Accurate classification of Related Lung conditions based on Spirometry Data

Submitted for Mini Project (MC470305) of 3rd Semester

MCA(AI & IOT)– CSE

Submitted By:
Prashant Kumar Mishra(2447021)

Under the Supervision of
Dr. Shyam Singh Rajput
CSE Department

Department of Computer Science & Engineering

NATIONAL INSTITUTE OF TECHNOLOGY PATNA

University Campus, Bihar – 800005

Jan 2023 – July 2023

INTRODUCTION

This project uses ML algorithms for precise classification of lung conditions in patients based on metrics extracted from spirogram time-series Pulmonary Function Test.

Motivation:

- COPD and lung diseases are a major life-threatening concern, globally as well as in India, and it is often linked to error-prone and inaccurate, creating the need for accurate diagnostic models.

Methods:

- LogisticRegression
- NaiveBayes
- KNN
- SVM
- DecisionTree
- NN

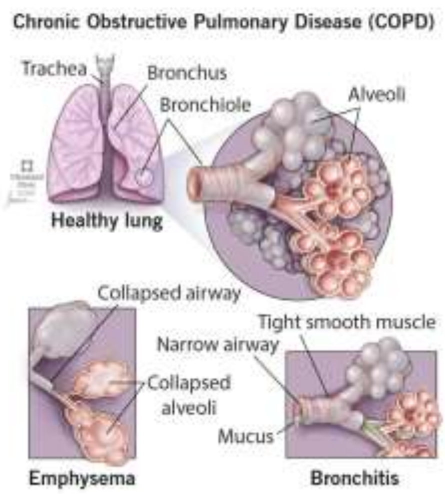


Fig 1: Blockage in Lung tissues

Objectives:

- Enhanced accuracy for classification of based on obstructive or nonobstructive types.

INPUT

The dataset contained spirometry investigation reports of 1314 patients from Institute of Pulmocare and Research (IPCR), Kolkata diagnosed with obstructive and non-obstructive diseases. The patients were divided in 2 groups - Group A and Group B consisting of 1163 and 151 patients respectively. The reports of the patients diagnosed with obstructive diseases were labelled as positive and those with non-obstructive diseases were labelled as negative.

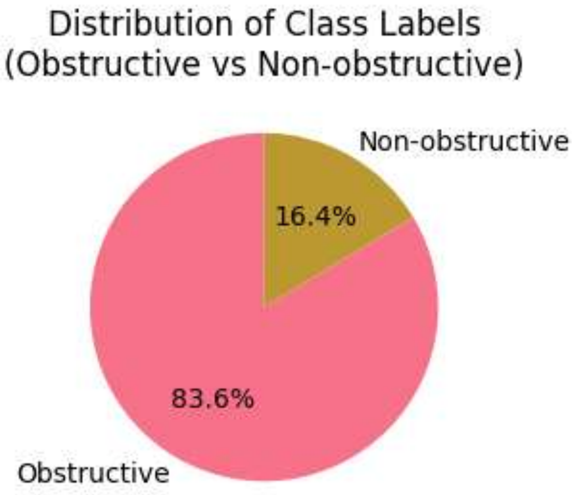


Fig 2 : Distribution of Class labels

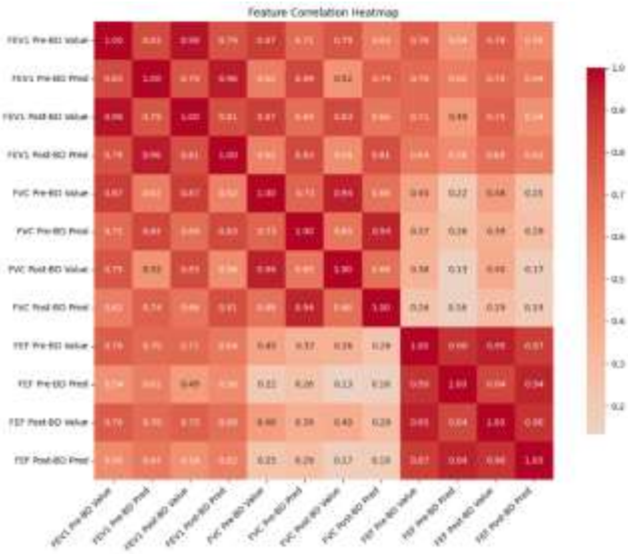


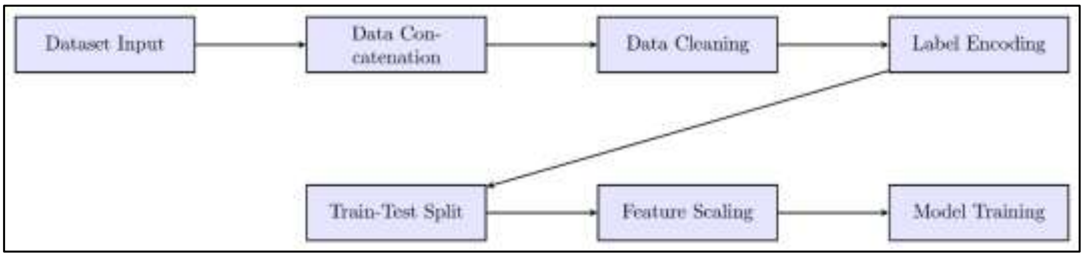
Fig 3 : Feature Corelation Heatmap

In spirometry, patients are asked to take a maximal inspiration and, then, expel the air forcefully as quickly as possible into a mouthpiece. The test is repeated following the administration of a bronchodilator. The pre and post bronchodilator values of the following three metrics were used as input:

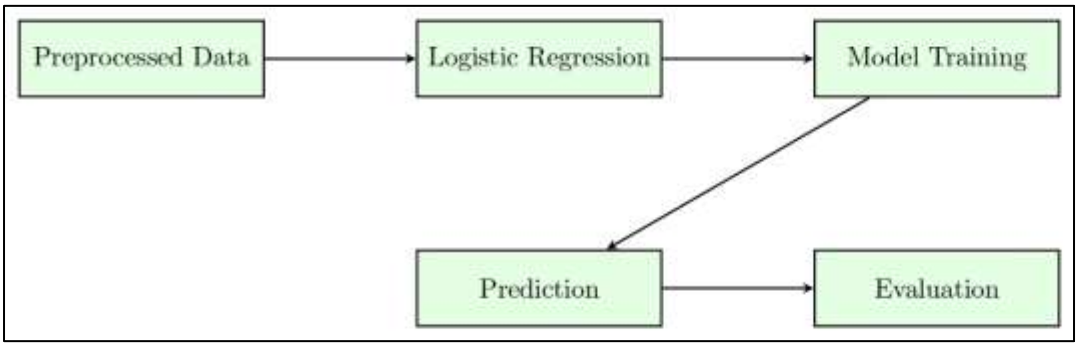
- **Forced Vital Capacity** (FVC): It is the volume of air exhaled forcefully and quickly after full inhalation.
- **Forced Expiratory Volume in one second** (FEV1): It is the volume of air expired during the first second of performing the FVC test.
- **Forced Expiratory Flow (FEF25-75):** It is the flow (or speed) of air coming out of the lung during the middle portion of a forced expiration. It is measured by taking the mean of the flow during the interval 25-75% of FVC.

Model Architecture

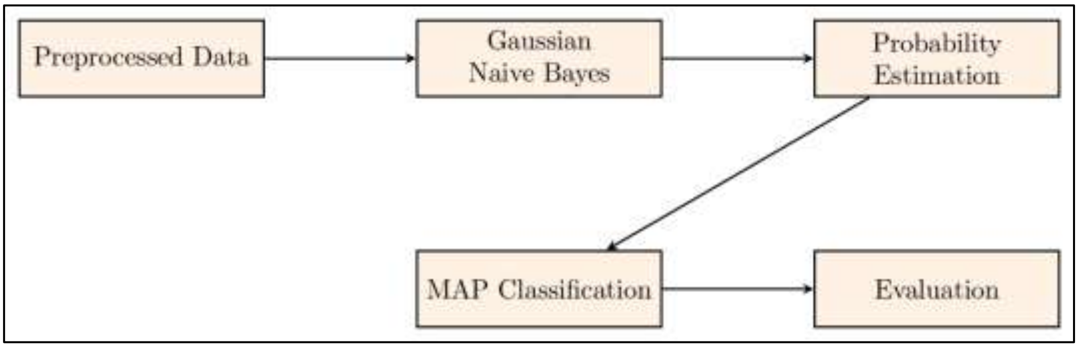
- Data Preprocessing



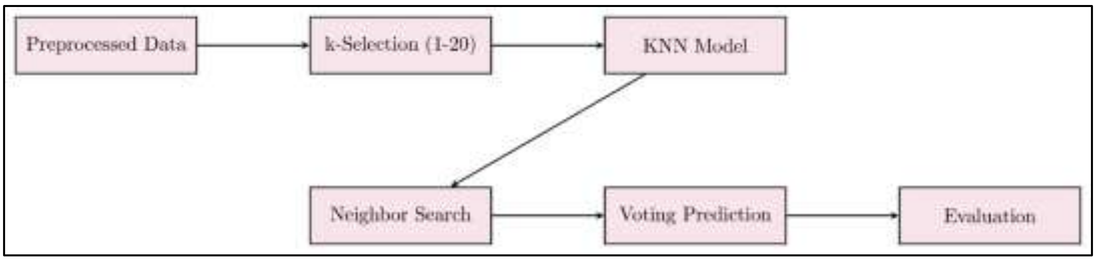
- Logistic Regression



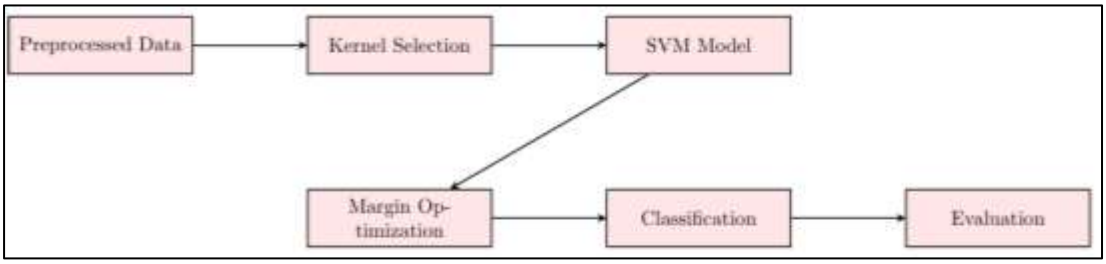
- Naive Bayes



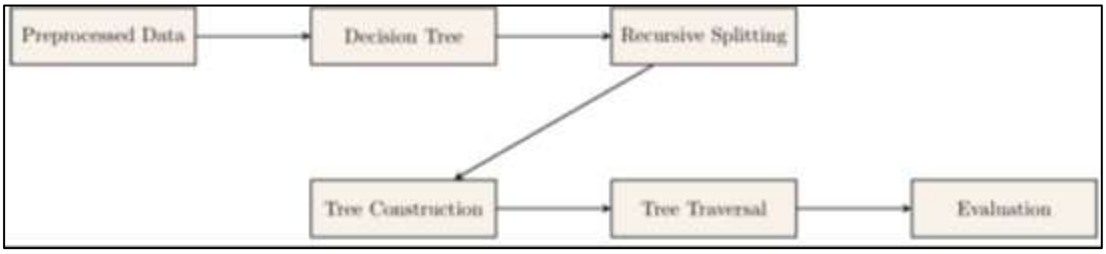
- KNN



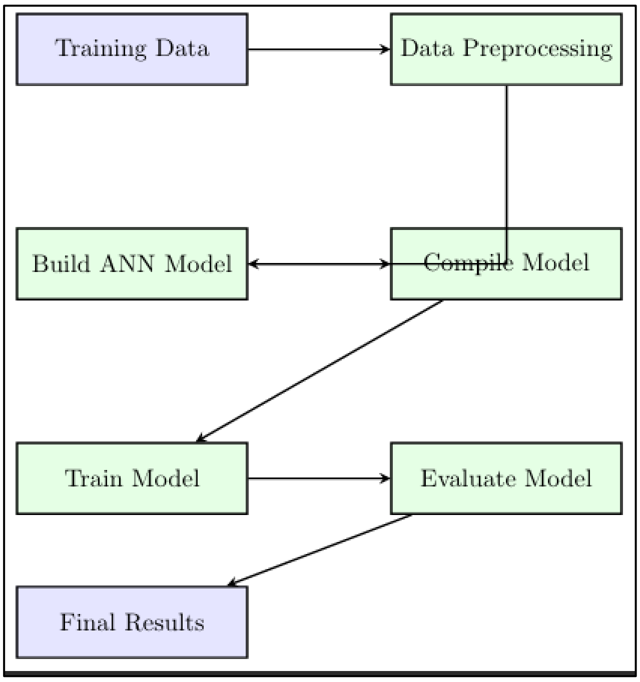
- Suppourt Vector Machine



- Decision Tree

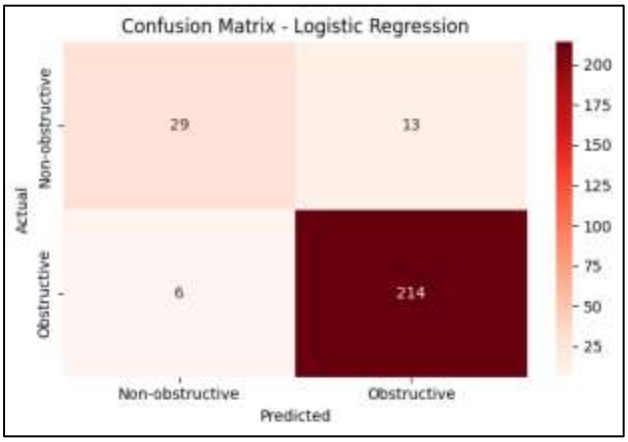


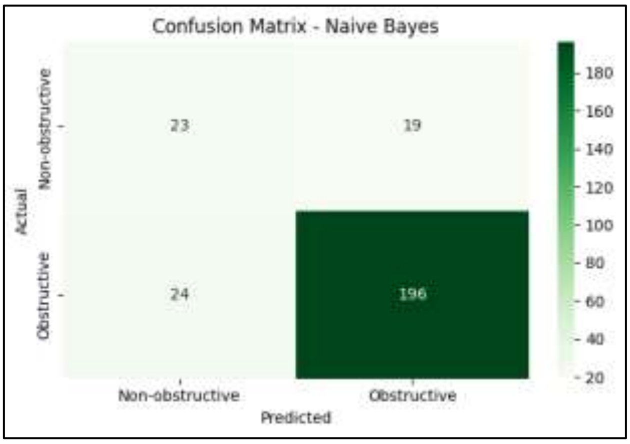
- Neural Network

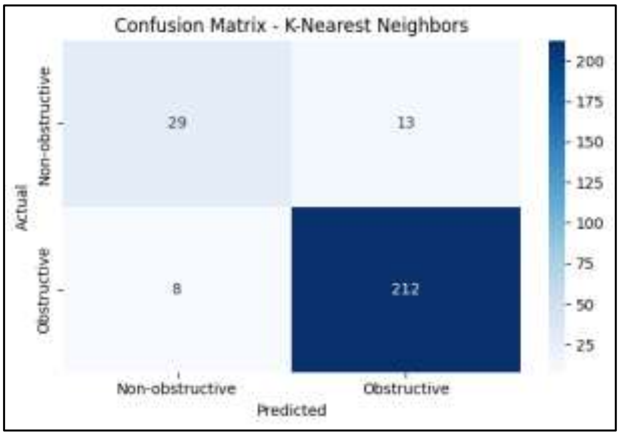


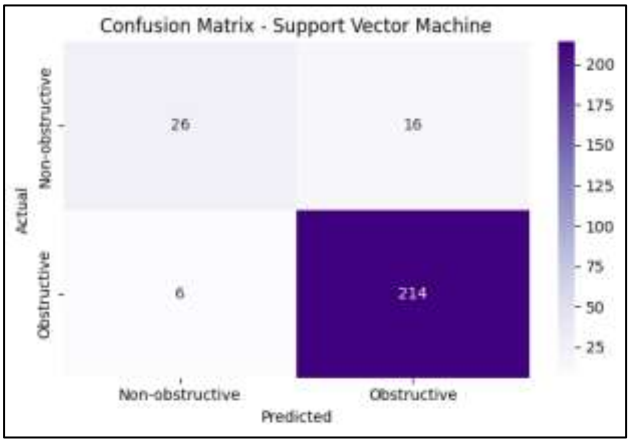
CONFUSION MATRIX

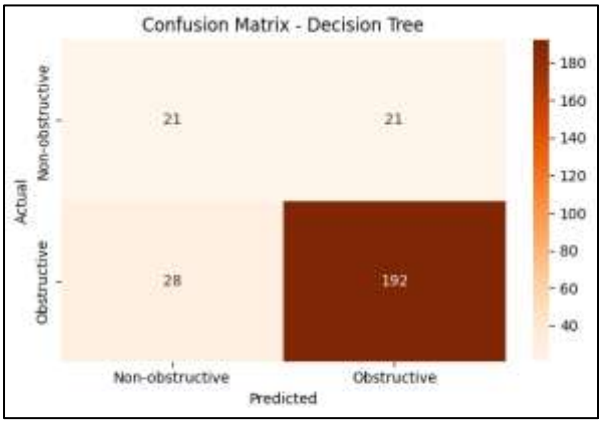
Confusion matrix is a table used to evaluate the performance of a machine learning model for classification problems. It compares the actual labels of a dataset with the predicted labels generated by the model.

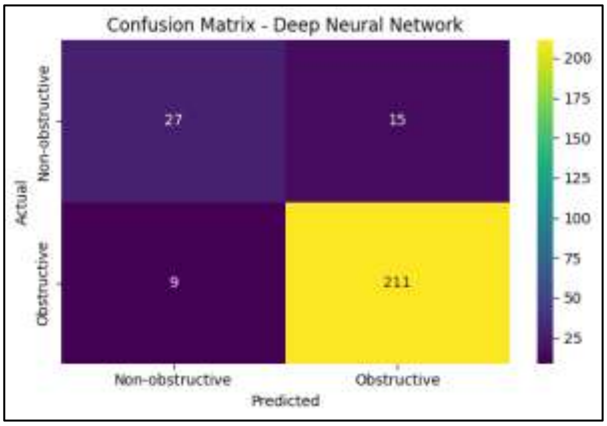










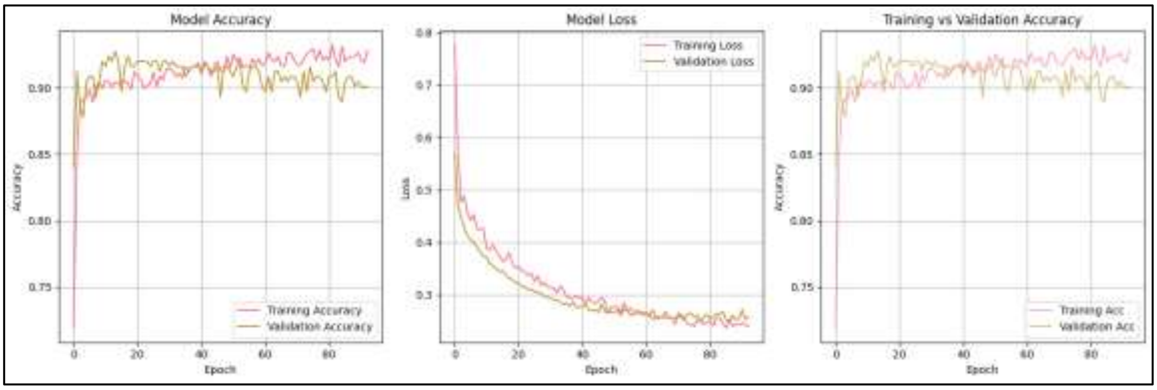


Results Comparision

- Accuracy Score Comparision

Logistic Regression Accuracy: 0.9275
Naive Bayes Accuracy: 0.8359
K-Nearest Neighbors Accuracy: 0.9198
Support Vector Machine Accuracy: 0.9160
Decision Tree Accuracy: 0.8130

- Neural Network Training And Validation Loss



REFERENCES

1. Fig 1 . URL: https://kongrewebcast.com/article/copd-causes-symptoms-diagnosis-treatment-prevention.
2. Datasets - PulmoPred
3. Machine Learning. Tom Mitchell, McGraw-Hill.
4. Machine learning: an algorithmic perspective. Marsland, Stephen. Chapman and Hall/CRC, 2011.
5. Introduction to artificial neural systems. Zurada, Jacek M. Vol. 8. St. Paul: West publishing company, 1992.
6. A Tutorial on Support Vector Machines for Pattern Recognition. Christopher Burges, Data Mining and Knowledge Discovery, 1998.
7. Bhattacharjee, S., Saha, B., Bhattacharyya, P., & Saha, S. (2022). Classification of obstructive and non-obstructive pulmonary diseases on the basis of spirometry using machine learning techniques. *Journal of Computational Science*, 63, 101768. https://doi.org/10.1016/j.jocs.2022.101768