iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.55770

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

Machine Learning Algorithms for COPD Patients Readmission Prediction

P. Abhishek Goud[1], Dr. M. Arathi[2]
[1]*Student, M.Tech (Data Sciences), Department of Information Technology, JNTU Hyderabad, Hyderabad, India*
[2]*Professor of CSE, Department of Information Technology, JNTU Hyderabad, Hyderabad, India*

Abstract: Patients' readmission might be seen as a crucial aspect in lowering costs while maintaining high-quality patient care. As a result, anticipating and reducing readmission rates for patients will considerably enhance healthcare delivery. The goal of this research is to use machine learning algorithms to predict readmission of COPD (Chronic Obstructive Pulmonary Disease) patients. The major metrics for measuring models' prediction capability in each time frame were Area under Curve (AUC) and Accuracy (ACC). Then, the factors' relevance for each result was clearly recognized, and specified key variables were discriminated. With%91 ACC, our research had the best accuracy in predicting readmission.
Keywords: Classification algorithms, COPD readmission, data mining, decision support systems, healthcare data analytics.

I. INTRODUCTION

Predictive analytics is a popular IT (Information Technology) tool in healthcare. For example, [1], [2] highlighted the prospect of employing electronic health records as a foundation for healthcare analysis for smart health. [3] Has also used predictive analytics and data mining to reduce the spread of chronic illnesses. In general, healthcare-related research in the last decade has concentrated on building and implementing IT models to satisfy the distinct and vital demands of healthcare systems [4]. The majority of these studies concentrate on analyzing large volumes of data in order to get significant knowledge and insights about the present and future behavior of the system under investigation. Chronic Obstructive Pulmonary Illness (COPD) is a lung disease characterized by restricted airflow [5]. COPD has been identified as one of the key reasons of increased mortality rates throughout the world. In 2016, the Global Burden of Disease Study anticipated 251 million new cases of COPD. COPD was also claimed to have caused 3.17 million fatalities in 2015 (i.e., 5% of all deaths that year) [6]. Between 1991 and 2000, the admission rate of COPD patients in the United Kingdom increased, and by 2000, it accounted for 1% of all hospital admissions [7]. The entire expenses of lung disorders in the EU (European Union) have been estimated to be about 6% of total healthcare expenditures, with COPD accounting for the largest proportion (%56) of these costs [5].

II. LITERATURE SURVEY

A. TaGiTeD: Predictive Task Guided Tensor Decomposition For Representation Learning From Electronic Health Records

With the increased availability of healthcare data, such as Electronic Health Records (EHR), an increasing number of data analytics approaches are being created with the goal of extracting insights from them in order to enhance the quality of care delivery. EHR analysis presents several problems, including as high dimensionality and event sparsity. Furthermore, unlike in other application fields, EHR analysis algorithms must be highly interpretable in order to be therapeutically relevant. As a result, representation learning from EHRs is critical. In this research, we present a method for analyzing EHRs termed Predictive Task Guided Tensor Decomposition (TaGiTeD). TaGiTeD, in particular, learns highly predictive event interaction patterns from EHRs via supervised tensor decomposition. TaGiTeD can develop good EHR representations in a more targeted manner than unsupervised approaches. This is critical since most medical issues have a small number of patient samples, which are insufficient for unsupervised algorithms to develop meaningful representations.

B. Deep Learning And Alternative Learning Strategies For Retrospective Real-World Clinical Data

In recent years, there has been a surge of interest in the healthcare research community for artificial intelligence to help with big data analytics and decision making. One of the primary reasons for this is the massive influence of deep learning on the use of complicated healthcare big data. Although deep learning is a strong analytic tool for the complex data included in electronic health records (EHRs), it has drawbacks that may make it unsuitable for specific healthcare applications. In this article, we provide a short summary of the limits of deep learning, as indicated through case studies completed over the years, with the goal of encouraging the consideration of alternate analytic methodologies for healthcare.

C. Predictive Analytics To Prevent And Control Chronic Diseases

Chronic illnesses are the leading cause of mortality in the world today. As a result, there is a significant increase in attention being devoted to individual wellness as a preventive tactic in healthcare. Creating and establishing a prediction model for chronic illnesses, on the other hand, is a remarkable shift in healthcare technology based on data analysis and decision-making. In this research, efficient techniques for chronic illness prediction were developed by mining data comprising previous health information. For the diagnosis of diabetes and heart disease, we employed Naive Bayes, Decision Tree, Support Vector Machine (SVM), and Artificial Neural Network (ANN) classifiers. We also give a comparison analysis of several classifiers to gauge performance based on accuracy rate in this research.

D. A Discrete Event Simulation Model For Waiting Time Management In An Emergency Department: A Case Study In An Egyptian Hospital

This work proposes a discrete event simulation model to aid in the improvement of emergency department services at a private hospital in Zigzag, Egypt. We build a patient flow division model by categorizing patients based on their severity degree. Although there is strong evidence that patient division and routing improve health care in terms of waiting times and length of stay (LoS), there is a lack of formal system assessment and implementation under this configuration. Based on system observations and interviews with health care providers, a thorough and clear image of the system was created, coupled with a conceptual model depicting possible patient flows across the investigated system. Using the acquired data, a discrete event simulation model of the Emergency Department is created. To explore the influence of patient flow division, multiple operational scenarios were evaluated against the baseline scenario, including varying staff capabilities and patient magnitudes. The results show that the suggested system design may greatly reduce waiting times and duration of patient stay.

E. Global Strategy For The Diagnosis, Management, And Prevention Of Chronic Obstructive Lung Disease: The Gold Science Committee Report 2019

Precision medicine is a patient-centered strategy that combines all relevant clinical, genetic, and biological data in order to optimize treatment benefit vs the likelihood of adverse effects for each person. In patients with chronic obstructive pulmonary disease (COPD), higher blood eosinophil levels are related with better efficiency of inhaled corticosteroids (ICSs). Blood eosinophil counts are a biomarker that has the potential to be utilized in clinical practice to assist target ICS medication with more accuracy in COPD patients who have had exacerbations despite adequate bronchodilator treatment. Based on the ABCD evaluation, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) 2017 pharmacological treatment algorithms may be administered very readily to treatment-naive people at first presentation. However, its usage during follow-up in patients who are already on maintenance medication is more challenging. A novel method is required to guide COPD pharmacological therapy throughout follow-up. Recent big randomized controlled studies have revealed critical new information on the therapeutic benefits of ICSs and long-acting bronchodilators on exacerbations. Changes in the GOLD pharmacological therapy guidelines were prompted by new findings addressing blood eosinophils and inhalation therapies, as well as the necessity to differentiate between initial and follow-up pharmacological care. The data and reasons for the GOLD 2019 pharmacological therapy guidelines are explained in this article.

III. PROPOSED METHODOLOGY

Predictive analytics is a common IT (Information Technology) tool in the healthcare business. For example, and have argued that electronic health records might be used as a basis for healthcare analysis for smart health. Data mining and predictive analytics have also been utilized by researchers to halt the spread of chronic illnesses. The bulk of recent healthcare-related research has focused on developing and implementing IT models to meet the specific and critical demands of healthcare systems. The bulk of these studies focus on analyzing huge amounts of data to acquire insight into the current and future behavior of the system under study.

A. Disadvantages

1) There is no effectiveness.

2) The need for trustworthy forecasting models

In this study, we are using the most powerful data mining techniques, such as Decision Trees (DT), Artificial Neural Networks (ANN), and Support Vector Machine (SVM), to create classification models that can identify the high-risk COPD patients who are most likely to be readmitted to the hospital within 30 days of discharge. Many researches have considered unplanned readmission.

However, there is still a clear lack of evidence of their usefulness. One of the reported explanations of the studies' inefficiency is that they spent time looking for the incorrect group of patients (i.e., patients with low risk of readmission). As a result, there is a great need for dependable prediction models capable of precisely identifying high-risk patients in the most efficient manner, enabling healthcare stakeholders to react appropriately.

B. Advantages

1) Identifying high-risk patients correctly
2) Most effectively, enabling healthcare stakeholders to react appropriately

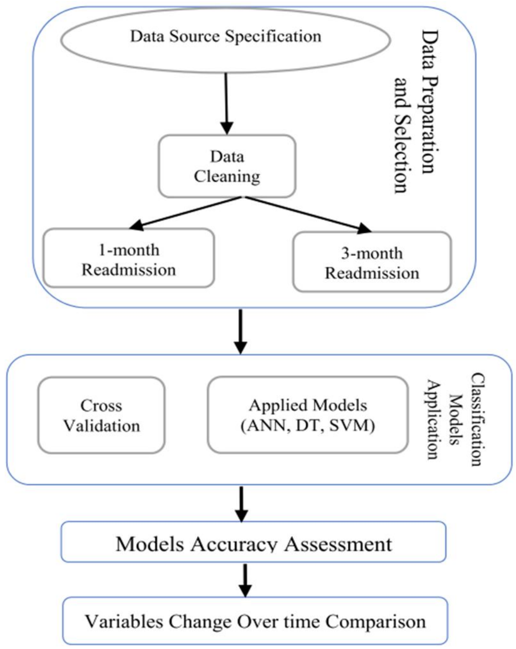


Fig.1: Architecture of the proposed system

IV. MODULES OF WORK

To carry out the aforementioned project, we created the modules listed below.

1) Data Exploration: we will put data into the system using this module.
2) Processing: we will read data for processing using this module.
3) Using this module, data will be separated into train and test groups.
4) Model generation: Support Vector Machine, Random Forest, Decision Tree, ANN, MLP, XGBoost, Voting Classifier.
5) User registration and login: Using this module will result in registration and login.
6) User input: Using this module will provide input for prediction
7) Prediction: the final projected value will be presented

V. IMPLEMENTATION

1) *Random Forest:* A Supervised Machine Learning Algorithm that is commonly utilized in Classification and Regression applications. It constructs decision trees from several samples and uses their majority vote for classification and average for regression.
2) *SVM:* Support Vector Machine (SVM) is a common Supervised Learning method that is used for Classification and Regression tasks. However, it is mostly utilized in Machine Learning for Classification difficulties. The SVM algorithm's purpose is to find the optimum line or decision boundary for categorizing n-dimensional space so that we may simply place fresh data points in the proper category in the future.
3) *Voting Classifier:* A machine-learning method that Kagglers often utilize to improve the performance of their model and climb the rank ladder. Voting Classifier may also be used to increase performance on real-world datasets, although it has significant limitations.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

4) *Decision Tree:* A decision tree is a non-parametric supervised learning technique that may be used for classification and regression applications. It has a tree structure that is hierarchical and consists of a root node, branches, internal nodes, and leaf nodes.
5) *ANN:* Artificial Neural Network (ANN) is a kind of neural network that harnesses brain processing to create algorithms that may be used to model complicated patterns and forecast issues.
6) *MLP:* Multilayer Perceptrons (MLPs) are the most common form of neural network. They are made up of one or more neuron layers. The input layer receives data, one or more hidden layers provide degrees of abstraction, and predictions are generated on the output layer, also known as the visible layer.

XGBoost (extreme Gradient Boosting): The XGBoost (extreme Gradient Boosting) technique is a popular and efficient open-source version of the gradient boosted trees algorithm. Gradient boosting is a supervised learning approach that combines an ensemble of estimates from a collection of simpler and weaker models to try to correctly predict a target variable.

VI. EXPERIMENTAL RESULTS

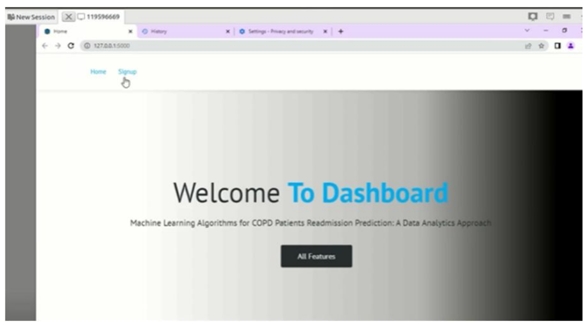


Fig.2: Home screen

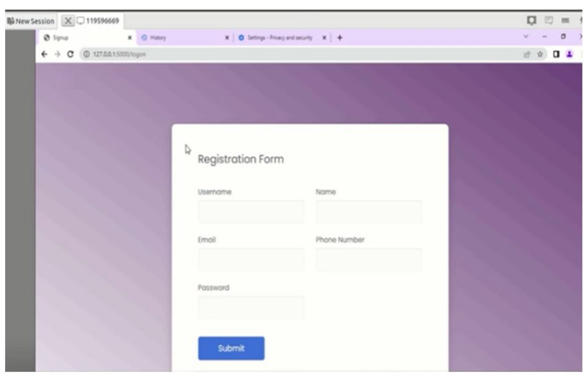


Fig.3: User registration

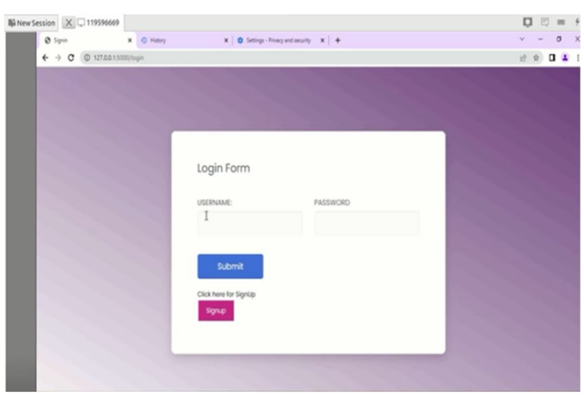


Fig.4: user login

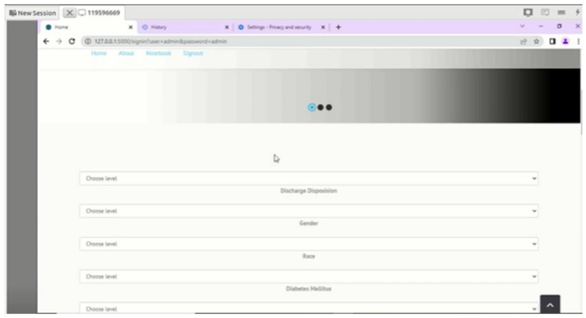


Fig.5: Main screen

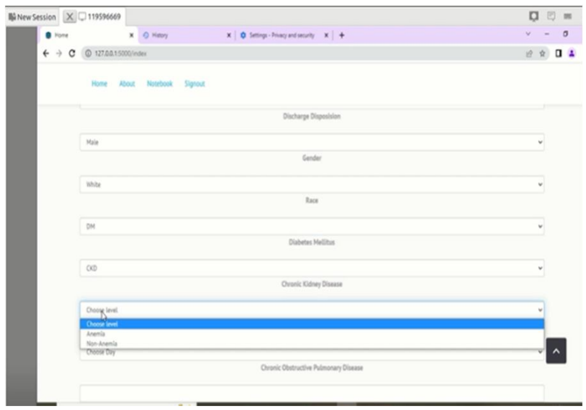


Fig.6: User input

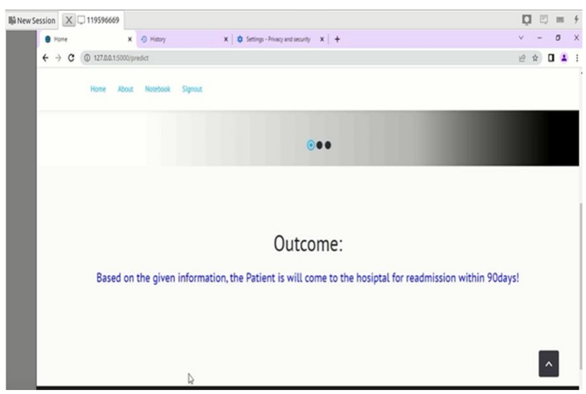


Fig.7: Prediction result

VII. CONCLUSION

The main contribution can be summarized as the use of machine learning algorithms and techniques to handle the class imbalance problem utilizing medical vector scattering to cope with the limited conventional readmission predictive models and hence improving predictability. We compare the different machine learning algorithms according to their predictability power of hospital readmission prediction. Nevertheless, our study still has some limitations due to the self-funding and limited budget. It was also very challenging to have a team of experts dedicated to our study needs (data collection, cleaning, and preparation). Another important point that limited our study was the limited number of COPD patients' records (195 records). Our future research direction is to study and investigate alternative classification techniques to further amend the classification model. We also plan to study more detailed predictions on hospital readmission, which has a higher effect in designing and building more efficient and effective post discharge models. For example, the probability of patient readmission within a specific time point is interesting to predict and study how this probability may be affected by earlier hospitalization events.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IX Sep 2023- Available at www.ijraset.com

REFERENCES

[1] K. Yang, X. Li, H. Liu, J. Mei, G. Xie, J. Zhao, B. Xie, and F. Wang, ''TaGiTeD: Predictive task guided tensor decomposition for representation learning from electronic health records,'' in Proc. 31st AAAI Conf. Artif. Intell. (AAAI), Feb. 2017, pp. 2824–2830.

[2] D. Chen, S. Liu, P. Kingsbury, S. Sohn, C. B. Storlie, E. B. Habermann, J. M. Naessens, D. W. Larson, and H. Liu, ''Deep learning and alternative learning strategies for retrospective real-world clinical data,'' NPJ Digit. Med., vol. 2, no. 1, pp. 1–5, Dec. 2019, doi: 10.1038/s41746-019-0122-0.

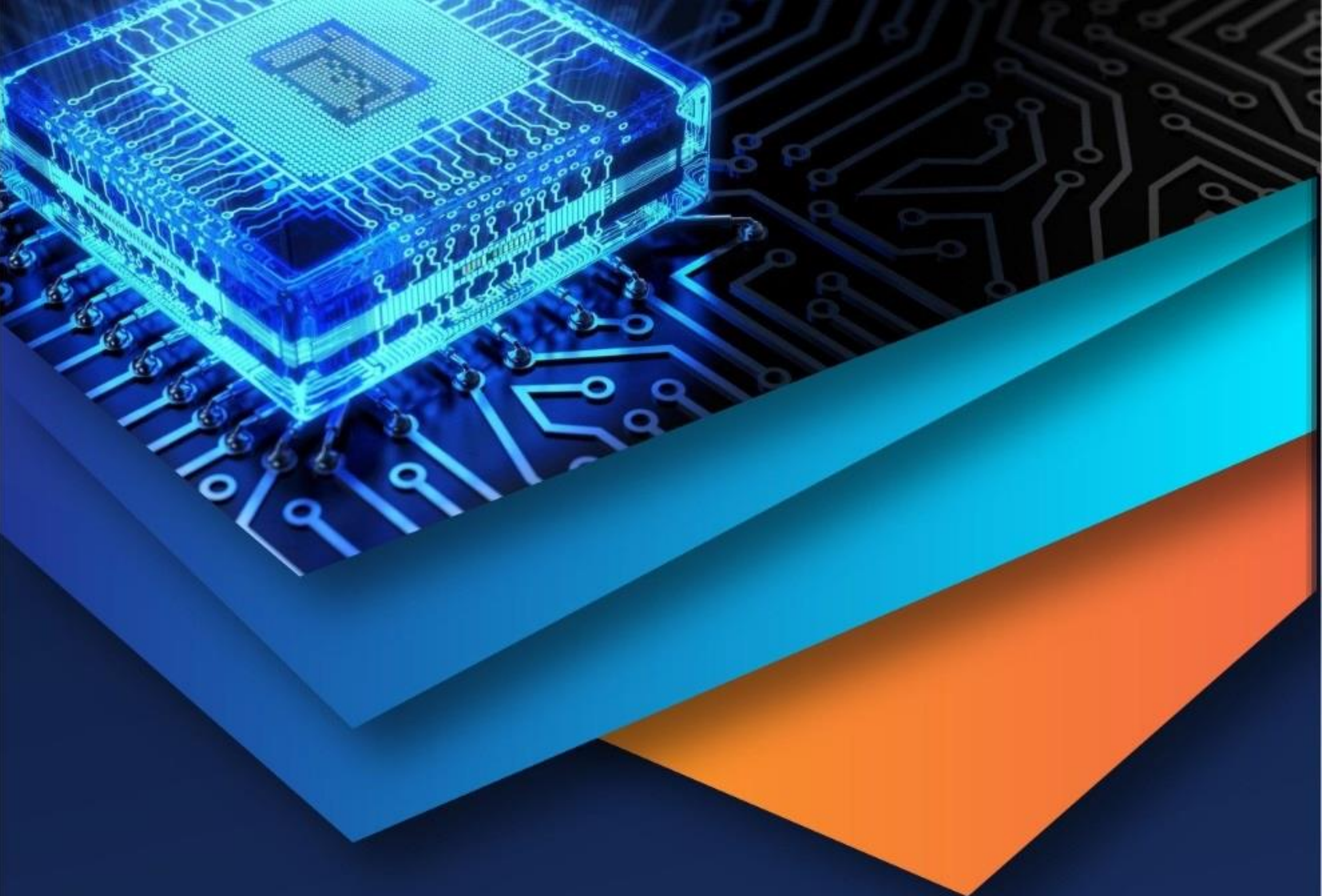
[3] K. Deepika and S. Seema, ''Predictive analytics to prevent and control chronic diseases,'' in Proc. 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. (iCATccT), 2016, pp. 381–386, doi: 10.1109/ ICATCCT.2016.7912028.

[4] I. Mohamed, ''A discrete event simulation model for waiting time management in an emergency department: A case study in an Egyptian hospital,'' Int. J. Model., Simul., Sci. Comput., vol. 12, no. 1, Feb. 2021, Art. No. 2050063, doi: 10.1142/S1793962320500634.

[5] D. Singh et al., ''Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease: The GOLD science committee report 2019,'' Eur. Respiratory J., vol. 53, no. 5, May 2019, Art. No. 1900164, doi: 10.1183/13993003.00164-2019.

[6] R. Rodriguez-Roisin, K. F. Rabe, J. Vestbo, C. Vogelmeier, and A. Augusta, ''all previous and current members of the science committee and the board of directors of GOLD (goldcopd.Org/committees/). Global initiative for chronic obstructive lung disease (GOLD) 20th anniversary: A brief history of time,'' Eur. Respiratory J., vol. 50, no. 1, Jul. 2017, Art. no. 1700671, doi: 10.1183/13993003.00671-2017.

[7] J. Menzin, L. Boulanger, J. Marton, L. Guadagno, H. Dastani, R. Dirani, A. Phillips, and H. Shah, ''The economic burden of chronic obstructive pulmonary disease (COPD) in a U.S. Medicare population,'' Respiratory Med., vol. 102, no. 9, pp. 1248–1256, Sep. 2008, doi: 10.1016/ j.rmed.2008.04.009.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY