Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023

PREDICTIVE MODEL FOR DETECTING COPD

Poorna Rajesh Pimpale, Suchetadevi Gaikwad, Dhanashree Salvi, Renuka Nagpure Assistant Professor Information Technology Engineering, Atharva College of Engineering Mumbai, India

Abstract : Chronic Obstructive Pulmonary Disease is a disease of lungs that involves gradual decrease of normal functioning of lungs. This paper aims at specifying the work done to predict chronic obstructive pulmonary disease at an early stage. As lung disease has become a very prevalent problem all over the world where the patients don't even come to know that they have severe lung disease at an early stage as there are no proper disease specific symptoms in an early stage it is quite difficult to predict the disease. If the problem is not detected in time it may prove to be fatal as there might be inflammation in lungs due to which the
patient might suffer severe breathlessness. This paper aims to specify the methods and techniques used previously to determine chronic obstructive pulmonary disease in its early stage so that the patient's quality of life can be improved with the proper medication, healthy diet and exercise. This paper intends to study technological advancement that includes machine learning algorithms to analyze the disease. It was observed that many machine learning algorithms are being used such as Convolutional neural network, K-nearest neighbor, random forest, regression, support vector machine and many more. The experts have used deep learning based ResNet with convolutional neural networks to classify the infected features in High Resolution Computed Tomography scans of lungs.

***IndexTerms*:** COPD, Exacerbation, Artificial Intelligence, Proposed System

I. INTRODUCTION

In the human body we have many vital organs out of which Lungs are too important as it is the vital part of the respiratory system. The most important function of lungs is to absorb oxygen from the outer world or environment and transfer it into the blood stream so as to transfer the oxygenated blood throughout the body and transfer the deoxygenated blood out of the body. Chronic Obstructive Pulmonary Disease is the disease where the lungs gradually decrease their regular functioning. If a person is suffering from COPD, It becomes difficult for him to breathe. In COPD due to inflammation of lungs the airflow is obstructed leading to breathlessness that might prove to be quite fatal. Chronic Obstructive Pulmonary Disease includes Asthma, Bronchitis, Pneumonia, Emphysema[5]. COPD is a progressive disease which worsens over a period of time. COPD can be treated if detected in early stages. There are four stages of COPD. If the disease is detected in early stages then it becomes curable, else at the last stage or the End Stage of COPD the patient might have frequent exacerbations which could be fatal. At the last stage the patient might have a condition called chronic respiratory failure, where the patient is unable to breathe due to various reasons such as inflammation of lungs, deposition of mucus in air sacs and many more[15]. If the condition is not detected and diagnosis is not done on time it would be difficult to cure the patient.
Thus it becomes very important to predict the occurrence of COPD at an early stage so that the patient may preserve the functioning of the lung and quality of life is also improved.
ARTIFICIAL INTELLIGENCE
It is a branch of computer science which has become an essential part in the technical industry which believes in creating the intelligence in machines. Artificial Intelligence is the branch of science which is highly technical and specialized for core
problems of intelligence such as Knowledge base, Reasoning, Problem Solving Capability, Perception i.e the way system perceives the problem, Learning from the incidences, ability to change or manipulate and move objects.

MOTIVATION

- Symptoms of Chronic obstructive pulmonary Disease are not detected in early stage
- At stage IV of Chronic Obstructive Pulmonary Disease it is difficult to cure the patient
- Spirometry is the test used to detect COPD which is too expensive.

OBJECTIVE

- Explore the state of art existing works on early detection of COPD can compare them in terms of accuracy, sensitivity, specificity and detection time
- To work on the features for COPD detection and do statistical correlation analysis to study the relation between features and COPD detection time over real data of patients[3].
- To find an appropriate predictive model to predict COPD based on multiple clinical parameters[4]. ● Test the effectiveness of proposed model for COPD detection among patients across different categories of age, sex , demographics etc
- To use data acquired from the fit bit used by patients and environment data from the various sensors put up in a particular area for early detection of COPD.

II. LITERATURE REVIEW

1.Title:Diagnosing asthma and chronic obstructive pulmonary disease with machine learning[15]
Published in year: Health Informatics Journal, SAGE, 2017

In this paper, author has considered the samples of about n=132 that identify the major factors that studies empirical pulmonology that contributes to diagnosis of the disease. It was observed that out of various machine learning algorithm Random Forest classifier proved to give 97.7 percent precision when it comes to detect the chronic obstructive pulmonary disease where the features or the attributes considered in this case were smoking, age of the patient, forced respiratory volume of the patient. It was also observed that the best precision for the asthama was achieved through Random Forest Classifier with the precision of 80.3 percent.

2.Title: Prediction of pulmonary pressure after Glenn shunts by Computed Tomography-based machine learning models Published in and year: Springer 2019[10]
In this paper, the author has trained the acquired dataset of about 100 samples with six machine learning algorithm such as Knearest neighbour, support vector machine, random forest, logical regression, Naïve Bayes and Linear Discriminant Analysis. Here the comparative analysis was done by the author for all the listed algorithms and its average performance was evaluated with respect to area under the curve, accuracy, sensitivity, and specificity.

3.Title:Machine Learning Characterization of COPD Subtypes Insights From the COPD Gene Study Published in and year: CHEST, 2019[9].
In this paper, the author has considered chest computed tomographic images and are trained through machine learning algorithms
to quantify COPD heterogeneity. Trajectory of lung growth is also depicted here.

4.Title: Comparison of Machine Learning Algorithms to Increase Prediction Accuracy of COPD Domain[8].
Published in and year: IEEE transactions, 2020
In this paper, the techniques used decision tree, naive Bayes and Bayesian network) based on ROC metric with 61 attributes were considered of affected patients. Theaccuracy achieved is only 76.8%

5.Title: Detection of Different stages of COPD Patients Using Machine Learning Techniques[3].

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 2, No. 1, February : 2023
Published in and year: IEEE transaction, 2021
In this paper, the author has recommended RFECV method which showed precision of about 90 percent with the data set.

6.Title: A Machine Learning Approach to the Interpretation of Cardiopulmonary Exercise Tests: Development and Validation[2].
Published in and year: Hindawai, April 2021.
In this paper, the author has used the dataset of 197 males and 37 females. The author has developed by using two novel approaches SVM and normalized percent of predicted normal for the interpretation process of CPET

III. EXISTING SYSTEM

Identifying patients who are at risk of acquiring this chronic disease is a challenge as there are many ambiguities when medicine and chronic disease comes into picture. There is a great impact on patients' lives when there is poor control and misdiagnosis occurs. As per World |Economic Forum, the global economic cost of chronic disease could reach up to 47$ trillion till 2030[9]. To overcome the challenges in the traditional healthcare system, Machine Learning algorithms are used to deal with uncertainty, accuracy, causal attributes, and their relationships. Exacerbation of Chronic Obstructive Pulmonary |Disease and its prediction is considered to be the challenge which can be met by using Machine Learning algorithms to predict the chances of occurrence of exacerbations at an early stage[3].
Existing system considers features from datasets acquired from blood reports, X-ray, Ultrasound, Computed Tomographic images. The dataset is trained using different machine learning algorithms like CNN,KNN,Random Forest, Naive Bayes, Regression etc. The results obtained from these algorithms are compared to predict Chronic Obstructive Pulmonary Disease[Fig.1]

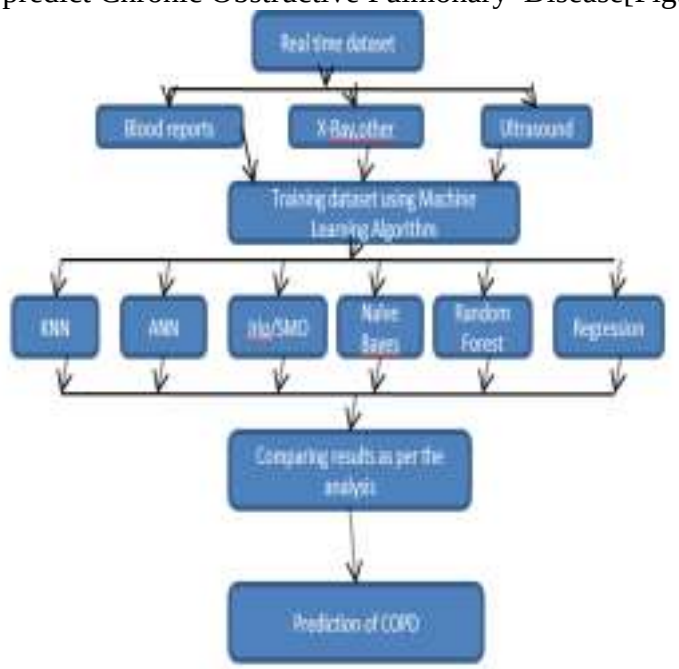

Figure.1. Existing System

IV. PROPOSED SYSTEM

Lot of research is already done in COPD, but

- Research papers seen so far have limited sample size and no real time samples.
- Limited papers on other input like CBC.

- Techniques used so far is only Random Forest, ANN,KNN,XGB,SVM,etc
- Alternative Techniques and different input types yet to be explored.

The proposed system aims to identify the chances of occurrence of chronic obstructive pulmonary disease at the early stages in order to improve the life of the patient by considering the real time datasets and training them against the competitive data analytics algorithm.
Explore efficient algorithms to correctly classify chronic obstructive pulmonary disease from standard/routine test reports to assist a doctor diagnose the disease faster/recommend additional testing.

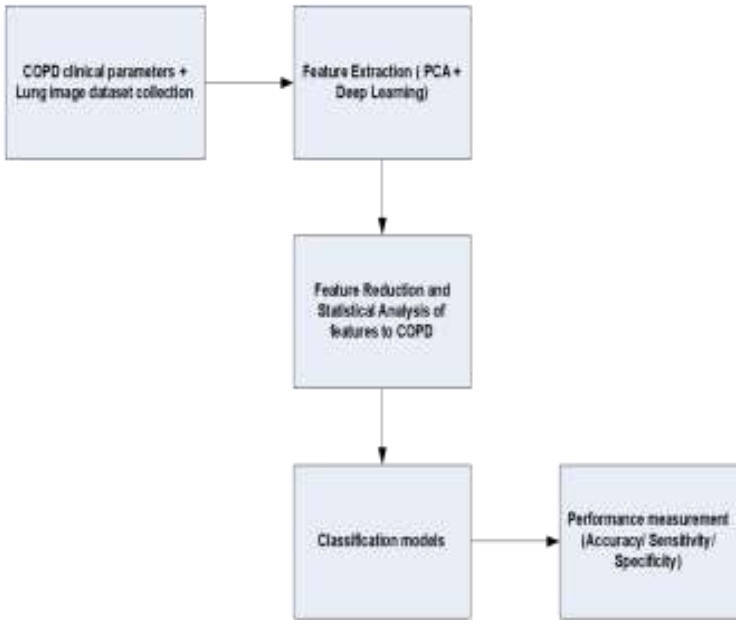

Figure.2. Proposed System

A.COPD Clinical Parameters[Fig.2]
In this module, data collection from hospitals or patients would be done. Data set required for analysis comprises of Physiological characters such as

- Diastolic Blood Pressure, Systolic Blood Pressure
- Age, Weight
- CBC report that comprises of all the blood components,
- SPO2 level
- Temperature

Data acquired from Fit Bit that comprises of

- heart rate,
- SPO2 level
- Temperature
- number steps taken in a day

HRCT(High resolution computed tomography) scans of lungs
X-Ray of Chest
Results of Spirometry Test
Answers to Questionnaire acquired from patients that includes questions related to their mood, temperament, emotional status, any symptoms
B.Feature Extraction (Deep Learning)[Fig.2]
To extract the feature from the data set acquired, ResNet50V2 can be used as it gives overall accuracy of about 98.49% which is observed from previous work.
Feature Pyramid Network can be used so as to classify the layers between the images obtained from HRCT[12] scans. Once obtained, dense neurons can be concatenated to obtain the final result.

C.Feature Reduction And Statistical Analysis Of Features[Fig.2]

After obtaining the required features, they must be reduced to the features that are quite essential to predict the occurrence of Chronic Obstructive Pulmonary Disease. Statistical Analysis of all the features that are reduced would be done here in order to find out what exactly is the result.

D.Classification Models[Fig.2]

There are various classification models such as Convoluted Neural Network, Support vector Machine, KNN, ANN and many more.

It is observed that the Convoluted Neural Network gives more accuracy and precision with respect to other classification models. Thus CNN can be used here as a classification model.

In most of the papers, CNN is used and it shows the maximum amount of accuracy. But the graph convolutional network method also proves to be best on non euclidean structure like graph[1]. The existing model of Graph convolutional network was applied on the dataset which is publicly available which had features like 320-dimensional feature vectors from computed topographical images of 300 patients suffering from COPD. In the proposed system the Graph Convolutional Network model can be used on real data set acquired from laboratories with desired features[1].

E.Performance Measurement[Fig.2]

In this module, after the application of classification and training the data set that is obtained, testing has to be done for the results that are obtained. Performance is measured with respect to Accuracy, Sensitivity and specificity of the results that are obtained. Performance measurement is quite important as here we determine how the modules have proved to be successful in prediction of occurrence of Chronic Obstructive Pulmonary Disease.

F.Feasibility Study

The major concern is to obtain the real time dataset. Existing dataset can be used to determine the performance. Convolutional Neural Network and Graph Convolutional Network models have to be applied on the real time dataset. Various laboratories are approached for the real time dataset and also hospitals that treat COPD patients are contacted to acquire the dataset. Local gyms are contacted to acquire the fitbit data.

IV. CONCLUSION

Thus, the proposed system will implement a system to identify whether a sample corresponds to a patient with chronic Obstructive Pulmonary Disease and the corresponding stage using a competitive data analytics algorithm with an industrially acceptable false positive rate and false negative rate. It would also explore efficient algorithms to correctly classify chronic obstructive pulmonary disease from standard/routine test reports to assist a doctor diagnose the disease faster/recommend additional testing.

References

[1]Early detection of COPD based on graph convolutional network and small and weakly labeled data , International Federation for Medical and Biological Engineering 2022

[2]A Machine Learning Approach to the Interpretation of Cardiopulmonary Exercise Tests: Development and Validation, Hindawai, April 2021.

[3]Detection of Different stages of COPD Patients Using Machine Learning Techniques, IEEE transaction, 2021 [4]Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs, *JAMA Netw Open.* 2020

[5]Deep Learning Enables Automatic Classification of Emphysema Pattern at CT, The Genetic Epidemiology of COPD (COPDGene) Investigator, 2020

[6]Deep neural network analyses of spirometry for structural phenotyping of chronic obstructive pulmonary disease, Sandeep Bodduluri,Purushotham V. Bangalore, Surya P. Bhatt JCI Insight. 2020

[7]Registration Based Lung Mechanical Analysis of Chronic Obstructive Pulmonary Disease

(COPD) Using A Supervised Machine Learning Framework, Springer, 2020
[8] In depth analysis of lung disease prediction using machine learning algorithms, International conference on machine learning, Image processing, network security and data sciences, Springer, 15 June2020.
[9] Comparison of Machine Learning Algorithms to Increase Prediction Accuracy of COPD Domain, IEEE transactions, 2020 [10]Prediction of pulmonary pressure after Glenn shunts by computed tomography–based machine learning models, Springer 2019
[11]An automatic system supporting clinical decision for chronic obstructive pulmonary disease, Springer, august 2019 [12]A Perlin Noise-Based Augmentation Strategy for Deep Learning with Small Data Samples of HRCT Images,Europe PMC, Dec 2018
[13]Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction,2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) April 4-7, 2018
[14]A prediction model for early death in non-small cell lung cancer patients following curative-intent chemo-radiotherapy, ACTA ONCOLOGICA, 2017
[15] Diagnosing asthma and chronic obstructive pulmonary disease with machine learning, Health Informatics Journal, SAGE, 2017