blockers have been proved to be beneficial in treating patients with CVD which has effectively reduced COPD deaths in patients and reduced hospitalization for them. Holm et al. [10] focuses on identifying a genetic deficiency Alpha-I Antitrypsin deficiency and its impact on COPD with growing age. With growing age, the effect of psychological and clinical conditions of patients has been studied with COPD arising due to genetic deficiency (AATD). 468 individuals were considered for study with the genetic deficiency within varying age groups of $32 - 84$ years. The individuals who were having severe AATD were found to be at greater risk of suffering from COPD. AATD is the genetic cause for the onset of COPD and this deficiency also aggravates smoking. The patients were studied for two years and from the study, it was observed that the younger generation was prone to anxiety, depression, and health issues regardless of their relationship status.

Fukuchi [8] significantly focuses on the growing age of individuals which is the factor of consideration for patients affected with COPD. Three models have been studied in which the animals were prematurely aged. Their lungs did not have any pathological changes like naturally aged lungs and were consistent in their function even after premature aging. The author has tried to state that the abnormal functioning of the lungs due to the increase in the size of the air spaces is not related to increasing age and the relationship between age and COPD is misleading. It has been suggested to further investigate the accelerated aging of lungs may be a factor to cause COPD but directly is not the cause.

The increased rate of smoking has also been greatly affecting patients and has been identified as the major cause of COPD. Laniado-Laborín [12] establishes the fact that smoking is the most important causal factor for patients suffering from COPD. The reduction of smoking in patients has been identified as a successful treatment for COPD patients. Different types of therapies i.e. both pharmacological and behavioral therapies have been suggested for stopping the progression of COPD in patients by controlling their smoking habits. However, studies have also suggested that the patients who refrained themselves from smoking after one year of follow-up were very low. Studies have also revealed that pharmacological therapies are more effective than placebo and 25-30% of people have abstained from smoking after taking these therapies. But, still smoking cessation remains a major challenge in the world and

patients continue to suffer from COPD due to their smoking habits. Khan et al. [11] describes in their paper the after-effects of smoking and how this creates abnormalities in lung function. It is responsible for the thickening of airways, dilation of air spaces with abnormal distension of alveoli. It is observed that almost all cigarette smokers have inflammation in their lungs. It has been summarized that tobacco inhalation active or passive results in abnormal inflammation, leads to tissue-damaging oxidants, a reduced level of antioxidants (for self-protection), and induced cell death.

Work has also been done to bring out the association of cor pulmonale with dysfunction of lungs by Shujaat et al. [23]. The earlier fact that the right ventricular dysfunction of the heart is due to enlargement of the tissue results because of left ventricular dysfunction adds up to the information that the right ventricle of the heart has an underlying cause of the malfunction of the lungs and its size. They have identified Pulmonary Hypertension as the underlying cause of right ventricle dysfunction resulting in heart failure. Cor pulmonale with pulmonary hypertension occurs due to various reasons and to treat the patients with the disease several treatments like inhalation of Nitric Oxide, usage of diuretics to remove excess water from the lungs and heart, giving a pulmonary vasodilator like sildenafil, reduction in hematocrit, and surgery for reducing lungs size have been suggested to reduce the cardiac arrests in patients suffering from COPD.

Quint [20] focuses on the fact that patients suffering from COPD have a higher risk of suffering from cardiovascular diseases. The author summarizes delayed identification of disease, late treatment given for reperfusion of STEMI (ST-Elevation Myocardial Infarction), and use of angiography after in-STEMI as causes of a gap for mortality of COPD patients suffering from Myocardial Infarction. Troponin has been identified as the direct indicator in patients suffering from COPD. The results revealed that the higher the presence of Troponin in cardiac patients, the longer they stay in hospitals and have less chance of survival.

The link to various diseases associated with COPD has also been previously studied. Feary et al. [6] has tried to summarize the diseases that are associated with inflammation in the lungs which include cardiovascular diseases and diabetes mellitus as the most common associated diseases. The study determines quantitatively the effect of cardiovascular diseases for patients suffering from COPD. The data of patients have been analyzed with logistic regression with multiple variables and Cox regression. The presence of cardiovascular diseases is found to be more in the young age group of COPD affected patients after considering several factors such as smoking and age strata of patients. It has been found that in most cases, patients suffering from COPD are already affected by myocardial infarction and diabetes.

Roever et al. [22] discussed the factors responsible for COPD and cardiovascular diseases. The authors have identified various factors such as sedentary lifestyle, systemic inflammation, improper function of skeletal muscle, etc for being responsible for cardiovascular diseases and which again becomes the major reason for patients suffering from COPD. The diseases like diabetes, hypertension, a metabolic syndrome that arises due to smoking, arterial fibrillation, Vitamin D deficiency, Congestive cardiac failure are among the several factors affecting patients suffering from heart disease. These patients are found to be vulnerable to COPD and in most cases are affected which is the ultimate factor for causing deaths. COPD patients mostly die of strokes and cardiovascular-related diseases have been identified as the major cause affecting these patients.

Esteban et al. [5] mentioned the factors responsible for the worsening of COPD. Using machine learning an early prediction system is designed which will warn about chronic obstructive pulmonary disease in three levels: red, yellow, and green. The model used Random forests for predicting the condition of COPD in patients. The attributes of the data set were obtained from the daily activities and questionnaires from the patients. The model was trained and tested using 10-fold cross-validation for improving the performance. The model achieved a ROC curve of 0.87 for forecasting whether a patient will suffer from COPD worsening within the next three days.

Peng et al. [19] developed a model for predicting acute illness in patients affected with Chronic Obstructive Pulmonary disease. The 28 important features were selected from watches, sphygmomanometers, thermometers, and routine clinical tests. They identified 410 records from the hospital database for the study and the trained: test ratio varied from 90:10 to 50:50. The best results were obtained with a split in the 80:20 ratio. The model used C4.5 and C5.0 decision trees for classification of the disease. ID3, CART, and C 5.0 classifiers have been compared to find out the best model. C5.0 with 80:20 train and test split gave 80.3% accuracy.

Xie et al. [26] tried to relate physiological homeostasis and the onset of COPD exacerbation. A regression model is built to study the patterns of variables extracted from patients and are evaluated longitudinally for all records to classify the nature of risk the patient is subjected to. The data set was obtained from a hospital in Sydney using TeleMedCare Health Monitor which included the attributes: weight, diastolic and systolic blood pressure (DBP, SBP, heart rate (HR), SpO2, and temperature. The model used Logistic regression with double loop 10- fold cross-validation and has yielded 79.27% accuracy with AUC of 0.84.

Attribute Name	Allowed Values	Description
Gender	1=Female, 0=Male	Gender of the patient
Lifestyle	1=Sedentary, 0=Active	Type of lifestyle which is adopted by the patient
Literacy	1:Illiterate, 0=Literate	The patients is literate or illiterate
Family History	1=Has family history, 0=does not have family history	Whether the patient has any family history of COPD.
Weight		The actual weight of the patient in Kgs.
Systolic Pressure		It's the blood pressure measures the pressure in your blood vessels when your heart beats
Diastolic Pressure		Blood pressure while the heart rests.
Alcohol	1=drinks, 0=does not drink	The patient consumes alcohol or not.
Smoking:	1=smokes, 0= does not smoke	The patient smokes or is a non-smoker.
Age		Current age if the patient.
Col Pulmonate	1=Present, 0=absent	Right side heart failure due to pulmonary hypertension.
Hemoptysis:	1=Present, 0=absent	Patient is coughing out with blood or not.
Type	1=COPD detected, 0= COPD not detected	Patient is affected with COPD or not.

Table 1: Data Set Detail Description

3 Materials/methods

3.1 Methods

3.1.1 Naïve Bayes classification

Naive Bayes classifiers are a throng of classification algorithms based on Bayes' Theorem. This classification technique makes an assumption of independence among predictors. In layman's terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Bayes' Thoerem [14] is based on probability theory:

$$P(A|B) = P(A)P(B|A)/P(B), \qquad (1)$$

where $P(A|B)$ is how often A happens given that B happens,
$P(B|A)$ is how often B happens given that A happens,
$P(A)$ is how likely A is on its own,
$P(B)$: is how likely B is on its own.

3.1.2 Support vector machine (SVM)

SVM is a set of learning methods that are supervised and used for classification and regression. An SVM model is a representation of the data as examples in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. This is done by identifying a hyperplane [27] which separates the classified data with maximum space between them and is determined in such a fashion that most of the points of one category fall on one side of the plane. SVM determines the best-fitted plane.

3.1.3 Decision trees

Decision Tree is a tool that represents nodes of the tree and helps take decisions depending upon the inputs of a node. It helps in giving a pictorial presentation of the consequences of a certain condition. It is used in classification and regression. Here the nodes represent the data [18] and not the decisions. Here a threshold value has to be given after which the algorithm will terminate. It is left with some points which could not be classified and this is called Gini impurity.

3.1.4 KNN classification

K Nearest Neighbor algorithm is a non-parametric method used for classification and regression [1]. This classification technique calculates the distance of a point with coordinates (x, y) from its neighbors. For example, when there are two sets of points in the space and a new point has to be plotted in the area, then the question is where it should be plotted and with which region to determine its basic characteristics that satisfy the classification correctly. The Euclidean distance is calculated from the point from its neighbors and finally, it is positioned in the area which is closest to its neighboring points.

3.1.5 Logistic regression

It is a statistical tool that is used for making decisions on the binary output of the testing condition. In the Linear model the equation used is:

$$y = b_0 + b_1(x), \qquad (2)$$

whereas logistic regression uses the equation:

$$P = \frac{1}{1 + e^{-(b_o + b_1 x)}} \qquad (3)$$

In the logistic regression, the constant (b_0) moves the curve left and right and the slope (b_1) defines the steepness of the curve. By simple transformation, the logistic regression equation can be written in terms of an odds ratio:

$$\frac{p}{1-p} = exp^{(b_0 + b_1 x)} \qquad (4)$$

3.1.6 Random forest

The random forest classifier [15] creates a set of decision trees instead of a single one which is generated by randomly taking one seed from a selected subset of the data of the training set. Subsequently, it determines the average of the results from different decision trees to determine the class to which the object belongs to. It is an ensemble method [4, 3] for classification as well as a regression that operates by constructing an assembly of decision trees at training time and finding out the class that is the most suitable for its outcome depending upon its predicted values.

4 Experiment

The following methodology has been adapted for the collection, study, and experimentation of the data which has been contributed by RMRC.

1. Carrying out of the survey of the 200 patients who volunteered to share their health cards and reports to carry out the research work.

2. Taking medical experts' opinions to identify which factors are related to the COPD occurrence.

3. Pre-processing of the raw data(imputing the missing values and dropping the features which are not directly relevant to the outcome).

4. Finding the correlation between all the risk factors with the presence or absence of COPD in the patient samples using heat-map. In this process, we can find the most relevant features associated with the outcome.

5. Splitting the entire dataset into training (80%) and testing sets (20%).

6. Using Supervised learning classifiers to classify the dataset in giving the prediction of the outcome (Gaussian Naïve Bayes, SVM, Decision Trees, Logistic Regression, Random Forest and KNN)
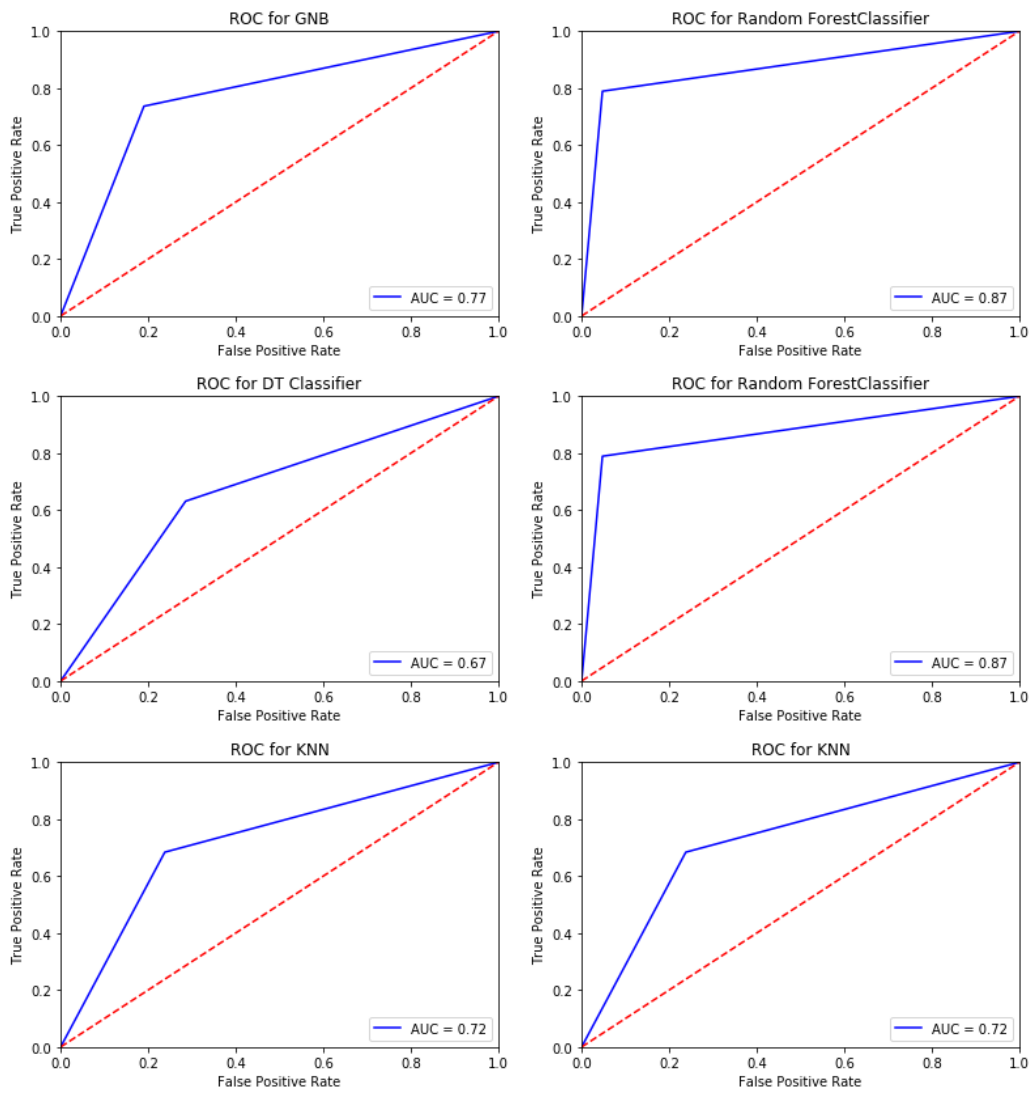
Figure 3: ROC Analysis of Gaussian Naive Bayes,Random Forest, Decision Trees, Logistic Regression, KNN and SVM Classifiers in which Random forest has given the best curve .The blue line aligning towards left represents the true positive rate of the predictive classifiers and the red line demarcates the threshold. The quality of the performance of the classifiers depends upon the nature of the blue curve.

5   Results

Our work has been implemented in Python to study the behavior of the supervised algorithms. The original data set has been collected from SCB Medical College through a survey conducted by RMRC. The dataset is novel and has been used for study only. The raw data was collected for 200 patients of Government Medical College to identify the critical connection between COPD and heart disease. Results of the six Classification Algorithms have been noted in Table 2 with their confusion matrix/accuracy in predicting the test samples.

Classifier	Accuracy	Precision	Recall	AUC
GNB	77.5	80.95	77.27	0.77
RF	87.5	95.23	90.90	0.87
DT	67.5	71.43	68.18	0.67
LR	82.5	85.71	81.81	0.82
KNN	72.5	76.19	72.72	0.72
SVM	77.5	80.95	77.27	0.77

Table 2: Performance Comparison of the Classifiers

The identification of important features [16] has been done using the heat map attached in Figure 4 and the factors which are mostly responsible for causing COPD have been determined as Cor Pulmonale, Age, and Smoking. These factors are then studied with several algorithms to analyze their performance in terms of accuracy, precision, and recall.

The performance of the classifiers have been studied by plotting the ROC for all methods as shown in Figure 3. The results show the best curve has been given by a random forest classifier which has the area under the curve 0.87. The true positives obtained for Random Forest are more than any other classifiers used. The Logistic Regression and SVM Classifiers have also given good results with AUC 0.77, which can be explicitly seen from the ROC curves.

From the data collected, the best results were produced by Random forest classifier(accuracy: 87.5%) with a Precision of 95.23% and a recall score is 90.90%.

6   Discussion

Various works have been carried out to identify the symptoms and factors affecting the health of COPD patients. A strong connection is found to be existing in patients with COPD, who are also affected by Cardiovascular disease [21, 2, 23]. Our paper also has stressed the explicit fact that these two diseases are interrelated and the onset of one disease causes the other disease to impact the patient soon.

The factors affecting the condition of COPD have been identified to be smoking, Cor pulmonate, age, and diabetes [23, 20, 6, 22], and other lifestyle-related factors. Similar facts have been established in our work, where the important factors obtained from the heatmap denote smoking,

coronary pulmonate, and age. Even to some extent literacy status and lifestyle played a vague role.

The papers which were used to study the performance of models for predicting COPD gave an accuracy of 79-80% [5, 19, 26] with max. ROC of 0.87. Our model has given superior results than other models and has given the highest accuracy of 87.5% with a ROC of 0.87. There is an improvement of almost 7% in the prediction of COPD patients using our suggested model.

7   Conclusion

Random forests being an ensemble classifier have given the best results for predicting the condition of a patient affected with COPD. The most important factors which have been identified as the causal ones include age, smoking, and Cor pulmonale. In addition to these Systolic and diastolic pressure also have been identified to have an impact on the underlying disease. Cor pulmonale is an abnormal enlargement of the heart due to infection in the lungs or any blood vessels. The above facts conclude that COPD is related to heart disease and gets worsened with the systolic and diastolic pressure of the patient. The systolic and diastolic pressure gets affected when the person is suffering from heart disease. With our data set, we have identified that these two diseases coexist, or both are interrelated to each other. Our work has been done with a novel data set of 200 patients. The work illustrates the fact that classification methods can be used to find the relationship between various diseases and their occurrences. The study can be extensively made for more number of patients so that the model behavior can be the same for new experimental case studies. The classification methods can be combined with optimization techniques for better prediction accuracy. Our model can be used by medical experts to determine heart disease well in advance and the symptoms of COPD can be interpreted by the model to predict occurrences of fatal diseases.

References

[1] G. E. Batista, R. C. Prati, and M. C. Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1):20–29, 2004. https://doi.org/10.1145/1007730.1007735.

[2] M. Cazzola, L. Calzetta, B. Rinaldi, C. Page, G. Rosano, P. Rogliani, and M. G. Matera. Management of chronic obstructive pulmonary disease in patients with cardiovascular diseases. *Drugs*, 77(7): 721–732, 2017. https://doi.org/10.1007/s40265-017-0731-3.

[3] S. R. Dash and S. Dehuri. Comparative study of different classification techniques for post operative patient dataset. *International Journal of Innovative Re-*
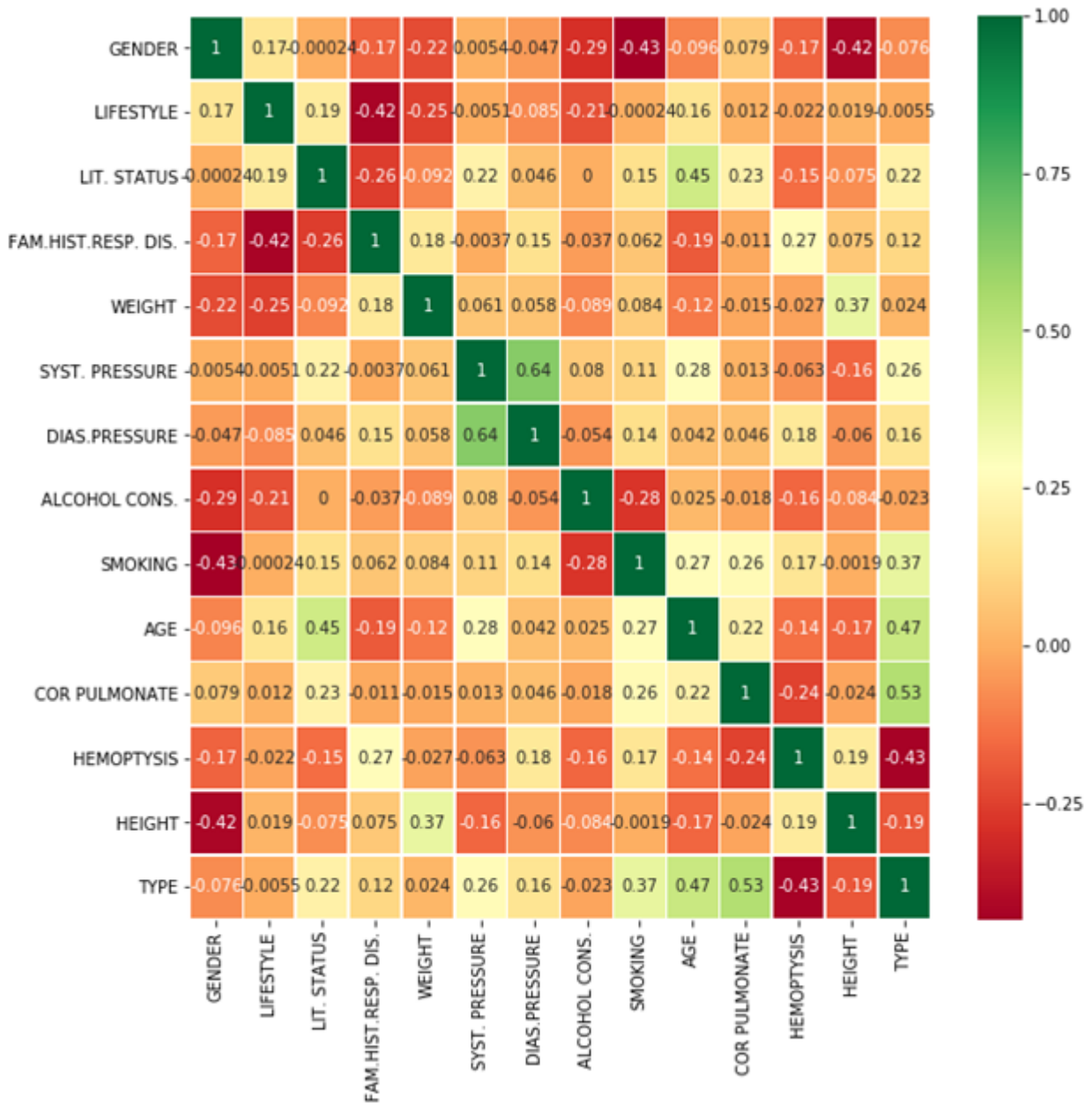
Figure 4: Heat Map determines Cor Pulmonale, age, and Smoking as Important Features. The Heat Map provides an important correlation between different features that are important about the classification process and assigns a degree of correlation, which is projected through the color mapping

search in Computer and Communication Engineering, 1(5):1101–1108, 2013.

[4] S. R. Dash and R. Ray. Predicting seminal quality and its dependence on life style factors through ensemble learning. *International Journal of E-Health and Medical Communications (IJEHMC)*, 11(2):78–95, 2020. https://doi.org/10.4018/ijehmc.2020040105.

[5] C. Esteban, J. Moraza, F. Sancho, M. Aburto, A. Aramburu, B. Goiria, A. Garcia-Loizaga, and A. Capelastegui. Machine learning for copd exacerbation prediction, 2015. https://doi.org/10.1183/13993003.congress-2015.oa3282.

[6] J. R. Feary, L. C. Rodrigues, C. J. Smith, R. B. Hubbard, and J. E. Gibson. Prevalence of major comorbidities in subjects with copd and incidence of myocardial infarction and stroke: a comprehensive analysis using data from primary care. *Thorax*, 65(11):956–962, 2010. https://doi.org/10.1136/thx.2009.128082.

[7] F. M. Franssen and C. L. Rochester. Comorbidities in patients with copd and pulmonary rehabilitation: do they matter?, 2014. https://doi.org/10.1183/09059180.00007613.

[8] Y. Fukuchi. The aging lung and chronic obstructive pulmonary disease: similarity and difference. *Proceedings of the American Thoracic Society*, 6(7):570–572, 2009. https://doi.org/10.1513/pats.200909-099rm.

[9] K. Ghoorah, A. De Soyza, and V. Kunadian. Increased cardiovascular risk in patients with chronic obstructive pulmonary disease and the potential mechanisms linking the two conditions: a review. *Cardiology in review*, 21(4):196–202, 2013. https://doi.org/10.1097/crd.0b013e318279e907.

[10] K. E. Holm, M. R. Plaufcan, D. W. Ford, R. A. Sandhaus, M. Strand, C. Strange, and F. S. Wamboldt. The impact of age on outcomes in chronic obstructive pulmonary disease differs by relationship status. *Journal of behavioral medicine*, 37(4):654–663, 2014. https://doi.org/10.1007/s10865-013-9516-7.

[11] S. Khan, P. Fell, and P. James. Smoking-related chronic obstructive pulmonary disease (copd). *Diversity and Equality in Health and Care*, 11(3-4):267–271, 2014.

[12] R. Laniado-Laborín. Smoking and chronic obstructive pulmonary disease (copd). parallel epidemics of the 21st century. *International journal of environmental research and public health*, 6(1):209–224, 2009. https://doi.org/10.3390/ijerph6010209.

[13] J. D. Maclay and W. MacNee. Cardiovascular disease in copd: mechanisms. *Chest*, 143(3):798–807, 2013. https://doi.org/10.1378/chest.12-0938.

[14] R. Mitchell, J. Michalski, and T. Carbonell. *An artificial intelligence approach*. Springer, 2013.

[15] D. Panda and S. R. Dash. Predictive system: Comparison of classification techniques for effective prediction of heart disease. In *Smart Intelligent Computing and Applications*, pages 203–213. Springer, 2020. https://doi.org/10.1007/978-981-13-9282-5_19.

[16] D. Panda, R. Ray, A. A. Abdullah, and S. R. Dash. Predictive systems: Role of feature selection in prediction of heart disease. In *Journal of Physics: Conference Series*, volume 1372, page 012074. IOP Publishing, 2019. https://doi.org/10.1088/1742-6596/1372/1/012074.

[17] A. I. Papaioannou, K. Bartziokas, S. Loukides, S. Tsikrika, F. Karakontaki, A. Haniotou, S. Papiris, D. Stolz, and K. Kostikas. Cardiovascular comorbidities in hospitalised copd patients: a determinant of future risk? *European Respiratory Journal*, 46(3):846–849, 2015. https://doi.org/10.1183/09031936.00237014.

[18] B. N. Patel, S. G. Prajapati, and K. I. Lakhtaria. Efficient classification of data using decision tree. *Bonfring International Journal of Data Mining*, 2(1):06–12, 2012. https://doi.org/10.9756/bijdm.1098.

[19] J. Peng, C. Chen, M. Zhou, X. Xie, Y. Zhou, and C.-H. Luo. A machine-learning approach to forecast aggravation risk in patients with acute exacerbation of chronic obstructive pulmonary disease with clinical indicators. *Scientific reports*, 10(1):1–9, 2020. https://doi.org/10.1038/s41598-020-60042-1.

[20] J. Quint. The relationship between copd and cardiovascular disease. *Tanaffos*, 16(Suppl 1):S16–S17, 2017.

[21] K. F. Rabe, J. R. Hurst, and S. Suissa. Cardiovascular disease and copd: dangerous liaisons? *European Respiratory Review*, 27(149), 2018. https://doi.org/10.1183/16000617.5057-2018.

[22] L. Roever et al. Translational medicine. 2015.

[23] A. Shujaat, R. Minkin, and E. Eden. Pulmonary hypertension and chronic cor pulmonale in copd. *International journal of chronic obstructive pulmonary disease*, 2(3):273–282, 2007.

[24] D. D. Sin and S. P. Man. Chronic obstructive pulmonary disease as a risk factor for cardiovascular morbidity and mortality. *Proceedings of the American Thoracic Society*, 2(1):8–11, 2005. `https://doi.org/10.1513/pats.200404-032ms`.

[25] A. Undas, P. Kaczmarek, K. Sladek, E. Stepien, W. Skucha, M. Rzeszutko, I. Gorkiewicz-Kot, and W. Tracz. Fibrin clot properties are altered in patients with chronic obstructive pulmonary disease. *Thrombosis and haemostasis*, 102(12):1176–1182, 2009. `https://doi.org/10.1160/th09-02-0118`.

[26] Y. Xie, S. J. Redmond, M. S. Mohktar, T. Shany, J. Basilakis, M. Hession, and N. H. Lovell. Prediction of chronic obstructive pulmonary disease exacerbation using physiological time series patterns. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6784–6787. IEEE, 2013. `https://doi.org/10.1109/embc.2013.6611114`.

[27] Y. Yang, J. Li, and Y. Yang. The research of the fast svm classifier method. In *2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pages 121–124. IEEE, 2015. `https://doi.org/10.1109/iccwamtip.2015.7493959`.