Journal of Medical and Health Studies
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| **RESEARCH ARTICLE**

Advancing Healthcare Outcomes with AI: Predicting Hospital Readmissions in the USA

Md Musa Haque[1], Shah Foysal Hossain[2], Sarmin Akter[3], Md Azharul Islam[4], Shahriar Ahmed[5], Irin Akter Liza[6] and Md Al Amin[7]

[1357]School of Business, International American University, Los Angeles, California, USA
[2]School of IT, Washington University of Science and Technology, Alexandria, Virginia, USA
[4]College of Business, Westcliff University, Irvine, California, USA
[6]College of Graduate and Professional Studies (CGPS), Trine University, Detroit, Michigan, USA
Corresponding Author: Md Musa Haque, **E-mail**: 199086@stu.hebut.edu.cn

| ABSTRACT

The issue of readmission rates in hospitals has been described as both a serious and perplexing problem in America's healthcare system. The high persistence of readmission rates underscores the urgent need for improvement in better tools and techniques for the forecasting and management of occurrences with efficiency. The chief objective of this research was to devise and ameliorate AI models that can effectively predict patient readmissions. Through machine learning and data analytics, this study worked toward developing tools that will highlight patients at a high risk of readmission, which can be targeted with interventions by healthcare providers. The hospital readmission dataset used in this study comprised a comprehensive collection of patient-related data aimed at understanding and predicting readmissions. The dataset was thereby developed using electronic health records which capture all clinical activities - diagnosis code treatment history, results of labs, and medication-related prescriptions. Demographic details related to patients will include: age, sex, and ethnic background - for contextualizing at the population level. This clinical information was complemented by unstructured data, such as clinical notes that give further detailed insight into patient conditions and advice on follow-up care. Several models were considered for classification tasks such as Random Forest Classifier, Logistic Regression, and XG-Boost Classifier. Some of the key metrics used to quantify the model's effectiveness included accuracy, precision, recall, F1-score, and ROC-AUC. Gradient Boosting had the highest scores on all four metrics and maximum accuracy and F1-score, showing the best all-rounded performance in prediction. Interpreting healthcare model outputs provides insightful predictions to inform clinical decisions. Care strategies have to be developed based on predictive insights and patient segmentation analysis to enhance the outcomes of patients. AI-driven insights will thus require a strategic approach to the integration of AI-driven models in the functioning of the hospital.

| KEYWORDS

Hospital readmissions, Healthcare outcomes, Artificial intelligence, U.S. healthcare system, Predictive modeling, Patient care, Cost reduction, Data analytics

| ARTICLE INFORMATION

ACCEPTED: 02 August 2023 **PUBLISHED:** 29 September 2023 **DOI:** 10.32996/jmhs.2023.4.5.13

I. Introduction

Context and Importance

The issue of readmission rates in hospitals has been described as both a serious and perplexing problem in America's healthcare system. Most definitions refer to hospital readmission as a patient who, after having been discharged from an earlier hospitalization, returns within some defined period, many times within 30 days of that previous discharge. The frequency with which patients are readmitted reflects quality in both the initial period of care and post-hospital management (Teo et al., 2023). Despite many attempts to help stem the tide, readmissions related to conditions of heart failure, chronic obstructive pulmonary disease, and diabetes remained alarmingly high. The economics of readmission to the hospital are huge. The Centers for Medicare and Medicaid Services estimated unplanned hospital readmissions cost the US health system more than $26 billion annually (Sarker et al., 2023). Besides the cost impact, frequent readmission has a detrimental effect on patient health: disruption to life, prolonged suffering, and erosion of trust in healthcare services. New approaches are needed that go beyond conventional methods of care management.

According to Medicaid Services, nearly 20% of Medicare beneficiaries are hospitalized again within 30 days of discharge, which has resulted in vital concerns about patient safety and quality health care. Frequent readmission does not only increase costs related to health care but crowding in hospitals increases the resources burden and lowers the quality level of care for all the patients. The ramifications extend beyond financial ones; patients are sicker because of the heightened health risks and complications from their initial diagnosis. Beyond the patient, readmission reverberations resound throughout the healthcare continuum for providers and payers alike. Under the HRRP, hospitals will be penalized if their rates of readmission exceed acceptable thresholds. This is in the form of reduced reimbursements under Medicare. This, therefore, creates a financial incentive for healthcare providers to reduce their readmission rates and necessitates them to identify mechanisms for risk-based identification and proactive intervention.

The adoption of AI into healthcare inspires a transformative opportunity to elevate patient care and significantly reduce healthcare costs. AI algorithms can process large volumes of data quickly and identify patterns and insights that might be missed by human analysts. Using machine learning techniques, healthcare providers will be able to develop predictive models identifying high-risk patients, thus enabling targeted interventions and personalized care plans. For instance, AI can flag those patients with certain comorbidities or socioeconomic factors associated with higher readmission rates, thus allowing for the taking of proactive measures such as increased follow-up care, patient education, and community support services.

Motivation

As per Farid et al. (2023), the high persistence of readmission rates underscores the urgent need for improvement in better tools and techniques for the forecasting and management of occurrences with efficiency. Traditional approaches involving clinical judgment and manual chart abstraction have, so far, proved ineffective in addressing the full range of such complexity: often poorly scalable, their precision to find early-at-risk patients for targeted interventions is too low. Artificial intelligence has emerged as a transformative force in healthcare, offering unparalleled capabilities for data processing and the ability to find patterns often invisible to human analysis. Healthcare providers will be able to predict the risk of readmission more precisely by using machine learning and other AI techniques, thus enabling targeted interventions and personalized care. This will improve patient outcomes and could greatly reduce healthcare costs by avoiding preventable readmissions. The potential of AI here is indicative of broader needs in health to better the care of patients, be judicious about the usage of resources, and move towards a model for health that's sustainable in these cases of chronic diseases. The predictions by AI will finally yield useful insights with which healthcare workers can assist hospitals in reconsidering their policies for post-discharge follow-up care and follow-up on their patients (Wang, 2022).

Research Objective

The key goal of this research is to devise and ameliorate AI models that can effectively predict patient readmissions. Through the power of machine learning and data analytics, this study will work toward developing tools that will highlight patients at a high risk of readmission, which can then be targeted with interventions by healthcare providers. Additionally, the research shall seek to provide actionable insights that will inform clinical decisions and further improve the quality of patient care. The following study integrates various data sources, including clinical data, patient history, and social determinants, to build better models that have their origins in the deep understanding of the needs of the patients.

Scope of this Research

This research will focus on readmission to hospitals, particularly in the USA because of various medical conditions, keeping in mind that different illnesses pose different challenges and risk factors. It also involves the analysis of information from various sources such as EHRs, insurance claims, and patient surveys so that the predictive models can be much more robust. It considers a wide range of variables, from clinical data of lab results or medication adherence to socioeconomic factors such as income and education level, to help capture the many facets of health outcomes. The ultimate aim is a framework that goes beyond prediction in the realms of readmissions toward reduction strategies that further improve health care across the United States.

II. Literature Review

Hospital Readmissions

According to Abisheganaden et al. (2023), readmissions to a hospital are among the very foundational issues of modern US healthcare, representing the meaning of a patient's admission to a hospital within a predetermined time, usually within 30 days after discharge. It has come to be an important pointer to the quality of provided care and evidence of fissures in post-discharge planning and management of patients. The issue of readmission can never be overestimated regarding the gravity of addressing it, since readmission is associated with poor clinical outcomes, increased morbidity, and a huge burden of health expenditure. Issues relating to the management of hospital readmissions are not singular. First, there is usually a lack of infrastructure on the part of health systems for monitoring patients incessantly post-discharge. Secondly, readmissions depend upon many factors, such as socioeconomic status, access to follow-up care, and adherence to their treatment schedules (Huang et al., 2021). These are the challenges that require comprehensive strategies regarding the prediction and prevention of readmissions effectively.

As per Croon et al. (2022), Hospital readmissions transcend beyond issues of individual patients as they reflect broader systemic problems in the healthcare framework. In this context, readmission rates serve as an effective indicator of quality for health institutions and are used as a basis for rating the hospitals and determining the extent of reimbursement from payers such as Medicare. Such facilities could be penalized and thus would be under even greater scrutiny for strategies aimed at reducing such incidents. Current challenges in managing readmissions are multifaceted and include poor discharge planning, poor patient education, and lack of coordination among care providers. Many patients leave the hospital with unclear instructions regarding medication management or follow-up appointments, leading to confusion and, ultimately, readmission. Besides, the social determinants of health-access transportation, housing stability, and social support play a very important role in whether a patient can adhere to a discharge plan. Overall, addressing these challenges will be important to improving patient outcomes and reducing the burden of readmissions on the healthcare system (Davis et al., 2022).

Traditional Methods

Traditionally, predictions of hospital readmission have relied on traditional clinical methods such as risk-scoring systems and manual chart reviews. Most of these methods use a very limited set of variables in estimating the likelihood of a patient's readmission, such as age, comorbidities, and length of stay. Although these methods are very useful, they are bound by the static nature of data and an inability to capture the dynamic nature of changes in the health status of patients (Hsu et al., 2022). Traditional approaches had serious limitations concerning scalability and accuracy. Resource-intensive manual processes are prone to human error, while simplistic risk scores often fail to capture the complex interplay of clinical, behavioral, and social factors at play in driving readmissions. These limitations further underscore the need for AI-driven approaches that can leverage a lot of data and provide nuanced insights (Amritphale, 2021).

As per Masum et al. (2022), the major limitation of these traditional methods of readmission prediction is their reliance on static data points to drive their predictions, which may not always catch the real-time dynamic nature of patient health. Most of the models also consider a very narrow range of variables and lack representations of the broader contextual features relative to a patient's social and environmental circumstances. Approaches using traditional methods might oversimplify complex interactions among many risk factors. As a result of this, many patients would not be identified who deserve every target given for interventions. This mismatch creates an urgent need to replace the simple analyses that have been developed so far with more sophisticated ones, including and driven by AI techniques processing large datasets and yielding detailed understandings of readmission risk. It will open up a whole range of possibilities whereby these sorts of methods will enable the integration of diverse data sources, such as EHRs, patient-reported outcomes, and social determinants of health, into the practice of healthcare to improve quality while reducing hospitalizations (Romero-Bruau et al. 2020).

Machine Learning in Healthcare

Mohanty et al. (2022), reported that Machine Learning is a revolutionary force in healthcare, providing a wide array of innovative applications that help improve diagnostic precision, make treatment personalized, and smoothen operational efficiency. The key applications of ML in health care include predictive analytics for patient outcomes, image recognition for radiology, natural language processing for clinical documentation, and personalized medicine tailored with genetic information. ML algorithms can analyze large datasets, like those in EHRs or even genomic sequences, and identify patterns and insights that might otherwise remain hidden using traditional approaches; this enables better clinical decisions.

Stachel (2020), argued that when comparing machine learning techniques with traditional clinical models, there are several distinctions apparent. Traditional clinical models often rely on linear relationships and predefined rules derived from expert knowledge, limiting their ability to adapt to complex nonlinear interactions among variables. This normally uses smaller sets of data and may be bound by human biases in the assumptions they make. Machine learning models can process massive volumes of big data to find intricate relationships implicitly without programming for the analysis; hence, there is greater flexibility and adaptability (Wang et al., 2022).

Salunkhe et al. (2022), indicated that other Machine Learning techniques also include decision trees, support vector machines, and deep learning, which can learn themselves from continuously changing data, with an increase in their predictive power. This capability is especially valuable in dynamic healthcare environments where the demographics are changing and also disease patterns alter. Although traditional models dwell on providing insights into a historical nature, machine learning allows for real-time decision-making that would work to improve the outcomes for patients by improving the efficiency of the delivery of care. The integration of machine learning will lie at the heart of future developments in patient care as the health sector rapidly embraces digital transformation (Saati, 2022).

Key Indicators of Readmission Risk

Clinical Indicators

Sarker (2023), asserted that the identification of patients who are at risk for hospital readmission is important and enables the effective implementation of interventions to improve overall healthcare outcomes. Several clinical indications have been identified as significant predictors of readmission risk; most of them are based on patient demographics, clinical diagnosis, and health-related factors. Understanding these indicators will place healthcare providers in a better position to allocate resources more appropriately and develop better strategies for the management of patients. The commonly used clinical prediction factors of readmission risk include demographic factors like age and gender, clinical characteristics regarding the presence of chronic diseases, and those relative to the index hospitalization. For example, older adults are more susceptible to readmission due to multiple comorbidities and physiological changes related to aging. Other chronic medical conditions include chronic heart failure, COPD, or diabetes, all of which are strong predictors because ongoing management is needed, at the failure of which one complication after another complication may ensue shortly. Other clinical indicators are the severity of the initial admission represented by length of stay and functional status at discharge. Teo et al. (2023), articulated that the longer the stays in hospitals, the graver one's health condition is, and therefore their likelihood of readmission may also increase. Furthermore, patients with a poor functional status upon discharge-usually characterized by limited mobility or inability to carry out activities of daily living-are most vulnerable to hospital readmission.

Croon (2022), contended that predicting hospital readmission requires the identification and analysis of major indicators that will show high risk. The clinical indicators include a patient's main diagnosis, comorbid conditions, medication adherence, and previous hospitalizations. Also, socio-economic factors like income, education level, and access to care are very critical in determining the likelihood of readmission. Recent literature has identified that the integration of traditional clinical indicators with newer data sources, such as patient-reported outcomes and social determinants of health, has resulted in improved performance. For example, the addition of patient mobility, home environment, and support systems has been shown to enhance predictive accuracy (Wang et al., 2022). These results point to the importance of comprehensive models that integrate a wide variety of data streams in capturing the complete spectrum of factors that drive readmissions.

III. Data Collection and Preprocessing

Dataset Description

The hospital readmission dataset used in this study comprised a comprehensive collection of patient-related data aimed at understanding and predicting readmissions. The dataset was thereby developed using electronic health records which capture all clinical activities - diagnosis code treatment history, results of labs, and medication-related prescriptions. Demographic details related to patients will include: age, sex, and ethnic background - for contextualizing at the population level. This clinical information was complemented by unstructured data, such as clinical notes that give further detailed insight into patient conditions and advice on follow-up care. The socioeconomic variables of income level, education, and access to follow-up care further enhance this dataset in ensuring that the holistic approach of modeling readmission risks is considered.

Data Preprocessing

The computed code snippet executed the pre-processing of data, which first started with the removal of irrelevant columns. Secondly, it is followed by the division of the data into features and the target variable. Thirdly, it separated numerical and categorical columns, creating a different pipeline for each of these types. Fourthly it then used a Standard-Scaler in the numeric pipeline to normalize the scale and a One-Hot-Encoder in the categorical one to turn the categorical variables into numeric format. Fifth, the column transformer was used to apply these pipelines to respective columns. Finally, splitting the data into training and test sets, applying preprocessing pipelines to both sets and turning back the preprocessed data into Data Frames for further analysis and modeling will be done. This ensures the data is in an appropriate format for the machine learning algorithms, as most of them require numeric input.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis formed the basis for most of the understanding that one would want to know in a dataset before the application of predictive models. This procedure included summarizing key statistics and understanding the distribution of data and anomalies or missing values. Essentially, EDA allows the analyst to discover the relationships between variables, detect the presence of outliers, and assess the quality of the data. In the context of hospital readmission, EDA described the trends in patient demographics and common risk factors; it also underlined the most important correlations between clinical indicators and readmission rates. Moreover, EDA provided extensive insight into the dataset, refined data preprocessing steps, selected features for machine learning algorithms, and increased the accuracy of predictions.

Age Distribution of Patients

The code snippet in Python was implemented to generate a histogram that can plot the age distribution of patients in a dataset. First, necessary libraries were imported: matplotlib.pyplot for plotting and seaborn for enhanced statistical graphics. Then, it sets up a figure with specified dimensions and uses Seaborn's histogram plot function to plot the age distribution. Here, 'KDE' is set to True to plot a kernel density estimation curve over the histogram for a smooth look at the data. Proper title and labels on this plot were set, adding gridlines to the y-axis and showing this histogram:

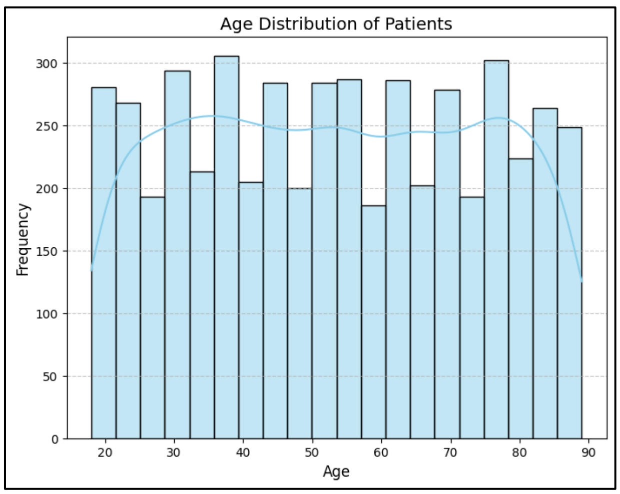

Figure 1: Exhibits Age Distribution of Patients

The above histogram is the distribution of the ages of the patients. The above histogram shows that the distribution is pretty uniform for most age groups, with a slight peak in the 30-40 age bracket. This observation is further supported by the overlaid KDE curve, showing a gradual increase in frequency up to the mid-30s and then a slow decline. The presence of gridlines and clear labels enhances the readability of the histogram. The overall distribution reflects that the patient population is relatively diverse in age, with no specific concentration in a particular age bracket.

Readmission Rate by Gender

The Python code script was used to visualize the readmission rate by gender for the provided dataset. The code started with the importation of libraries, mainly matplotlib.pyplot for plotting and seaborn for enhanced statistical graphics. Further, the code proceeds to create a figure of certain dimensions, and utilizing Seaborn's count plot functionality, it created a visualization that represents the count of patients according to gender while using different colors for their readmission status:

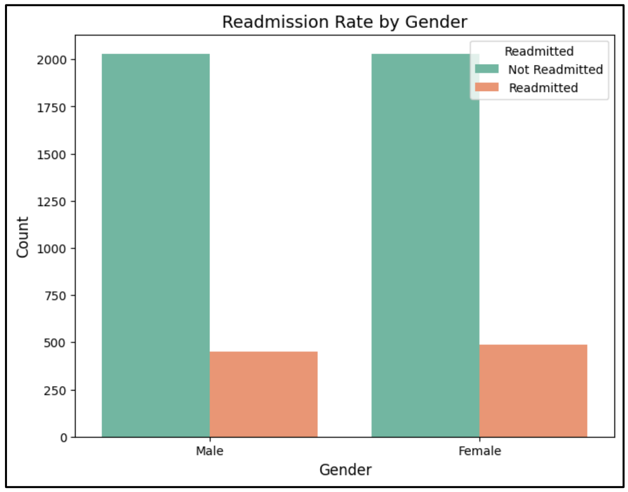

Figure 2: Visualizes Readmission Rate by Gender

The bar plot depicts the readmission rates for male and female patients. There are two bars for each gender: one for those patients not readmitted and one for those who were readmitted. From the plot, it is observable that both males and females have a higher count of patients who were not readmitted compared to those who were readmitted. More precisely, for males, about 2000 patients were not readmitted, while about 500 were readmitted. For females, the numbers are slightly higher with about 2000 not readmitted and about 500 readmitted. The plot suggests that there is a similar proportion of patients who are readmitted between males and females.

Top 10 Primary Diagnoses

This Python code fragment develops a bar plot to identify the top 10 primary diagnoses depending on the whole dataset. This code first imports necessary libraries; among them is matplotlib. pyplot, which is called for plot functionality, and seaborn for statistical graphics. Further, the code develops a figure with parameters and uses a seaborn bar plot function to build a horizontal bar plot where the y-axis represents the top 10 primary diagnoses and the x-axis the frequency. Here, the palette parameter specifies the color for the plot. Finally, it provides appropriate titles and labels for the plot and returns the resulting bar plot.

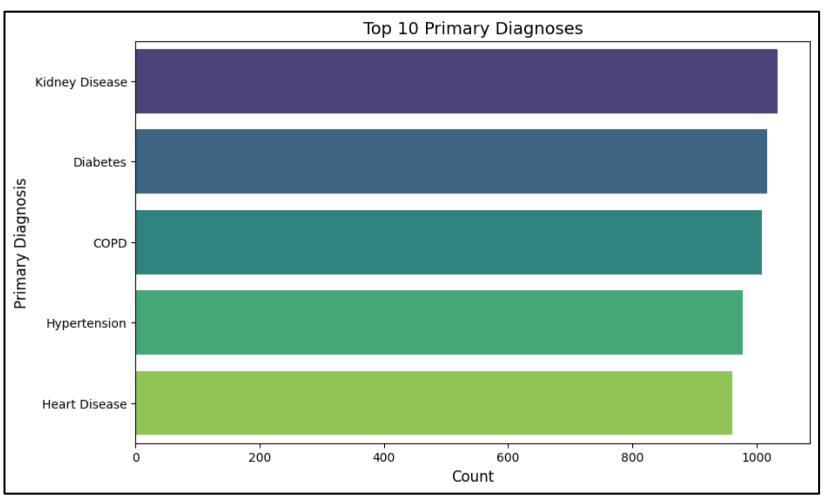

Figure 3: Illustrates Top 10 Primary Diagnoses

This histogram shows the top 10 primary diagnoses in patients and brings high-intensity health concerns to the data. Precisely, "Kidney Disease" has the highest frequency and is almost close to 1,000 diagnoses, making this a critical diagnosis that healthcare should act upon. Coming close in diagnosis are "Diabetes" and "COPD" conditions, both indicating high counts and, thus, important pathological conditions to also be attended to. "Hypertension" and "Heart Disease" further explain the importance of cardiovascular health, as they also appear with very high incidence rates at the level of the hospital. This distribution underlines a targeted health approach in conditions management through a prioritization strategy focused on prevention with its effective management to have a low readmission rate and improved results for the patients.

Days in Hospital vs. Readmission

The Provided Python code snippet was executed for constructing a boxplot showing the relationship between Days In Hospital by readmission status, which is produced by importing the common library, matplotlib. pyplot to create plotting functionalities and another statistical graphics Library named seaborn that is built on matplotlib. It then creates a figure with the specified width and height, using Seaborn's boxplot to create a boxplot, where the x-axis gets the readmission status 0 or 1, respectively readmitted/readmitted, while the y-axis shows the days in the hospital. Here, 'palette' is utilized to define the color for the plot. Finally, this code fragment provided the appropriate plot title and labels, after which it showcased the resultant boxplot.

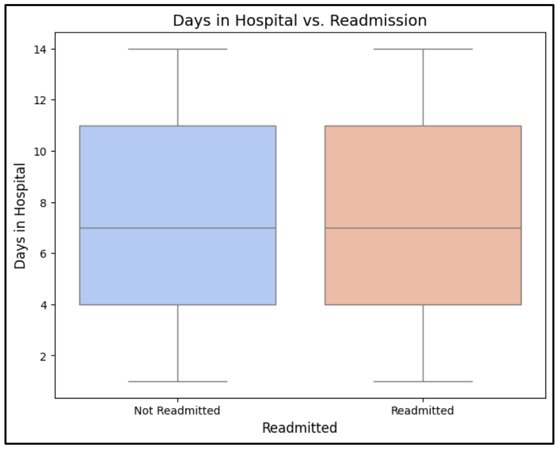

Figure 4: Depicts Days in Hospital vs. Readmission

The above boxplot illustrates the relationship between days in the hospital and readmission status. The x-axis represents the readmission status, where 0 = not readmitted and 1 = readmitted, and the y-axis represents the days in hospital. From the boxplot, one can observe that the readmitted patients stayed longer in the hospital compared to the ones who were not readmitted. It is observed that the median length of stay for readmitted patients is higher compared to the non-readmitted patients, and the IQR for readmitted patients is wider, indicating greater variability in the length of stay within this group. This plot suggests that length of hospital stay may be a factor of readmission risk.

Comorbidity Score Distribution by Readmission Status

The code snippet in Python created a kernel density plot to depict the distribution of comorbidity scores for patients who were not readmitted versus those who were readmitted. Import necessary libraries: matplotlib.pyplot for plotting and seaborn for enhanced statistical graphics. It created a figure with specified dimensions, and then Seaborn's kdeplot was used to plot the KDE for two groups while assigning color and label parameters to each group. The shading parameter was set to True so that the area beneath the curves of the kernel density estimate is filled out, thus viewed a bit better. The code finally sets proper titles and labels for the plot, adds a legend that allows distinguishing between the two groups, and displays the resulting KDE plot:

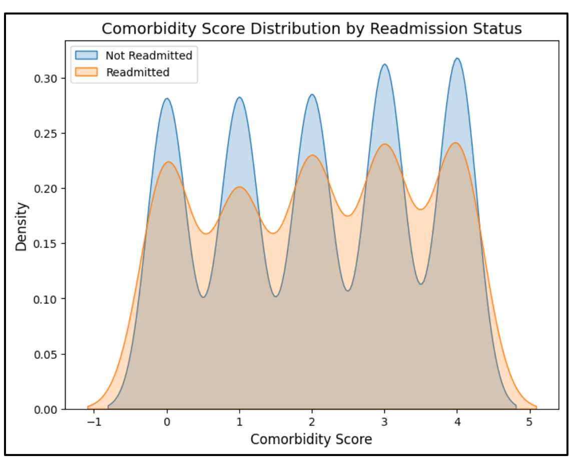

Figure 5: Portrays Comorbidity Score Distribution by Readmission

The histogram displays the distribution of the comorbidity score by readmission status. It is evident from the pattern that there is a difference between readmitted and not-readmitted patients. In the "Not Readmitted" patients, the distribution is higher across the low values of the comorbidity score, peaking around 0 and 1, suggesting that the lesser the number of comorbidities, the less likely a patient will be readmitted. In contrast, the "Readmitted" group represented in orange has a wider distribution with higher peaks at scores of 2 and 3, indicating that medium-level comorbidity patients are more likely to be readmitted. Therefore, from this analysis, there is indeed an association between levels of comorbidity and readmission, targeted intervention might be required for those patients who present higher scores of comorbidity to reduce possible readmission. The overlapping densities also suggest that though comorbidity is a significant factor, other variables may influence readmission status.

Distribution of Discharge Destination

The Python code snippet was utilized to create a bar plot of the distribution of discharge destinations in the provided dataset. It first imported necessary libraries such as matplotlib.pyplot for plotting and seaborn for improved statistical graphs. A figure was created with specified dimensions, and then Seaborn's barplot function was utilized to create a bar plot such that the x-axis represents the discharge destinations, while the y-axis represents the frequency. The palette parameter sets color variably for the plot. Finally, code-appropriate titles and labels were set up for this plot, rotated the x-axis labels for better readability, and then displayed the resulting bar plot:

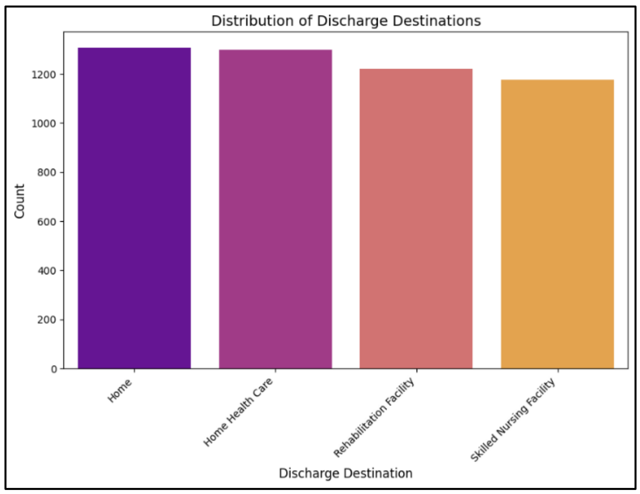

Figure 6: Showcases Distribution of Discharge Destinations

This histogram represents the distribution of discharge destinations for patients, such as Home, Home Health Care, Rehabilitation Facility, and Skilled Nursing Facility. All those categories are represented reasonably, with counts above 800 discharges; each category has more than 800 discharges, showing the variety of post-discharge options utilized by the patients. Also notice that "Home" and "Skilled Nursing Facility" are much higher, with over 1,200 discharges, indicating these are the most prominent pathways for patients post-hospitalization. High rates of discharge to "Home Health Care" and "Rehabilitation Facilities" indicate a high need for continuing care and rehabilitation services post-discharge. This distribution underlines how important it is to learn the needs and preferences of the patients concerning discharge plans and also how resource investment in such a care setting has the potential to be helpful in recovery and continuity of care.

IV. Methodology

Feature Engineering

Feature engineering is one of the major steps in data preprocessing, particularly in a hospital readmission dataset. The raw data was transformed into a meaningful feature that will enhance the predictive power of machine learning models in this step. Interaction features are some of the effective techniques that capture the relationship between different variables. For example, combining age and comorbidities can yield insights into how these factors jointly influence readmission risk. Temporal features like the time since the last admission or lengths of stay can also be engineered from date fields for finding temporal patterns. Other techniques include the normalization/standardization of continuous variables; which ensures that each such variable contributes equally to a model, and the encoding of categorical variables by methods like one-hot encoding or label encoding to make the data usable for algorithms that take data in numerical form.

The selection of relevant features was also fundamental and was guided by both clinical knowledge and the characteristics of the dataset. The interaction with healthcare professionals provided insights on clinically relevant variables: history of previous admissions, compliance to medication regimen, and social determinants like socioeconomic status. This outcome, therefore, was achieved through statistical methods involving correlation analysis and recursive feature elimination, which identified those features most predictive by eliminating those features that are not meaningful in contributing to the model. In addition, some techniques like the ranking of feature importance from tree-based models have indicated which features drive the predictions. This integration of clinical insights into data-driven approaches made the ensuing features robust, relevant, and therefore far more powerful in increasing a model's ability to correctly predict hospital readmission.

Model Selection and Development

Credible algorithms in machine learning were selected for the analysis of the dataset of hospital readmission to appraise their performance in predicting readmission. Logistic Regression is a baseline model because the features and classes are usually interpretable, simple, and effective in binary classification problems. Random Forest was also considered, as it represents one of the key ensemble methods, built on a combination of decision trees. It uses averaging to avoid overfitting and improve predictive accuracy and robustness. Gradient Boosting, another ensemble technique, builds models greedily, optimizing the errors of previously built models and probably improving performance.

The choice of these algorithms was motivated by several factors specific to the hospital readmission context. Logistic regression was preferred for its simplicity and interpretability, which is quite practical for healthcare providers to comprehend how each feature affects readmission. Random Forest overcomes some of the limitations of decision trees by the combination of results across the individual trees, thereby making it rather robust and reliable. This is very important in the medical field, where false positives and negatives are very costly. Gradient Boosting was opted for because of its high performance, especially when the data contains complex feature interactions, which improve predictive accuracy for the identification of high-risk patients.

Training and Evaluation

The data was divided into training and testing datasets, usually in an 80/20 or 70/30 proportion, to ensure good model training. This split allowed one part of the data subset to train the model while using another, unseen subset for assessing its performance, to avoid overfitting. Cross-validation methods, including k-fold cross-validation, are used in enhancing model robustness; this includes further dividing the training data into k subsets, training the model k times, each time with a different subset used for validation, and the rest used for training. This procedure was more realistic not only to estimate the model performance but also to help choose among several candidates for the best model. Some of the key metrics used to quantify the model's effectiveness included accuracy, precision, recall, F1-score, and ROC-AUC. While accuracy gives a general feel for how correct the predictions are, precision and recall give more fine-grained information regarding the positive cases identified by the model. The F1-score weighs precision and recall, while ROC-AUC summarizes the trade-off between true positive rates and false positive rates, hence giving a broader overview of the predictive powers of a model.

Hyperparameter Tuning

Tuning hyperparameters is one of the most crucial steps toward model performance optimization because this usually modifies the outside settings by which machine learning algorithms work while being trained. Deployed methods of hyperparameter optimization including grid search and random search. Grid search works to perform a systematic search in a range of pre-specified values for each hyperparameter against other possible hyperparameter values to obtain an optimal combination of the same. Though highly systematic and comprehensive, the method had ground-level computationally costly approaches in which all the combinations under consideration are studied over their certain finite space. On the other hand, random search sampled a portion of the hyperparameter space and can be an efficient way out with competitive results. Techniques such as Bayesian optimization and genetic algorithms were also considered advanced ways to perform hyperparameter tuning in a way that better sets of parameters are found while keeping computational costs lower. This hyperparameter enhanced accuracy, increased generalization capability, and improved performance on unseen data.

V. Model Evaluation and Comparison

Performance Metrics

a) Logistic Regression Modelling

Logistic Regression Modeling in Python for Binary Classification. The code snippet performed a Logistic Regression model in Python for binary classification. First of all, it imported the libraries that may be used during model building and evaluation. Then, it instantiated a Logistic Regression model, including a random state for the model reproducibility. Finally, it trained the model on the preprocessed training data and respective target variable y_train. The model performed predictions on the preprocessed test data X_test_preprocessed using the fitted model. Finally, using different metrics of performance that included accuracy, confusion matrix, and classification report, a check is provided about the performance on how well or accurately and precisely the model does in respect to recall or F1-score:

Output:

Table 1: Portrays the Logistic Regression Results

```
Logistic Regression Model Accuracy: 0.501
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.50      0.62       812
           1       0.19      0.49      0.27       188

    accuracy                           0.50      1000
   macro avg       0.50      0.50      0.45      1000
weighted avg       0.69      0.50      0.55      1000
```

The above table shows the performance metrics of the logistic regression model. Overall, the model correctly predicts the outcome with an accuracy of 0.501, which is 50.1%. The classification report gives a more detailed breakdown of the model's performance. The precision regarding class 0 is 0.81, which means 81% of the instances predicted as class 0 are class 0. Recall for class 0 is 0.50; that is, 50% of actual class 0 instances were correctly identified. The f1-score for class 0 is 0.62, being the balance of precision and recall. Class 1 has a precision of 0.19, a recall of 0.49, and an f1-score of 0.27. Both the macro average and weighted average give the result across the two classes; the former giving equal weight to each class, the latter giving more weight to the largest class. This gives a low score for macro average and weighted average in comparison with accuracy, indicating that the model is imbalanced across different classes in performance.

b) Random Forest Modelling

The Random Forest Classifier model in Python started by importing the necessary library Random Forest Classifier from sklearn. Ensemble. Afterward, it instantiated a Random Forest model with a random state for reproducibility. Then, the model was trained on preprocessed training data and the target variable. The trained model was used to make predictions on the preprocessed test data. In the end, the performance of the model was evaluated on various matrices such as accuracy, a confusion matrix, and classification reports, which can give information about the model's accuracy, precision, recall, and F1 score:

Output:

Table 2: Random Forest Results

```
Random Forest Model Accuracy: 0.803
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.99      0.89       812
           1       0.15      0.01      0.02       188

    accuracy                           0.80      1000
   macro avg       0.48      0.50      0.46      1000
weighted avg       0.69      0.80      0.73      1000
```

The above table represents some performance metrics of the Random Forest model. The overall accuracy of the model is 0.803, which means that it correctly predicts the outcome in 80.3% of cases. The classification report gives further detail concerning the performance of the model. Precision is 0.81 for class 0-for instances, 81 % have been correctly predicted by the model when it was an actual predicated case of class 0. Whereas recall for Class 0 0.99 means 99 percent times the model finds the correct original data whose class was supposed to be ' 0'. Precision for class 1 is 0.15, recall 0.01 and f1 0.02. The macro and weighted average measures give the overall performances across both classes. The macro average is based on taking each class as equally important, while the weighted average takes a more weighted consideration towards the larger class. For this case, lower scores in macro average and weighted average compared to accuracy could reflect an imbalance in model performance across classes.

c) XGB-Classifier Modelling

The Python code snippet implemented the XG-Boost Classifier model. The code began with importing the needed library XGB-Classifier from the xg-boost package. An instance of an XG-Boost model was created using the random state for reproducibility, the use-label-encoder parameter is set to False, and eval_metric to log-loss. Then it went ahead with the training on the preprocessed train data – X-train-preprocessed, with its respective target variable y-train. At last, the preprocessed test data, X_test_preprocessed, is used for prediction with the trained model. The performance of this model is evaluated using different metrics: accuracy, confusion matrix, and classification report, which allow us to gauge the performance based on accuracy, precision, recall, and F1-score.

Table 3: Portrays Gradient Boosting Model Results

```
Gradient Boosting Model Accuracy: 0.811
Classification Report:
              precision    recall  f1-score   support

           0       0.81      1.00      0.90       812
           1       0.00      0.00      0.00       188

    accuracy                           0.81      1000
   macro avg       0.41      0.50      0.45      1000
weighted avg       0.66      0.81      0.73      1000
```

The overall table presents the performance measures of the Gradient Boosting model. The overall accuracy of the model is 0.811, which can be interpreted as an 81.1% chance that the model will predict the correct outcome. A more detailed report is given by the classification report for the model's performance. Precision for the Class 0 is 0.81. This implies that 81% of the instances the model predicted as a class 0 case are class 0. Recall for a class 0 is 1.00, implying 100% of instances of actually being class 0 were correctly detected by the classifier. The f1-score regarding class 0, giving both precision and recall balanced importance, is 0.90, while class 1 has 0.00 regarding precision and recall and regards the value of the f1-score. The macro average and weighted average metrics summarize performance across both classes. The macro average gives equal weight to each class, whereas the weighted average assigns a higher weight to the larger class. The macro average and weighted average scores are lower than the accuracy in this case, which may indicate that the model is imbalanced across classes.

Comparison of All Models

The code in Python compared the performances of the Algorithms of Logistic Regression, Random Forest, and XG-Boost; it provided a dictionary that held the models' names and their corresponding accuracy results using the function accuracy score with the library sklearn. The dictionary fed the data into a pandas Data Frame for the sake of better visualization. Lastly, the code plotted a bar plot using seaborn for clear comparisons of the accuracy of each model. It was customized with a title and labels and colored appropriately for better readability.

Output:

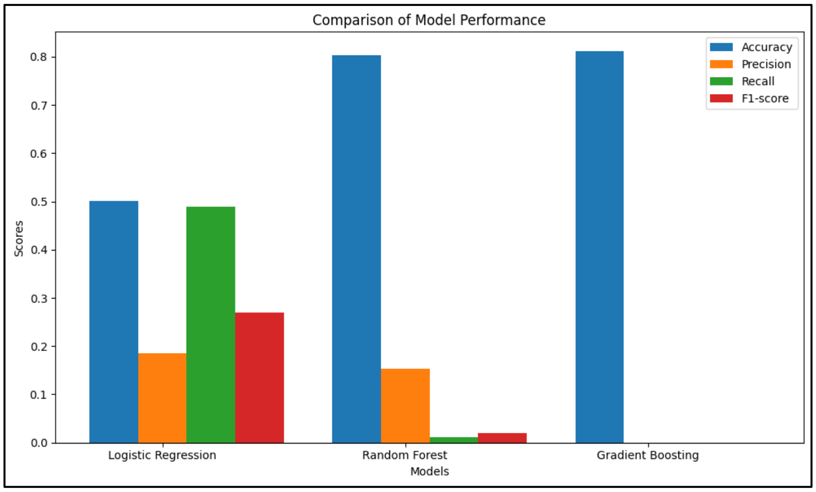

Fig 7: Depicts Model Accuracy Comparison

The table above compares and contrasts the performance of three machine learning algorithms—Logistic Regression, Random Forest, and Gradient Boosting—across four key metrics: Accuracy, Precision, Recall, and F1-score. Among the three models, Gradient Boosting had the highest scores on all four metrics and maximum accuracy and F1-score, showing the best all-rounded performance in prediction. Random Forest has presented moderated performance, especially concerning Precision and Recall, which allows it to do an effective job of identifying true positives while maintaining a good balance between false positives and negatives. Simultaneously, Logistic Regression is the worst among all metrics, meaning it may not work for this data set. The study showed the superiority of Gradient Boosting for the given task; therefore, model selection should take into consideration such metrics for optimal prediction performance.

Validation Techniques

Some of the most proven techniques for checking the robustness and generalizability of machine learning models are validation techniques. The most helpful way to perform this was through adopting a method called k-fold cross-validation. In this protocol, the dataset was divided into k subsets or folds. A model will be trained on k-1 folds and validated on the remaining one. This process was repeated k times; that is, train and validate the model k times, each time on a different subset. The risk of overfitting decreases due to the model's performance was estimated across the more comprehensive segments. Cross-validation also used stratification techniques at the time of splitting to keep target class distribution balanced across different folds. This procedure helped separate cases of class imbalance. Depending on dataset size and particular requirements for modeling, the application of other validation methods, such as LOOCV or bootstrapping, may enormously improve the reliability of results.

Evaluating algorithm consistency and stability across different datasets was equally paramount for ensuring that the model can perform well in real-world scenarios. This process was done through methods such as external validation, where the model was tested on an entirely separate dataset that had not been used either in training or in the initial validation. In this way, its performance was evaluated more objectively, helping to find potential biases and weaknesses. Besides, the performance metrics of the model, such as accuracy, precision, recall, and F1-score on different subsets or external datasets, indicated its robustness. Another way to check stability was by sensitivity analysis, which examined the change in predictions of the model with changes in input data. This holistic validation of the model not only ascertained the correctness of the model but also instilled confidence in its applicability across varied contexts, leading to better decision-making with improved outcomes in practical applications.

VI. Insights for Healthcare Providers

Predictive Insights

Interpreting healthcare model outputs provides insightful predictions to inform clinical decisions. For instance, machine learning models can detect high-risk patients for complication or readmission and thus allow providers to take necessary precautions to prevent such eventualities. Health professionals can therefore make a discriminated analysis of the drivers of patient outcomes based on socio-economic status, pre-existing conditions, and treatment adherence by carefully looking into the model outputs. This understanding can be taken one step further through scenario analysis, which can simulate different patient segments to demonstrate how various interventions might affect such specific groups. For instance, a model could show that the risk of readmission is higher for patients who are elderly and have chronic illnesses. Providers can consider strategies for better management of these patients to optimize resource utilization and improve patient outcomes.

Patient Segmentation Analysis

Patient segmentation analysis is one of the most important tools that could help in identifying high-risk patient groups, thus enabling healthcare providers to concentrate their resources in the required direction. Using clustering techniques for patients' data, providers will, therefore, be able to group them into distinct segments according to various attributes, such as age, medical history, and treatment responses. For example, the segment may become overweight and diabetic among the younger adult population, a cohort that absolutely would benefit most from targeted interventions. Understanding those at high risk allows for tailor-made care planning, ensuring coverage in support and monitoring needed among the vulnerable populations. This addresses health policy and resource distribution concerning health care, aiming to improve overall care quality.

Personalized Care Strategies

Care strategies have to be developed based on predictive insights and patient segmentation analysis to enhance the outcomes of patients. Health providers design targeted interventions using the insights from models that are well-fitted to the needs of each particular group of patients. Examples include structured follow-ups for high-risk patients that are identified through analytics, including regular check-ins, consultations via telehealth, and lifestyle modification programs. These personalized strategies can increase the likelihood of patient involvement in their care and treatment plans and, by extension, improve health outcomes. Monitoring the effectiveness of these interventions allows for the establishment of a system that encourages continued refinement of the care plans to keep them current and relevant.

Operationalizing AI-Driven Insights

AI-driven insights will thus require a strategic approach to the integration of AI-driven models in the functioning of the hospital. First is the need to constitute a cross-functional team consisting of data scientists, clinicians, and operational leaders who would guide this process. Such a team will have to study the present infrastructure and then come out with the need for technological upgrades on EHR systems and advanced data analytics platforms. With the foundational elements set, a framework for real-time prediction of readmission and further interventions would be possible. This will include protocols for data collection, analyses, and interpretation that will help the healthcare providers easily understand it at the point of care. These AI-driven insights, after being operationalized by the hospitals, will help further enhance the existing capability of correct estimation of the patient outcomes, workflow, and overall performance of care, thereby leading to better quality, improvement in patient experience, and value creation from available healthcare resources.

VII Case Study: Implementing AI-Driven Predictive Modeling within a Specific US Hospital

Background

The problem of readmission has posed a big challenge to many hospitals in the United States during recent years and adds to healthcare costs while straining hospital resources. Corewell Health has been at the frontline in solving such challenges using AI-driven predictive modeling. Corewell Health operates several hospitals and outpatient facilities within its system. However, large as the healthcare system it operates, realizes that readmission to hospitals can be pretty burdensome in affecting patients' outcomes, let alone being expensive for the health system itself. Leadership has identified a need to refresh traditional approaches to the management of post-hospital discharge patients to ensure innovation in patient monitoring and strategies for further interventions.

The major issue that was at stake was readmission within 30 days of discharge, particularly those patients with complex health needs. Most of these patients had several co-occurring issues with interrelated behavioral health, complicated clinical comorbidities, and social determinants of health clearly against them in recovery. Corewell Health aimed at developing more proactive patient care to lower the rates of readmission with the use of advanced analytics and machine learning techniques applied.

Implementation

To address the matter of readmissions, Corewell Health implemented an AI-powered predictive model designed to identify patients at high risk of returning to the hospital shortly after discharge. The initiative involved several sources of data analysis: electronic health records, previous histories, and demographic information. The predictive modeling system applies machine learning algorithms, in patterns and risk factors related to readmission. The implementation process involved the following steps:

Step 1: Data Integration-Corewell Health integrated high volumes of data from its EHR system to build a robust dataset that could be analyzed for predictive modeling.

Step 2: Machine Learning Model Development. The team created models using historical data to allow the machine learning algorithms to highlight patients who are at higher risk, based on variables such as prior admissions, chronic conditions, and social determinants like housing stability and access to care.

Step 3: Interdisciplinary Care Teams. Once the predictive model had identified potential candidates for readmission, interdisciplinary care teams were mobilized to address identified risks. The teams thus focused on three pillars: behavioral health support, management of the clinical challenge, and social determinants of health.

Step 4: Real-time Alerts. The predictive analytics tool offered real-time alerts to the healthcare providers in case any patient was flagged as high risk, thus enabling them to make timely interventions before complications could arise.

Results

The deployment of AI-powered predictive modeling at Corewell Health attained significant results in minimizing hospital readmissions. Particularly, in just one year of this deployment, the hospital reported the prevention of about 200 readmissions, thus saving an estimated $5 million associated with avoided hospital stays and treatments. The key outcomes of this case study are:

Readmission Rates: A corresponding decrease in 30-day readmission rates for high-risk patients because of targeted interventions.

Improved Patient Outcomes: Patients receiving proactive care management had better health outcomes and quality of life because the timely interventions address their unique needs.

Better Resource Utilization: Corewell Health was able to optimize resource utilization by identifying high-risk patients much in advance, ensuring that care teams focused their efforts where they were most needed.

VIII Discussion

Implications for the US Health Care System

The use of AI-powered readmission prediction models demonstrates great potential and advantages to the hospitals in the US healthcare system. Indeed, with the help of advanced analytics, healthcare professionals easily detect patients who have a high readmission risk and thus provide timely interventions that can prevent unnecessary hospital admissions. These predictive insights, beyond improving quality in patient care, may further have the potential to reduce health care costs by reducing the rates of readmission, which under certain federal programs are financially penalized. Those hospitals that will succeed in implementing these models may also see improved patient satisfaction and outcomes since targeted interventions may result in more personalized care and management of chronic conditions. In addition, fewer readmissions can ease the strain on facilities so that staff and other resources can be freed up for prevention programs and other essential services.

Retrospectively, in the implementation of these AI-driven models, the real hurdles lie in integrating this into established workflow systems or pipelines that may also be very demanding in staff training and a change in their approach/culture at work. Resistance to change may reduce adoption, and hence, such solutions are necessary: extensive training, establishing the potential for AI tools, and establishing an innovative environment. Further, the models are effective when the data is accurate and consistent across all systems. This can be partly solved by streamlining the collection process with collaboration from IT departments and ensuring a standardized protocol of some sort. Ultimately, overcoming these challenges requires an all-out effort by all the constituencies involved in the healthcare ecosystem: providers, administrators, and policymakers.

Ethical and Privacy Considerations

AI-driven insights raise numerous ethical and privacy considerations, especially as far as patient data are concerned. First is the matter of ensuring confidentiality for patients, followed by the ethics concerning sensitive health information. While security will be a primary factor given the ever-present risk of data breaches and unauthorized access to personal health information, this is where the healthcare organization would need to resort to rigid data governance policies describing the manner of collection, storage, and utilization of patient data. This means anonymizing data when possible and making it clear how that information may be used in predictive modeling. Ethical considerations also include issues of bias in the data themselves, which can perpetuate unequal treatment outcomes across demographic groups. Those biases would have to be mitigated through representative data if equity were to be maintained within the applications of AI.

Other than that, regulatory compliance is another important aspect of the implementation of AI models in healthcare. The organizations will have to be compliant with regulations such as the Health Insurance and Accountability Act of 1996, commonly known as HIPAA, which dictates the privacy and security of patient information. Best practices for compliance shall include periodic audits of data handling practices, training of staff regarding policy on privacy, and well-documented protocols on data sharing. Furthermore, the continuous development and evolution of AI technologies themselves will require that regulatory frameworks adjust to these changes in trying to address new challenges. This will also engage organizations in regulatory bodies for discussion on best practices to stay ahead in compliance while keeping ethical considerations atop in the deployment of AI.

Limitations

Despite the luring promises of AI-driven readmission prediction models, several limitations need to be acknowledged. Probably one of the main limitations relates to data quality since successful performance is highly dependent on the availability of good-quality and comprehensive data. Many times, this data may be incomplete, incoherent, or biased, influencing the performance and generalization capability of the models. Apart from this, healthcare data carries lots of contextual dependencies, which means a model trained on one population just cannot perform with equal skill if applied to a different patient group and setting. The above weakness puts a premium on further striving to develop better means of data collection and also ensuring that data should mirror the diversity in the patient population of any given healthcare provider.

For future research, the integration of real-time data streams, such as wearable technology and telehealth monitoring, may further improve predictive power. Understanding the impact of social determinants of health on readmission can provide a more nuanced understanding of the complex drivers of patient outcomes. Additionally, the research agenda would be channeled into the development of methods for routine updating and validation of the predictive models, so that they remain relevant as health practices and patient populations change. Overcoming these limitations and exploring new areas of research will help healthcare realize significant advances in the effectiveness and reliability of AI-driven insights that should improve patient care and, ultimately, health outcomes.

IX. Conclusion

The key objective of this research was to devise and ameliorate AI models that can effectively predict patient readmissions. Through the power of machine learning and data analytics, this study worked toward developing tools that will highlight patients at a high risk of readmission, which then can be targeted with interventions by healthcare providers. The hospital readmission dataset used in this study comprised a comprehensive collection of patient-related data aimed at understanding and predicting readmissions. The dataset was thereby developed using electronic health records which capture all clinical activities - diagnosis code treatment history, results of labs, and medication-related prescriptions. Demographic details related to patients will include: age, sex, and ethnic background - for contextualizing at the population level. This clinical information was complemented by unstructured data, such as clinical notes that give further detailed insight into patient conditions and advice on follow-up care. Several models were considered for classification tasks such as Random Forest Classifier, Logistic Regression, and XG-Boost Classifier. Some of the key metrics used to quantify the model's effectiveness included accuracy, precision, recall, F1-score, and ROC-AUC. Gradient Boosting had the highest scores on all four metrics and maximum accuracy and F1-score, showing the best all-rounded performance in prediction. Care strategies have to be developed based on predictive insights and patient segmentation analysis to enhance the outcomes of patients. AI-driven insights will thus require a strategic approach to the integration of AI-driven models in the functioning of the hospital.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Abisheganaden, J., Lee, K. H., Low, L. L., Shum, E., Goh, H. L., Ang, C. G. L., ... & Miller, S. M. (2023). Lessons learned from the hospital-to-home community care program in Singapore and the supporting AI multiple readmissions prediction model. *Health Care Science*, *2*(3), 153-163.

[2] Aluru, K. S. Transforming Healthcare: The Role of AI in Improving Patient Outcomes.

[3] Amritphale, A., Chatterjee, R., Chatterjee, S., Amritphale, N., Rahnavard, A., Awan, G. M., ... & Fonarow, G. C. (2021). Predictors of 30-day unplanned readmission after carotid artery stenting using artificial intelligence. *Advances in therapy*, *38*, 2954-2972.

[4] Croon, P. M., Selder, J. L., Allaart, C. P., Bleijendaal, H., Chamuleau, S. A. J., Hofstra, L., ... & Winter, M. M. (2022). The current state of artificial intelligence-based algorithms for hospital admission prediction in patients with heart failure: a scoping review. *European Heart Journal-Digital Health*, *3*(3), 415-425.

[5] Davis, S., Zhang, J., Lee, I., Rezaei, M., Greiner, R., McAlister, F. A., & Padwal, R. (2022). Effective hospital readmission prediction models using machine-learned features. BMC Health Services Research, 22(1), 1415.
[6] Farid, F., Bello, A., Ahamed, F., & Hossain, F. (2023). The roles of AI technologies in reducing hospital readmission for chronic diseases: a comprehensive analysis.
[7] Hilton, C. B., Milinovich, A., Felix, C., Vakharia, N., Crone, T., Donovan, C., ... & Nazha, A. (2020). Personalized predictions of patient outcomes during and after hospitalization using artificial intelligence. *NPJ digital medicine*, *3*(1), 51.
[8] Hsu, J. C., Wu, F. H., Lin, H. H., Lee, D. J., Chen, Y. F., & Lin, C. S. (2022). AI models for predicting readmission of pneumonia patients within 30 days after discharge. *Electronics*, *11*(5), 673.
[9] 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, 1-14.
[10] Liu, W., Stansbury, C., Singh, K., Ryan, A. M., Sukul, D., Mahmoudi, E., ... & Nallamothu, B. K. (2020). Predicting 30-day hospital readmissions using artificial neural networks with medical code embedding. *PloS one*, *15*(4), e0221606.
[11] Masum, S., Hopgood, A., Stefan, S., Flashman, K., & Khan, J. (2022). Data analytics and artificial intelligence in predicting length of stay, readmission, and mortality: a population-based study of surgical management of colorectal cancer. *Discover Oncology*, *13*(1), 11.
[12] Mohanty, S. D., Lekan, D., McCoy, T. P., Jenkins, M., & Manda, P. (2022). Machine learning for predicting readmission risk among the frail: Explainable AI for healthcare. Patterns, 3(1).
[13] Romero-Brufau, S., Wyatt, K. D., Boyum, P., Mickelson, M., Moore, M., & Cognetta-Rieke, C. (2020). Implementation of artificial intelligence-based clinical decision support to reduce hospital readmissions at a regional hospital. Applied clinical informatics, 11(04), 570-577.
[14] Saati, A. (2022). *An analysis of the effectiveness of a commercial hospital artificial intelligence system in reducing readmissions in high-risk medicare patients* (Doctoral dissertation, Rutgers University-School of Health Professions).
[15] Salunkhe, V., Chintha, V. R., Pamadi, V. N., Jain, A., & Goel, O. (2022). AI-Powered Solutions for Reducing Hospital Readmissions: A Case Study on AI-Driven Patient Engagement. *International Journal of Creative Research Thoughts*, *10*(12), 757-764.
[16] Stachel, A. (2020). Development of a hospital readmission reduction program for patients discharged to skilled nursing facilities: An application of artificial intelligence and machine learning techniques.
[17] Sarker, M. (2023). Assessing the Integration of AI Technologies in Enhancing Patient Care Delivery in US Hospitals. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 338-351.
[18] Teo, K., Yong, C. W., Chuah, J. H., Hum, Y. C., Tee, Y. K., Xia, K., & Lai, K. W. (2023). Current trends in readmission prediction: an overview of approaches. Arabian journal for science and engineering, 48(8), 11117-11134.
[19] Wang, L., Li, G., Ezeana, C. F., Ogunti, R., Puppala, M., He, T., ... & Wong, S. T. (2022). An AI-driven clinical care pathway to reduce 30-day readmission for chronic obstructive pulmonary disease (COPD) patients. *Scientific reports*, *12*(1), 20633.