PH240C Final Project
Predicting 30-Day Readmissions in Diabetic Patients Across Demographic Groups

Yuanyuan Gao, Kathy Li

May 13, 2025

Abstract

Hospital readmissions are a major challenge for healthcare systems, especially among patients with chronic conditions like diabetes. This project investigates which patient factors most strongly predict 30-day hospital readmission and how these predictors vary across demographic subgroups. Using a dataset of over 100,000 diabetic encounters, we trained and evaluated three models including logistic regression, XGBoost, and a feedforward neural network. XGBoost achieved the best overall performance and was selected as the final model. Feature importance analysis revealed that prior hospitalizations, emergency visits, and discharge settings were key predictors of readmission risk. Subgroup analyses uncovered demographic variations in predictive factors, emphasizing the need for future research to address structural disparities and improve equitable healthcare outcomes.

1 Introduction

Hospital readmissions within 30 days of discharge are widely regarded as a critical quality-of-care metric and a major driver of healthcare costs. Among patients with chronic illnesses such as diabetes, the risk of readmission is especially elevated, highlighting the need for accurate prediction models to guide clinical interventions. In this project, we aim to identify key demographic and clinical factors that predict 30-day hospital readmission among diabetic patients and investigate how these predictive factors vary across subgroups defined by race, gender, and age.

The dataset used in this project is the **Diabetes 130-US hospitals for years 1999–2008** dataset, compiled and analyzed by Strack et al. (2014) and publicly available through the UCI Machine Learning Repository (Strack et al., 2014; UCI Machine Learning Repository, 2008). The dataset contains more than 100,000 hospital admissions for diabetic patients across 130 U.S. hospitals, covering a wide range of demographic, clinical, and hospital-related variables. It includes patient demographics, diagnosis codes, laboratory results, medications, procedures, and hospital outcomes such as readmission status.

Our main scientific question is: ***Which patient factors most strongly predict 30-day hospital readmission among diabetic patients, and how do these predictive factors vary across race, gender, and age groups?***

The motivation for this work stems from two central considerations. First, early identification of patients at high risk of readmission can enable targeted clinical interventions, improving patient outcomes and reducing healthcare costs. Second, understanding whether predictors differ across demographic groups can inform more equitable healthcare delivery and guide clinical decision-making. Prior work by Strack et al. (2014) primarily focused on the association between HbA1c measurement and readmission risk. In contrast, our study expands the analysis by identifying a broader range of predictors and examining how their importance varies across subgroups.

To address our research questions, we apply a range of machine learning methodologies. We employ logistic regression as a baseline linear model due to its interpretability and well-established use in clinical prediction tasks. We also use tree-based ensemble model, specifically random forests and XGBoost, to capture nonlinear relationships and assess feature importance. Finally, we implement feedforward neural networks to model complex, high-dimensional patterns in the data. By comparing these modeling approaches, we aim to identify robust predictors of readmission across different algorithmic frameworks.

The remainder of this report is organized as follows. Section 2 describes the dataset, including subsetting procedures, feature selection, and formal problem formulation. Section 3 outlines the machine learning models and statistical methods used, along with their justification and tuning strategies. Section 4 presents the results of our real data analysis, including model evaluation and interpretation of key predictors. Finally, Section 5 concludes with a discussion of our findings and potential future directions.

2 Dataset Description

The dataset used in this study is the **Diabetes 130-US hospitals for years 1999–2008** dataset, sourced from the UCI Machine Learning Repository (UCI Machine Learning Repository, 2008) and originally analyzed by Strack et al. (2014) (Strack et al., 2014). It contains detailed clinical and administrative information from over 100,000 hospital admissions for patients with diabetes across 130 hospitals in the United States. The dataset includes demographic characteristics, diagnosis codes, laboratory test results, medication usage, procedures, and hospital outcomes such as readmission status.

To align the dataset with the objectives of our analysis, we performed several subsetting steps. We excluded patient encounters where the discharge disposition indicated death, as readmission is undefined for deceased patients. We also removed records with missing or unknown values for critical demographic variables including race, gender and age, since our analysis involves comparisons across these subgroups. Additionally, we filtered the dataset to retain only encounters with clearly documented 30-day readmission status. These steps ensure consistency in outcome definition and

comparability across demographic groups.

After subsetting, the final dataset contains approximately **100,011** patient encounters, each described by **46 features**. These features include demographic variables (race, gender, age), hospital admission and discharge details, clinical measurements (e.g., HbA1c levels), medication records, and diagnosis codes. Race is categorized into five groups (Caucasian, African American, Asian, Hispanic, and Other), gender is labeled as male or female, and age is binned into 10-year intervals. Some variables exhibit substantial missingness; for instance, weight is missing for approximately 97% of records and payer code is missing for about 40%. Features with extremely high missingness, such as `weight`, `payer_code`, and `medical_specialty`, were dropped. For other variables containing missing or unknown values denoted by `?`, we standardized them by replacing with the label `Unknown`. We also excluded patient records with discharge dispositions indicating death or hospice care, as well as those with invalid gender labels.

2.1 Scientific Objective and Mathematical Formulation

We formalize our scientific question as a binary classification task to predict the probability of hospital readmission within 30 days, conditional on patient features. ***Can we predict the probability that a patient will be readmitted to the hospital within 30 days, given their demographic and clinical features?***

Mathematically, we model the conditional probability of readmission as $P(Y = 1 \mid X)$, where $Y \in \{0, 1\}$ indicates whether a patient is readmitted within 30 days (1 = yes, 0 = no), and $X \in \mathbb{R}^p$ denotes the vector of observed patient features. Our objective is to learn a predictive function $f(X; \theta)$ that estimates $\hat{y}_i = f(X_i; \theta) \approx P(Y_i = 1 \mid X_i)$, by minimizing the binary cross-entropy loss:

$$\min_{\theta} \ \mathcal{L}(\theta) = -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right],$$

where y_i is the observed label and $\hat{y}_i$ is the predicted probability for patient i. To evaluate model performance, we consider four key metrics. First, we use the Area Under the Receiver Operating Characteristic Curve (AUC) to quantify the model's overall ability to discriminate between patients who are readmitted within 30 days and those who are not. To assess the stability and reliability of these AUC estimates, we compute 95% confidence intervals using bootstrap resampling. Additionally, we evaluate each model's practical classification performance using sensitivity (true positive rate) and specificity (true negative rate), both measured at a fixed probability threshold of 0.5.

We further examine whether predictive patterns vary across demographic subgroups. Specifically, we conduct subgroup-specific performance assessments and feature importance analyses by race, gender, and age to identify potential disparities in prediction.

3 Methodology

To address our question, we apply three predictive modeling approaches: logistic regression, random forests (implemented via XGBoost), and feedforward neural networks. Each model was selected to balance considerations of interpretability, flexibility, and predictive capacity within a healthcare data context.

3.1 Proposed Models and Rationale

Logistic regression is used as a baseline model due to its widespread use in clinical prediction studies and its interpretability. It models the log-odds of the outcome as a linear function of patient features, allowing direct interpretation of each variable's contribution to readmission risk. Given that many features in the dataset, such as demographic variables (race, gender, age), admission type, and discharge disposition, are categorical or ordinal, logistic regression offers a straightforward framework to quantify their effects. Furthermore, the interpretability of logistic regression is critical for one of our key research goals: identifying and explaining the relative importance of predictors across demographic groups.

Random forests implemented through XGBoost, are included to capture complex nonlinear relationships between patient characteristics and readmission risk. In contrast to logistic regression, random forests are capable of modeling interactions among features without requiring explicit specification. This is particularly advantageous in our dataset, where combinations of factors (e.g., age and comorbidities, or race and discharge disposition) may interact in non-additive ways. Additionally, XGBoost handles missing values internally and efficiently processes datasets with a mix of categorical and continuous features, which matches the structure of our dataset. The ability to extract feature importance scores from tree-based models also aligns with our objective of determining which factors contribute most significantly to readmission risk.

We also incorporate **Neural Network (MLP)** networks to explore whether modeling high-dimensional patterns can further improve prediction accuracy. Our dataset contains both clinical measurements (HbA1c levels, number of diagnoses) and administrative features (admission type, discharge disposition, time in hospital). Many of these variables may interact in ways that are not easily captured by linear or tree-based models. For example, the effect of time in hospital on readmission risk may vary depending on patient age or discharge destination. Neural networks is able to capture such interactions without requiring explicit specification. Although they are less interpretable compared to logistic regression or random forests, neural networks can identify subtle patterns associated with readmission risk. Additionally, the large size of our dataset is suitable for training neural networks without substantial risk of overfitting.

3.2 Statistical Methods and Implementation

For model training, we use an 80%-20% train-test split with 5-fold cross-validation on the training set to select hyperparameters. Hyperparameter tuning is performed with grid search to optimize

model performance. For the logistic regression model, we tune the regularization parameter C; for XGBoost, we tune learning rate, maximum tree depth, number of estimators, and subsample ratio; and for the neural network, we tune the number of hidden units, activation function, regularization strength, and initial learning rate. All models were optimized using binary cross-entropy loss, which is suitable for binary classification tasks. To address the class imbalance in our dataset, we applied model-specific balancing strategies: `class_weight='balanced'` for logistic regression, the `scale_pos_weight` parameter for XGBoost, and SMOTE (Synthetic Minority Over-sampling Technique) prior to training the neural network. Model performance was evaluated using four key metrics: **Area Under the Receiver Operating Characteristic Curve (AUC)** to assess discriminative ability; **95% Confidence Intervals (CIs)** for AUC computed via bootstrap resampling to assess statistical reliability; and **Sensitivity** and **Specificity** at a fixed threshold of 0.5 to evaluate practical classification performance in a clinical context.

3.3 Comparison to Prior Work

Compared to existing literature, our methodological framework expands on prior analyses by incorporating modern ensemble and deep learning methods alongside traditional logistic regression. In particular, Strack et al. (2014) focused primarily on logistic regression and standard decision tree models to predict hospital readmissions among diabe21wsaztic patients. Their study emphasized the role of HbA1c measurement and hospitalization history as key predictors but did not evaluate more advanced machine learning techniques such as boosted trees or neural networks. Subsequent studies, such as Futoma et al. (2015) and Rajkomar et al. (2018), have shown that ensemble methods and neural networks can substantially improve predictive performance in clinical datasets. Motivated by these developments, our method expands on traditional models by including XGBoost and feedforward neural networks, enabling a broader comparison of predictive modeling strategies and a more detailed investigation into how feature importance varies across demographic subgroups.

Regarding statistical validity, logistic regression has well-established properties: under correct model specification, the maximum likelihood estimates are consistent, asymptotically normal, and efficient. Random forests and XGBoost have been shown to be consistent under certain conditions, particularly when tree depth is constrained and ensemble size is large. Neural networks are capable for universal approximation; however, their statistical guarantees regarding convergence and generalization depend heavily on network architecture, regularization, and the volume of training data. By applying cross-validation and regularization, we aim to ensure robust and generalizable model estimates across all methodologies.

4 Real Data Analysis

In this section, we apply our full modeling pipeline to the diabetic readmission dataset to identify key predictors of 30-day hospital readmissions. We begin with data cleaning and exploratory

analysis, followed by feature preprocessing and class balancing. We then describe model tuning and selection, concluding with performance evaluation and feature importance analysis at both global and subgroup levels.

4.1 Data Preprocessing

Prior to model development, we performed a series of data cleaning steps to ensure consistency and alignment with our research objectives. From the original 50 features, we removed columns with limited clinical relevance or excessive missingness, specifically `encounter_id`, `patient_nbr`, `weight`, `payer_code`, and `medical_specialty`. Missing values in categorical variables, initially encoded as `?`, were replaced with the label `Unknown` to retain as much information as possible. We excluded encounters where the `discharge_disposition_id` indicated death or hospice care (values 11, 19, 20, or 21), as these cases are not eligible for readmission. We retained only encounters with valid `readmitted` labels (`NO`, `<30`, or `>30`) and constructed a binary target variable, `readmit_binary`, where 1 represents readmission within 30 days and 0 otherwise. Additionally, patient encounters with invalid gender entries (`Unknown/Invalid`) were removed to maintain consistency in demographic variables. The cleaned dataset was saved to `../data/diabetic_data_cleaned.csv` for reproducibility and future modeling steps.

4.2 Exploratory Data Analysis

To understand the demographic and clinical structure of our dataset, we visualized the distributions of key variables related to patient outcomes and characteristics. Figure 1 summarizes the distributions of 30-day readmission status, race, gender, and age groups after data cleaning. The dataset is notably imbalanced with only 11.3% of patients readmitted within 30 days, highlighting the challenge of predicting rare adverse outcomes. This imbalance motivates the use of class balancing strategies during model training. In terms of race, the majority of patients are Caucasian (74.8%), followed by African American (18.9%), with smaller proportions of Hispanic, Asian, and Other groups. The gender distribution is relatively balanced, with 53.8% female and 46.2% male patients. The age distribution reveals that the majority of patients are between 60 and 90 years old, with the highest concentration in the 70–80 age group (25.5%), followed by 60–70 (22.2%) and 80–90 (16.7%). These trends are consistent with the epidemiology of diabetes-related hospitalizations, which are more prevalent in older populations. This exploratory analysis informed both our preprocessing decisions and the downstream modeling approach, particularly with respect to handling imbalanced outcomes and evaluating fairness across demographic subgroups.

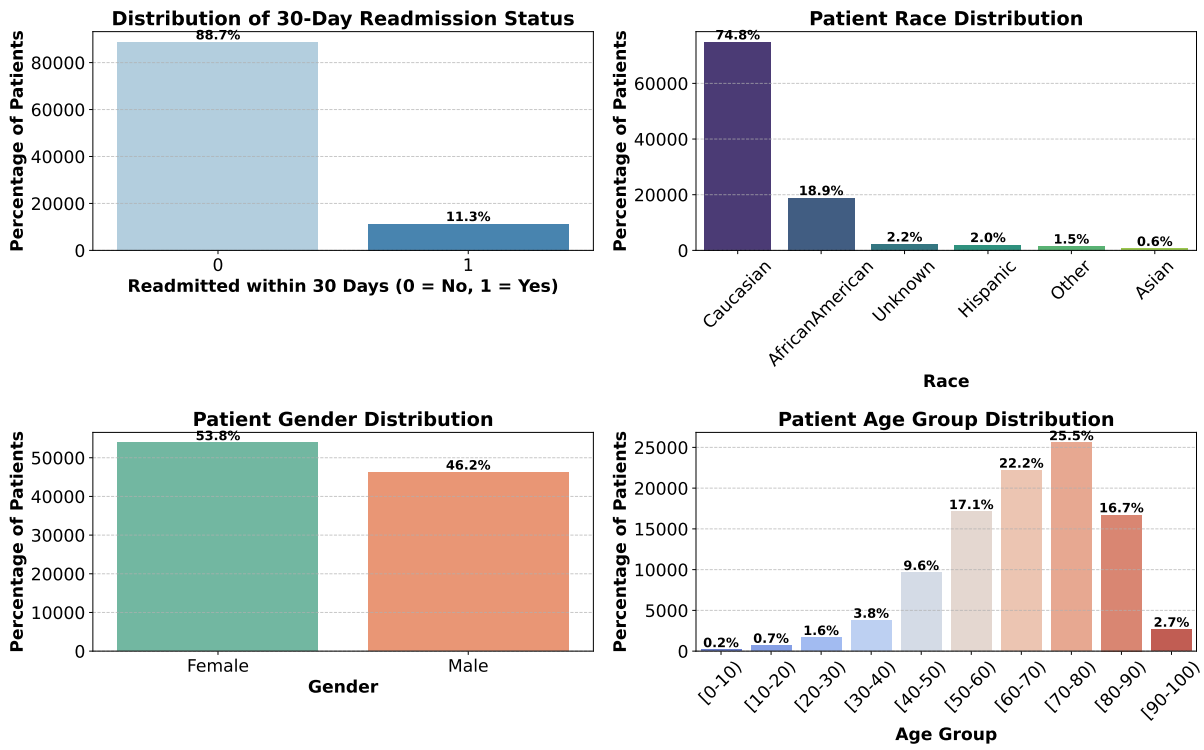
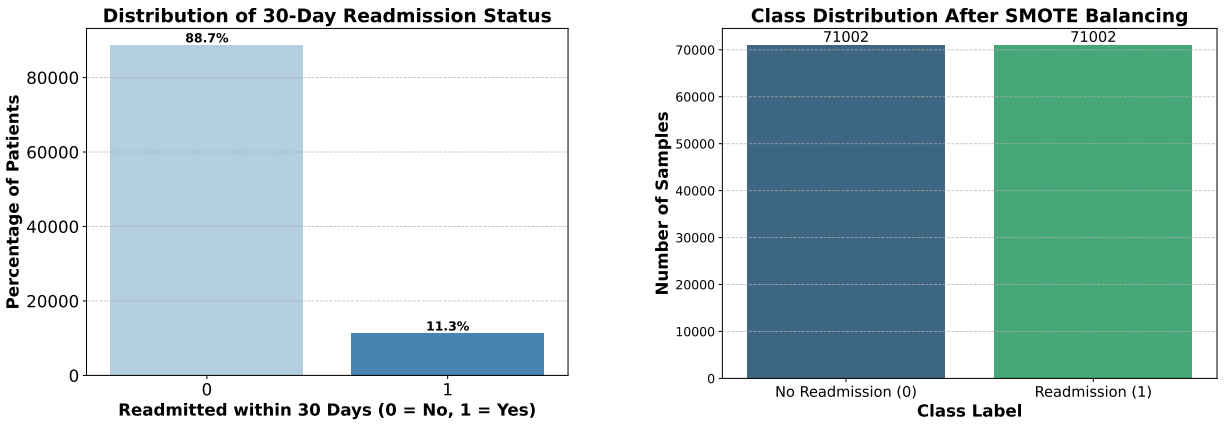


Figure 1: Distributions of (a) readmission status, (b) patient race, (c) patient gender, and (d) patient age group after data cleaning.

4.3 Feature Engineering and Preprocessing Pipeline

To prepare the cleaned dataset for modeling, we organized the variables into categorical and numerical groups and constructed a preprocessing pipeline accordingly. Categorical features encompassed demographic information (such as race, gender, and age group), hospital admission characteristics (including admission type, admission source, and discharge disposition), laboratory test results (e.g., `max_glu_serum` and `A1Cresult`), medication usage flags, and insulin-related medication patterns. Numerical features captured patient utilization metrics, including `time_in_hospital`, `num_lab_procedures`, `num_procedures`, `num_medications`, `number_outpatient`, `number_emergency`, `number_inpatient`, and `number_diagnoses`.

To ensure compatibility with machine learning models, preprocessing steps were applied based on feature type. Missing values in categorical variables were imputed with the constant `"Unknown"` and transformed using `one-hot encoding`. Numerical features were imputed with the median and standardized using `StandardScaler` to normalize their scales. These transformations were implemented using a `ColumnTransformer` embedded within a scikit-learn `Pipeline` to ensure reproducibility and modular design.



(a) Original class distribution.

(b) Class distribution after applying SMOTE.

Figure 2: **Comparison of original and balanced training set distributions.** (a) The original dataset exhibits a strong class imbalance with only 11.3% positive cases. (b) SMOTE balancing generated an equal number of samples for both readmission and non-readmission classes, enabling more robust model training.

Given the class imbalance where only 11.3% of patients were readmitted within 30 days, we adopted model-specific strategies to mitigate bias. For the neural network, we applied the Synthetic Minority Oversampling Technique (SMOTE) to generate a fully balanced training set with 71,002 samples per class (Figure 2). Logistic regression was trained with `class_weight='balanced'`, while XGBoost used the `scale_pos_weight` parameter, which we implemented manually as the ratio of negative to positive cases in the training data: `np.sum(y_train == 0) / np.sum(y_train == 1)`. This value was chosen to directly align with XGBoost's internal mechanism for adjusting the loss function in imbalanced binary classification settings. These strategies helped each model learn from imbalanced data without disproportionately favoring the majority class.

After preprocessing, the data was split into 80% training and 20% testing subsets using stratified sampling to preserve the original distribution of the target variable. The resulting feature matrix included 155 variables, combining standardized numerical features and expanded one-hot encoded categorical variables.

4.4 Hyperparameter Tuning and Optimization

To achieve optimal model performance, we conducted hyperparameter tuning for all three classifiers—Logistic Regression, XGBoost, and a Neural Network (MLP) using 5-fold cross-validation with grid search. Table 1 summarizes the full hyperparameter search grids and the optimal values selected for each model.

Model	Hyperparameter	Search Grid	Optimal
Logistic Regression	Regularization Strength (C)	{0.01, 0.1, 1, 10}	0.1
XGBoost	Learning Rate	{0.01, 0.1, 1, 10}	0.1
	Max Depth	{3, 5, 7}	3
	Number of Estimators	{100, 200}	200
Neural Network	Activation Function	{relu, tanh}	`tanh`
	Regularization (α)	{0.0001, 0.001}	0.0001
	Hidden Layer Sizes	{(50,), (50,30), (100,)}	(100,)
	Learning Rate Init	{0.001, 0.01}	0.001

Table 1: Grid search hyperparameter configurations and optimal selections for each model. SMOTE was applied before neural network tuning.

4.5 Model Evaluation

We evaluated model performance on the held-out test set using four key metrics: Area Under the Receiver Operating Characteristic Curve (AUC), 95% confidence intervals (CI), sensitivity, and specificity. These metrics collectively assess both the discriminative ability and the practical utility of the models in identifying patients at risk of 30-day hospital readmission.

Our primary metric was the Area Under the Receiver Operating Characteristic Curve (AUC), which summarizes the model's ability to distinguish between positive and negative cases across all possible classification thresholds. In this case, it is to differentiate between readmitted and non-readmitted patients. To assess the reliability of AUC estimates, we computed 95% confidence intervals using bootstrap resampling. In addition, we reported sensitivity and specificity based on a fixed probability threshold of 0.5 to evaluate each model's practical classification performance.

Sensitivity (true positive rate) measures the model's ability to correctly identify patients who were readmitted within 30 days. In clinical settings, high sensitivity is critical to avoid missing high-risk individuals, which could result in inadequate follow-up care and adverse health outcomes. **Specificity** (true negative rate) indicates the model's capacity to correctly identify patients unlikely to be readmitted. High specificity minimizes false positives, helping hospitals avoid unnecessary interventions and optimize limited resources.

Given the clinical importance of both identifying true readmissions and avoiding false alarms, a model with well-balanced and sufficiently high sensitivity and specificity is best suited for real-world deployment in hospital settings, where both under- and over-prediction can carry substantial consequences.

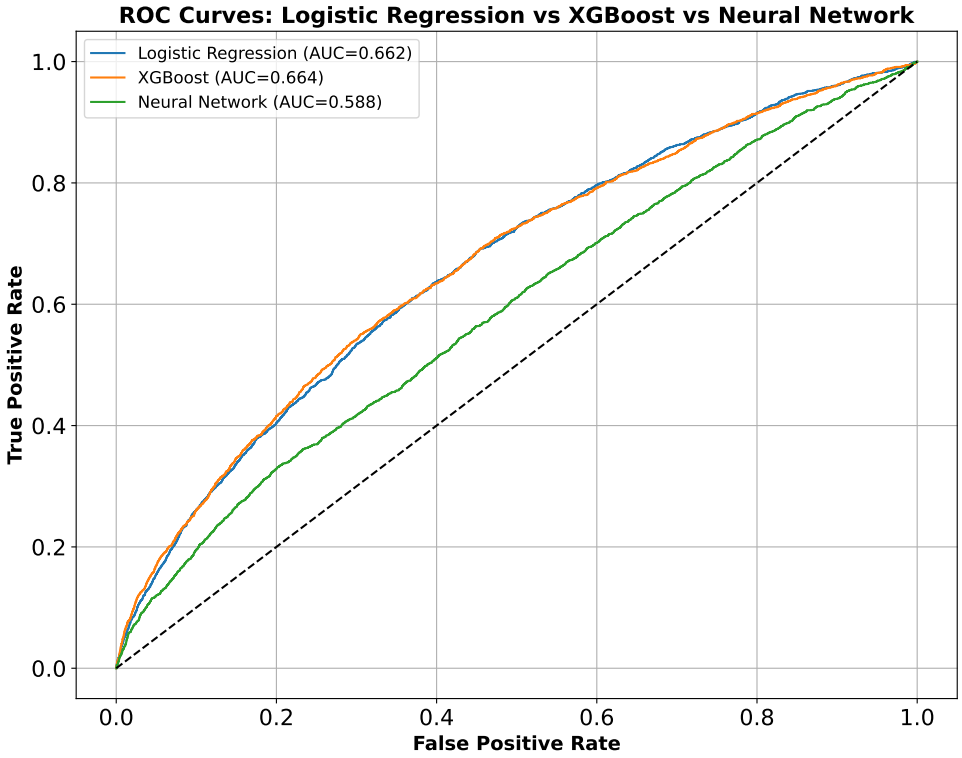


Figure 3: ROC curves on the test set

Table 2: Comparison of Model Evaluation Metrics

Model	**AUC**	**95% CI**	**Sensitivity**	**Specificity**
Logistic Regression	0.6621	(0.6501, 0.6739)	0.5350	0.6984
XGBoost	**0.6640**	(0.6525, 0.6756)	**0.5883**	0.6538
Neural Network (MLP)	0.5881	(0.5753, 0.6005)	0.2404	**0.8685**

Based on the evaluation results, we selected XGBoost as the final model for downstream analysis. It achieved the highest AUC (66.40%) and sensitivity (58.83%), which reflects strong discriminative power and reliable identification of patients at risk of 30-day readmission. In clinical settings, high sensitivity is essential to avoid missing high-risk individuals and to enable timely interventions. Although the neural network produced the highest specificity (86.85%), its low sensitivity (24.04%) limited its ability to identify true positives, reducing its value for proactive care. Logistic regression offered interpretable coefficients and a comparable AUC of 66.21%, but its sensitivity (53.50%) and specificity (69.84%) fell slightly behind those of XGBoost. Given its balanced performance, high AUC, and capability for feature importance analysis, XGBoost emerged as the most appropriate model for our research goals, including subgroup evaluation and the identification of key predictors for readmission risk.

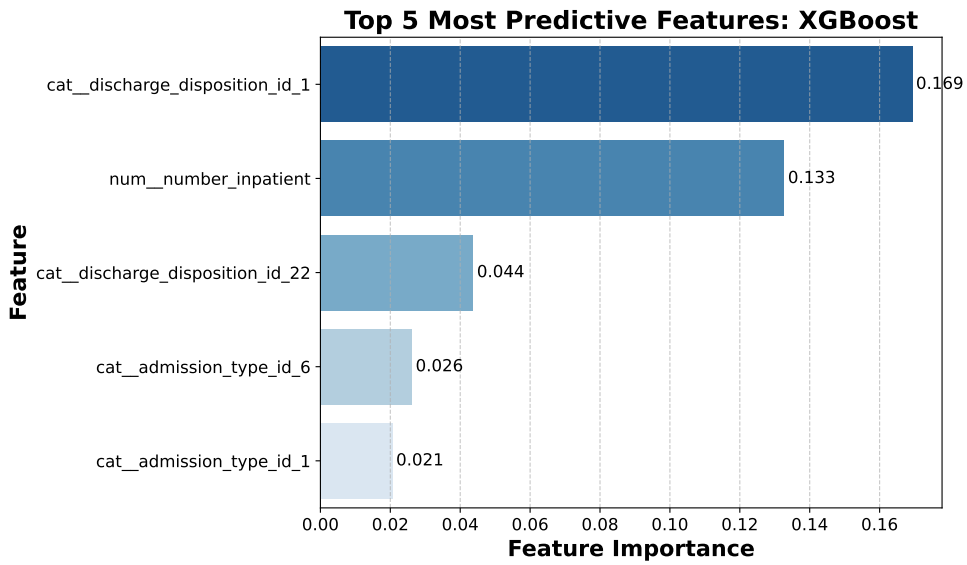


Figure 4: Top 5 global features ranked by XGBoost importance. The most influential predictors include prior inpatient visits, discharge disposition, and emergency department usage. These variables reflect patients' healthcare utilization patterns and care transitions, which are critical indicators of 30-day readmission risk. Feature importance was derived from the trained XGBoost model using gain-based metrics.

After selecting XGBoost as the final model, we examined its feature importance scores to address our first research question: ***Which patient factors most strongly predict 30-day hospital readmission among diabetic patients?*** The top-ranked features also reflect critical aspects of patients' clinical history and care transitions.

The **number of inpatient visits** emerged as the most important predictor. This variable reflects disease severity or instability, as patients with multiple hospitalizations often face complex or poorly managed conditions. The **discharge disposition**, especially cases involving transfer to another facility or provision of home care, ranked second. Such outcomes often indicate limited independence or a continued need for medical support, both of which heighten the risk of readmission. The **number of emergency visits** captured the extent of acute care usage and suggested challenges in managing chronic conditions through outpatient care. Similarly, the **number of diagnoses** revealed the burden of comorbidities, with higher values pointing to more complex treatment paths. Lastly, the **diabetes medication status**, particularly when no medication was prescribed, highlighted potential treatment gaps or non-adherence, which can lead to poor glycemic control and elevated readmission risk.

These features illustrate patterns of high healthcare utilization and clinical complexity. They help the model identify patients vulnerable to adverse outcomes after discharge and in need of closer follow-up. To address our second research question: ***How do predictor importance patterns vary across demographic groups?*** We conducted subgroup-specific analyses by race, gender, and age. This allowed us to explore how clinical and contextual drivers of readmission risk differ

across populations and to identify group-specific disparities that may inform more equitable and targeted.

4.6 Subgroup Feature Importance Analysis

Across all subgroups, **number of prior inpatient visits** ranked as the most important predictor. This confirms that recent hospitalization history is a strong signal for readmission risk.

However, the relative importance of other predictors varied. For racial groups, features related to discharge settings and length of stay stood out for African American and Hispanic patients. For Asian patients, age and medication use played larger roles, suggesting different treatment patterns. In gender subgroups, both men and women showed similar patterns, but with slight differences. Men had higher feature importance for comorbidity indicators, while medication-related features were more influential for women. Age-related differences were more pronounced. In younger adults (ages 20–40), medication and admission source were more predictive. In older groups (60+), prior inpatient use and discharge conditions dominated, reflecting higher medical complexity.

We also evaluated predictive performance by subgroup. As shown in Table 3, AUC and sensitivity varied across groups. Younger adults (20–30) had the highest AUC (0.803) and sensitivity (73.08%). In contrast, the model struggled to identify positive cases in the 10–20 group (sensitivity 23.08%) despite high accuracy. Gender differences were modest, with slightly better overall metrics for males. Racial subgroup AUCs were generally consistent, with the best performance in the Hispanic groups.

Our findings partially align with prior research by Strack et al. Strack et al. (2014), which emphasized the relationship between HbA1c measurement and hospital readmission rates among diabetic patients. Their study highlighted that HbA1c testing was performed infrequently but was associated with improved outcomes, suggesting that greater attention to glycemic control could reduce readmissions. While our analysis did not focus solely on HbA1c, we expanded the scope to examine a broader set of demographic and clinical predictors. Importantly, discharge disposition, prior inpatient visits, and admission types emerged as strong predictors in our model, supporting the general conclusion that proactive and coordinated care reduces readmission risk. Thus, our results align with the broader message of the existing literature: structured attention to a patient's chronic condition and care transitions plays a critical role in reducing early hospital returns.

Subgroup	Samples	Accuracy	AUC	Sensitivity	Specificity
Race Group					
African American	3,743	65.56%	0.670	57.27%	66.65%
Asian	124	79.03%	0.647	55.56%	80.87%
Caucasian	15,032	63.82%	0.660	59.46%	64.39%
Hispanic	427	67.21%	0.707	57.14%	68.31%
Other	284	66.20%	0.580	45.83%	68.08%
Unknown	413	77.97%	0.753	58.62%	79.43%
Gender Group					
Female	10,757	63.08%	0.658	59.12%	63.60%
Male	9,266	66.44%	0.671	58.48%	67.44%
Age Group					
[10–20)	130	87.69%	0.661	23.08%	94.87%
[20–30)	346	71.97%	0.803	73.08%	71.77%
[30–40)	692	72.69%	0.697	53.75%	75.16%
[40–50)	1,938	71.88%	0.719	62.62%	72.98%
[50–60)	3,418	73.49%	0.687	49.35%	75.90%
[60–70)	4,443	65.90%	0.632	53.61%	67.41%
[70–80)	5,123	60.06%	0.656	64.85%	59.40%
[80–90)	3,354	54.47%	0.617	60.33%	53.62%
[90–100)	548	56.20%	0.629	64.47%	54.87%

Table 3: XGBoost model performance metrics by demographic subgroups on the test set. Metrics include Area Under the ROC Curve (AUC), accuracy, sensitivity, and specificity.

5 Conclusion

In this study, we built models to predict 30-day readmission among diabetic patients. XGBoost outperformed logistic regression and neural networks. It showed strong accuracy, sensitivity, and specificity. The most influential predictors included the number of prior inpatient visits, emergency encounters, and discharge to skilled nursing facilities. These factors reflect disease severity, care transitions, and limited patient independence, which were well-established risk factors in clinical

literature.

Subgroup analyses showed differences in predictor importance across race, gender, and age groups. This highlights the presence of underlying disparities in healthcare access and outcomes. However, healthcare inequity is a deeply rooted and complex issue. Algorithmic models alone cannot resolve structural bias. Future work should integrate social determinants of health, expand subgroup evaluation, and promote fairness-aware model design. Additional efforts in data collection, policy support, and interdisciplinary collaboration are needed to advance health equity.

References

Joseph Futoma, John Morris, and Jacob Lucas. A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics*, 56:229–238, 2015.

Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M. Dai, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1):18, 2018.

Beata Strack, Jonathan P DeShazo, Chris Gennings, Juan L Olmo, Sebastian Ventura, Krzysztof J Cios, and John N Clore. Impact of hba1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. *BioMed Research International*, 2014:1–11, 2014.

UCI Machine Learning Repository. Diabetes 130-us hospitals for years 1999–2008 data set, 2008. URL https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008. Accessed: 2025-04-20.