Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 44 (2023), No. 6, pp. 1097–1114
DOI : 10.47974/JIOS-1444

A comparative study of metaheuristic-based machine learning classifiers using non-parametric tests for the detection of COPD severity grade

Akansha Singh *
Guru Gobind Singh Indraprastha University
University School of Information, Communication & Technology
New Delhi
India

Nupur Prakash
Department of Computer Science and Engineering
Northcap University
Gurgaon
Haryana
India

Anurag Jain
Guru Gobind Singh Indraprastha University
University School of Information, Communication & Technology
New Delhi
India

Abstract

In this paper, a comparative study between five different Machine Learning (ML) classifiers has been performed for the detection of Chronic Obstructive Pulmonary Disease (COPD) severity grade. There are various existing studies that provide various computer-aided frameworks for the diagnosis of COPD. However, such existing approaches have certain limitations such as lower performance, imbalanced data, and missing data. Also, no statistical tests were performed that could justify the validity of the proposed techniques. Hence, to provide the best model for the detection of COPD severity grade, a conceptualized framework has been provided. The performance of ML classifiers has been optimized by utilizing the Correlation-based Elephant search algorithm (CFS+ESA) and InformationGain and GainRatio feature selection techniques (FS). Experimental work carried out on the COPD patient dataset has been validated using statistical tests. The Wilcoxon test and Friedman test

* *E-mail:* **akansha.trar03@gmail.com**

© Taru Publications

were utilized to compare the FS techniques and ML techniques respectively. The performance comparison has been carried out across different training-testing criteria namely, 10-fold cross-validation, 70%-30%, 75%-25%, and 80%-20%. The results from Wilcoxon tests showed that CFS+ESA has given the best results across all the training-testing criteria. Similarly, the results from Friedman's tests demonstrated that the CFS+ESA-based Logistic Regression classifier trained using 10-fold cross-validation has outperformed the rest combinations of classifiers and training-testing criteria by achieving the highest rank.

Subject Classification: *62M20, 91B82, 60-08.*

Keywords: *Meta-heuristic optimization, SMOTE, EM imputation, Feature selection, Machine learning, Wilcoxon test, Friedman test.*

1. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a Chronic Respiratory disease (CRD) that causes airway inflammation making it difficult for a person to breathe. It is the third deadliest disease in the world, causing millions of deaths worldwide. It has been estimated that by 2050, the prevalence of COPD will reach 592 million people. It is generally caused by smoking, but other factors also include air pollution, chemical fumes, and alpha anti-trypsin deficiency. The symptoms of COPD include short breaths, chest tightness, mucus production, and fatigue [1]. It is not a fully reversible disease but with proper treatment and management, the complexity of COPD can be reduced further improving the quality of life of COPD patients. Morbidity and mortality of COPD patients can only be

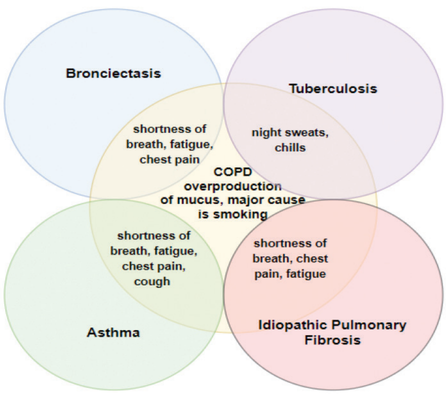


Figure 1

Euler diagram depicting the similar and overlapping symptoms between COPD and other Chronic respiratory diseases.

reduced if detected early. However, the early diagnosis of COPD is quite complicated because of its similarity with other CRDs in terms of similar and overlapping symptoms as shown in Figure 1.

The major problem leading to the fast worsening of COPD patients is acute exacerbations. These exacerbations can be mild, moderate, severe, and very severe. The management of exacerbations depends on the correct detection of their severity grade. The tools available for diagnosing COPD such as Spirometry [2], Chest X-Rays, and Peak expiratory flow rate (PEFR) [3] are not reliable for diagnosing COPD especially when considered standalone. Also, the tools necessary for diagnosing COPD are not available in every clinic. Hence, there is a need for computer-aided framework such as Machine learning (ML) [4-8] and Deep Learning (DL) [9] that can diagnose COPD accurately without relying on any expensive tools. In this regard, various research have been conducted for diagnosing COPD and the associated factors with it such as predicting mortality rate [10], hospitalization rate [11] and exacerbations [12] etc. However, the existing methods encompass certain limitations such as lower performances, imbalanced data, and missingness problems. Also, no statistical tests were performed to justify the validity of the proposed models.

Also, lots of research were conducted to determine the severity of COPD [13,14]. However, only the technique proposed by [14] has determined COPD severity grades using the same dataset that has been utilized in this study. The limitations of this study are that the dataset consisted of missing values that were left untreated, and the dataset was imbalanced which further affected the results. Hence, the main aim of this study is to provide a framework for diagnosing the COPD severity grade as per GOLD standard efficiently and accurately. It also aims to solve the data imbalance and missingness problem. Therefore, this study utilizes different ML classifiers for the detection of COPD severity grade. The performance of ML classifiers has been enhanced by deploying a metaheuristic [15,16] and different non-metaheuristic techniques. In addition, to select the best classifier for COPD severity grade detection, a comparison has been made between metaheuristic-based ML classifiers. Furthermore, to validate the resulted performance of classifiers, two non-parametric statistical tests have been utilized namely, Wilcoxon [17] and Friedmann tests [18].

The rest of the paper is organized as follows: Section 2 discusses the relevant related work in the field of detecting COPD severity grade using ML techniques while Section 3 elaborates on the material and methods used in this study. Section 4 illustrates the experimental work carried out on the deployed dataset using different FS and ML techniques. It further validates the performance of different FS and ML models using various statistical tests. The conclusion and future work related to the research are discussed in Section 5.

2. Literature Review

This section discusses the previous research in diagnosing COPD using machine learning models. Based on such data, the researchers have predicted the mortality, hospitalization, exacerbation rates, etc. The authors [19] utilized different machine-learning algorithms for predicting all-cause mortality across two COPD cohorts namely COPDGene and ECLIPSE from which 2,632 and 1,268 participants were selected respectively. The proposed model MLMP-COPD outperformed the other existing models for predicting all-cause mortality and achieved a C index in both cohorts. Whereas in [20], the authors collected 5807 case data from a COPD surveillance program to identify the people at risk of COPD. The unbalanced data were treated using a cost-sensitive algorithm and SMOTE. The authors utilized seven different machine learning algorithms to predict COPD risk and concluded that among all the algorithms LR performed better than other algorithms. Similarly, five different machine learning models namely support vector machines (SVM), Naïve Bayes (NB), Boosted Decision Tree (BDT), Probabilistic Neural Network (PNN), and Logistic Regression (LR) were utilized for the prediction of COPD severity stages [21]. The performance of all the models was compared using different performance metrics such as accuracy, specificity, sensitivity, etc. The results showed that among all the models, PNN gives the best performance. The study [22] utilized a support vector machine to differentiate COPD and asthma using quantitative CT imaging. 95 patients were recruited from Heidelberg University Hospital. The result showed that the SVM models achieved accuracy and F1-score of 80% and 81% respectively in differentiating COPD and asthma. Also, the study listed seven top features of different COPD and asthma as inner area RB1, wall thickness RB1, outer area LB1, outer area LB1, TAC LB10, average outer/inner perimeter, Pi10, and TAC. Similarly, various machine learning classifiers such as NB, LR, NN, SVM, KNN, DT, and RF were compared for diagnosing asthma and COPD [23]. The result showed that Random Forest outperformed the other models by achieving a precision of 97.7% for COPD and 80.3% for asthma.

3. Materials and Methods

This section discusses the materials (e.g., dataset) and methods utilized in this study. It further explores the conceptual workflow designed for the detection of COPD severity grade. The workflow comprises of four main stages, namely, preprocessing, feature selection, classification, and evaluation.

3.1 *Dataset*

In this study, a COPD patient dataset has been deployed for the detection of COPD severity grade. It has been collected from the Kaggle repository. The dataset consists of 101 instances and 23 attributes. The dataset obtained from the repository has some limitations such as small size, missingness, and class imbalance problem. The dataset contains information regarding COPD patients who were asked to participate in a rehabilitation program to manage COPD. All these patients are classified into four categories based on their COPD severity grade i.e., mild, moderate, severe, and very severe whose distribution is shown in Figure 2.

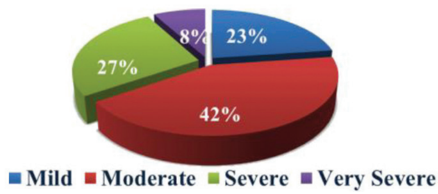


Figure 2

Distribution of COPD patients in the dataset based on the COPD Severity Grade i.e., mild, moderate, severe, and very severe.

3.2 *Conceptual framework for the detection of COPD severity grade*

In this section, a conceptual framework for the detection of COPD severity grade has been visualized in Figure 3. The framework comprised three phases, i.e., data preprocessing, feature selection, and classification. The first phase, i.e., the data preprocessing phase focuses on improving the raw original data in terms of quantity and quality. The second phase i.e., feature selection deals with the procedure of finding the minimal optimal feature subset. In this regard, two sets of techniques have been utilized i.e., metaheuristic-based FS and nonmetaheuristic-based FS. The complete procedure of FS has been discussed in the subsection. The minimal optimal feature subset from phase two has been then passed onto

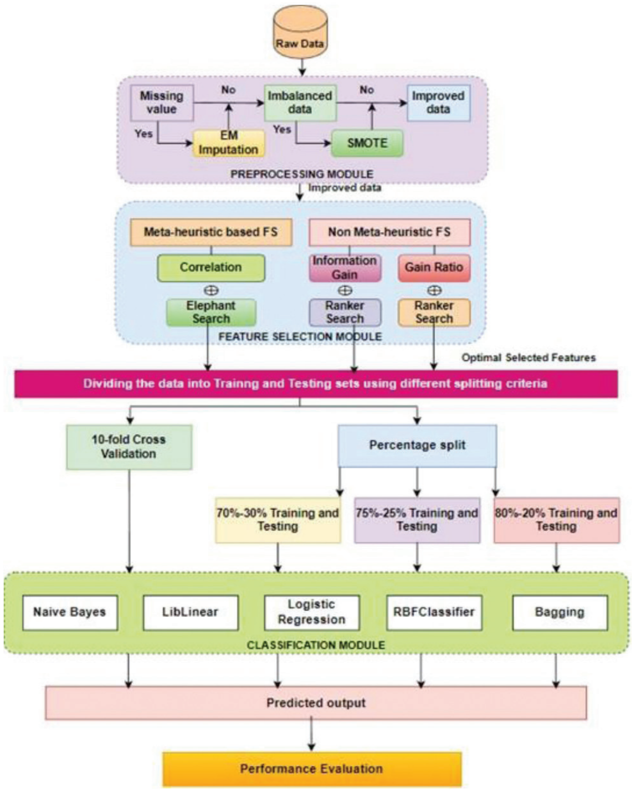


Figure 3

A conceptual framework for the detection of COPD severity grade

the third phase for classification. But before training, the dataset has been split into training and testing sets using different methods, i.e., 10-fold cross-validation, 70%-30%, 75%- 25%, and 80%-20% ratios. Then for each training and testing set, five different ML classifiers have been compared. The main objective is to find the best model for the detection of COPD severity grade.

3.2.1 Preprocessing Techniques

As mentioned in the above section, the dataset used in this study suffers from a class imbalance and missingness problem. Both problems are handled by different techniques as discussed in the following subsection.

- **Missing value problem**: The dataset has 4 missing values. To tackle the missing value problem, the Expectation Maximization (EM) Imputation [24] technique has been used. This technique finds the maximum log-likelihood estimates of the parameters of the model. It is an iterative method consisting of two major steps, namely, Expectation and Maximization. The first step finds the distribution of the unknown value from the known values of the observed data and the parameter's current estimates [25]. The second step maximizes the complete-data log-likelihood from the first step. The algorithm for the EM Imputation method has been shown in Algorithm 1.
- **Class Imbalance problem**: The dataset utilized in this study suffers from a class imbalance problem. To tackle this problem, SMOTE [12, 26] technique has been utilized. The algorithm and working of SMOTE technique are shown in Algorithm 2.

Algorithm 1: Expected Maximization Imputation

Data: Dataset $(Z) = (X, Y), mean\ vector\ (\mu), \Sigma n$
Result: Y: missing values
for *each* i **do**
  while $E_{new} = E\_old$ **do**
    Calculate $E_{old} = Y\_imiss | Y_{iobs}, \mu, \Sigma$;
    /* Calculate the EM and covariance of the missing value */
    $E_{new} = \max(E\_old)$; /* maximizing the old Expectation */
  end
end

Algorithm 2: Synthetic Minority Oversampling Technique

Data: Minority Class A consisting of x_k instances
Result: Set A_1 consisting x_m of synthetic instances
for *each* $x \in A$ **do**
  Find k nearest neighbors;
  Set percentage of the number of synthetic instances to be created;
  Select N random examples from k nearest neighbors;
  $A_1 \leftarrow N$;
  for *each* x_{m1} *belongs to* A **do**
    Calculate $x_{m1} = x + rand(0,1) * |x - x_k|$;
  end
end

3.2.2 Feature Selection Techniques

In this study, three different FS techniques, namely, Information Gain, Gain Ratio, and Elephant search have been utilized to find out the best minimal optimal feature subset. The first two techniques, Information Gain and Gain Ratio are based on filter methods that rank each feature based on some univariate metric[1]. The last technique is the Correlation feature selection [27]-based Elephant search algorithm (CFS+ESA). It is a wrapper method that finds a subset of features using ESA. ESA is a metaheuristic (bio-inspired) optimization algorithm that is inspired by the behavioral characteristics of elephant herds as shown in Figure 4.

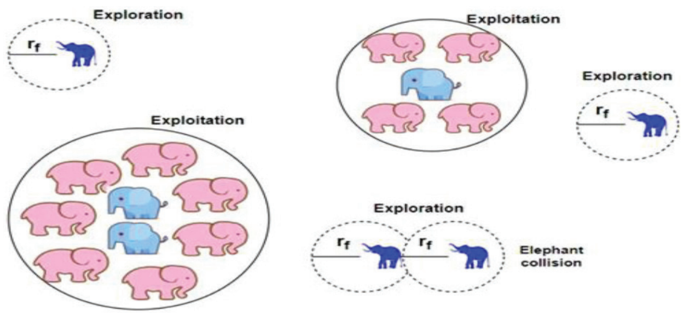


Figure 4

Diagrammatic representation of Elephant search algorithm (ESA).

In ESA, the local and global search is done by two search agents: male and female elephants. A can be seen in the figure, pink female elephants along with blue baby calves form a herd and always move in concentric circles doing a local search, i.e., exploitation. Whereas the blue male elephants move in solidarity performing a global search, i.e., exploration. Their main objective is to do a global search and find the best optimal solutions. Other than these objectives, some of the other objectives also include avoiding local optima and well covered search space. The working of ESA has been shown in Algorithm 3.

3.2.3 ML Classifiers

After performing pre-processing and FS, the dataset is divided into training and testing sets using K-fold cross-validation where the K value is set to 10. The training data is fed into different supervised classifiers,

[1] Filter methods, accessed on 22/06/2023

Algorithm 3: Elephant Search Algorithm (ESA)

Data: Set of features $E = [E_m, E_f]$ where E_m and E_f corresponds to male and female elephants population size respectively, R_m: Radius of male elephant search range, A:age of elephant, α: maximum age limit, b: born elephant, X: position of Male Elephant, Y: position of female elephants
Result: best global fitness value: g_{best}, Feature subset: E'

```
Initialize;                                                /*Initialization of all ESA parameters */
c = 0.4;                                                   /* chaotic coefficient */
M_p = 0.01;                                                /* mutation probability */
p = 22;                                                    /* Population size */
t = 1;                                                     /* Number of iterations */
While t < 20 do
    for each e_m in E_m do                                 /*Exploration by Male agents */
        X_new,mi(t) = X_mi(t) + c_m * r_m(X_best,mi - X_mi(t));
        /* generate new locations   for male agents */
        If (|R(e_mi) - R(e_mj)| < R(e_m)) && (f(e_mi) > f(e_mj)) then
            Remove e_mj;                                   /* Collision of two male elephants */
        end
    end
    for each e_f in E_f do                                 /*Exploitation by Female agents */
        E_f(leader) ← max[A(e_f1, e_f2, ... ... ... e_fn)];
        X_new,fi(t) = X_fi(t) + c_f * r_f(X_best,fi - X_fi(t));
        /* generate new locations   for female agents */
        Update f(E_f);
        f_best(E_f);
    end
    for all elephants e_m in E_m & e_f in E_f do
        E_f(leader) ← max[A(e_f1, e_f2, ... ... ... e_fn)];
        If A(E) > α then
            Replace A(E) ← b(e');
            /* Evolution of previous generation with mutation probability 0.01 */
        end
    end
    If f_best(E_m) ≥ f_best(E_f) then
        f_best(E_f) ← f_best(E_m);
    /* Female herd will move towards male agent search space */
    end
    t++;
    Return g_best, E';
end
```

namely, Naïve Bayes (NB), LibLinear, Logistic Regression (LR), RBFClassificer, and Bagging. The dataset is a labeled dataset hence supervised classifiers have been used for classification. The hyperparameters that were set for all these classifiers are shown in Table 1.

Table 1

Hyperparameters set corresponding to different ML classifiers.

Classifier	Hyperparameters
Naïve Bayes	batchSize=100, supervised_discretion='false'
LibLinear	SVM Type: L2-regularized L2-loss SVM, BatchSize=100, bias=1.0, Max. no. of iterations=1000
Logistic Regression	BatchSize=100, estimator=ridge
RBF Classifier	Seed =1, batchSize=100, number of basis function to use=2, ridge penalty factor for output layer=0.01
Bagging	bagSizePercent=100, batchSize=100, seed=1, and number of iterations=10

4. Result and analysis

In this section, the series of experiments performed on the COPD patient dataset have been discussed.

4.1 *Experimental setup*

All the experiments have been performed on a Windows machine with AMD Ryzen 5 4600H with a Radeon Graphics processor and 24 GB RAM. This study has utilized the WEKA tool for data mining procedure, jupyter notebook for visualization and exploratory data analysis (EDA), and Statistical Package for Social Sciences (SPSS) tool for statistical testing.

4.2 *Comparison and analysis of feature selection techniques*

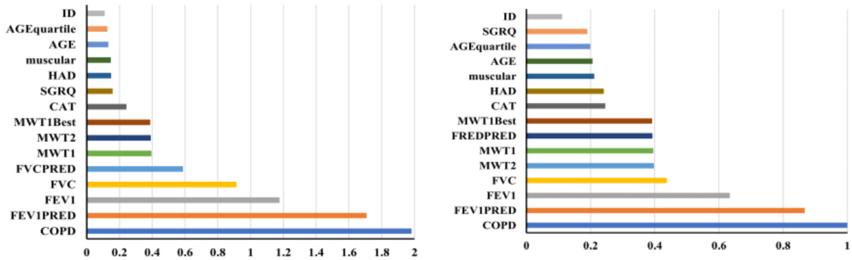
This section discusses the experiments carried out for choosing the best optimal feature set that can aid in improving classification accuracy. As mentioned earlier, three different types of FS techniques have been considered, namely, CFS+ESA, Info- Gain+Ranker search, and GainRation+Ranker search. After running all the FS tech- niques on the COPD dataset, the resulting number of features is shown in Table 2.

The number of features selected by CFS+ESA is way less as compared to other non-metaheuristic-based techniques, i.e., InfoGain+Ranker search and GainRa- tion+Ranker search. Also, the number of features selected by non-metaheuristic-based techniques is the same. However, the rank obtained for the selected features is different as visualized in Figure 5.

Table 2

The number of features selected by different feature selection algorithms.

Feature Selection Techniques	No. of Features Selected	Selected Features
Correlation with Elephant search	6	MWT1, FEV1, FEV1PRED, FVCPRED, HAD, copd
Information Gain	15	Copd, FEV1PRED, FEV1, FVC, FVCPRED, MWT1, MWT1Best, MWT2, CAT, SGRQ, HAD, muscular, AGE, AGEquartile, ID
Gain Ratio	15	Copd, FEV1PRED, FEV1, FVC, MWT1, MWT1Best, MWT2, FVCPRED, CAT, HAD, muscular, AGE, AGEquartile, SGRQ, ID



(a) Information Gain (b) Gain Ratio

Figure 5

Ranking of features on the basis of their importance calculated using (a) InformationGain and (b) Gain- Ratio metric.

4.3 *Wilcoxon statistical test for Pairwise FS techniques*

It is evident from Table 2 that out of all the FS techniques, CFS+ESA has provided a minimal feature set. However, to justify that the minimal feature set obtained from CFS+ESA is optimal as compared to other non-metaheuristic FS techniques, this study has utilized the Wilcoxon test [17]. The main objective is to determine how two or more sets of pairs are significantly apart from one another. The FS techniques have been compared based on classification accuracy. The results obtained after performing the Wilcoxon test are shown in Table 3. In the table, Negative rank indicates CFS+ESA> InformationGain/GainRatio+Ranker search and positive rank indicates CFS+ESA< InformationGain/ GainRatio+Ranker search.

Table 3

Wilcoxon test statistics and ranks for pairwise comparison between CFS+Elephant searcg and (InformationGain/GainRatio)+Ranker search across different training and testing percentages.

Wilcoxon Test Parameters	**CFS+Elephant search and (InfoGain/ Gainratio)+Ranker search**			
	10-fold cross validation	**70%-30%**	**75%-25%**	**80%-20%**
Sum of positive ranks	2	6	4	5
Sum of negative ranks	8	9	6	10
Z	-1.095	-0.405	-0.365	-0.674
p-value	0.273	0.686	0.715	0.5

The results have evidently shown that Metaheuristic based FS technique i.e., CFS+ESA has performed better across all the different training and testing per- centages. The negative ranks indicated that CFS+ESA has outperformed the other Non-metaheuristic FS technique. Furthermore, among the Cross-validation and different training-testing percentages, CFS+ESA has performed the best for 80%-20% training-testing percentage.

4.4 *Comparison analysis of ML Classifiers*

This section discusses the series of classification experiments performed on the COPD patient dataset using the minimal optimal feature set obtained from the CFA+ESA FS technique. The results obtained after the classification for different evaluation metrics such as Accuracy, Kappa statistics (KS), ROC, F-measure, Matthew's correlation coefficient (MCC), Mean Absolute Error (MAE), and Root mean square error (RMSE) are depicted in Table 4. The classification results have shown that among all the ML classifiers, LibLinear has given the worst performance in terms of all evaluation metrics and also across different training- testing criteria. The rest of the classifiers have somewhat similar performances across all the training-testing criteria for all the evaluation metrics. Hence, to find out the best model with the best training-testing criteria, further statistical testing has been done as discussed in the subsection below.

Table 4

Classification Performance of ML classifiers across different training-testing criteria based on Accuracy(%), Kappa statistics(%), ROC(%), F-measure(%), MCC,MAE, and RMSE metric.

Training-Testing ratio	Performance metric	ESA + NB	ESA + LibLinear	ESA + LR	ESA + RBFClassifier	ESA + Bagging
10-fold CV	Accuracy	98.06	57	100	97.1	98.5
70%-30%		98.38	41.9	96.77	98.3	98.3
75%-25%		98.07	28.84	100	98.07	98.07
80%-20%		97.56	41.46	97.56	92.6	97.56
10-fold CV	Kappa statistics	97.4	42.04	100	96.1	98
70%-30%		97.8	13.02	95.62	97.8	97.8
75%-25%		97.34	98.9	100	97.34	97.34
80%-20%		96.61	20.06	96.61	89.81	96.61

Contd...

10-fold CV	ROC	99.6	71.4	100	99.2	100
70%-30%		99.8	56.1	99.5	100	100
75%-25%		99.99	55.7	100	100	100
80%-20%		99.99	60	100	100	100
10-fold CV	F-measure	98.6	53.1	100	97.1	98.5
70%-30%		98.4	38.1	96.8	98.4	98.4
75%-25%		98.1	28.8	100	98.1	98.1
80%-20%		97.5	36.4	97.5	92.5	97.5
10-fold CV	MCC	97.5	41.6	100	96.2	98
70%-30%		97.8	42.7	95.5	97.8	97.8
75%-25%		97.4	44.6	100	97.4	97.4
80%-20%		96.8	34.5	96.8	90.8	96.8
10-fold CV	MAE	0.0114	0.215	0.0001	0.1328	0.0088
70%-30%		0.0106	0.2903	0.0162	0.1572	0.0069
75%-25%		0.0123	0.3558	0.0001	0.1483	0.0108
80%-20%		0.016	0.2927	0.0116	0.1447	0.0101
10-fold CV	RMSE	0.0914	0.4637	0.002	0.1936	0.0605
70%-30%		0.0907	0.5388	0.127	0.2207	0.0605
75%-25%		0.0979	0.59965	0.0008	0.2081	0.0636
80%-20%		0.1116	0.541	0.1052	0.207	0.0674

4.5 *Friedman's statistical test for multiple ML classifiers*

Friedman is also a non-parametric test to compare multiple results [18]. In this study, this test has been utilized for comparing the classification performance of ML classifiers for different evaluation metrics across different training-testing criteria. It assigns ranks to individual classifiers determining how well that classifier has performed by setting the different Friedman's statistics as shown in Table 5. In this Table, N defines the total number of individuals considered, i.e., 10-fold cross validation, 70%-30%, 75%-25%, and 80%-20%.

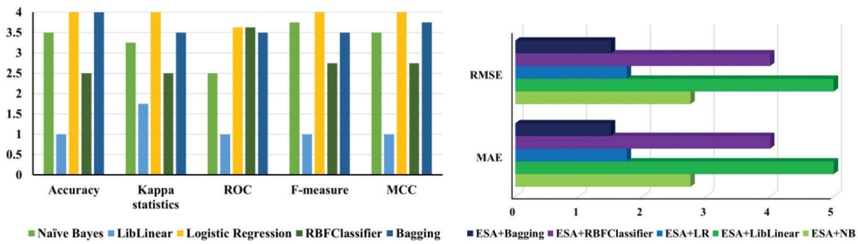
The ranks obtained after performing Friedman's tests are shown in Figure 6.

It is evident from Friedman's test results, that among all the ML classifiers, LR has shown the best performance as it has obtained the highest rank among all the classifiers across all the evaluation metrics.

Table 5

Friedman's test statistics for comparing ML classifiers across different training-testing criteria for different evaluation metrics.

Statistics	Accuracy	KS	ROC	F-measure	MCC	MAE	RMSE
N	4	4	4	4	4	4	4
Chi-square	11.556	5.882	12.057	11.059	11.059	14.200	14.200
Degree of freedom	4	4	4	4	4	4	4
P-value	0.021	0.208	0.017	0.026	0.026	0.007	0.007



(a) Accuracy, Kappa statistics, ROC, F-measure, MCC (b) MAE and RMSE

Figure 6

Comparison of ML classifiers on the basis of ranks obtained through Friedman's testing different metric (a) and (b).

Moreover, in terms of error rate, Bagging exhibits the least error rate and LR exhibits the second least error rate. However, the error difference between LR and Bagging is only 0.25. Hence, it can be said that the overall best performance has been given by LR. The reason behind the outstanding performance of LR as compared to other classifiers could be its ability to deal with overfitting. Furthermore, another Friedman's test has been performed for comparing the training-testing criteria and the test statistic setting done for this experiment has been depicted in Table 6.

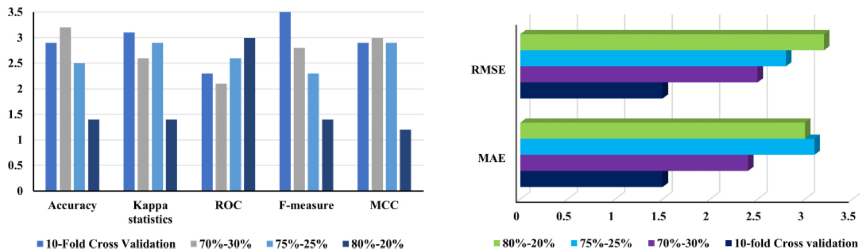
In the Table, N defines the total number of individuals considered, i.e., Naïve Bayes, LibLinear, Logistic Regression, RBFClassifier, and Bagging. The ranks obtained after performing Friedman's test on different training-testing criteria are shown in Figure 7.

The results obtained from Friedman's test indicated that for Kappa statistic and F-measure, 10-fold Cross-validation, for accuracy and MCC, 70%-30% ratio, and for ROC, ratio 80%-20% has given the best results. However, it has also been observed that among all the training-testing criteria, 10-fold Cross-validation has the least error rate. The excellent

Table 6

Friedman's test statistics for comparing different training-testing criteria across ML classifiers for different evaluation metrics.

Statistics	Accuracy	KS	ROC	F-measure	MCC	MAE	RMSE
N	5	5	5	5	5	5	5
Chi-square	5.694	5.327	2.226	7.163	6.918	4.959	4.837
Degree of freedom	3	3	3	3	3	3	3
P-value	0.127	0.149	0.527	0.67	0.75	0.175	0.184



(a) Accuracy, Kappa statistics, ROC, F-measure, MCC (b) MAE and RMSE

Figure 7

Comparison of different training-testing criteria based on ranks obtained through Friedman's testing using different metric (a) & (b).

performance of cross-validation has been supported by the fact that it works really with a small dataset. Instead of completely partitioning the data into train-test splits, Cross-validation makes a number of folds and randomly chooses a different train-test set. This helps in avoiding overfitting as it works on different subsets of data. Therefore, this study concludes that for the detection of COPD severity grade, the CFS+ESA-based Logistic regression model has given the best performance using the 10-fold cross-validation criteria.

5. Conclusion and future work

The main objective of this study was to provide a framework for the detection of COPD severity grade. For this purpose, five different ML classifiers were compared across different training-testing criteria. To further enhance the classification performance, both metaheuristic and non-metaheuristic-based FS techniques were utilized. In addition, missing values and class imbalance problems were treated using EM imputation and SMOTE. Furthermore, Wilcoxon and Friedman tests were conducted to determine the best FS and ML techniques. The results showed that for ESA, all the ML classifiers performed the best. Also, the LR classifier performed the best in terms of Accuracy, Kappa statistic, F-measure, ROC, and MCC by achieving a rank of 4. Also, the error rate was lower for 10-fold cross validation criteria. Hence, this study concludes that for the detection of COPD severity grade, Elephant search-based Logistic regression model has given the best performance by utilizing 10-fold cross-validation.

In the future, this study aims to provide a framework for the diagnosis of COPD as well as other Chronic respiratory diseases by utilizing various unstructured data i.e., images, sound data, etc.

Conflicts of Interest

The authors declare that they have no competing interests.

References

[1] Singh, A., Prakash, N., & Jain, A. : A Review on Prevalence of World-wide COPD Situation. Proceedings of Data Analytics and Management: ICDAM 2022, 391-405 (2023).

[2] Perret, J., Yip, S. W. S., Idrose, N. S., Hancock, K., Abramson, M. J., Dharmage, S. C., Waidyatillake, N. : Undiagnosed and 'overdiagnosed'COPD using postbronchodilator spirometry in primary healthcare settings: a systematic review and meta-analysis. *BMJ Open Respiratory Research*, 10(1), e001478 (2023).

[3] Sorge, R., DeBlieux, P. : Acute exacerbations of chronic obstructive pulmonary disease: a primer for emergency physicians. *The Journal of Emergency Medicine*, 59(5), 643-659 (2020).

[4] Panda, H., Das, M., & Sahu, B. : Received signal strength prediction model for wireless underground sensor networks using machine learning algorithms. *Journal of Information and Optimization Sciences*, *43*(5), 949-962 (2022).

[5] Khan, Tayyab, and Karan Singh. "Resource management based secure trust model for WSN." *Journal of Discrete Mathematical Sciences and Cryptography* 22.8: 1453-1462 (2019).

[6] Joshi, Ashish, and Amar Kumar Mohapatra. "Authentication protocols for wireless body area network with key management approach." *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, no. 2, pp. 219-240, (2019).

[7] Khan, T., Singh, K., Shariq, M., Ahmad, K., Savita, K. S., Ahmadian, A., ... & Conti, M. : An efficient trust based decision-making approach for WSNs: Machine learning oriented approach. *Computer Communications*, *209*, 217-229 (2023).

[8] Kumar, A., Singh, K., & Khan, T. : L-RTAM: Logarithm based reliable trust assessment model for WBSNs. *Journal of Discrete Mathematical Sciences and Cryptography*, 24(6), 1701-1716 (2021).

[9] Bathla, Gourav, Himanshu Aggarwal, and Rinkle Rani. "Using deep learning to improve recommendation with direct and indirect social trust." *Journal of Statistics and Management Systems* 22.4 : 665-677 (2019).

[10] Liao, K. M., Liu, C. F., Chen, C. J., Shen, Y. T. : Machine Learning Approaches for Predicting Acute Respiratory Failure, Ventilator Dependence, and Mortality in Chronic Obstructive Pulmonary Disease. *Diagnostics*, 11(12), 2396 (2021).

[11] Goto, T., Jo, T., Matsui, H., Fushimi, K., Hayashi, H., & Yasunaga, H. : Machine learning-based prediction models for 30-day readmission after hospitalization for chronic obstructive pulmonary disease. *COPD: Journal of Chronic Obstructive Pulmonary Disease*, 16(5-6), 338-343 (2019).

[12] Hussain, A., Choi, H. E., Kim, H. J., Aich, S., Saqlain, M., Kim, H. C. : Forecast the exacerbation in patients of chronic obstructive pulmonary disease with clinical indicators using machine learning techniques. *Diagnostics*, 11(5), 829 (2021).

[13] Choi, E. A., Kim, J. W., Nah, G., Kim, W., Yoo, K., Kim, Y. Y., Yoon, D. : Prediction of COPD severity based on clinical data using Machine Learning. In 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1646- 1648 (2021, December). IEEE.

[14] Dritsas, E., Alexiou, S., & Moustakas, K. : Copd severity prediction in elderly with ml techniques. In Proceedings of the 15th International Conference on PErvasive Technolo- gies Related to Assistive Environments, pp. 185-189 (2022, June).

[15] Singh, A., & Jain, A. : Cost-sensitive metaheuristic technique for credit card fraud detection. *Journal of Information and Optimization Sciences*, 41(6), 1319-1331 (2020).

[16] Kumar, K. V., Kumari, P., Rao, M., & Mohapatra, D. P. : Metaheuristic feature selection for software fault prediction. *Journal of Information and Optimization Sciences*, 43(5), 1013-1020 (2022).

[17] Balamurugan, R., Ratheesh, S., Venila, Y. M. : Classification of heart disease using adaptive Harris hawk optimization-based clustering

algorithm and enhanced deep genetic algorithm. *Soft Computing*, 1-17 (2022).

[18] Gupta, A., Kumar, R., Arora, H. S., & Raman, B. : C-CADZ: computational intelli- gence system for coronary artery disease detection using Z-Alizadeh Sani dataset. *Applied Intelligence*, 52(3), 2436-2464 (2022).

[19] Wang, X., Ren, H., Ren, J., Song, W., Qiao, Y., Ren, Z., ... Qiu, L. : Machine learning- enabled risk prediction of Chronic Obstructive Pulmonary Disease with unbalanced data. *Computer Methods and Programs in Biomedicine*, 107340 (2023).

[20] Moll, M., Qiao, D., Regan, E. A., Hunninghake, G. M., Make, B. J., Tal-Singer, R., ... Cho, M.H. : Machine learning and prediction of all-cause mortality in COPD. *Chest*, 158(3), 952-964 (2020).

[21] Mohamed, I. : Prediction of Chronic Obstructive Pulmonary Disease Stages Using Ma- chine Learning Algorithms. *International Journal of Decision Support System Technology (IJDSST)*, 14(1), 1-13 (2022). http://doi.org/10.4018/IJDSST.286693.

[22] Moslemi, A., Kontogianni, K., Brock, J., Wood, S., Herth, F., Kirby, M. : Differenti- ating COPD and asthma using quantitative CT imaging and machine learning. *European Respiratory Journal*, 60(3) (2022).

[23] Spathis, D., Vlamos, P. : Diagnosing asthma and chronic obstructive pulmonary disease with machine learning. *Health Informatics Journal*, 25(3), 811-827 (2019).

[24] Nekoee Zahraei, H., LOUIS, R., Donneau, A. F. : Comprehensive Clustering Analysis for Incomplete COPD Dataset. *In Journ´ee des Doctorants de l'Ecole Doctorale SPSS* (2019, December).

[25] Romanova, A. : Predicting Length of Stay for COPD Patients with Generalized Linear Models and Random Forests. In Advances in Computer Vision and Computational Biol ogy: Proceedings from IPCV'20, HIMS'20, BIOCOMP'20, and BIOENG'20, pp. 449-457 (2021). Cham: Springer International Publishing.

[26] Mahmudah, K. R., Purnama, B., Indriani, F., Satou, K. : Machine Learning Algorithms for Predicting Chronic Obstructive Pulmonary Disease from Gene Expression Data with Class Imbalance. In *BIOINFORMATICS*, pp. 148-153 (2021, February).

[27] Singh, A., Jain, A. : Hybrid bio-inspired model for fraud detection with correlation based feature selection. *Journal of Discrete Mathematical Sciences and Cryptography*, 24(5), 1365-1374 (2021).