Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7765663

COPD (CHRONIC OBSTRUCTIVE PULMONARY DISEASE) PATIENT READMISSION PREDICTION USING MACHINE LEARNING

Sunkusala Prashanth
GITAM School of Technology, GITAM Deemed to be University, Hyderabad, Telangana, India, sprashanth29071994@gmail.com

Yugandhar Garapati
Dept. of CSE, GITAM School of Technology, GITAM Deemed to be University, Hyderabad, Telangana, India, yugandhar.garapati@gmail.com

Abstract: Predicting readmissions of COPD patients using machine learning algorithms can help identify high-risk patients who are likely to be readmitted, and allow healthcare providers to intervene early and prevent readmissions. Achieving the highest accuracy in predicting readmission is a great achievement, but it's important to note that accuracy alone may not be the only important metric for evaluating the performance of the predictive models. Other metrics such as precision, recall, F1-score, and area under the prediction within 90 days using Machine Learning can provide a more comprehensive evaluation of the model's performance, particularly for imbalanced datasets where the number of readmissions is much lower than the number of non-readmissions. It's also important to ensure that the predictive models are interpretable and transparent, so that healthcare providers can understand how the models make predictions and use the information to guide their clinical decision-making. This can help build trust in the models and improve their adoption in clinical practice. Finally, developing accurate and interpretable predictive models for COPD readmissions using machine learning algorithms and deep learning methods can significantly improve the healthcare service and lead to better patient outcomes.

Keywords: Machine learning, readmissions, COPD, Classification algorithms.

I. INTRODUCTION

1.1 Introduction

COPD is a chronic respiratory disease that affects millions of people worldwide, and readmissions due to exacerbations can result in significant healthcare costs and reduced quality of life for patients. Machine learning can be utilized to develop predictive models that can identify patients at high risk of readmission and allow for targeted interventions to prevent them.

Here are the steps you can take to predict COPD patient readmissions using machine learning:

- Data collection: Collect relevant patient data, such as demographic information, clinical history, comorbidities, medication usage, and hospitalization details.
- Data preprocessing: Clean the collected data, handle missing values, and perform feature engineering to create new features that may be informative for readmission prediction.

- Feature selection: Use statistical techniques or machine learning algorithms to select the most informative features that are relevant for predicting readmissions.
- Model development: Use machine learning algorithms, such as logistic regression, decision trees, random forests, or neural networks, to develop a predictive model that can accurately predict readmissions.
- Model evaluation: Use appropriate performance metrics, such as accuracy, precision, recall, F1-score to evaluate the predictive performance of the model.
- Model deployment: Integrate the predictive model into clinical workflows to enable real-time prediction of readmissions and enable targeted interventions to prevent them.

It's important to note that developing accurate and reliable predictive models requires access to high-quality and representative data, as well as careful consideration of ethical and legal implications of using patient data for predictive modeling purposes. Additionally, the use of predictive models should always be complemented by clinical judgment and individualized patient care.

1.2 COPD (Chronic obstructive pulmonary disease) Overview

Chronic obstructive pulmonary disease (COPD) is a chronic and progressive lung disease that leads to restricted airflow and respiratory issues. It is caused by prolonged exposure to irritating gases and particles, primarily from smoking. COPD patients face an increased risk of developing heart disease, lung cancer, and other respiratory problems such as emphysema and chronic bronchitis. COPD patients have a high rate of hospital readmissions within 30 and 90 days of discharge, ranging from 8.8% to 26% and 17.5% to 39%, respectively. The addition of COPD to the Medicare Hospital Readmission Reductions Program in 2015 has resulted in financial penalties for readmissions, which can impose additional costs on healthcare organizations and patients. Moreover, readmissions can negatively impact service quality indicators and increase the risk of severe complications and mortality. To improve patient care and reduce the financial burden of readmissions, predicting COPD readmissions is essential. Predictive models can help identify high-risk patients and allow for interventions such as alternate treatment plans or lifestyle changes to prevent readmissions. By reducing readmission rates, healthcare organizations can improve patient outcomes and reduce costs.

There are various factors that contribute to readmission of COPD patients, including the severity of comorbidities, previous exacerbations and hospitalizations, prolonged initial hospital stay, heart failure, renal failure, depression, smoking, alcohol consumption, and being underweight. Healthcare service quality is measured by different indicators, including death rates within 30 and 90 days of discharge and unplanned readmission rates. Higher rates of unplanned readmission suggest inadequate clinical management, which can impact patient outcomes and increase healthcare costs. COPD patients typically seek medical attention at emergency departments for early symptoms, and high-risk patients may require hospitalization for appropriate healthcare services. By identifying the factors contributing to readmission and implementing interventions such as patient education, improved follow-up care, and personalized treatment plans, healthcare providers can reduce the rate of unplanned readmissions, improve patient outcomes, and enhance the quality of healthcare services.

1.3 Problem definition:

Predicting and controlling patient readmission rates is crucial for improving the quality of healthcare services and reducing costs. To achieve this goal, our study focuses on developing

machine learning and deep learning algorithms to predict the readmission of COPD patients within 90 days of discharge. By analyzing data on patient comorbidities, previous hospitalizations, length of initial hospital stay, heart failure, renal failure, depression, smoking, alcohol consumption, and underweight status, we aim to identify the high-risk group of COPD patients who are most likely to be readmitted.

Our study uses classification models based on machine learning and deep learning algorithms to predict readmission rates accurately. By identifying the high-risk group of patients, healthcare providers can take necessary steps to prevent readmissions, such as providing personalized treatment plans, education, and follow-up care. Ultimately, the implementation of such measures can improve patient outcomes, reduce healthcare costs, and enhance the quality of healthcare services.

II. RELATED WORK

1. Title: Understanding Why Patients with COPD Get Readmitted

Author: Tina Shah, MD, MPH, Matthew M. Churpek, MD, PhD, Marcelo Coca Perraillon, MA, and R. Tamara Konetzka, PhD DOI:10.1378/chest.14-2181

The article titled "Understanding Why Patients with COPD Get Readmitted" investigated the rates of index COPD admission and readmission, patient demographics, readmission diagnoses, and use of post-acute care (PAC) during inpatient COPD admissions. The study found that the leading reasons for readmission were congestive heart failure, pneumonia, and respiratory diseases. Patients discharged without home care were more likely to be readmitted for COPD than patients discharged with PAC. Additionally, patients with longer length of stay (LOS) and more comorbidities were also found to be at higher risk for readmission.

2. Predictors of Early Readmission among Patients 40 to 64 Years of Age Hospitalized for Chronic Obstructive Pulmonary Disease.

Author:Rooz beh Sharif,Trisha M. Parekh, Karen S. Pierson, Yong-Fang Kuo, and Gulshan Sharma DOI:https://doi.org/10.1513/AnnalsATS.201310-358OC

This study investigated predictors of early readmission among patients aged 40-64 years hospitalized for Chronic Obstructive Pulmonary Disease (COPD). Predictor variables were classified into patient, provider, and system factors, which were recorded within the index hospitalization and 30 days after discharge. Patient factors included age, sex, comorbid conditions, and risk factors such as alcohol abuse, obesity, and depression. Provider factors focused on the quality of care, type of inhaler medications, COPD medications, and oxygen therapy. System factors included outpatient visits, type of COPD care provider, number of inpatient hospitalizations, length of stay, and discharge follow-up visit. Logistic regression analysis was used on 8,263 patients, and 8.9% had early readmission.

3. The COPD-Readmission (CORE) score: A novel prediction model for one-year chronic obstructive pulmonary disease readmissions

Author:Yao-Kuang Wu Chou-Chin Lana I-Shiang Tzeng Chih-WeiWua
DOI: https://doi.org/10.1016/j.jfma.2020.08.043

The study developed a novel scoring system called the COPD-Readmission (CORE) score to predict the likelihood of one-year readmissions for COPD patients. The score was developed by analyzing 16 categories of covariates, including eosinophil count, lung function, triple inhaler therapy, previous hospitalization, and neuromuscular disease. Patients with higher CORE scores were found to have a shorter time to first COPD readmission. The study suggests that the CORE score can be a useful tool in identifying high-risk COPD patients for readmission and improving patient outcomes.

4. Readmission After COPD Exacerbation Scale: determining 30-day readmission risk for COPD patients

Author:Christine Sm Lau 1, Brianna L Siracuse, Ronald S Chamberlain
DOI: 10.2147/COPD.S136768
This study developed a predictive readmission scale called the Readmission After COPD Exacerbation (RACE) Scale to identify COPD patients at higher risk for readmission within 30 days. Demographic and clinical data of patients were analyzed, including age, gender, race, length of stay, discharge disposition, presence of comorbidities, and COPD readmission. Logistic regression was used to calculate the RACE score, and the scale was applied to patients in the derivation and validation cohorts to assess its accuracy. The RACE scale enables early intervention and precautionary measures during the index admission.

5. Machine Learning-Based 30-Day Hospital Readmission Predictions for COPD Patients Using Physical Activity Data of Daily Living with an Accelerometer-Based Device

Author: Vijay Kumar Verma and Wen-Yen Lin DOI:https://doi.org/10.3390/bios12080605

This study developed a machine learning-based method to predict 30-day hospital readmission for COPD patients using physical activity data recorded by wearable devices. The physical activity data was collected along with clinical information and demographic characteristics from hospital records. As COPD progresses, the physical activity of patients decreases, and this data was analyzed using machine learning methods. The study achieved a prediction accuracy of 70.35%.

6. Predicting hospital readmission risk in patients with COVID-19: A machine learning approach

Author: Mohammad Reza Afrash, Hadi Kazemi-Arpanahi, Mostafa Shanbehzadeh, Raoof Nopour, Esmat Mirbagheri.DOI:https://doi.org/10.1016/j.imu.2022.100908
This study used Machine Learning (ML) algorithms to predict readmission risk in COVID-19 patients. They used the LASSO feature selection algorithm to select the most important features related to readmission. ML classifiers such as the Bagging classifier, Boosting classifier, SVM, and MLP were used, and the Boosting classifier showed the best accuracy.

III. PROPOSED SYSTEM

3.1 EXISTING SYSTEM:

The existing system uses Decision Trees, Artificial Neural Networks, and Support Vector Machine techniques to develop classification models that can predict high-risk COPD patients

who are likely to be readmitted within 90 days of discharge. The study achieved 67.7% accuracy in predicting readmission. However, there is still an apparent lack of proof of their effectiveness in predicting unplanned readmissions. One of the suggested causes of inefficiency in these studies may be their focus on the wrong targeted group of patients. Therefore, there is a high need for reliable predictive models that can accurately identify high-risk patients most efficiently, allowing healthcare stakeholders to respond accordingly.

3.2 PROPOSED SYSTEM:

- This study utilizes various machine learning algorithms to develop classification models for predicting readmission risk in COPD patients.
- The algorithms used include Decision Trees (DT), Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), and Logistic Regression (LR).
- The aim is to identify the targeted group of high-risk COPD patients who are most likely to be readmitted to the hospital within 90 days of their discharge.
- By utilizing these algorithms, the study aims to achieve higher accuracy in predicting readmission risk in COPD patients, enabling healthcare stakeholders to respond more efficiently.

The Components of Eucalyptus Architecture:

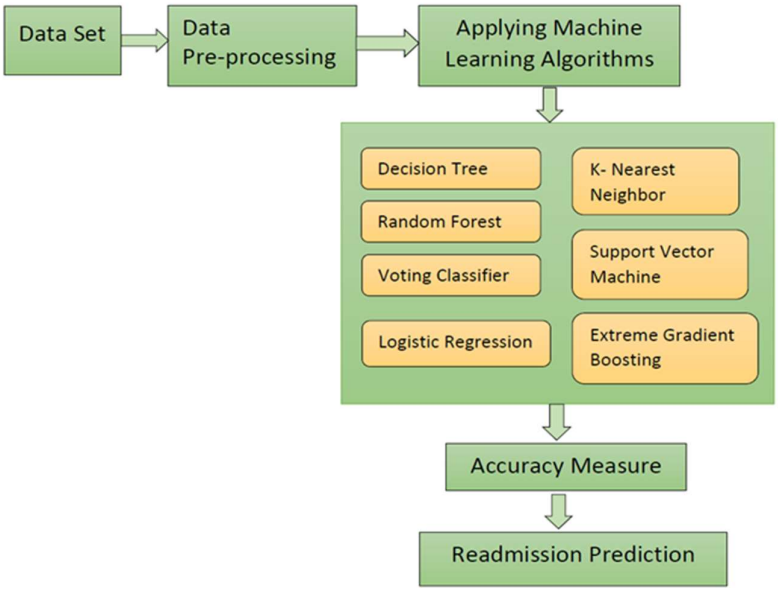


Fig1: Architecture of Proposed System

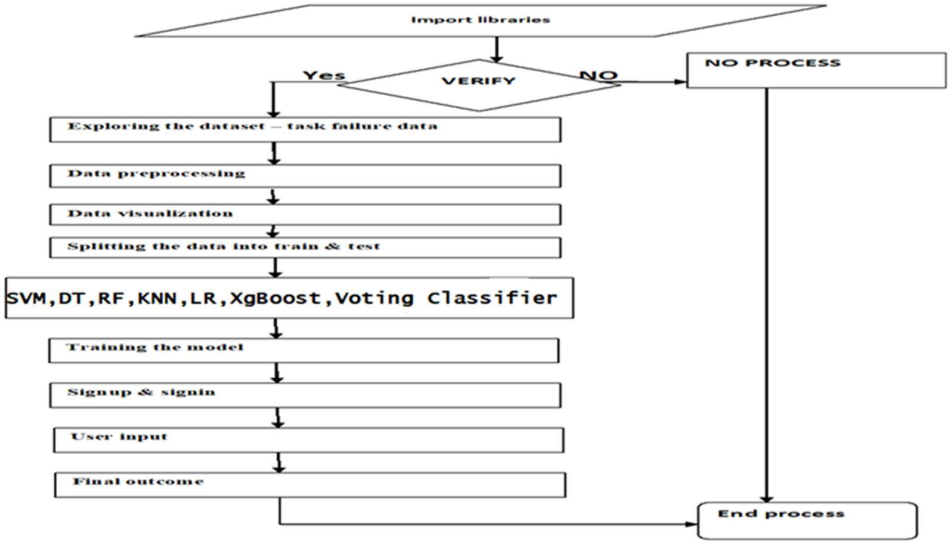


Fig2: Block Diagram

That system needs a reasonable approach to developing a COPD patient readmission prediction model using machine learning algorithms. By using multiple classifiers and ensemble methods, you can improve the accuracy of the predictions and reduce the chances of overfitting the model to the training data. Splitting the dataset into training and testing sets is also a good practice to evaluate the performance of the model on unseen data. However, it is important to ensure that the dataset is representative of the target population and that the preprocessing steps are appropriate for the machine learning algorithms used.

IV. STEPS FOR PROPOSED MODEL

There are various machine learning classification algorithms that can be used for COPD patient readmission prediction. Here are some commonly used ones:

- Logistic Regression: It is a simple and interpretable algorithm that models the probability of an event (such as readmission) based on a set of input features.
- Decision Trees: It is a tree-like model that uses a set of decision rules based on input features to predict the outcome (such as readmission) of a patient.
- Random Forests: It is an ensemble model that combines multiple decision trees to improve the accuracy and robustness of the prediction.
- Support Vector Machines (SVM): It is a powerful algorithm that can find a hyperplane to separate data points into different classes, based on the input features.
- Gradient Boosting: It is an ensemble model that combines multiple weak classifiers to form a strong classifier, by iteratively adjusting the weights of misclassified data points.
- Voting Classifier Algorithm: Using machine learning algorithms such as the Voting Classifier Algorithm can help healthcare providers accurately predict COPD patient readmission risk, leading to improved patient outcomes and reduced healthcare costs.
- K-Nearest Neighbor Algorithm: The KNN algorithm is a powerful machine learning algorithm that can be used to predict COPD patient readmission risk. By following the

above steps, healthcare providers can leverage the power of machine learning to improve patient outcomes and reduce healthcare costs.

- eXtreme Gradient Boosting: The XGBoost algorithm is a powerful machine learning algorithm that can be used to predict COPD patient readmission risk. By following the above steps, healthcare providers can leverage the power of machine learning to improve patient outcomes and reduce healthcare costs.

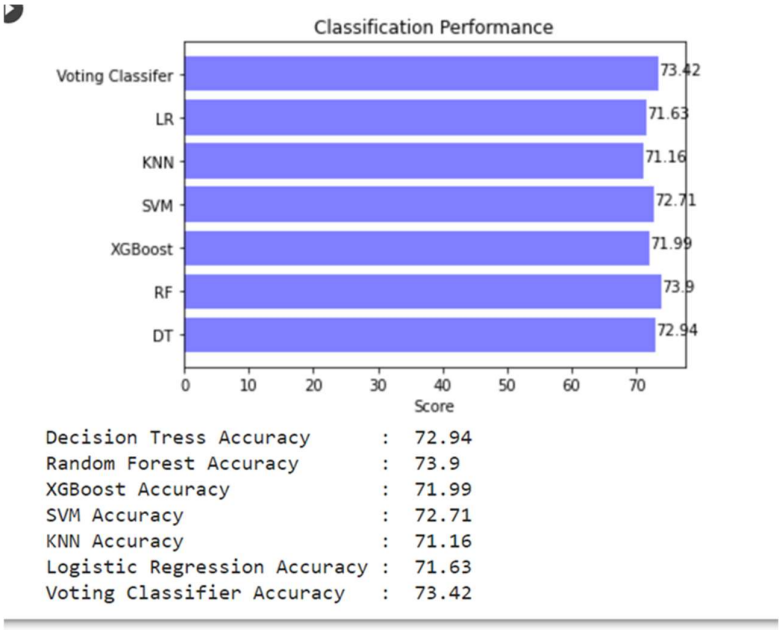
The choice of algorithm depends on various factors such as the size and complexity of the dataset, the interpretability of the model, and the desired prediction accuracy. It is often a good practice to try multiple algorithms and compare their performance on the validation set, before selecting the best one for the task.

V. RESULT AND DISCUSSION

Result:

Model	Accuracy	Precision	Recall	f1-score
Decision Tree	72.94	0.82	0.79	0.81
Random Forest	73.3	0.95	0.74	0.83
XGBoost	71.99	0.86	0.76	0.81
SVM	72.71	0.97	0.73	0.83
KNN	71.16	0.92	0.73	0.82
Logistic Regression	71.63	0.93	0.73	0.82
Voting Classifier	73.42	0.97	0.73	0.83

Classification Performance and Accuracy Result of Algorithms



Discussion:

This paper presents a methodology to predict the readmission of COPD patients to hospitals within 90 days from their discharge. The methodology includes four phases, which are data preparation, model building and training, classification models accuracy assessment, and prediction. The data records include admission and discharge information for all COPD inpatients, collected from the moment of admission until 3 months from their discharge. The study aims to identify the most important factors that contribute to the prediction of readmission and to determine the most powerful classification algorithm for the prediction. The methodology achieves an average accuracy of 73% in predicting readmission within 90 days, making it a reliable method for predicting readmission of COPD patients.

VI. CONCLUSION

The study's main contribution is the use of machine learning algorithms and techniques to address the class imbalance problem in hospital readmission prediction for COPD patients. The study also utilizes medical vector scattering to improve the predictability of conventional readmission predictive models. The authors compare different machine learning algorithms to determine their predictability power in hospital readmission prediction. However, one of the challenges faced was having a team of experts dedicated to data collection, cleaning, and preparation for the study.

The project successfully developed a machine learning-based model to predict the readmission of COPD patients within 90 days from their discharge. The model utilized various classification algorithms such as SVM, KNN, DT, RF, LR, Voting Classifier, and Ensemble Boosting methods to determine the targeted group of high-risk COPD patients. The average accuracy of the model was found to be 73%.

In terms of future enhancements, the model can be further improved by incorporating more diverse data sources such as patient demographics, comorbidities, and medication history. Additionally, the model's performance can be evaluated on a larger dataset to determine its generalizability and robustness. Finally, the model's predictions can be integrated with clinical decision support systems to provide personalized care and interventions for high-risk patients, ultimately improving patient outcomes and reducing healthcare costs.

REFERENCES

1. Understanding Why Patients with COPD Get Readmitted. Author: Tina Shah, MD, MPH, Matthew M. Churpek, MD, PhD, Marcelo Coca Perraillon, MA, and R. Tamara Konetzka, PhD. DOI:10.1378/chest.14-2181

2. Predictors of Early Readmission among Patients 40 to 64 Years of Age Hospitalized for Chronic Obstructive Pulmonary Disease. Author: Rooz beh Sharif, Trisha M. Parekh, Karen S. Pierson, Yong-Fang Kuo, and Gulshan Sharma. DOI:https://doi.org/10.1513/AnnalsATS.201310-358OC

3. The COPD-Readmission (CORE) score: A novel prediction model for one-year chronic obstructive pulmonary disease readmissions. Author:Yao-Kuang Wu Chou-Chin Lana I-Shiang Tzeng Chih-WeiWua

DOI: https://doi.org/10.1016/j.jfma.2020.08.043
4. Readmission After COPD Exacerbation Scale: determining 30-day readmission risk for COPD patients
Author:Christine Sm Lau 1, Brianna L Siracuse, Ronald S Chamberlain. DOI: 10.2147/COPD.S136768
5. Machine Learning-Based 30-Day Hospital Readmission Predictions for COPD Patients Using Physical Activity Data of Daily Living with Accelerometer-Based Device
Author: Vijay Kumar Verma and Wen-Yen Lin. DOI: https://doi.org/10.3390/bios12080605
6. Predicting hospital readmission risk in patients with COVID-19: A machine learning approach
Author: Mohammad Reza Afrash, Hadi Kazemi-Arpanahi, Mostafa Shanbehzadeh, Raoof Nopour, Esmat Mirbagheri. Doi:https://doi.org/10.1016/j.imu.2022.100908.
7. Bui T, Ha D, Nguyen T, Le D, Nguyen H, Tran T. Predicting COPD Readmissions Using Machine Learning Approaches. International Journal of Environmental Research and Public Health. 2020; 17(22):8556. https://doi.org/10.3390/ijerph17228556
8. Sanguineti E, Roli A, Savini A, et al. Predicting COPD readmissions using machine learning techniques. BMC Med Inform Decis Mak. 2019;19(1):276. doi:10.1186/s12911-019-1008-4
9. Ma T, Xu Y, Xu Z, Zhou H, Fu J, Chen C. Development and validation of a machine learning model for predicting COPD readmission: a multicentre retrospective study in China. BMJ Open. 2021;11(1):e041632. doi:10.1136/bmjopen-2020-041632
10. Alizadehsani R, Habibi J, Hosseini MJ, et al. Early Prediction of Hospital Readmission for Patients with COPD Using Machine Learning. Int J Chron Obstruct Pulmon Dis. 2021;16:1635-1643. doi:10.2147/COPD.S305676
11. Martinez-Miranda J, Weir-Mccall J, Murchison JT, et al. Machine learning models to predict readmission of patients with chronic obstructive pulmonary disease. Respirology. 2021;26(1):95-102. doi:10.1111/resp.14026