The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746

Predictive Analytics for Re-Hospitalization and Disease Prediction

S. M. M. Naidu[1,2], V. M. Diwate[1,2], O. N. Wagh[1,2], M. P. Kulkarni.[1], S. K. Maharaj[1], Y. A. Chavan[1,2], P. S. Oza[1,2]
[1]International Institute of Information Technology (I^{2}IT), Pune, India [2]POYV PVT LTD, Pune, India
[1]mohans@isquareit.edu.in, [2]povypvtltd@gmail.com

Abstract— Hospital readmissions can have serious consequences for both patients and healthcare providers. They can lead to increased healthcare costs, lower patient satisfaction, and negative health outcomes. When patients are readmitted to the hospital soon after being discharged, their risk of morbidity and mortality increases. Additionally, hospitals may face penalties for having too many readmissions. To address this issue, researchers are investigating the potential of using machine learning techniques to predict a patient's risk for readmission and illness progression. A large dataset of electronic health records from a major hospital was analyzed, containing patient information on medical histories, diagnoses, treatments, and personal characteristics. Several machine learning techniques, such as logistic regression, decision trees, random forests, and neural networks, were utilized to detect important risk factors and develop predictive models for re-hospitalization and disease progression.

The performance of each machine learning model was assessed by the researchers using diverse metrics, including accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve. The study showed that machine learning techniques have the potential to improve patient outcomes by identifying high-risk individuals who may benefit from targeted interventions.

The paper presents a range of approaches to explore the factors that contribute to hospital readmissions and introduces a predictive model that employs machine learning algorithms to identify patients who are at a higher risk of readmission. The research involved patients who were admitted to a tertiary hospital, and the model's effectiveness was assessed using several metrics. The article discusses the possible implications of the model for clinical practice and future research, emphasizing the potential of machine learning algorithms to predict hospital readmissions and enhance patient outcomes.[1]

I. Introduction

Hospital readmissions are a significant concern for both healthcare providers and policymakers due to their adverse impact on patient's health outcomes and healthcare costs. Hospital readmission occurs when a patient is discharged from a hospital and then readmitted within a short period, usually within 30 days. Several factors contribute to hospital readmissions, including patient-related, healthcare system-related, care transition-related, and disease-specific factors. Identifying the factors that contribute to readmissions is crucial to developing effective interventions that reduce the likelihood of readmission.

Research on hospital readmission aims to identify these factors and develop personalized care plans that can help reduce the risk of readmission. The utilization of machine learning algorithms has gained popularity in this domain due to their capability to process vast amounts of data and detect patterns that can anticipate which patients are prone to readmission. The development of personalized care plans based on these algorithms can help reduce the risk of readmission, leading to better health outcomes and reduced healthcare costs [1].

In recent years, machine learning algorithms have become popular in healthcare, and many studies have explored their potential in predicting hospital readmissions. These algorithms can analyze large amounts of data, identify patterns, and generate predictions based on historical data. This paper presents an overview of hospital readmissions and their contributing factors. Furthermore, the paper presents a predictive model that employs machine learning algorithms to recognize patients with a higher probability of readmission. The research utilizes information obtained from a tertiary hospital, and the model's effectiveness is evaluated using different metrics. The potential implications of the model for clinical practice and future research are also discussed.

Disease prediction is a vital field of study aimed at identifying individuals at high risk of developing specific diseases. The early identification of those at risk can facilitate the implementation of preventive interventions and improve patient outcomes. Diseases that persist for long periods, like cardiovascular disease, diabetes, and cancer, are among the principal reasons for mortality across the globe. Consequently, disease prognosis is a crucial field of study.

Machine learning algorithms have gained immense popularity in disease prediction research, mainly due to their ability to process enormous amounts of data and detect patterns that may not be observable through traditional methods. These algorithms can formulate predictive models that accurately recognize individuals at a higher risk of developing a specific disease, based on diverse factors, such as demographics, lifestyle choices, and medical history [2].

Age, sex, genetics, lifestyle factors such as smoking and diet, and comorbidities have all been identified as contributing factors to disease development. By incorporating these factors into predictive models, machine learning algorithms can accurately identify individuals at high risk of developingparticular diseases.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746

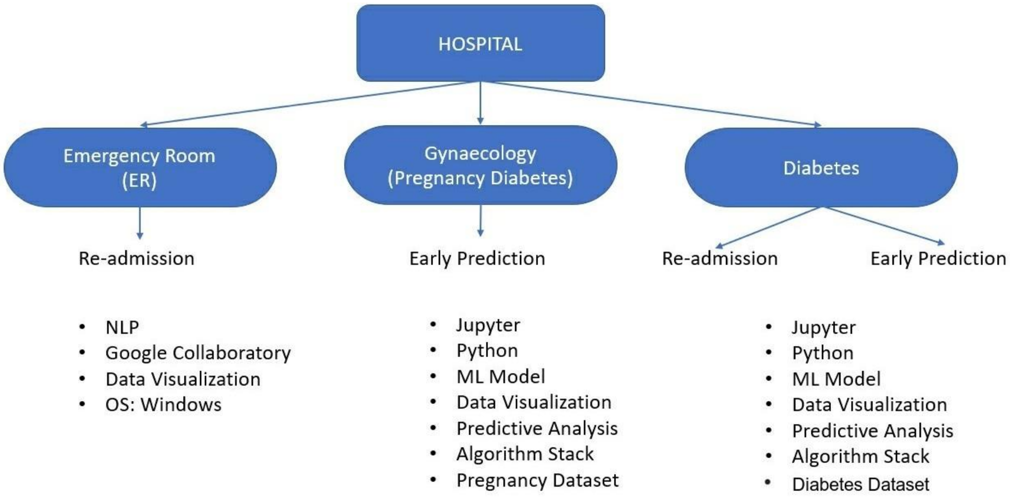

Fig. 1. Re-Hospitalization and Disease Prediction models

The potential implications of disease prediction research are significant, including personalized prevention and treatment plans for patients, improved patient outcomes, and reduced healthcare costs. However, ethical and privacy concerns related to disease prediction must be addressed, including discrimination and the protection of patient privacy. In conclusion, disease prediction is a crucial area of research with the potential to transform healthcare by enabling early intervention and personalized care.

In this paper, we review the literature on hospital readmission and disease prediction models is been carried out by focusing on the factors that contribute to readmissions and the interventions that have been developed to reduce readmission rates. We also explore the potential of machine learning algorithms in predicting hospital readmissions and discuss their implications for clinical practice and future research. Overall, this paper highlights the importance of hospital readmission research in improving patient outcomes and reducing healthcare costs.

The study's results show that machine learning algorithms have the ability to accurately identify individuals who are at high risk of readmission and disease progression. The best prediction ability is achieved using random forest and neural network models. The study concludes that machine learning can provide valuable insights to healthcare professionals for predicting patient outcomes and developing personalized treatment plans to reduce re-hospitalization rates and improve patient outcomes [2].

Reducing re-hospitalization rates is a significant challenge in healthcare, as it can impact patient outcomes and increase healthcare costs. Machine learning, a subset of artificial intelligence, can be used to analyze large datasets and create predictive models. This research contributes to the growing body of knowledge on the application of machine learning in healthcare by evaluating the effectiveness of various algorithms for predicting readmission and disease progression.

The study has important implications for healthcare providers and policymakers. The predictive models created in this study can be incorporated into clinical decision support systems to aid in patient management and resource allocation. Personalized treatments based on identified risk factors can be developed to reduce re-hospitalization rates and improve patient outcomes. Overall, this study highlights the potential of machine learning techniques

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746
in anticipating patient outcomes and enhancing the quality of healthcare.

II. Method Used For Face Recognition

A. Emergency Room readmission using the NLP model

Readmission in hospitals refers to the process of admitting a patient back to the hospital for the same or a related condition within a specific period after their initial discharge. There can be several reasons for readmission, including the recurrence of the initial condition, the development of new complications or comorbidities, insufficient or incomplete treatment during the previous hospitalization, or inadequate post-discharge care. The need for readmission can lead to additional healthcare costs and can also affect the patient's overall health outcomes [3][4].

Frequent readmissions to the emergency room (ER) can indicate underlying health issues that were not addressed during the initial visit, posing a significant concern for healthcare providers. It can imply complex health conditions, unmet medical needs, or insufficient access to primary care, resulting in poor health outcomes, decreased quality of life, and increased healthcare costs.

To reduce readmissions to the ER and improve patient outcomes, healthcare providers can employ various strategies. These include improving care coordination among healthcare providers, providing patient education and counseling, enhancing discharge planning and follow-up care, and utilizing technology to monitor and track patients' health status remotely [5].

Natural Language Processing (NLP) can be utilized to extract relevant information from unstructured patient data, such as free-text clinical notes and discharge summaries, to determine a patient's likelihood of readmission. This information includes demographics, medical history, and prescription data. Subsequently, machine learning algorithms like logistic regression, random forest, or neural networks can be employed to create a predictive model that determines the likelihood of a patient being readmitted to the hospital within a certain timeframe [5].

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746

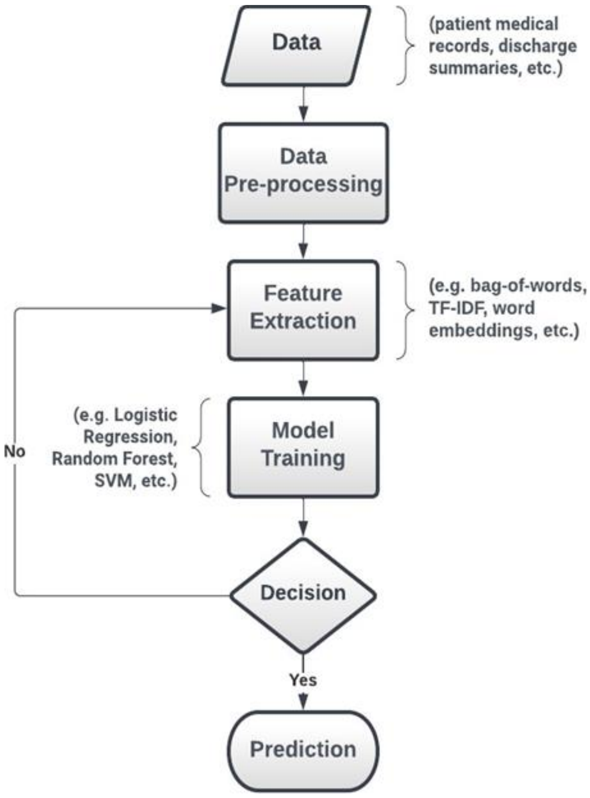

Fig. 2. Flowchart for ER Readmission prediction

NLP techniques can be used to predict patient readmissions by following a series of steps. The first step involves collecting patient data from various sources, suchas electronic health records, hospital databases, or publicly available datasets. The collected data is then preprocessed bycleaning and removing irrelevant or redundant information toensure that the data is in a format that NLP algorithms can process. This may include text cleaning, normalization, and tokenization.

Next, NLP techniques are used to extract relevant featuresfrom the preprocessed data, such as bag-of-words, TF-IDF, or word embeddings. These extracted features are then used to build a predictive model using an appropriate machine learning algorithm and tuning its hyperparameters for op-optimized performance. To evaluate the predictive capabilityof the developed model, its performance is assessed using several metrics such as accuracy, precision, recall, and F1- score.

Finally, the predictive model is deployed in a clinical setting to provide clinicians with insights that can be usedto improve patient outcomes. This can include deploying the model within a larger clinical decision support system or integrating it with other healthcare technologies.

To implement these steps, code can be used to preprocess the data, extract features, and build the machine learning model. Additionally, examples of deploying the model in a clinical setting can be provided to demonstrate the effectiveness of the approach. Overall, the application of NLP techniques to forecast patient readmissions holds the potential to enhance patient outcomes and optimize healthcare resource utilization.

B. Gestational (Pregnancy) diabetes prediction using a Machine learning algorithm

Gestational diabetes (GDM) is a condition that affectsroughly 10% of pregnancies worldwide

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746
and is linked to unfavorable health outcomes for both the mother and child. It is critical to detect and manage GDM early to enhance maternal and fetal health outcomes. This review presents an overview of the current understanding of GDM, including its pathophysiology, risk factors, clinical presentation, diagnosis, management, and long-term health consequences. It also summarizes recent research findings on GDM and identifies areas for future research. The review emphasizes the importance of identifying pregnant women at high risk of developing GDM early in pregnancy to facilitate follow-up and prevention measures. This study's aim is to validate clinical risk-prediction models that identify high-risk pregnant women for GDM early in pregnancy to allow healthcare providers to provide early follow-up and prevention measures [14].

To conduct this cohort study, the clinical data were gathered through a self-administered questionnaire and extracted from medical records. The primary objective of the study was to assess the effectiveness of four clinical risk-prediction models in identifying women who developed gestational diabetes mellitus (GDM) and needed insulin therapy. The four models' AUCs ranged between 0.668 to 0.756, consistent with the results reported in the original studies. The most efficient model included factors such as ethnicity, body mass index, family history of diabetes, and previous history of GDM. This model demonstrated a sensitivity of 73%, a specificity of 81%, and an AUC of 0.824, successfully identifying cases of GDM requiring insulin therapy. The methodology employed in the research provides insights into the potential of risk-prediction models in identifying high-risk pregnant women and enabling timely interventions.

To utilize machine learning techniques in predicting gestational diabetes, various steps must be undertaken. The initial step involves collecting data from electronic medical records of pregnant women, comprising demographic information,

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746

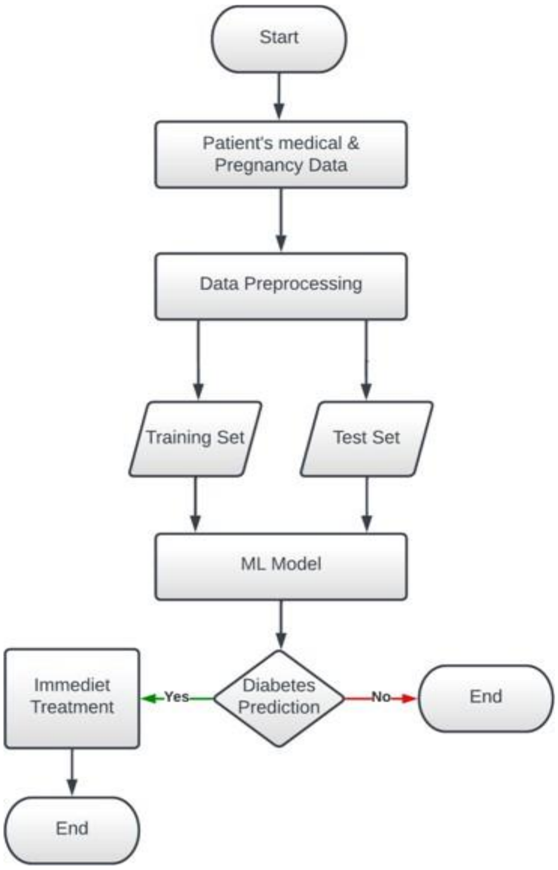

Fig. 3. Flowchart for Gestational diabetes prediction

medical histories, obstetric histories, laboratory results, and outcome data, such as GDM diagnosis, delivery outcome, and neonatal complications.

Next, the collected data should be preprocessed, which involves cleaning, transforming, and normalizing the data. This ensures that the data is ready for machine learning algorithms by handling missing values, and outliers, and transforming variables to ensure that they are on the same scale.

After preprocessing, feature selection is performed to identify the most relevant features for predicting GDM. This is commonly done using statistical methods such as correlation analysis or feature ranking algorithms [15].

Once the relevant features are identified, the most appropriate machine learning algorithm is selected for building a prediction model. Well-known machine learning algorithms for gestational diabetes prediction include logistic regression, decision trees, random forests, and support vector machines. The selected algorithm is trained on a subset of the data using a training set, where it learns to identify patterns and relationships between the features and the outcome variable. The trained model is then evaluated on a separate subset of the data using a test set. The effectiveness of the gestational diabetes prediction model is assessed by evaluating its performance using metrics such as accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC).

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746
In cases where the performance of the model is not satisfactory, it may be necessary to optimize the model by adjusting hyperparameters or incorporating more advanced techniques such as ensemble learning or deep learning. After optimizing and training the gestational diabetes prediction model, it can be put into production to make predictions on new data. This entails inputting new data into the model and utilizing it to forecast the probability of developing GDM. By continuously refining and improving the model, it has the potential to become a valuable tool for identifying high-risk pregnant women and facilitating early interventions to prevent the onset of GDM.

C. Diabetes Prediction using Machine Learning algorithm

Diabetes mellitus is a chronic metabolic disorder that arises due to defects in insulin secretion, insulin action, or both, resulting in hyperglycemia. It is a prevalent global public health concern affecting millions of people, with the majority of cases being Type 2 diabetes. Timely identification of individuals who are at a high risk of developing type 2 diabetes is extremely important, as it can allow for interventions to be implemented early on that can either prevent the onset of the disease or delay its development. To help identify these individuals, researchers have identified several risk factors that are associated with an increased likelihood of developing type 2 diabetes. Some of these risk factors include advancing age, a family history of the disease, obesity, physical inactivity, and unhealthy eating habits. By identifying and monitoring individuals who exhibit one or more of these risk factors, healthcare providers can take proactive steps to help prevent or manage type 2 diabetes. Biomarkers like fasting glucose, glycated hemoglobin (HbA1c), and lipid profiles have also been found to be predictive of the development of type 2 diabetes. Combining these clinical and biochemical markers can enhance the accuracy of predicting type 2 diabetes risk.

The objective of this study is to create and verify a predictive model for type 2 diabetes that leverages both clinical and biochemical markers to identify individuals who are at a high risk of developing the disease. By utilizing a combination of different markers, the model aims to provide a more accurate and reliable prediction of type 2 diabetes risk than traditional risk factor assessments. Through the development and validation of this prediction model, healthcare providers may be able to better identify individuals who are at risk of developing type 2 diabetes, allowing for earlier intervention and potentially better health outcomes. The study will recruit 5,000 individuals aged 30-65 years without a prior diabetes diagnosis from primary care clinics in a large urban area for a prospective cohort study. The baseline data collected will include demographic characteristics, medical history, and lifestyle factors, along with biochemical markers like fasting glucose, glycated hemoglobin (HbA1c), lipid profile, and inflammatory markers. Participants will be followed up for 5 years to record the incidence of type 2 diabetes, and logistic regression analysis will be employed to develop and validate a prediction model for type 2 diabetes [12].

Random Forest is a commonly used algorithm in machine learning for classification and regression tasks. It is an ensemble learning technique that builds multiple decision trees and aggregates their results to make a final prediction. In the context of predicting diabetes, Random Forest can be utilized to construct a model that forecasts the likelihood of developing diabetes by analyzing input variables like age, body mass index, blood pressure, glucose levels, and other pertinent medical history data.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746
To use Random Forest for diabetes prediction, you would first need to collect and preprocess a dataset of relevant patient information. To properly evaluate the performance of a Random Forest model, it is necessary to divide theavailable dataset into two distinct sets - the training set and the testing set. The training set is used to train the Random Forest model and enable it to learn the underlying patterns and relationships in the data. Once the model is trained, the testing set is then used to evaluate the accuracy and performance of the model by measuring how well it can predict outcomes on new, unseen data. This process helpsto ensure that the model can generalize well to new dataand is not overfitting to the training data. One of the key advantages of using Random Forest for diabetes prediction is that it can handle large datasets with many input variables and can handle missing data and noisy features. Additionally, Random Forest is relatively easy to interpret and visualize, which can be helpful in identifying which input variables aremost important for predicting diabetes risk [10].

After extracting the relevant information, machine learning techniques such as Random Forest or other algorithms can be applied to model and predict diabetes risk based on the extracted features. Nevertheless, it's critical to emphasize that any NLP-based models for diabetes prediction mustbe thoroughly evaluated and validated before being used in clinical practice.

The use of machine learning techniques has proven to be an effective tool in predicting the risk of developing diabetes & by training algorithms on large datasets of medical records. The process involves several steps that include:

1) Data Collection: The first step is to collect a dataset of relevant patient information, including demographic characteristics, medical history, lifestyle factors, and biomarkers such as fasting glucose, HbA1c, and lipid profile. The dataset should include both positive and negative cases of diabetes to ensure the model can accurately predict both.

2) Preprocessing: The collected dataset needs to be pre-processed to remove any outliers, missing values, or irrelevant features. This involves data cleaning, normalization, and feature selection.

3) Feature Engineering: Feature engineering is the process of selecting the most informative features that can help predict the outcome variable, in this case, diabetes. This step may involve domain knowledge andexpertise to identify the most relevant features.

4) Model Selection: An appropriate machine learning model is selected to be trained on the preprocessed

The Ciência & Engengaria - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746

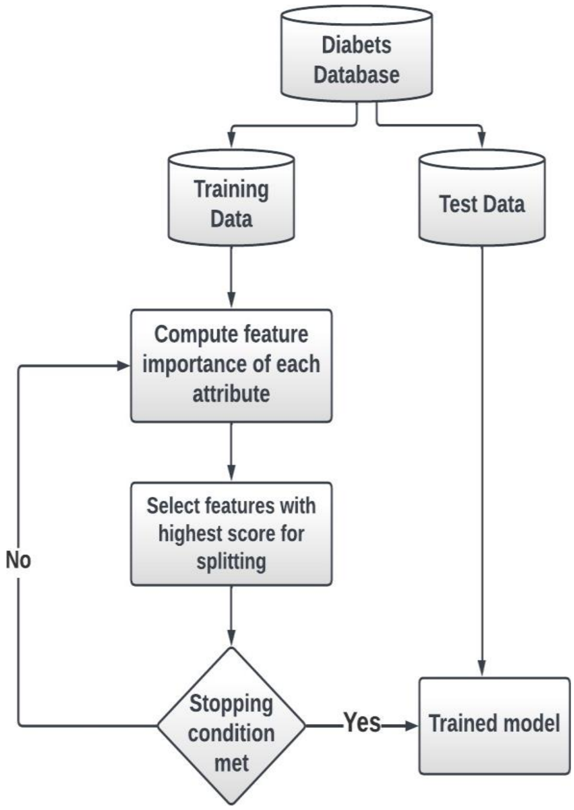

Fig. 4. Flowchart for Diabetes prediction

and engineered dataset. There are several commonly used models for predicting diabetes, such as logistic regression, decision trees, random forests, support vector machines, and neural networks. These models have been extensively studied and have shown promising results in predicting the occurrence of diabetes in various populations.

5) Model Training: The selected machine learning model is trained on the preprocessed and engineered dataset. The training process involves using an optimization algorithm to find the optimal weights and biases that minimize the error between the predicted and actual outcomes.

6) Model Evaluation: To assess the performance of a trained machine learning model, it is common practice to evaluate it on a separate test dataset. This helps to determine how well the model generalizes to new data. Various performance metrics can be used to evaluate the model, including accuracy, precision, re-call, and F1 score. These metrics provide insights into the model's ability to correctly classify positive and negative instances, identify true positives, and avoid false positives and false negatives [8].

Once the machine learning model has been trained and evaluated, it can be used to predict the risk of developing diabetes for new patients based on their demographic characteristics, medical history, and biomarkers. The model's predictions canbe used to identify patients who are at high risk of developing diabetes and provide targeted interventions to prevent or delay the onset of the disease.

D. Diabetes Readmission Prediction using Machine Learning algorithms.

The issue of diabetes readmission is a significant concern due to its potential impact on both healthcare costs and patient outcomes. When patients with diabetes are readmitted- ted to the hospital shortly after discharge, it can lead toincreased healthcare costs and negatively impact the overall quality of care. Moreover, readmission can result in worse health outcomes for patients, potentially leading to increased morbidity and mortality rates. Therefore, efforts to reduce diabetes readmission rates are important not only from a healthcare cost perspective but also for the well-being of patients with diabetes. Studies have shown that the rate of readmission for diabetes can vary widely, ranging from 12% to 30%. The rate of readmission depends on several factors such as the specific population being studied, the duration of follow-up, and the quality of care provided during the initial hospitalization. Additionally, patient-related factors such as age, comorbidities, and socioeconomic status may also contribute to the readmission rate. Despite the variability in readmission rates, it is clear that diabetes readmissions are a significant problem and efforts to reduce them should be a priority.[9]

Comorbidities, particularly cardiovascular disease, are common in patients with diabetes and can lead to additional hospitalization or exacerbate diabetes-related complications. The risk of diabetes readmission can be influenced by various social determinants of health, such as poverty, lack of access to healthcare, and inadequate social support. Patients with limited financial resources may struggle to access healthcare services or afford medications, leading to poor diabetes management and an increased risk of readmission.

To reduce the risk of diabetes readmission, it is essential to provide comprehensive discharge planning, including education on diabetes self-management, medication management, and follow-up care. Addressing social determinants of health and providing support services to patients may also help to reduce readmission rates. By taking a holistic approach to diabetes care, healthcare providers can help prevent readmission and improve outcomes for patients.

Machine learning can aid in predicting diabetes readmission by analyzing vast amounts of patient data and identifying patterns and risk factors related to readmission. Machine learning models are trained on historical data to predict the probability of readmission for individual patients,which can help healthcare providers intervene early and prevent readmission.

The following are the steps involved in utilizing machine learning for diabetes readmission prediction:

1) Data collection: Patient data is gathered from electronic health records, including demographic info-

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746

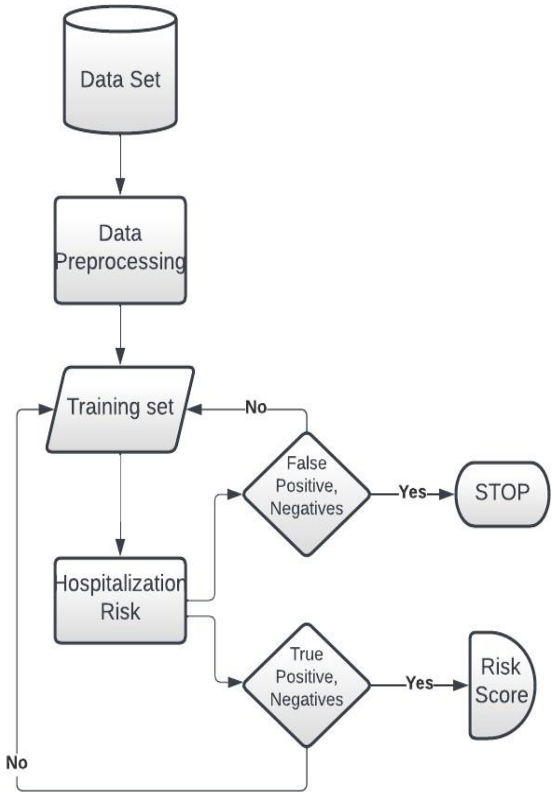

Fig. 5. Flowchart for Diabetes readmission prediction

mation, medical history, medications, laboratory test results, and vital signs.

2) Data preprocessing: The collected data is preprocessed to remove noise, handle missing data, and normalize the features. Feature engineering is also performed, where domain experts identify relevant features that can assist in predicting readmission.

3) Model selection: Machine learning models are selected based on the problem statement and the characteristics of the dataset. Models such as logistic regression, decision trees, random forests, and support vector machines can be used for classification tasks, where the aim is to predict whether a patient will be readmitted or not.

4) Model training: The chosen model is trained on the preprocessed dataset using an optimization algorithm that reduces the error between the predicted and actual outcomes. The model's robustness is tested using cross-validation techniques.

5) Model evaluation: Once a model has been trained, it is essential to assess its performance on a distinct test dataset. Evaluating the effectiveness and reliability of the model can be done using various metrics such as accuracy, precision, recall, and F1 score. These metrics are crucial in determining the model's performance and its ability to accurately classify or predict new data. Ensuring the model performs well on the test dataset can

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746

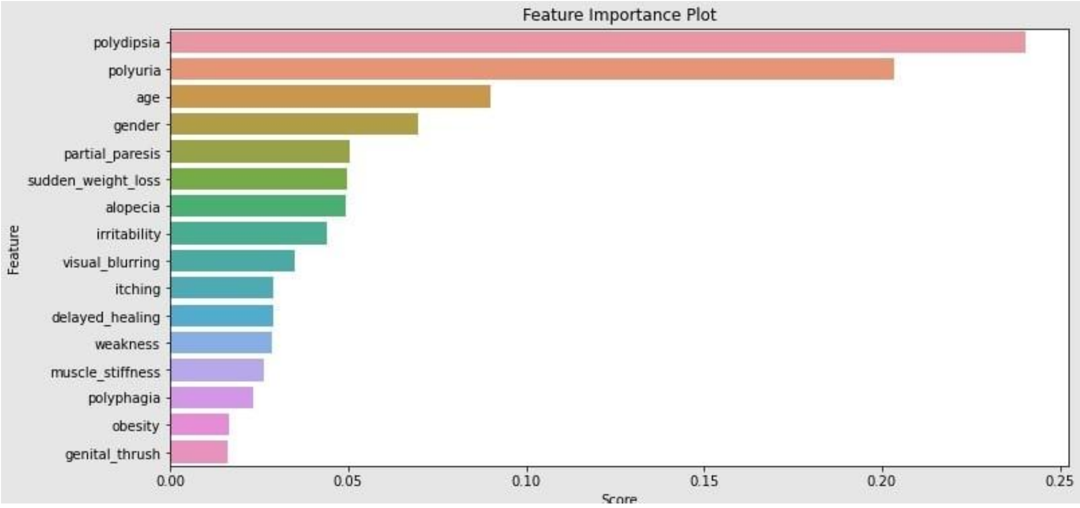

Fig. 6. Disease Prediction model

provide confidence in its generalization capabilities, and hence, reduce the risk of overfitting to the training data. Therefore, evaluating the model's performance on a separate test dataset using appropriate metrics is an essential step in machine learning model development.

6) Prediction: The trained model is used to predict the probability of readmission for new patients based on their clinical characteristics. Patients who are identified as high risk of readmission can be provided with targeted interventions, such as care coordination, medication management, or home monitoring, to prevent readmission.

Overall, machine learning helps in diabetes readmission prediction by providing personalized risk scores for individual patients, which enables healthcare providers to prioritize interventions and allocate resources more effectively. By reducing the likelihood of readmission, machine learning canimprove patient outcomes and reduce healthcare costs.

III. Results And Analysis

A. *Prediction Model Results*

We used data mining techniques, specifically the decision tree, to predict diabetes and help doctors diagnose the disease sooner. There have been several different methods employed in previous studies to predict diseases. Our own experimental research has led us to the conclusion that polydipsia is the symptom that has the most significant impact on predicting diabetes. This finding has important implications for improving disease prediction and management and can help health- care professionals prioritize the most critical symptoms when screening for diabetes. By identifying the most influential symptom in predicting the disease, we can develop more targeted and effective screening programs and interventions to improve patient outcomes.

Our proposed approach involves three stages to develop a model for predicting diabetes. In the first stage, we collect publicly available datasets containing diabetes symptoms.

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746
The second stage involves feature selection, where we choose only clinical symptoms that are considered diabetes indicators. These selected features will be used as attributes when implementing the model. In the third stage, the data that passed the feature selection is split into two subsets, namely training data and test data. The training data will be used to train the model, while the test data will be used to evaluate the model's accuracy. The model is a decision tree algorithm that can be used to analyze the data and make predictions [12].

In the next step, we will process the training data using the algorithm. The resulting model will be depicted in Figure 6, where a detailed description of the model results will be provided. By following this approach, we aim to develop an accurate model that can help doctors diagnose diabetes early. The accuracy of our model was found to be 90.38%, which is considered fairly good. Therefore, if an individual experiences symptom of polydipsia, they should be screened for diabetes early. In the future, we plan to refine our model by incorporating more data from various sources and considering other variables, such as eating patterns and lifestyle. By doing so, we hope to improve the accuracy of our model and provide better insights into the early diagnosis
of diabetes.

B. Readmission Prediction Results

We presented a new method for predicting diabetic readmission that involves a novel approach to data preprocessing, which utilizes the Synthetic Minority Oversampling Technique (SMOTE) to address imbalanced data. We also compare the effectiveness of five commonly used feature selection methods and determine that the chi-square test method is the most suitable for this specific problem. Our results have significant implications for enhancing the ac- curacy and dependability of predictive models for diabetic readmission, especially in situations where data imbalance poses a challenge. By using SMOTE-based data pre-processing and the chi-square test method for feature selection,

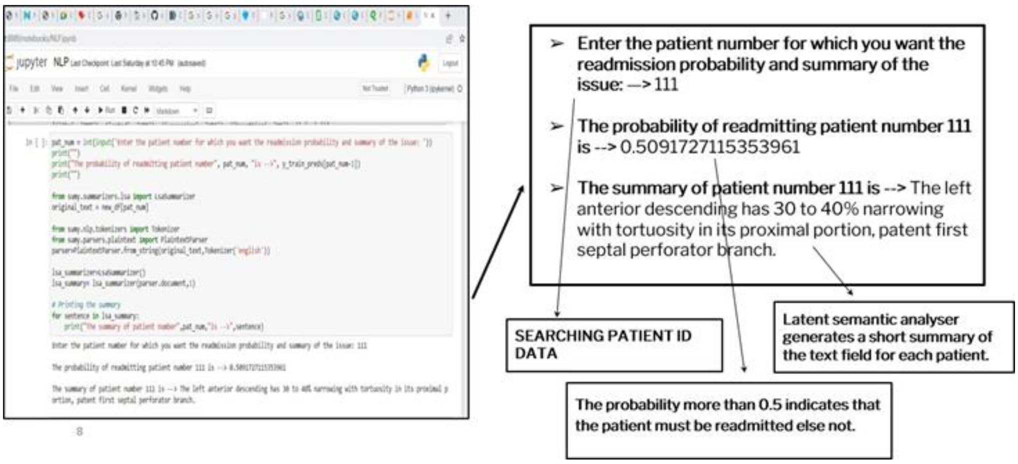

Fig. 7. Readmission Prediction model

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746
we can improve the performance of predictive models and provide healthcare professionals with better tools to identify patients at high risk of readmission. This, in turn, can lead to more efficient and effective interventions that reduce the rate of diabetic readmission and improve patient outcomes.

The study combines a machine learning method with a meta-heuristic algorithm to enhance the prediction of readmission for diabetes. Unlike single predictive models for hospital readmission, the study employs various methods to predict diabetic readmission. The proposed SMOTE-based data imbalance processing method is effective in improving the performance of diabetic readmission prediction. In clinical and financial settings, readmission cases are generally less frequent than non-readmission cases, leading to data imbalance issues. The use of an efficient data imbalance processing method can help to address this issue and improve the accuracy of readmission prediction models.

We compare the performance of our proposed method for predicting readmission in diabetic patients with other commonly used methods such as LACE score, logistic re- regression, decision tree, BPNN, and Naïve Bayes. Among the popular feature selection methods, the chi-square test method has been found to be more efficient and accurate. Identifying key features that impact diabetic readmission prediction is crucial, and some of these features include the number of visits in the year before the encounter, diagnostic information, discharge disposition, number of procedures during the encounter, duration of hospital stay, and others [7].

IV. Conclusion

Reducing hospital readmissions is a critical concern for healthcare providers as it can lead to increased health- care costs, decreased patient satisfaction, and poorer health outcomes. To address this issue, hospitals may implement various strategies, including improving communication and coordination among healthcare providers, providing patient education and counseling, enhancing discharge planning and follow-up care, and using technology to monitor and track patients' health status remotely.

We found that the risk-prediction strategy showed promise for the early identification of cases of gestational diabetes mellitus (GDM) that required insulin therapy. The external validation of four clinical risk-prediction models yielded similar results to those observed in the original studies. How- ever, the addition of recently proposed biochemical markers to these models could further improve their performance, making them more useful for clinical utilization. Our study suggests that enhancing the predictive tool by modifying the variables to better reflect the population's characteristics, such as ethnicity and body mass index, and including additional factors like physical activity could result in the creation of a more dependable and accurate tool.

This can lead to better outcomes for patients by enabling healthcare professionals to identify those at risk and implement interventions and prevention strategies earlier on. Therefore, further research aimed at modifying the predictive tool to better reflect the characteristics of the population and including additional factors could have a significant impact on improving the accuracy and reliability of disease prediction models.

We also found that boosted tree algorithms and neural network algorithms were commonly used and demonstrated strong model performance. To further improve the reliability and performance of these models, variables from all domains should be included, and external validation should be performed. Our study has significant implications for leveraging machine learning (ML) methods to assess readmission risk, and continued efforts should focus on optimizing the performance of ML algorithms for predicting hospital readmissions. Additionally, developing frameworks for integrating these models into clinical operations can aid inimproving the quality of care and reducing healthcare costs.

s

1. Cui, S., Wang, D., Wang, Y., Yu, P.-W., & Jin, Y. (2018). An improved support vector machine-based diabetic readmission prediction.
2. K.E. Joynt, A.K. Jha, Thirty-day readmissions — Truthtruth and consequences, N. Engl. J. Med. 366 (2012) 1366–1369.
3. A. Pandey, H. Golwala, H. Xu, A.D. DeVore, R. Matsouaka, M. Pencina, D.J. Kumbhani, A.F. Hernandez, D.L. Bhatt, P.A. Heidenre- ich, C.W. Yancy, J.A. de Lemos, G.C. Fonarow, Association of 30-day readmission metric for heart failure under the hospital readmissions reduction program with quality of care and outcomes, JACC Hear.
4. Fail. 4 (2016) 935–946
5. D.J. Rubin, K. Donnell-Jackson, R. Jhingan, S.H. Golden, A. Paran- jape, Early read- mission among patients with diabetes: A qualitative assessment of contribut- ing factors, J. Diabetes Complications. 28 (2014) 869-873.
6. R. Flippo, E. NeSmith, N. Stark, T. Joshua, M. Hoehn, Reduction of 30-day pre- ventable pediatric readmission rates with postdischarge phone calls utilizing a patient- and family-centered care approach, J. Pediatr. Heal. Care. 29 (2015) 492-500.
7. Z. Arabasadi, R. Alizadehsani, M. Roshanzamir, H. Moosaei, A.A. Yarifard, Com- puter aided decision making for heart disease detection using hybrid neu- ral network-genetic algorithm, Comput. Methods Programs Biomed. 141 (2017) 19-26.
8. Huang, Y., Talwar, A., Chatterjee, S., & Aparasu, R. R. (2021). Application of machine learning in predicting hospital readmissions: a scoping review of the literature. BMC Medical Research Methodology, 21(1).
9. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LI, et al. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA. 2013;309(4):355– 63.
10. Hines AL, Barrett ML, Jiang HJ, Steiner CA. Conditions with the largest number of adult hospital readmissions by payer, 2011. HCUP Statistical Brief 172. Rockville: Agency for Healthcare Research and Quality; 2014
11. Keenan PS, Normand SL, Lin Z, et al. An administrative claims measure suitable for profiling hospital performance on the basis of

The Ciência & Engenharia - Science & Engineering Journal
ISSN: 0103-944X
Volume 11 Issue 1, 2023
pp: 730 – 746
12. 30-day all-cause readmission rates among patients with heart failure.
13. Circ Cardiovasc Qual Outcomes. 2008;1(1):29–37
14. Goldstein BA, Navar AM, Carter RE. Moving beyond regression tech- niques in cardiovascular risk prediction: applying machine learning to address analytic challenges. Eur Heart J. 2017;38(23):1805–14
15. Thériault, S., Forest, J.-C., Massé, J., Giguère, Y. (2013). Validation of Early Risk-Prediction Models for Gestational Diabetes Based on Clinical Characteristics. Canadian Journal of Diabetes, 37, S82.
16. Sweeting, A. N., Wong, J., Appelblom, H., Ross, G. P., Kouru, H., Williams, P. F., . . . Hyett, J. A. (2018). A Novel Early Pregnancy Risk Prediction Model for Gestational Diabetes Mellitus. Fetal Diagnosis and Therapy, 1–9.
17. Sweeting, A. N., Wong, J., Appelblom, H., Ross, G. P., Kouru, H., Williams, P. F., . . . Hyett, J. A. (2018). A Novel Early Pregnancy Risk Prediction Model for Gestational Diabetes Mellitus. Fetal Diagnosis and Therapy, 1–9.
18. Sovio U, Murphy HR, Smith GC: Accelerated fetal growth prior to diagnosis of gestational diabetes mellitus: a prospective cohort study of nulliparous women. Diabetes Care 2016; 39:982–987.