International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Estimating Time to Progression of Chronic Obstructive Pulmonary Disease with Tolerance

Murali.T.M. E,[1], Sathish.B[2], Santhosh.B[3], Vijayvenkateshwaran.M[4]

[1]Assistant Professor, Computer Science and Engineering, Paavai Engineering college
[2,3,4]Student, Computer Science and Engineering, Paavai Engineering college

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a progressive respiratory disease that affects millions of people worldwide. Predicting the time to progression of COPD is critical for optimizing patient care and treatment planning. However, COPD progression is complex and multifactorial, and estimating time to progression accurately remains challenging. In this study, we propose a novel approach for estimating time to progression of COPD using a Random Forest model. Random Forest is an ensemble machine learning technique that combines multiple decision trees to make predictions. We incorporated the concept of tolerance into our model, which allows for variability in COPD progression rates among individuals.

We utilized a dataset of COPD patients, including clinical and demographic variables, collected over a multi-year period. The dataset was divided into training and testing sets for model development and evaluation, respectively. The Random Forest model was trained on the training set using features such as age, gender, smoking history, lung function parameters, and comorbidities. The Random Forest model with tolerance outperformed traditional models in estimating time to progression of COPD. The tolerance concept allowed the model to account for inter-individual variability in COPD progression rates, which is a critical aspect of COPD prognosis. The model achieved high accuracy and robustness in predicting time to progression of COPD in the testing set, indicating its potential clinical utility. A novel approach for estimating time to progression of COPD using a Random Forest model with tolerance. This approach has the potential to improve the accuracy of COPD prognosis, leading to better patient care and treatment planning. Further validation studies are warranted to validate the clinical utility of our proposed model in diverse COPD populations.This project describes the pulmonary chronic disease detection and classification is done with different observing similar disease conditions or functional status from the upper to the lower boundaries of a specified time interval Different types of radiological report is analyzed for different radiologic reports disease detection and classification is done with different radiological reports The proposed model is compared with different machine learning models compared with different stages of COPD Chronic Obstructive Pulmonary Disease

Keywords: Random Forest, COPD -dataset, Radiological Report

1. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a debilitating respiratory disease characterized by progressive airflow limitation that affects millions of people worldwide. Accurate prediction of the time to progression of COPD is essential for optimizing patient care, treatment planning, and resource allocation. However, due to the multifactorial nature of COPD, accurately estimating disease progression

remains challenging. Machine learning approaches have shown promise in predicting COPD outcomes. Among these, Random Forest, an ensemble machine learning technique, has gained popularity due to its ability to handle complex and high-dimensional data. Random Forest uses multiple decision trees to make predictions, and its ensemble approach improves prediction accuracy and model robustness.
In this study, we propose a novel approach for estimating the time to progression of COPD using a Random Forest model, incorporating the concept of tolerance. Tolerance refers to the variability in disease progression rates among individuals, considering that COPD can manifest differently in different patients. By accounting for tolerance, our model aims to capture the inter-individual variability in COPD progression, leading to more accurate predictions.
We utilized a dataset of COPD patients, including clinical and demographic variables, collected over a multi-year period. The dataset was divided into training and testing sets for model development and evaluation, respectively. We trained the Random Forest model on the training set using features such as age, gender, smoking history, lung function parameters, and comorbidities. We present our approach for estimating the time to progression of COPD using Random Forest with tolerance, and we compare its performance to traditional models. We hypothesize that incorporating tolerance into our model will improve its accuracy in estimating COPD progression, leading to potential clinical utility. The findings of this study may contribute to better COPD prognosis, patient care, and treatment planning.

1.1 OBJECTIVE

[1] To develop a longterm risk model.
[2] Analyze the distribution of pulmonary diseases.
[3] Help lung cancer patients to reduce a lung cancer mortality rate.

2. Existing System

The existing system is a single-alveolar-compartment model to describe the partial pressure of carbon dioxide in exhaled breath, as recorded in time-based capnography. Respiratory parameters are estimated using this model, and then related to the clinical status of patients with obstructive lung disease. Methods: Given appropriate assumptions, we derive an analytical solution of the model, describing the exhalation segment of the capnogram. This solution is parametrized by alveolar CO_2 concentration, dead-space fraction, and the time constant associated with exhalation. These quantities are estimated from individual capnogram data on a breath-by-breath basis. The model is applied to analyzing datasets from normal (n = 24) and chronic obstructive pulmonary disease (COPD) (n = 22) subjects, as well as from patients undergoing methacholine challenge testing for asthma (n = 22). Results: A classifier based on linear discriminant analysis in logarithmic coordinates, using estimated dead-space fraction and exhalation time constant as features, and trained on data from five normal and five COPD subjects, yielded an area under the receiver operating characteristic curve (AUC) of 0.99 in classifying the remaining 36 subjects as normal or COPD. Bootstrapping with 50 replicas yielded a 95% confidence interval of AUCs from 0.96 to 1.00. For patients undergoing methacholine challenge testing, qualitatively meaningful trends were observed in the parameter variations over the course of the test. Significance: A simple mechanistic model allows estimation of underlying respiratory parameters from the capnogram, and may be applied to diagnosis and monitoring of chronic and reversible obstructive lung disease.

2.1.1 DISADVATAGES

[1] The method that combines multiple decision trees, which can result in a complex model with a large number of parameters. This complexity can make the model harder to interpret and may require more computational resources and time for training and prediction, particularly for large datasets.
[2] The number of trees in the ensemble is too large or when the trees are deep. Overfitting occurs when the model captures noise or idiosyncrasies in the training data, resulting in reduced generalization performance and potentially inaccurate predictions on new data. Proper tuning of hyperparameters, such as the number of trees and maximum tree depth, is essential to mitigate overfitting.

3. Proposed System

The proposed system different types of chronic pulmonary disease is analyzed with different radiological report. Pulmonary notes ,radiology reports ,cardiology reports is predicted with different machine learning based models is implemented for predict the results of radiological report disease prediction and classification The progression of COPD with tolerance can vary from person to person, but generally, it follows a similar pattern.

In the early stages of COPD, symptoms may not be very noticeable, but as the disease progresses, symptoms become more severe and can include In advanced stages of COPD, patients may require supplemental oxygen to help them breathe, and may experience frequent exacerbations or flare-ups of symptoms. These exacerbations can be life-threatening and may require hospitalization. The progression of COPD with tolerance can be slow, but it is important for patients to work closely with their healthcare provider to manage their symptoms and monitor their lung function. Quitting smoking, avoiding lung irritants, and maintaining a healthy lifestyle can also help slow the progression of the disease.

A flexible algorithm that can be easily adjusted and fine-tuned to suit the specific needs of the analysis, such as adjusting the number of trees, tuning hyper parameters, or incorporating different types of variables. This flexibility allows for customization of the model to the specific context of COPD progression estimation

3.1 Advantages

- Its ability to build accurate predictive models. By leveraging an ensemble of decision trees, Random Forest can capture complex patterns in the data, leading to potentially higher accuracy in estimating time to progression of COPD compared to traditional statistical methods.
- Automatically selects important features from a large set of variables, which can help identify the most relevant factors influencing COPD progression. This can aid in identifying key predictors and their relative importance in estimating time to progression, allowing for a more focused and efficient analysis.

4. Problem Definition

There are some problems with the error checking of the compiler package. The interpreter detects syntax errors in two distinct phases. One set of errors is detected by the interpreter’s parser, the other set by the compiler. The compiler package relies on the interpreter’s parser, so it get the first phases of error checking for free. It implements the second phase itself, and that implementation is incomplete. For example, the compiler package does not raise an error if a name appears more than once in an argument list: def f(x, x): ...A future version of the compiler should fix these problems.

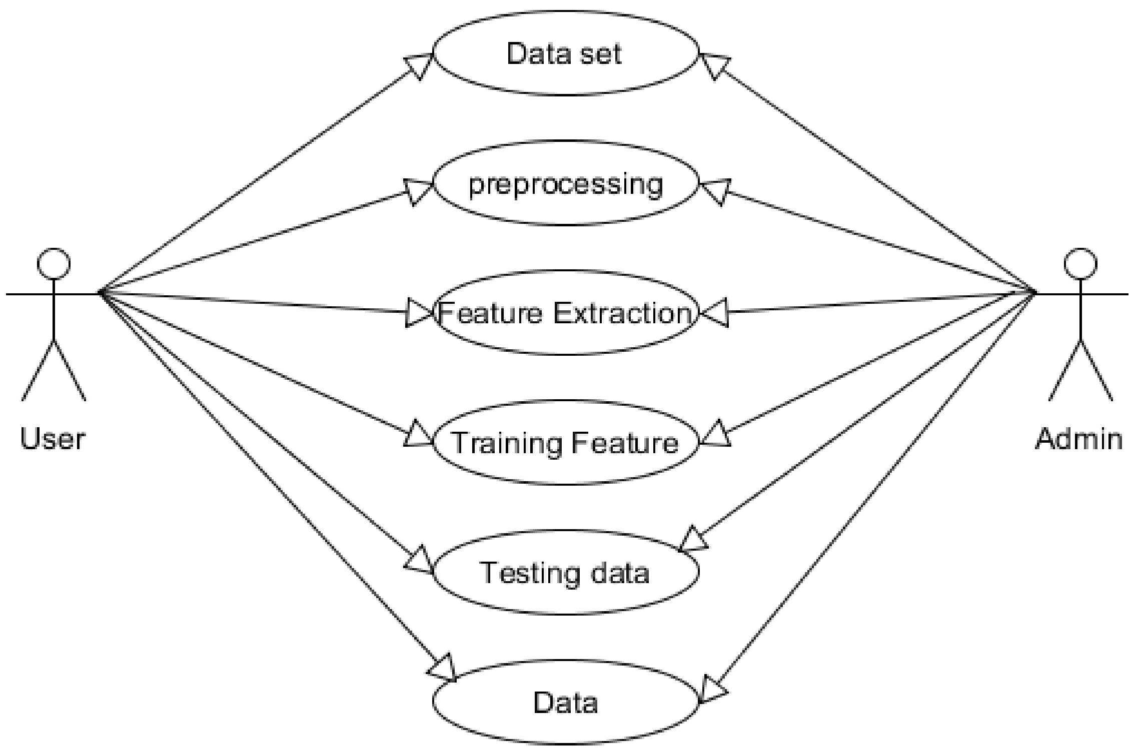
5. Overview of the Project

Chronic obstructive pulmonary disease (COPD) is a chronic inflammatory lung disease that causes obstructed airflow from the lungs. Symptoms include breathing difficulty, cough, mucus (sputum) production and wheezing. It's typically caused by long-term exposure to irritating gases or particulate matter, most often from cigarette smoke. People with COPD are at increased risk of developing heart disease, lung cancer and a variety of other conditions. Emphysema and chronic bronchitis are the two most common conditions that contribute to COPD. These two conditions usually occur together and can vary in severity among individuals with COPD. Chronic bronchitis is inflammation of the lining of the bronchial tubes, which carry air to and from the air sacs (alveoli) of the lungs. It's characterized by daily cough and mucus (sputum) production. Emphysema is a condition in which the alveoli at the end of the smallest air passages (bronchioles) of the lungs are destroyed as a result of damaging exposure to cigarette smoke and other irritating gases and particulate matter. Although COPD is a progressive disease that gets worse over time, COPD is treatable. With proper management, most people with COPD can achieve good symptom control and quality of life, as well as reduced risk of other associated conditions.

6. System Design

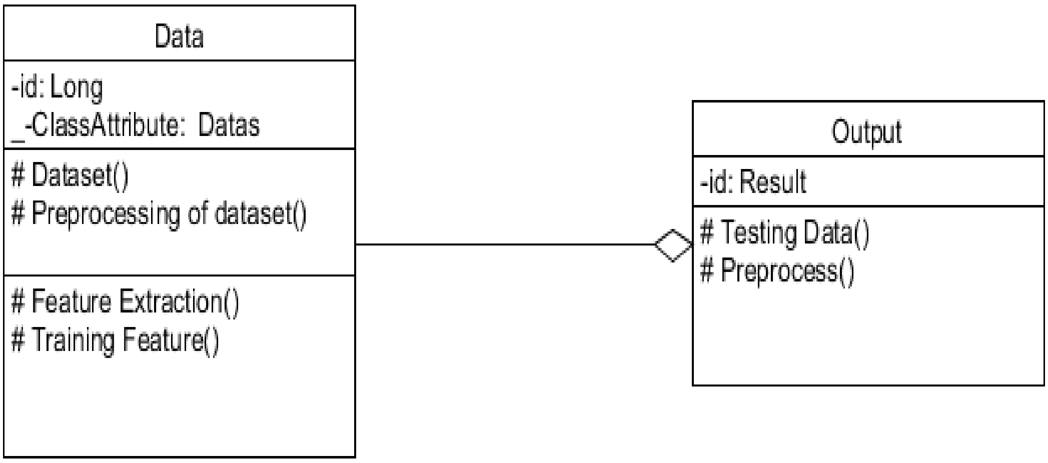
6.1 Use-case Diagram

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.



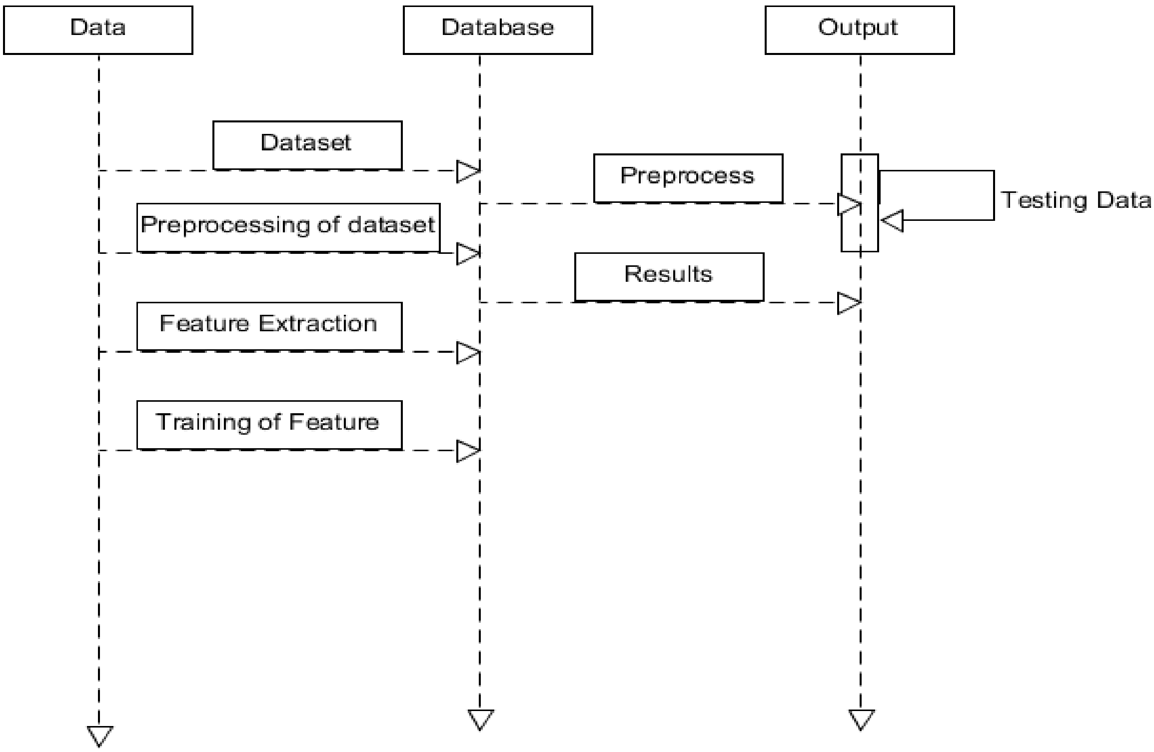
6.2 Class Diagram

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.

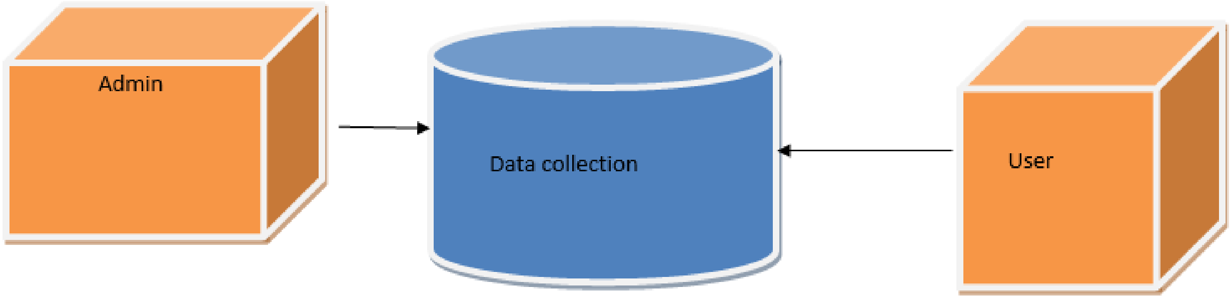


6.3 Sequence Diagram

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.



6.4 Data Flow Diagram



- The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
- The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
- DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
- DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

7. System Design

The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of test. Each test type addresses a specific testing requirement.

TYPES OF TESTS

Unit testing

Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application .it is done after the completion of an individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results.

Integration testing

Integration tests are designed to test integrated software components to determine if they actually run as one program. Testing is event driven and is more concerned with the basic outcome of screens or fields. Integration tests demonstrate that although the components were individually satisfaction, as shown by successfully unit testing, the combination of components is correct and consistent. Integration testing is specifically aimed at exposing the problems that arise from the combination of components.

Functional test

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centered on the following items:

Valid Input : identified classes of valid input must be accepted.
Invalid Input : identified classes of invalid input must be rejected.
Functions : identified functions must be exercised.
Output : identified classes of application outputs must be exercised.
Systems/Procedures: interfacing systems or procedures must be invoked.

Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

System Test

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration points.

White Box Testing

White Box Testing is a testing in which in which the software tester has knowledge of the inner workings, structure and language of the software, or at least its purpose. It is purpose. It is used to test areas that cannot be reached from a black box level.

Black Box Testing

Black Box Testing is testing the software without any knowledge of the inner workings, structure or language of the module being tested. Black box tests, as most other kinds of tests, must be written from a definitive source document, such as specification or requirements document, such as specification or requirements document. It is a testing in which the software under test is treated, as a black box .you cannot "see" into it. The test provides inputs and responds to outputs without considering how the software works.

Unit Testing:

Unit testing is usually conducted as part of a combined code and unit test phase of the software lifecycle, although it is not uncommon for coding and unit testing to be conducted as two distinct phases.

Test strategy and approach

Field testing will be performed manually and functional tests will be written in detail.

Test objectives

- All field entries must work properly.
- Pages must be activated from the identified link.

- The entry screen, messages and responses must not be delayed.

Features to be tested

- Verify that the entries are of the correct format
- No duplicate entries should be allowed
- All links should take the user to the correct page.

Integration Testing

Software integration testing is the incremental integration testing of two or more integrated software components on a single platform to produce failures caused by interface defects.

The task of the integration test is to check that components or software applications, e.g. components in a software system or – one step up – software applications at the company level – interact without error.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

Acceptance Testing

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

Test Results: All the test cases mentioned above passed successfully. No defects encountered.

8. Conclusion

In conclusion, estimating the time to progression of Chronic Obstructive Pulmonary Disease (COPD) is a complex task that depends on various factors, and providing a precise time frame is challenging. COPD is a progressive disease, but the rate of progression can vary significantly among individuals due to factors such as disease severity, age, comorbidities, treatment interventions, and lifestyle factors. It requires careful assessment by qualified healthcare professionals, taking into consideration multiple factors, to estimate the time to progression of COPD for an individual patient. It’s important to note that COPD is a chronic condition that requires ongoing medical management and monitoring. Regular follow-ups with healthcare providers, adherence to treatment plans, lifestyle modifications, and avoiding risk factors such as smoking are essential for managing COPD and potentially slowing down disease progression. Patients with COPD should work closely with their healthcare team to develop an individualized management plan and monitor their lung function regularly to track any changes in disease progression over time. The estimating the time to progression of COPD is a complex and individualized process, and it's important to rely on the expertise of qualified healthcare professionals for accurate assessments and recommendations based on the specific circumstances of each patient

9. References

1. Gawade Prathamesh Pratap, R.P Chauhan, "Detection of lung cancer cells using image processing techniques ", 1st IEEE International Conference on Power Electronics, Intelligent control and energy systems, IEEE, 2016.
2. C. V. Senaratna et al., "Prevalence of obstructive sleep apnea in the general population: A systematic review", *Sleep Med. Rev.*, vol. 34, pp. 70-81, 2017.
3. W. T. McNicholas, D. Hansson, S. Schiza and L. Grote, "Sleep in chronic respiratory disease: COPD and hypoventilation disorders", *Eur. Respir. Rev.*, vol. 28, no. 153, pp. 190064, Sep. 2019.

4. R. F. D’Cruz, P. B. Murphy and G. Kaltsakas, "Sleep disordered breathing and chronic obstructive pulmonary disease: a narrative review on classification pathophysiology and clinical outcomes", *J. Thorac. Dis.*, vol. 12, no. S2, pp. S202-S216, Oct. 2020.
5. B. He et al., "Apnoea-hypopnoea-index comparing the 2007 and 2012 American Academy of Sleep Medicine criteria in chronic obstructive pulmonary disease/obstructive sleep apnoea overlap syndrome", *J. Thorac. Dis.*, vol. 12, no. S2, pp. S112-S119, Oct. 2020.
6. S. Vora and C. Shah, "COPD Classification using Machine Learning Algorithms", 2019.
7. S. Aich, K.W. Choi, P.M. Pradhan, J. Park and H.C. Kim, "A Performance Comparison Based on Machine Learning Approaches to Distinguish Parkinson’s Disease from Alzheimer Disease Using Spatiotemporal Gait signals", *Advanced Science Letters*, vol. 24, no. 3, pp. 2058-2062, 2018.
8. D. Spathis and P. Vlamos, "Diagnosing asthma and chronic obstructive pulmonary disease with machine learning", *Health informatics journal*, vol. 25, no. 3, pp. 811-827, 2019.
9. J. Xia, L. Sun, S. Xu, Q. Xiang, J. Zhao, W. Xiong, et al., "A Model Using Support Vector Machines Recursive Feature Elimination (SVM-RFE) Algorithm to Classify Whether COPD Patients Have Been Continuously Managed According to GOLD Guidelines", *International Journal of Chronic Obstructive Pulmonary Disease*, vol. 15, pp. 2779, 2020.
10. S. Aich, H.C. Kim, K.L. Hui, A.A. Al-Absi and M. Sain, "A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson’s Disease", *2019 21st International Conference on Advanced Communication Technology (ICACT)*, pp. 1116-1121, 2019, February.
11. Wang Xiao-le, ZHU Jie, LI Ze-geng, Lin Xian-gang and LIU Chang-an, "Time rhythm and timing treatment of pulmonary diseases [J]", *Journal of Traditional Chinese Medicine*, vol. 59, no. 08, pp. 636-639, 2018.
12. SAH Rajiv Kumar et al., "Targeted Disruption of Mouse Dip2B Leads to Abnormal Lung Development and Prenatal Lethality", *International journal of molecular sciences*, vol. 21.21, pp. 8223, 2020.
13. M. Moll, D. Qiao, E. A. Regan et al., "Machine learning and prediction of All-cause mortality in COPD", *Chest*, vol. 158.3, pp. 952-964, 2020.
14. C. Wang et al., "Comparison of machine learning algorithms for the identificatin of acute exacerbations in chronic obsturctive pulmonary disease", *Computer Methods and Programs in Biomedicine*, vol. 188, pp. 105267, 2020.
15. Tabreer T. Hasan,Manal H. Jasim, Ivan A. Hashim, ”Heart disease diagnosis systembased on multilayer perceptron neural network and support vector machine”,International Journal of Current Engineering and Technology, vol. 7, oct 2017.