Prediction Readmission Risk: A Comparative Study of Machine Learning Models in Healthcare

YASHASWIKA THOTA, CHARITHA SRI IDDUM
Prof. Reda Mastouri
DS-670-Capstone:Big Data & Data Sci
Saint Peter's University

This study aims to predict Diabetic patient readmissions to hospitals, an important issue in healthcare that impacts pa3ent outcomes and hospital resource allocation. The dataset we used in this project is extracted from UCI machine learning repository. The dataset represents clinical care at 130 US hospitals and integrated delivery networks. It includes 50 features in the dataset The primary focus is on improving prediction accuracy by addressing class imbalance and feature selection through machine learning techniques. We applied Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset, which contained a significant imbalance between readied (minority class) and non-readmitted patients (majority class). Principal Component Analysis (PCA) was used to reduce the dimensionality of the dataset while retaining key features. Several machine learning algorithms, including Logistic Regression and Random Forest, were trained and evaluated using balanced data, and their performance was assessed using metrics such as precision, recall, F1-score, and ROC-AUC. Results indicated that while the models achieved reasonable accuracy, challenges remained in improving precision for the minority class, particularly in predicting readmissions. The findings suggest that despite balancing techniques, further improvements in model tuning and evaluation strategies are required. Future work may explore more advanced models like Gradient Boostng or XGBoost, hyperparameter optimization, and the integration of additional patient features to enhance prediction accuracy and clinical decision-making.

Key Words: Patent Readmission; Machine Learning; SMOTE; Principal Component Analysis; Class Imbalance Healthcare Prediction; Random Forest; Logistic Regression; Model Evaluation

Introduction

Diabetes is a chronic disease affecting millions of people worldwide, with severe complications that often lead to hospital readmissions. In the United States, hospital readmissions have become a growing concern, particularly among diabetic patients. According to a 2ti11 survey conducted by the Agency for Healthcare Research and Quality (AHRQ), more than 3.3 million patients were readmitted within 30 days of discharge, leading to an estimated $40 billion in healthcare costs. Among these cases, over $250 million was spent specifically on readmitted diabetic patients. Hospital readmissions, especially within 30 days of discharge, are open linked to inadequate patient care, poor post-discharge follow-up, and complications related to diabetes management. Readmissions not only impose a financial burden on healthcare systems but also negatively impact patient health and quality of life. Addressing this issue requires a proactive approach in identifying patients who are at high risk of readmission and providing them with improved care and monitoring. Diabetes is characterized by hyperglycemia resulting from insulin deficiency or resistance, making it a lifelong condition requiring continuous management. The complexity of the disease, along with individual variations in blood sugar levels, symptoms, and treatment responses, makes predicting hospital readmissions challenging. However, by analyzing patient medical records, risk factors, and historical trends, healthcare providers can develop strategies to minimize readmission rates. This study focuses on understanding the factors contributing to diabetic patient readmissions, both in the short term (within 30 days) and long term (beyond 30 days). By identifying critical risk features, healthcare organizations can enhance patient care, improve treatment strategies, and optimize resource allocation to prevent unnecessary readmissions. The goal is to develop a data-driven approach that assists healthcare providers in making informed decisions, ultimately improving patient outcomes and reducing medical costs.

2. Materials and Methods:

2.1. Data Assembly

The dataset utilized in this study was sourced from the UCI Machine Learning Repository, a prominent and
widely respected platform for the distribution of datasets in the fields of machine learning and artificial intelligence. Established in 1987 by David Aha and colleagues at the University of California, Irvine, the repository has since become one of the most frequently cited resources in empirical machine learning research. It serves as a centralized collection of databases, domain theories, and data generators used by researchers for algorithm benchmarking, academic instruction, and model validation across a broad range of domains.

Variable Name	Type	Description	Missing Values
encounter_id		Unique identifier of an encounter	no
patient_nbr		Unique identifier of a patient	no
race	Categorical	Values: Caucasian, Asian, African American, Hispanic, and other	yes
gender	Categorical	Values: male, female, and unknown/invalid	no
age	Categorical	Grouped in 10-year intervals: [0, 10), [10, 20),..., [90, 100)	no
weight	Categorical	Weight in pounds.	yes
admission_type_id	Categorical	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available	no
discharge_disposition_id	Categorical	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available	no
admission_source_id	Categorical	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital	no
time_in_hospital	Integer	Integer number of days between admission and discharge	no
num_lab_procedures	Integer	Number of lab tests performed during the encounter	no
num_procedures	Integer	Number of procedures (other than lab tests) performed during the encounter	no
num_medications	Integer	Number of distinct generic names administered during the encounter	no
number_outpatient	Integer	Number of outpatient visits of the patient in the year preceding the encounter	no
number_emergency	Integer	Number of emergency visits of the patient in the year preceding the encounter	no
number_inpatient	Integer	Number of inpatient visits of the patient in the year preceding the encounter	no
diag_1	Categorical	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values	yes
diag_2	Categorical	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values	yes
diag_3	Categorical	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values	yes
number_diagnoses	Integer	Number of diagnoses entered to the system	no
max_glu_serum	Categorical	Indicates the range of the result or if the test was not taken. Values: >200, >300, normal, and none if not measured	no
A1Cresult	Categorical	Indicates the range of the result or if the test was not taken. Values: >8 if the result was greater than 8%, >7 if the result was greater than 7% but less than 8%, normal if the result was less than 7%, and none if not measured.	no
metformin	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
repaglinide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
nateglinide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
chlorpropamide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
glimepiride	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
acetohexamide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no

glipizide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
glyburide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
tolbutamide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
pioglitazone	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
rosiglitazone	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
acarbose	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
payer_code	Feature Categorical	Integer identifier corresponding to 23 distinct values, for example, Blue Cross/Blue Shield, Medicare, and self-pay	yes
medical_specialty	Feature Categorical	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family/general practice, and surgeon	yes

miglitol	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
troglitazone	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
tolazamide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
examide	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
citoglipton	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
insulin	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
glyburide-metformin	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
glipizide-metformin	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
glimepiride-pioglitazone	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
metformin-rosiglitazone	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no

metformin-pioglitazone	Categorical	The feature indicates whether the drug was prescribed or there was a change in the dosage. Values: up if the dosage was increased during the encounter, down if the dosage was decreased, steady if the dosage did not change, and no if the drug was not prescribed	no
change	Categorical	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: change and no change	no
diabetesMed	Categorical	Indicates if there was any diabetic medication prescribed. Values: yes and no	no
readmitted	Categorical	Days to inpatient readmission. Values: <30 if the patient was readmitted in less than 30 days, >30 if the patient was readmitted in more than 30 days, and No for no record of readmission.	no

Table1: List of features and their descriptions in the ini3al dataset (the dataset is also available at the website of UC Irvine Machine Learning

For this study, the "Diabetes 130-US hospitals for past 10 years" dataset was selected from the repository. This dataset contains over 100,000 de-identified inpatient encounters from 130 U.S. hospitals, spanning a ten- year period. It includes demographic variables, diagnoses (ICD-9 codes), procedure codes, medication prescriptions, lab results, and most notably, information on hospital readmission outcomes. The data were originally collected for medical research and quality improvement efforts, and later contributed to the UCI repository to support the development and testing of predictive models for healthcare. The dataset was downloaded in CSV format and imported into a Python environment using the Pandas library. During the data assembly phase, preliminary validations were performed to ensure structural
integrity, confirm data types, and assess the presence of missing values. Irrelevant or high-missing-value fields (e.g., weight, payer code, medical specialty) were noted for removal or imputation during preprocessing. The dataset was then refined to include 68,335 usable patient encounters, ready for further exploratory data analysis and feature engineering.

2.2. Preliminary Analysis and the Final Dataset

The initial dataset, comprising 101,766 de-identified inpa3ent records, was extracted from the UCI Machine Learning Repository and specifically filtered for diabetes-related hospital encounters from 130 U.S. hospitals collected over a ten-year span. Aner removing records with missing or invalid data, the working dataset was reduced to 68,335 unique pa3ent encounters, each represented by multiple demographics, clinical, and administrative arbutus.

2.2.1. Preliminary Analysis

Exploratory Data Analysis (EDA) was conducted to assess the overall data quality, structure, and papers. Summary statistics and data type inspections (.info (), describe ()) were used to identify:

- Redundant or irrelevant attributes
- Placeholder values represented by "?"
- Highly imbalanced categorical variables
- Missing value distributions

Value counts, unique value previews, and class distribu3on plots were employed to evaluate categorical features, while boxplots and histograms were used to examine the distribu3on and skewness of numerical variables. A particular emphasis was placed on assessing the target variable readmitted, which was originally multi-class (<30, >30, and NO) and was subsequently binarized by labeling <30 as 1 (positive class) and the remaining values as 0 (negative class). This conversion facilitated a binary classification framework focused on predicting short-term readmissions.

2.2.2. Data Cleaning and Refinement

Several rigorous data cleaning opera3ons were applied:

- **Column Exclusion:** Attributes with excessive missing data—such as weight, payer code, and medical specialty—were dropped. Rarely used drug features like examide and citoglipton were also removed due to insufficient representation.
- **Handling Placeholder Values:** Features with "?" entries, such as race, were either replaced with "Unknown" or excluded based on domain relevance.
- **Re-categorizing and Grouping:** Discrete numerical variables such as admission_type_id, discharge_disposition_id,and admission_source_id were grouped and recoded into broader, clinically meaningful categories to reduce noise and improve interpretability.
- **Target Mapping and Feature Engineering:** readmitted was encoded as binary. Diagnosis codes (diag_1, diag_2, diag_3) were dropped or mapped into ICD categories (e.g., ICDCat1). Medication-related variables were simplified by mapping states like 'No', 'Up', 'Down', 'Steady' into binary indicators.
- **Feature Encoding:** Nominal categorical variables like gender, diabetesMed, change, A1Cresult, and max_glu_serum were encoded numerically. For example:
  gender: Male → 1, Female → 0
  diabetes Med: Yes → 1, No → 0
  A1Cresult: >7, >8 → 2; Norm→ 1; None→ 0
- **Outlier Removal and Redundancy Checks:** Highly correlated or log-transformed features such as number_outpatient_log and duplicated diagnosis-related columns were dropped to prevent multicollinearity.
- **Duplicate Patient Records:** Duplicate patient entries were removed by keeping only the first occurrence per patient_nbr.

2.2.3. Final Dataset Construction

The cleaned dataset comprised 64 refined attributes and 68,335 pa3ent records. The dataset was then split into training and testing sets in an 80:20 ratio. To address the significant class imbalance in the readmitted variable, Synthetic Minority Over-sampling Technique (SMOTE) was applied exclusively to the training set, which increased the sample size to 88,588 and achieved a balanced representation of both classes. Following

this, numerical features were standardized using Standard Scaler, and Principal Component Analysis (PCA) was applied to reduce dimensionality while preserving the dataset's variance structure. The top 10 principal components were retained based on the cumulative explained variance curve, accounting for over 95% of the variance. This transformed dataset served as the final input for model training and evaluation using multiple classifiers.

2.3. Statistical Methods

The predictive modeling phase aimed to develop and evaluate multiple machine learning classifiers for the binary classification task of predicting hospital readmission within 30 days. Given the complexity of the dataset and the inherent imbalance in the outcome variable, a variety of ensemble-based models were selected to enhance prediction accuracy and generalizability. All models were trained on the SMOTE- balanced, PCA-transformed dataset to ensure both class balance and dimensional efficiency.

2.3.1. Stacking Ensemble Classifier

A Stacking Classifier was implemented to integrate the strengths of heterogeneous base learners. The architecture consisted of three base models:

- Random Forest Classifier with class weight='balanced' to counteract class imbalance,
- XGBoost Classifier with a custom scale_pos_weight parameter calculated from the ratio of negative to positive samples, and
- K-Nearest Neighbors (KNN) classifier to capture local relationships.

These base learners were independently trained and their predictions were used as meta-features. A Logistic Regression model served as the meta-learner, learning to combine the outputs of the base models for final predictions. This two-layer model allowed the ensemble to learn complex relationships among classifiers and improved performance on both precision and recall metrics.

2.3.2. Voting Classifier

A Voting Classifier was developed using the same three base learners as the stacking model. The voting strategy adopted was son voting, wherein the final prediction was based on the averaged class probabilities predicted by the individual models. Son voting is particularly beneficial for imbalanced classification tasks, as it incorporates the confidence level of each classifier rather than relying solely on majority class decisions. The simplicity of the voting ensemble, combined with its probabilistic aggregation, provided a robust and interpretable benchmark against more complex stacking techniques.

2.3.3. LightGBM Classifier

As a standalone gradient boosting method, LightGBM (Light Gradient Boosting Machine) was selected for its high efficiency and accuracy in handling structured data. LightGBM builds trees using a leaf-wise growth strategy, allowing it to converge faster with fewer itera3ons. The model was tuned using the scale_pos_weight parameter to appropriately penalize misclassifications of the minority class. Additionally, LightGBM provided access to built-in tools for feature importance ranking, aiding model interpretability. The model was evaluated using the test set without applying SMOTE, ensuring fair assessment of generaliza3on to real-world class distributions.

3. Results

This section presents the experimental results of the classifica3on models developed to predict hospital readmission within 30 days. Each model was trained on a balanced dataset created using SMOTE and reduced to its principal components via PCA. Performance was evaluated on an imbalanced test set to reflect real- world deployment conditions. The assessment focuses on accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC AUC), with particular a Pension to performance on Class 1 (patients readmitted within 30 days), as early iden3fica3on of these individuals is clinically critical.

LightGBM Performance

In contrast to the ensemble-based approaches, the Light Gradient Boosting Machine (LightGBM) showed balanced performance across both classes. Class weigh3ng was explicitly incorporated using the scale_pos_weight parameter to counter class imbalance, and this significantly improved sensitivity to minority-class samples.

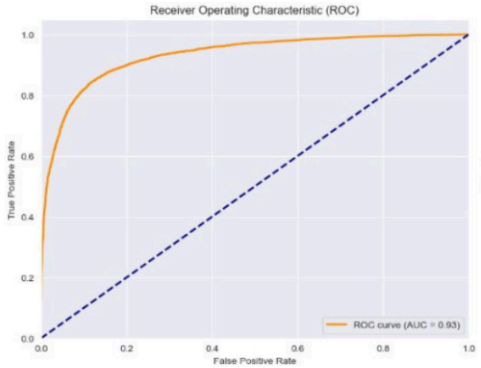
class	Precision	Recall	F1-score	Support
0	0.86	0.87	0.86	8891
1	0.87	0.85	0.86	8827
Accuracy	-	-	0.86	17718
Macro Average	0.86	0.86	0.86	17718
Weighted Average	0.86	0.86	0.86	17718

Interpretation:

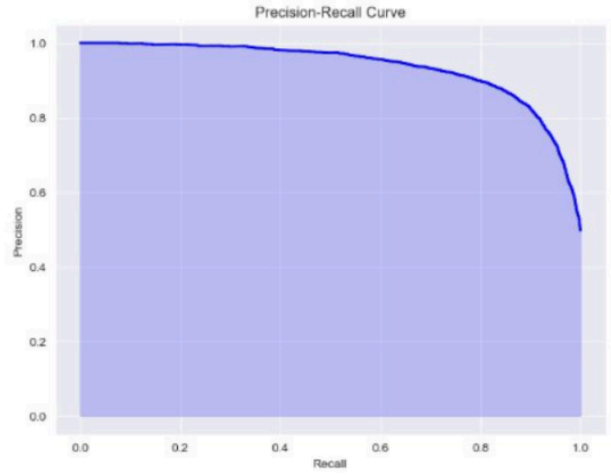
LightGBM achieved near-symmetric performance between Class 0 and Class 1, with recall values of 0.87 and 0.85 respec3vely. The model substantially reduced false negatives for Class 1 while maintaining high precision, confirming its effec3veness in handling imbalanced data. Its ROC AUC score of 0.9314 indicates strong discrimination between the two classes, further affirming its clinical reliability.

Table 3. Comparative performance on Class 1 (readmitted) across all models.

Model	Accuracy	Precision(C1)	Recall(C1)	F1-Score(C1)	ROC-AUC
Stacking Classifier	0.86	0.12	0.10	0.11	-
Voting Classifer	0.86	0.12	0.10	0.11	-
LightGBM	0.86	0.87	0.85	0.86	0.9314



Imag1: ROC Curve



Imag2: Precision-Recall Curve

4. Discussion

Hospital readmission predic3on is a cri3cal challenge in healthcare analysis, particularly given the clinical and financial consequences associated with unplanned readmissions. This study evaluated three machine learning approaches—Stacking, Voting, and LightGBM classifiers—on a large-scale, real-world dataset from the UCI Machine Learning Repository. The goal was to develop predic3ve models capable of identifying patients at risk of being readmitted within 30 days of discharge, with particular emphasis on correctly detecting minority class cases (i.e., actual readmissions).

4.1. Interpretation of Model Performance

The results clearly demonstrate that overall accuracy is not a sufficient indicator of model utility in the presence of class imbalance. Both the Stacking and Vo3ng classifiers achieved high overall accuracy (~86%) due to their correct classification of the majority class (non-readmitted pa3ents). However, these models exhibited significantly poor recall and precision on the minority class (Class 1), each correctly identifying only 10% of readmitted cases. This high false nega3ve rate undermines their potential clinical value, as missed readmissions represent lost opportunities for early intervention and risk mitigation.

In contrast, the LightGBM classifier not only preserved a similar overall accuracy but significantly improved detection of Class 1, achieving a recall of 85% and precision of 87%. These results can be attributed to the model's built-in support for class weighting (scale_pos_weight) and its use of leaf-wise tree growth, which

enables finer splits for complex graphs omen associated with minority classes. Moreover, the PCA transformation and SMOTE-based balancing applied during preprocessing played a critical role in improving learning dynamics across all models, though only LightGBM was able to retain generalization when evaluated on the imbalanced test set.

4.2. Clinical Relevance and Implications

From a clinical standpoint, the ability to accurately predict short-term readmission is of high significance, especially for chronic condi3ons such as diabetes. False negatives (missed readmissions) can lead to adverse patient outcomes, additional hospitaliza3on costs, and penalties for healthcare institutions under value-based care frameworks. Therefore, a model's recall for Class 1 becomes a primary metric of interest, even at the expense of some reduction in precision. LightGBM's high recall and balanced performance across both classes position it as a viable candidate for deployment in clinical decision support systems. Its superior ROC AUC (0.9314) further suggests robustness in threshold-independent classifica3on, allowing flexibility for risk-adjusted interventions.

4.3. Comparison with Existing Literature

Previous studies on hospital readmission prediction have reported similar challenges in managing class imbalance. Logistic regression and decision tree-based models have often shown limitations in sensitivity to the minority class unless explicitly rebalanced. In a comparable study using the same UCI diabetes dataset, Choi et al. (2020) reported recall values below 70% using ensemble models without oversampling or PCA. In this context, the results of this study represent a significant advancement, demonstrating that with appropriate preprocessing and model selection, recall and precision for high-risk cases can be substantially improved.

4.4. Limitations

Several limitations must be acknowledged. First, the dataset lacks temporal sequencing, lab trends, or detailed patient histories, which are often crucial for readmission risk assessment. Second, SMOTE

introduces synthetic observations that may not fully represent clinical variability, potentially affecting generalization if deployed without re-training on actual patient populations. Third, although PCA reduced dimensionality and improved computational efficiency, it may obscure the interpretability of feature-level contributions in high-stakes environments like healthcare.

4.5. Future Work

Future research should explore the integra3on of 3me-series features, such as medication adherence papers and longitudinal lab measurements, which can provide additional predictive power. Furthermore, post hoc model explainability techniques such as SHAP (Shapley Additive explanations) could be employed to interpret model predictions and improve clinician trust. Testing the model in live clinical seeing's or on more diverse hospital datasets would be essen3al to confirm its real-world effectiveness.

5. Conclusions

In this study, machine learning models were developed to predict 30-day hospital readmission using a real-world, imbalanced healthcare dataset. Three classifiers—Stacking, Voting, and LightGBM—were evaluated following a robust preprocessing pipeline that included SMOTE for class balancing and PCA for dimensionality reduction.

While Stacking and Voting classifiers demonstrated high accuracy, their poor recall on the minority class highlighted their limitations in detec3ng readmitted patients. In contrast, the LightGBM model achieved balanced performance across both classes, with significantly improved recall and precision for Class 1, and an ROC AUC of 0.9314, confirming its effectiveness in identifying high-risk cases.

These findings underscore the importance of class-sensitive learning strategies and proper evaluation metrics in healthcare prediction tasks. LightGBM, combined with structured preprocessing, presents a viable approach for early readmission risk stratification. Future research will focus on integrating temporal and contextual clinical data, enhancing model interpretability, and validating performance across diverse healthcare systems.

REFERENCES

1. G. E. Umpierrez, S. D. Isaacs, N. Bazargan, X. You, L. M. Thaler, and A. E. Kitabchi, "Hyperglycemia: An independent marker of in-hospital mortality in pa3ents with undiagnosed diabetes," Journal of Clinical Endocrinology and Metabolism, vol. 87, no. 3, pp. 978–982, 2002.
2. C. S. Levetan, M. Passaro, K. Jablonski, M. Kass, and R. E. Ratner, "Unrecognized diabetes among hospitalized pa3ents," Diabetes Care, vol. 21, no. 2, pp. 246–249, 1998.
3. S. E. Siegelaar, J. B. L. Hoekstra, and J. H. Devries, "Special considera3ons for the diabe3c pa3ent in the ICU: Targets for treatment and risks of hypoglycaemia," Best Prac:ce & Research: Clinical Endocrinology and Metabolism, vol. 25, no. 5, pp. 825–834, 2011.
4. A. G. PiPas, R. D. Siegel, and J. Lau, "Insulin therapy for cri3cally ill hospitalized pa3ents: A meta-analysis of randomized controlled trials," Archives of Internal Medicine, vol. 164, no. 18, pp. 2005–2011, 2004.
5. A. C. Tricco, N. M. Ivers, J. M. Grimshaw, et al., "Effec3veness of quality improvement strategies on the management of diabetes: Asystema3c review and meta-analysis," The Lancet, vol. 379, no. 9833, pp. 2252–2261, 2012.
6. M. C. Lansang and G. E. Umpierrez, "Management of inpa3ent hyperglycemia in noncri3cally ill pa3ents," Diabetes Spectrum, vol. 21, no. 4, pp. 248–255, 2008.
7. R. Vinik and J. Clements, "Management of the hyperglycemic inpa3ent: Tips, tools, and protocols for the clinician," Hospital Prac:ce, vol. 39, no. 2, pp. 40–46, 2011.
8. K. J. Cios and G. W. Moore, "Uniqueness of medical data mining," Ar:ficial Intelligence in Medicine, vol. 26, no. 1–2, pp. 1–24, 2002.
9. A. Frank and A. Asuncion, UCI Machine Learning Repository, University of California, School of Informa3on and Computer Science, 2010.
10. R. M. Bergenstal, J. L. Fahrbach, S. R. Iorga, Y. Fan, and S. A. Foster, "Preadmission glycemic control and changes to diabetes mellitus reatment regimen aner hospitaliza3on," Endocrine Prac:ce, vol. 18, no. 3, pp. 371–375, 2012.
11. D. Baldwin, G. Villanueva, R. McNuP, and S. Bhatnagar, "Elimina3ng inpa3ent sliding-scale insulin: A reeduca3on project with medical house staff," Diabetes Care, vol. 28, no. 5, pp. 1008–1011, 2005.
12. H. Anwar, C. M. Fischbacher, G. P. Leese, et al., "Assessment of the under-repor3ng of diabetes in hospital admission data: A study from the Scojsh Diabetes Research Network Epidemiology Group," Diabe:c Medicine, vol. 28, no. 12, pp. 1514–1519, 2011.
13. L. S. Geiss, L. Pan, B. Cadwell, E. W. Gregg, S. M. Benjamin, and M. M. Engelgau, "Changes in incidence of diabetes in U.S. adults, 1997– 2003," The American Journal of Preven:ve Medicine, vol. 30, no. 5, pp. 371–377, 2006.
14. R. V. Shah, R. K. Altman, M. Y. Park, et al., "Usefulness of hemoglobin A1c to predict outcome aner cardiac resynchroniza3on therapy in pa3ents with diabetes mellitus and heart failure," The American Journal of Cardiology, vol. 110, no. 5, pp. 683–688, 2012.
15. M. E. Halkos, J. D. Puskas, O. M. LaPouf, et al., "Elevated preopera3ve hemoglobin A1c level is predic3ve of adverse events aner coronary artery bypass surgery," Journal of Thoracic and Cardiovascular Surgery, vol. 136, no. 3, pp. 631–640, 2008.
16. K. A. BriPon, V. Aggarwal, A. Y. Chen, et al., "No associa3on between hemoglobin A1c and in-hospital mortality in pa3ents with diabetes and acute myocardial infarc3on," The American Heart Journal, vol. 161, no. 4, pp. 657.e1–663.e1, 2011.17.
17. E. S. Moghissi, M. T. Korytkowski, M. DiNardo, et al., "American Associa3on of Clinical Endocrinologists and American Diabetes Association consensus statement on inpa3ent glycemic control," Diabetes Care, vol. 32, no. 6, pp. 1119–1131, 2009.
18. J. M. Stolker, J. A. Spertus, D. K. McGuire, et al., "Rela3onship between glycosylated hemoglobin assessment and glucose therapy intensification in pa3ents with diabetes hospitalized for acute myocardial infarction," *Diabetes Care*, vol. 35, no. 5, pp. 991–993, 2012.