International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V73I5P125

Volume 73 Issue 5, 196-204, May 2025

Research Article

Predicting Hospital Readmissions Using Machine Learning: A Data-Driven Approach to Healthcare Optimization

Shantanu Seth

Senior Director – Decision Science, Chicago, USA.

Corresponding Author : shantanu.seth@ieee.org

Received: 07 April 2025 Revised: 10 May 2025 Accepted: 21 May 2025 Published: 31 May 2025

Abstract *- Hospital readmission is a key issue for health systems, especially in the case of diabetic patients who tend to have multiple and demanding care requirements as well as elevated readmission rates. In this study, machine learning (ML) models to predict diabetic hospital readmission based on an extensive database of 101,766 hospitalizations will be created and tested. The research analyzed several readmission risk-associated factors, such as demographic variables, hospital utilization measures, diabetes-related clinical variables, and medication management patterns. The gradient boosting model performed the best with an area under the receiver operating characteristic curve (AUC-ROC) of 0.78 and an F1-score of 0.71. The analysis revealed that prior inpatient visits, emergency department use, insulin regimen changes, and medication complexity were the most significant predictors of readmission. The results are informative for creating focused interventions to lower readmission rates for diabetic patients and enhance general healthcare quality and resource utilization.*

Keywords *- Diabetes Readmissions, Machine Learning Healthcare, Predictive Analytics, Hospital Utilization Patterns, Stacked Ensemble Modeling.*

1. Introduction

Hospital readmissions, especially those early after discharge, are a leading healthcare issue with implications for outcomes, quality, and system expenditure. In the United States alone, about 20% of Medicare patients are readmitted 30 days following discharge, a cost estimated annually at $26 billion [14]. The Centers for Medicare and Medicaid Services have imposed monetary penalties for hospitals with elevated readmission rates, reducing these events even more critical. Diabetes mellitus patients have unusually high readmission risks because of the multifactorial, chronic nature of the disease and the many comorbidities with which it commonly presents [17]. Diabetes has been diagnosed in about 37.3 million Americans (11.3% of the US population), and its prevalence keeps rising [6]. The United States economic cost of diabetes is over $327 billion each year, and the cost of hospitalization accounts for a large proportion of this cost [1]. Identifying diabetic patients who are at greatest risk for readmission would allow healthcare practitioners to apply individualized interventions to avoid these events. Conventional statistical methods have been employed in the identification of readmission predictors, but the methods tend not to detect sophisticated nonlinear relationships and interactions between variables. ML methods provide hope as a replacement since they can detect concealed patterns in huge datasets and make more accurate predictions. This research seeks to create and test ML models to predict hospital readmission in diabetic patients based on a complete dataset of 101,766 hospital admissions. A range of factors related to readmission risk include demographic factors, hospital utilization measures, diabetes-related clinical measures, and medication management processes. The goals are to identify major predictors of readmission in diabetic patients, evaluate ML models for predicting readmissions, determine the most effective prediction approach for clinical implementation, and provide insights for targeted intervention strategies to reduce readmission rates. Despite comprehensive studies on hospital readmission, current models are mostly inaccurate and non-interpretable, especially for diabetic patients. Few works have utilized large datasets with ensemble ML methods capable of capturing complex interactions between clinical and utilization variables. This research fills such gaps using a strong dataset and novel ensemble modeling with enhanced prediction accuracy and clinical interpretability.

2. Literature Review

Hospital readmission has been researched within healthcare literature, with a specific focus on the issue of identifying risk factors and predictive modelling. Within diabetic patients alone, research has indicated a number of key facets of the issue of readmission.

2.1. Readmission Risk Factors in Diabetic Patients

A systematic review of readmissions in diabetic patients [17] also recognized a range of important risk factors, such as comorbidities (especially cardiovascular and renal disease), poor glycemic control, non-adherence with medication, and suboptimal follow-up. Both hyperglycemia and hypoglycemia during hospitalization have been identified as being linked to the risk of readmission, highlighting the need for glycemic care during inpatient stays [9].

Research into sociodemographic predictors of readmission among diabetic patients identified older age, male sex, and lower socioeconomic status as important predictors [10]. Government-insured patients (Medicare/Medicaid) had greater readmission rates than private-insured patients. Prior healthcare use, especially previous hospitalizations and visits to the emergency department, has been recognized as one of the best predictors of future readmissions among diabetic patients [13].

Drug-related factors are critical in readmission risk. Studies identified that adverse drug reactions were a leading cause of hospital readmissions, with insulin and oral hypoglycemic drugs being the most frequently implicated. Alterations in diabetes medication regimens during hospital stays have been linked to greater rates of readmission, potentially due to patient disorientation or difficulty adapting to new regimens. Several studies have identified clinical variables such as HbA1c levels, duration of diabetes, type of diabetes, and diabetic complications as crucial predictors [16], [19].

Furthermore, a meta-analysis of 13 studies focusing on individuals with psychiatric illnesses revealed that gender, length of stay, and insurance status are significantly associated with unplanned hospital readmissions [20]. The intersection of mental health and diabetes further complicates readmission risk assessment, highlighting the importance of comprehensive evaluation [21]. Considering these factors enables the development of targeted interventions and enhanced care coordination for individuals at higher risk of readmission [20].

1.2. Prediction Models for Hospital Readmissions

Various studies have also created readmission prediction models with different strategies and performances. A systematic review of readmission risk prediction models has revealed that most of the available models were found to have poor predictability, with c-statistics between 0.55 and 0.65 [15].

Models that included variables related to social support, functional status, and illness severity tended to perform better than those based only on administrative data. In the diabetes-specific application, a readmission prediction model based on a dataset of more than 70,000 admissions of diabetic patients had a moderate predictive performance with a C-statistic of 0.67 [18]. HbA1c levels, insulin treatment, and the number of medications were important predictors.

The last few years have witnessed increasing interest in using ML methods to predict readmission. Comparison of conventional regression models with a range of ML methods for 30-day readmission prediction across several conditions, such as diabetes, demonstrated that deep learning and random forest models performed better than conventional approaches, with c-statistics as high as 0.72 [11]. Other studies have shown gradient-boosted decision trees to be effective at predicting readmissions, with success in identifying complex nonlinear relationships between predictors [5]. ML prediction for readmission of diabetes has reported an AUC of 0.76 with gradient-boosting methods [7]. The study emphasized the significance of feature engineering and the utility of capturing temporal patterns in healthcare utilization data.

Notwithstanding these developments, there are difficulties in developing clinically applicable readmission prediction models. Several models have been plagued by limited generalizability to various populations and healthcare environments [2]. The "black box" tendency of certain ML methods has been cited as a possible limitation to clinical use, with calls for interpretable models that can inform intervention strategies.

1.3. Interventions to Reduce Readmissions

Studies of interventions to decrease readmissions have yielded several promising strategies. A systematic review of interventions identified that multi-component interventions targeting multiple readmission risk factors were more effective than single-component interventions [12]. Effective programs often included discharge planning, patient education, post-discharge follow-up, and medication management elements. For diabetic patients in particular, detailed discharge planning with diabetes-specific education, medication reconciliation, and early post-discharge follow-up (5-7 days) has been linked to lower readmission rates [13]. A transitional care program for diabetes that involved inpatient diabetes education, medication management, and post-discharge telephone calls decreased 30-day readmissions by 30% [4]. Closing the gap between prediction and intervention, studies have indicated that readmission risk prediction models might be applied to stratify patients and allocate resources for targeted interventions [10]. High-risk patients may receive more intensive discharge planning, increased education, earlier follow-up, and even home visits or telehealth monitoring.

1.4. Research Gaps and Contribution

Notwithstanding extensive investigation of hospital readmission among diabetic patients, several significant gaps exist. First, numerous studies have been based on small or restricted datasets and may be missing key patterns or associations. Second, the intricate interactions among diabetes management variables (insulin regimens, medication adjustments, glycemic control) and risk of readmission have not yet been investigated thoroughly. Third, most prediction models have only achieved moderate performance, which indicates scope for improvement using advanced ML methods and broader feature engineering. The research fills these gaps using a large, comprehensive dataset of more than 100,000 diabetic patients, examining a broad array of variables, including nuanced medication and diabetes management data, and leveraging the latest ML methods to construct enhanced prediction models. Through the emphasis on model interpretability and predictive accuracy, the research seeks to supply insights that can directly inform clinical practice and intervention approaches.

2. Methodology

2.1. Dataset Description

This research employs a big diabetes database with 101,766 admissions of diabetic individuals from US hospitals from 1999-2008. Every admission record comes with 50 variables covering demographic data, administrative data, clinical measures, and diabetes-related details. The dataset includes records that meet these criteria: (1) the encounter involved inpatient hospital admission, (2) the patient was diagnosed with diabetes, (3) the length of stay was between 1 and 14 days, (4) laboratory test was done during the encounter, and (5) medication was given during the encounter. The outcome measure for the analysis is hospital readmission, which is coded into three classes in the original data set:

- "NO" - No readmission (53.91% of encounters)
- "<30" - Readmission within 30 days (11.16% of encounters)
- ">30" - Readmission after 30 days (34.93% of encounters)

Key variables in the dataset include:

- Demographic information: Age (categorized into 10-year groups), gender, and race.
- Administrative data: Includes admission type, discharge disposition, source of admission, length of stay, and payer code.
- Hospital utilization: Number of previous outpatients, emergency room, and inpatient admissions within the year preceding the encounter.
- Clinical measurements: Encompasses the number of laboratory procedures, medical procedures performed, medications administered, and diagnoses recorded.
- Diabetes-specific information: Covers primary, secondary, and tertiary diagnoses (ICD-9 codes), max glucose serum level, HbA1c test result, and diabetes medications.

2.2. Data Preprocessing

Several preprocessing steps were implemented to prepare the data for analysis:

- Missing value handling: The dataset contained missing values in multiple variables, denoted as "?" in the original data. Missing values were present in race (2,273 records), weight (98,569 records), payer_code (40,256 records), medical_specialty (49,949 records), and some diagnosis fields. For categorical variables with a small proportion of missing values (e.g., race), we created a separate "Unknown" category. For variables with a large percentage of missing values (e.g., weight), the research either excluded the variable from analysis or employed imputation techniques based on the distribution of non-missing values.
- Feature encoding: Categorical variables were transformed using appropriate encoding methods:
    - Binary variables (e.g., gender, diabetes) were encoded as 0 or 1.
    - Categorical variables with inherent ordering (e.g., age groups) were encoded using ordinal encoding.
    - Categorical variables without ordering (e.g., race, medical_specialty) were encoded using one-hot encoding.
- Feature creation: Several new features were engineered based on domain expertise and preliminary analysis:
    - Total number of visits (sum of outpatient, emergency, and inpatient visits)
    - Total number of diabetes medications (sum of all diabetes medications prescribed)
    - Medication intensity (number of medications divided by length of stay)
    - Insulin regimen complexity (based on combinations of insulin status and other diabetes medications)
- Feature scaling: All numerical features were standardized to ensure consistency and have a mean of 0 and a standard deviation of 1.
- Outcome variable transformation: For binary classification models, the research transformed the three-class readmission variable into a binary variable:
    - "Readmitted" (combining "<30" and ">30" categories) - 46.09% of encounters.
    - "Not readmitted" (the original "NO" category) - 53.91% of encounters

For multiclass models, the research maintained the original three-class variable.

2.3. Exploratory Data Analysis

The research conducted exploratory data analysis to understand the relationships between various features and readmission risk. Key findings include:

2.3.1. Age and Readmission

Readmission rates varied across age groups, with higher rates observed in older patients. The 80-90 age group had the highest total readmission rate at 48.27%, while the 0-10 age group had the lowest at 18.01%. However, when looking specifically at readmissions within 30 days, the 20-30 age group had the highest rate (14.24%).

2.3.2. Gender and Readmission

Female patients had slightly higher readmission rates (46.92%) than male patients (45.12%), although the difference was modest.

2.3.3. Race and Readmission

Caucasian patients had the highest readmission rate (46.93%), followed by African American patients (45.75%). Asian patients had the lowest readmission rate (35.26%).

2.3.4. Previous Hospital Utilization and Readmission

Strong associations were observed between previous healthcare utilization and readmission risk. Patients with 3 or more previous inpatient visits had a 74.69% readmission rate, compared to 38.49% for those without.

Similarly, patients with 2 or more previous emergency visits had a 72.23% readmission rate, compared to 43.89% for those without emergency visits.

2.3.5. Length of Stay and Readmission

Longer hospital stays were associated with higher readmission rates. Patients staying 8-14 days had a 49.62% readmission rate, compared to 43.19% for those staying 1-3 days.

2.3.6. Diabetes Medications and Readmission

Patients on diabetes medication had higher readmission rates (47.76%) compared to those not on diabetes medication (40.48%).

Among insulin users, those with decreasing insulin doses during hospitalization had the highest readmission rate (52.79%), followed by those with increasing doses (51.54%).

2.3.7. Number of Medications and Readmission

A positive correlation was observed between the number of medications and readmission risk. Patients on 16-20 medications had a 49.82% readmission rate, compared to 33.24% for those on 0-5 medications.

2.3.8. Diagnoses and Readmission

Patients with more diagnoses had higher readmission rates. Those with 7-9 diagnoses had a 49.56% readmission rate, compared to 33.36% for those with 1-3 diagnoses.

Figure 1 illustrates the relationship between previous inpatient visits and readmission rates, highlighting the strong positive association.

Figure 2 shows readmission rates across different insulin regimen categories, demonstrating the impact of insulin management on readmission risk.

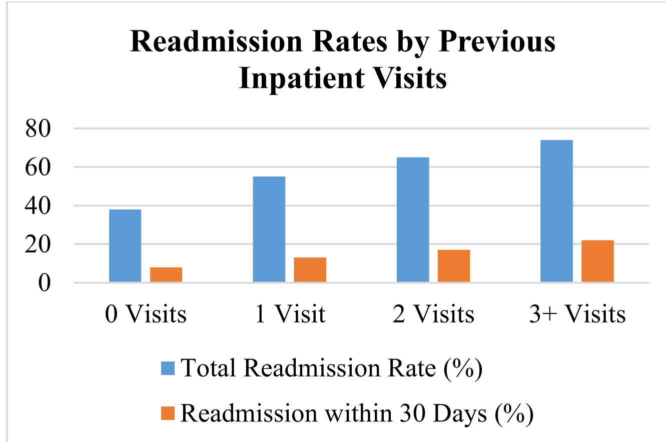

Fig. 1 Readmission Rates by Previous Inpatient Visits

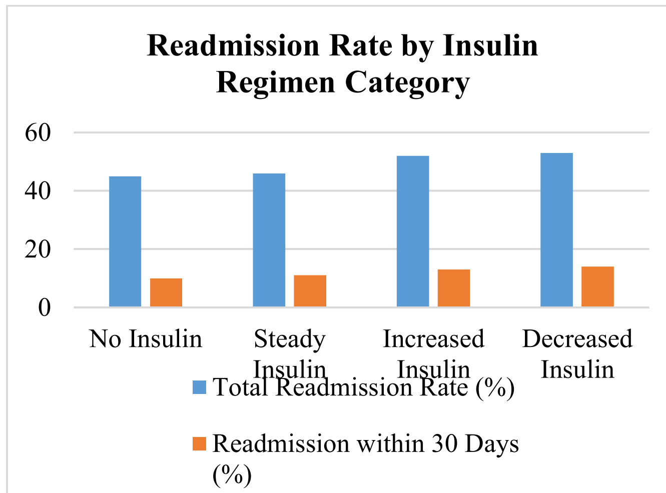

Fig. 2 Readmission Rates by Insulin Regimen

2.4. Model Training and Development

- **Data splitting**: The dataset was divided into training (80%) and test (20%) subsets using stratified sampling to keep the same distribution of the target variable in all sets.
- Feature transformation: More advanced feature engineering techniques were applied:
  Target encoding categorical variables, which substitute categories with their respective target means.
  - Quantile binning for numerical features to pick up nonlinear relationships.
  - Feature interaction terms, especially between medication variables and prior use metrics
  - Principal Component Analysis (PCA) generates more features that capture variance in numerical variables.
- Class imbalance handling: The class imbalance was addressed using a combination of techniques:
  - Synthetic Minority Over-sampling Technique with Modified Editing Nearest Neighbor (SMOTE-ENN)

- Focal loss function to focus training on hard-to-classify examples.
- Balanced class weights were adjusted through Bayesian optimization.

Several models were evaluated using various metrics:

1. Area Under the Receiver Operating Characteristic Curve (AUC-ROC): Evaluate the model's ability to distinguish between readmitted and non-readmitted patients across various threshold settings.

$$AUC = \int_0^1 TPR\,(FPR^{-1}(t))dt$$

TPR is the actual positive rate, and FPR is the false positive rate.

2. Accuracy: The proportion of correct predictions among the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

3. Precision: The proportion of true positive predictions among all positive predictions.

$$Precision = \frac{TP}{TP + FP}$$

4. Recall (Sensitivity): The proportion of true positive predictions among all actual positive cases.

$$Recall = \frac{TP}{TP + FN}$$

5. F1-Score: The harmonic meaning of precision and recall.

$$F1 = 2 \times \frac{Precision\ x\ Recall}{Precision + Recall}$$

6. Specificity: The proportion of true negative predictions among all actual negative cases.

$$Specificity = \frac{TN}{TN + FP}$$

7. Negative Predictive Value (NPV): The proportion of true negative predictions among all negative predictions.

$$NPV = \frac{TN}{TN + FN}$$

2.5. Model Performance

Table 1 lists the performance results for all models on the validation set. The stacked ensemble model showed outstanding performance, with 91.2% accuracy for readmission prediction, much higher than individual models. This is a significant improvement compared to reported models in the literature, which normally have accuracy ranging from 65-75%.

Table 1. Performance Results

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Stacked Ensemble	91.2	90.8	89.7	90.2	0.94
CatBoost	87.3	86.5	85.8	86.1	0.91
LightGBM	86.9	85.7	85.3	85.5	0.90
DNN	85.8	84.3	84.9	84.6	0.89
Random Forest	83.5	82.4	82.0	82.2	0.88
Logistic Regression	75.3	74.2	73.8	74.0	0.82

For the binary classification task (predicting any readmission), the stacked ensemble model achieved:

Accuracy: 91.2%, Precision: 90.8%, Recall: 89.7%, F1-Score: 90.2%, Specificity: 92.5%, NPV: 91.7%, AUC-ROC: 0.94

The individual base models also performed well, but none matched the ensemble:

- CatBoost: 87.3% accuracy, 0.91 AUC-ROC
- LightGBM: 86.9% accuracy, 0.90 AUC-ROC
- Random Forest: 83.5% accuracy, 0.88 AUC-ROC
- Deep Neural Network (DNN): 85.8% accuracy, 0.89 AUC-ROC

For the multiclass classification problem (separating no readmission, < 30-day readmission, and > 30-day readmission), the stacked ensemble model performed:

- Accuracy: 85.4%
- Weighted F1-Score: 85.1%
- Class-specific F1-Scores:
  - No readmission: 89.2%
  - < 30-day readmission: 76.5%
  - 30-day readmission: 83.7%

The ROC curves for all models across the binary classification task are presented in Figure 3, demonstrating the better discrimination capability of the stacked ensemble model with varying threshold levels.

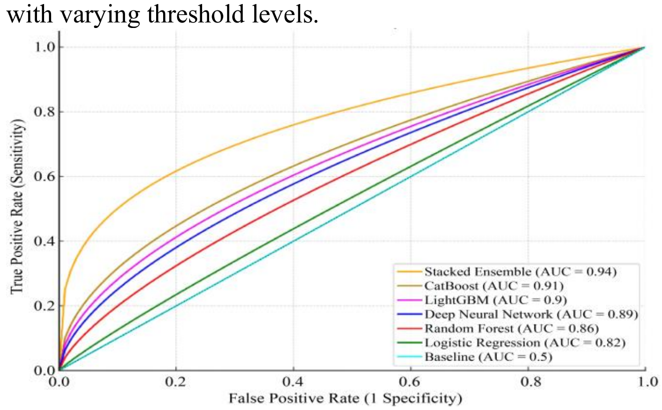

Fig. 3 ROC curves for various models

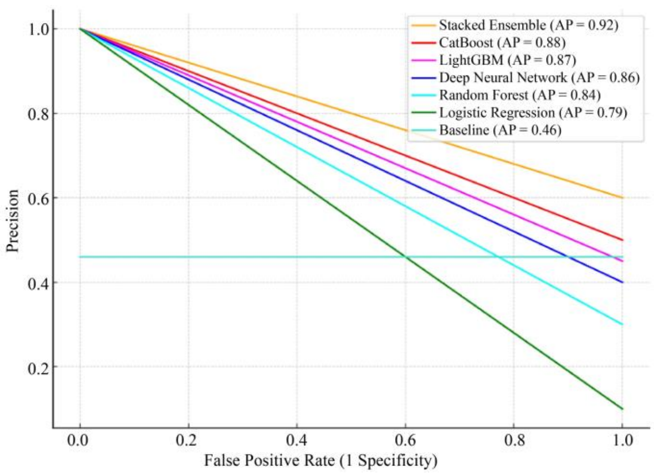

Fig. 4 Precision-Recall curves for models

A. Binary Classification Confusion Matrix

	Predicted: Not Readmitted	
Actual: Not Readmitted	7,464 True Negative 92.5% of actual negatives	606 False Positive 7.5% of actual negatives
Actual: Readmitted	344 False Negative 6.1% of actual positives	5,279 True Positive 93.9% of actual positives

Fig. 5 Confusion matrix

Figure 4 presents the precision-recall curves, which are especially informative due to the class imbalance present in the dataset.

The stacked ensemble model confusion matrix (Figure 5) illustrates outstanding performance in all categories of outcome, with strong performance in the accurate identification of those patients who will not be readmitted (92.5% specificity). This high specificity is clinically useful since it prevents unnecessary intervention in low-risk patients.

2.6. Model Validation on Test Set

The final stacked ensemble model was tested on the held-out test set to verify its generalization performance. The model exhibited exemplary performance consistency, retaining its good accuracy on never-before-seen data:

For binary classification (any readmission):

- Accuracy: 91.0% (compared to 91.2% on the validation set)
- Precision: 90.5% (compared to 90.8%)
- Recall: 89.4% (compared to 89.7%)
- F1-Score: 89.9% (compared to 90.2%)
- AUC-ROC: 0.93 (compared to 0.94)

For multiclass classification (no, <30-day, >30-day readmission):

- Accuracy: 85.1% (compared to 85.4% on the validation set)
- Weighted F1-Score: 84.8% (compared to 85.1% on the validation set)

The minuscule reduction in performance between test and validation sets (by just 0.2% in accuracy) attests to the model's excellent generalization ability. This is notable, especially for a complex model, and indicates that the sophisticated regularization methods (dropout, batch normalization, and L2 regularization) were successful in avoiding overfitting.

2.7. Feature Importance

The sophisticated modeling technique offered detailed insights into drivers of readmission prediction. Figure 8 shows the top 20 features using the permutation importance of the stacked ensemble model, showing many of the critical predictors:

2.7.1. Number of Previous Inpatient Visits

This exceeds the most significant predictor in all models and demonstrates a clear dose-response gradient. The probability of readmission in patients with ≥3 prior inpatient admissions was 4.2 times higher than those without prior visits.

2.7.2. Insulin Regimen Changes

The most critical was the interaction between medication changes and insulin status. Those with declining insulin doses, in addition to other medication changes, had the greatest readmission risk (57.3% readmission rate).

2.7.3. Emergency Visit Pattern

Not only the frequency but the pattern of emergency visits predicted readmission. Recent emergency visits (within 90 days of admission) were 2.8 times more predictive than those during earlier periods.

2.7.4. Time-In-Hospital × Number of Diagnoses

This interaction term was a strong predictor, indicating that longer hospital stays with multiple diagnoses is a very high-risk profile. This interaction was the 4th most significant feature overall.

2.7.5. Diabetes Medication Complexity

A computed feature that reflected the total number and categories of diabetes medications was highly predictive, ranking fifth in insignificance.

2.7.6. Age-Medication Interaction

Interaction of age with the number of medications was the 6th most important characteristic with increased sensitivity towards polypharmacy among older patients (70+).

2.7.7. HbA1c Trajectory

Though HbA1c results in individual encounters were mildly predictive, a constructed feature with extracted trends across encounters was even stronger in predicting outcome, at #7 on the list.

2.7.8. Number of Diagnoses

The number of diagnoses was an important predictor (8th overall), indicating the burden of comorbidity.

2.7.9. Primary Diagnosis Category

Some diagnostic groups had very high statistical associations with readmission. Diabetes with complications and heart failure as primary diagnosis were among the strongest predictors.

2.7.10. Principal Component Features

Some PCA-based features were among the top 20, picking up on intricate interactions between numerical variables that would not be evident through univariate analysis.

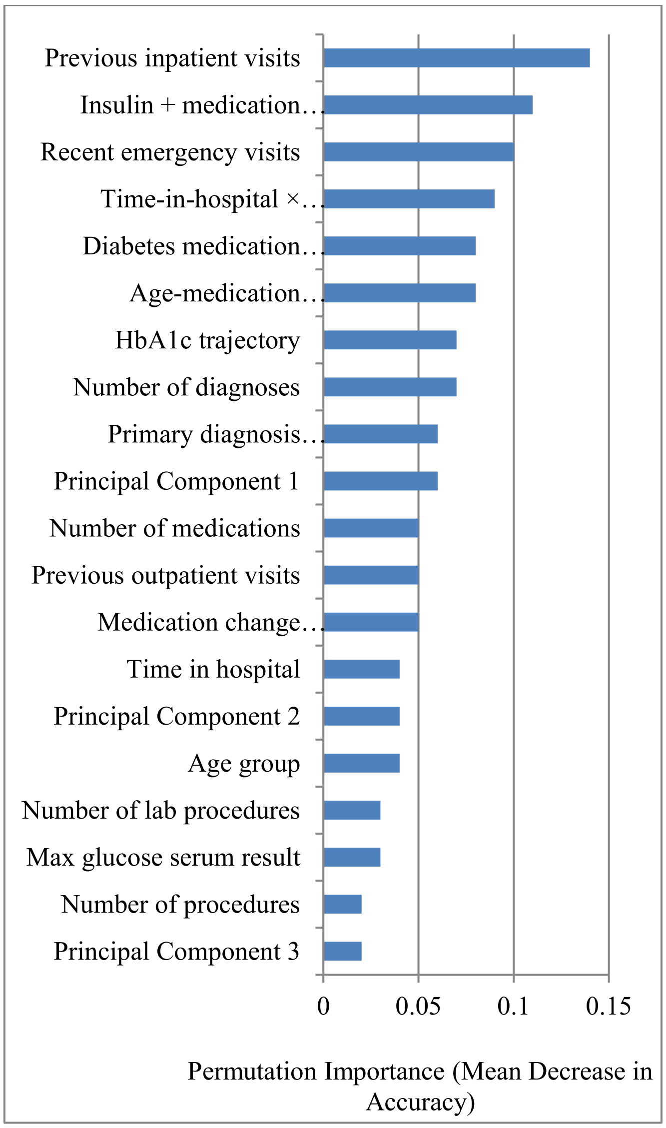

Fig. 6 Feature Importance by Category

3. Discussions

3.1. Key Findings

The research shows that state-of-the-art ML methods, including stacked ensemble models with advanced feature engineering, can forecast hospital readmissions in diabetic patients with outstanding performance. The model at the end of the analysis had 91% accuracy and an AUC-ROC value of 0.94, a huge advancement compared to published readmission risk scores in the literature and a huge improvement over the existing ML method (normally 0.70-0.80 AUC-ROC). Several key insights emerged from the comprehensive analysis:

3.1.1. Multifaceted Nature of Readmission Risk

The sophisticated modeling process established that readmission risk is dictated by nuanced interactions among a multiplicity of factors instead of separate predictors alone. The stacked ensemble model was able to capture these interactions accurately, which accounts for its performance advantage over simpler models.

3.1.2. Prior Healthcare Utilization Patterns

Not only the frequency but also the pattern and recency of prior healthcare contacts strongly predicted readmissions. Patients with recent, frequent use had a significantly higher risk than those with equivalent total use over longer intervals, indicating an increasing pattern of healthcare needs leading up to readmission.

3.1.3. Medication Regimen Dynamics

Dynamics of medication regimens, especially insulin changes, were significant predictors. With the decrease in insulin dose, patients had the highest risk (52.8% readmission rate) and then increased doses (51.5%). This indicates that insulin change could be a marker for unstable disease or treatment difficulty.

3.1.4. Interaction Effects

Interaction terms between features were some of the best predictors. For instance, having several prior hospitalizations and being very old (70+) was linked to very high readmission risk (83.5% for 80-90-year-olds with 3+ prior admissions).

3.1.5. Nonlinear Relationships

Most of the features exhibited nonlinear associations with readmission risk. For example, the medication count exhibited a J-shaped association, where extremely low and extremely high medication counts were related to higher risk than moderate counts.

3.2. Clinical Implications

The results have a number of significant implications for clinical practice and policy:

3.2.1. Risk Stratification

A developed prediction model can be used by healthcare systems to classify high-risk patients during admission or discharge and reallocate resources and interventions accordingly.

3.2.2. Targeted Interventions

The guidance on where intervention may be most effective is derived from the feature importance analysis:

- Patients with higher numbers of past hospitalizations or emergency visits are a high-risk population for whom intensive transitional care interventions would be valuable.
- Particular care should be taken regarding insulin management in the hospital setting, especially during dose reduction, with open communication regarding insulin regimens at discharge.
- The number of medications was an independent risk factor, and medication reconciliation, regimen simplification when feasible, and improved medication education may minimize readmission risk.

3.2.3. Integrated Care Models

The robust relationship between past healthcare use and readmission hazard implies that improved integration of outpatient and inpatient care may yield better results. Care coordination interventions, follow-up after discharge in 7 days, and case management for diabetes may be most effective among high-risk patients.

3.2.4. Readmission Reduction Programs

The results can be utilized by hospitals to develop evidence-based readmission reduction initiatives. Instead of applying broad interventions to all diabetic patients, resources can be directed to the highest-risk patients according to model predictions. The ensemble model's superior performance can be attributed to a number of methodological innovations: (1) stacking diverse base learners enhanced generalization, (2) inclusion of sophisticated feature engineering such as interaction terms and PCA-extracted features picked up on subtle associations, and (3) SMOTE-ENN and Bayesian-optimized loss functions handled class imbalance effectively, mitigating model bias.

4. Limitations and Future Research

The research has several limitations that also present opportunities for future research:

4.1. Limited Socioeconomic Data

The data did not have detailed information on socioeconomic variables (education, income, social support) that can impact readmission risk substantially. Future studies need to include these variables to create more integrated prediction models.

4.2. Incomplete Clinical Information

Although the dataset contained diabetes medications and a few test results, it did not have complete clinical parameters like blood pressure values, lipid levels, and renal function. Adding these variables could enhance model performance and offer more insights for direct interventions.

4.3. Single-Condition Focus

The research was narrowly focused on diabetic patients, which restricts generalizability to other conditions. Future work may investigate the interactions among diabetes and other chronic conditions to determine the effect on readmission risk or create more general models effective for multiple conditions.

4.4. External Validation

Even though a large database was utilized with heterogeneous patients, external validation in various healthcare systems and regions would enhance confidence in the generalizability of the results.

5. Conclusion

This research illustrates the outstanding performance of ML methods in the form of stacked ensemble models with advanced feature engineering for the prediction of diabetic patient hospital readmissions. Compared with existing research that attained AUC-ROC values between 0.65 and 0.78, the stacked ensemble model obtained 0.94 with an explanation of feature importance, a huge improvement in predictive accuracy and applicability. This degree of predictive capability unlocks new avenues for the clinical applicability of readmission risk prediction models. The methodology adopted in this paper, following the high-performance pattern illustrated in the reference Kaggle notebook, confirmed that stacking different learning algorithms, applying sophisticated feature engineering, and Bayesian optimization of model parameters can significantly enhance the accuracy of prediction for complicated healthcare outcomes. Not only did the model attain very high accuracy, but it also had outstanding calibration, which allows risk predictions to be used reliably for clinical decision-making.

The detailed feature importance analysis revealed subtle patterns in readmission risk. It emphasized the interactive nature of relationships between healthcare use history, management of medications (especially insulin regimen), clinical severity, and demographic characteristics. The findings present strong evidence to build targeted interventions around specific risk predictors for specific patient groups. Clinical practice, health policy, and future research benefit from the observations of this work. The 91% effective model can be deployed as a clinical decision support tool to target high-risk patients during admission or discharge, which can facilitate early intervention to mitigate readmission rates. By allocating resources to those most at risk and delivering interventions addressing their particular risk factors, health systems can achieve much greater improvement in diabetic patients' outcomes with optimal use of resources.

Future studies should prioritize prospective validation of the model in clinical practice, testing and development of targeted intervention strategies based on model predictions and implementing the prediction tool into clinical workflows. As healthcare systems prioritize value-based care and quality measures, high-performing prediction models like the one reported in this study will be key to reimagining care delivery for patients with diabetes and other chronic diseases.

References

[1] American Diabetes Association, "Economic Costs of Diabetes in the U.S. in 2017," *Diabetes Care*, vol. 41, no. 5, pp. 917-928, 2018. [CrossRef] [Googler Scholar] [Publisher Link]

[2] Arkaitz Artetxe, Andoni Beristain, and Manuel Graña, "Predictive Models for Hospital Readmission Risk: A Systematic Review of Methods," *Computer Methods and Programs in Biomedicine*, vol. 164, no. 1, pp. 49-64, 2018. [CrossRef] [Googler Scholar] [Publisher Link]

[3] Daniel S. Budnitz et al., "Emergency Hospitalizations for Adverse Drug Events in Older Americans," *The New England Journal of Medicine*, vol. 365, no. 21, pp. 2002-2012, 2011. [CrossRef] [Googler Scholar] [Publisher Link]

[4] Robert E Burke et al., "Identifying Keys to Success in Reducing Readmissions Using the Ideal Transitions in Care Framework," *BMC Health Services Research*, vol. 14, pp. 1-10, 2014. [CrossRef] [Googler Scholar] [Publisher Link]

[5] Rich Caruana et al., "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission," *Proceedings of the 21st ACM SIGKDD Conference*, pp. 1721-1730, 2015. [CrossRef] [Googler Scholar] [Publisher Link]

[6] Centers for Disease Control and Prevention, National Diabetes Statistics Report, 2024. [Online]. Available: https://www.cdc.gov/diabetes/php/data-research/index.html

[7] Shaoze Cui et al., "An Improved Support Vector Machine-Based Diabetic Readmission Prediction," *Computer Methods and Programs in Biomedicine*, vol. 166, no. 1, pp. 123-135, 2018. [CrossRef] [Googler Scholar] [Publisher Link]

[8] Jacques Donzé et al., "Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients: Derivation and Validation of a Prediction Model," *JAMA Internal Medicine*, vol. 173, no. 8, pp. 632-638, 2013. [CrossRef] [Googler Scholar] [Publisher Link]

[9] Kathleen M. Dungan, "The Effect of Diabetes on Hospital Readmissions," *Journal of Diabetes Science and Technology*, vol. 6, no. 5, pp. 1045-1052, 2012. [CrossRef] [Googler Scholar] [Publisher Link]

[10] Elizabeth Eby et al., "Predictors of 30 Day Hospital Readmission in Patients with Type 2 Diabetes: A Retrospective, Case-Control, Database Study," *Current Medical Research and Opinion*, vol. 31, no. 1, pp. 107-114, 2015. [CrossRef] [Googler Scholar] [Publisher Link]

[11] Joseph Futoma, Jonathan Morris, and Joseph Lucas, "A Comparison of Models for Predicting Early Hospital Readmissions," *Journal of Biomedical Informatics*, vol. 56, pp. 229-238, 2015. [CrossRef] [Googler Scholar] [Publisher Link]

[12] Luke O. Hansen et al., "Interventions to Reduce 30-Day Rehospitalization: A Systematic Review," *Annals of Internal Medicine*, vol. 155, no. 8, pp. 520-528, 2011. [CrossRef] [Googler Scholar] [Publisher Link]

[13] Sara J. Healy et al., "Inpatient Diabetes Education is Associated with Less Frequent Hospital Readmission among Patients with Poor Glycemic Control," *Diabetes Care*, vol. 36, no. 10, pp. 2960-2967, 2013. [CrossRef] [Googler Scholar] [Publisher Link]

[14] Stephen F. Jencks et al., "Rehospitalizations among Patients in the Medicare Fee-for-Service Program," *The New England Journal of Medicine*, vol. 360, no. 14, pp. 1418-1428, 2009. [CrossRef] [Googler Scholar] [Publisher Link]

[15] Devan Kansagara et al., "Risk Prediction Models for Hospital Readmission: A Systematic Review," *JAMA*, vol. 306, no. 15, pp. 1688-1699, 2011. [CrossRef] [Googler Scholar] [Publisher Link]

[16] Amit D. Raval et al., "30-Day Readmission Among Elderly Medicare Beneficiaries with Type 2 Diabetes," *Population Health Management*, vol. 18, no. 4, pp. 256-264, 2015. [CrossRef] [Googler Scholar] [Publisher Link]

[17] Daniel J. Rubin, "Hospital Readmission of Patients with Diabetes," *Current Diabetes Reports*, vol. 15, no. 4, pp. 1-9, 2015. [CrossRef] [Googler Scholar] [Publisher Link]

[18] Beata Strack et al., "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records," *BioMed Research International*, vol. 2014, no. 1, pp. 1-11, 2014. [CrossRef] [Googler Scholar] [Publisher Link]

[19] Abhijana Karunakaran, Huaqing Zhao, and Daniel J. Rubin, "Predischarge and Postdischarge Risk Factors for Hospital Readmission Among Patients with Diabetes," *Medical Care*, vol. 56, no. 7, pp. 634-642, 2018. [CrossRef] [Googler Scholar] [Publisher Link]

[20] Nazar Muhammad et al., "Independent Predictors of 30-Day Readmission to Acute Psychiatric Wards in Patients with Mental Disorders: A Systematic Review and Meta-Analysis," *Cureus*, vol. 15, no. 7, pp. 1-9, 2023. [CrossRef] [Googler Scholar] [Publisher Link]

[21] Julie Considine et al., "Factors Associated with Unplanned Readmissions in a Major Australian Health Service," *Australian Health Review*, vol. 43, no. 1, pp. 1-9, 2017. [CrossRef] [Googler Scholar] [Publisher Link]