International Journal of Sustainable Applied Science and Engineering (IJSASE)

Contents lists available at IJSASE
International Journal of Sustainable Applied Science and Engineering
journal homepage: IJSASE
Volume 3, No. 2, 2026

Predicting 30-Day Hospital Readmission in Patients with Diabetes Using Explainable Machine Learning: A Comparative Study of Missing-Data Strategies

Delaram Alizadeh [a], Arman Kavoosi Ghafi [b]

[a] *Department of Computer Engineering, Ahvaz Women's College, National University of Skills, Ahvaz, Iran.*
[b] *Department of Computer Engineering, Isl.C., Islamic Azad University, Islamshahr, Iran.*

ARTICLE INFO

Received: 2026/08/28

Revised: 2026/09/10

Accepted: 2026/10/29

Keywords:

Hospital readmission; diabetes mellitus; machine learning; CatBoost; probability calibration; explainable artificial intelligence; electronic health records.

ABSTRACT

Predicting 30-day hospital readmission among patients with diabetes is challenging because of class imbalance and heterogeneous clinical data. This study evaluated machine-learning models with attention to discrimination, calibration, and explainability. We analyzed 99,343 eligible encounters from 69,990 patients in the Diabetes 130-US Hospitals dataset collected between 1999 and 2008. Logistic regression, XGBoost, and CatBoost were compared using three-fold patient-grouped cross-validation in the development set and alternative missing-data strategies. The primary model was selected based on average precision, and its scores were calibrated using out-of-fold development predictions. Performance was evaluated on a patient-disjoint internal test set that had been viewed during an earlier project iteration. Patient-level bootstrap confidence intervals and SHapley Additive exPlanations (SHAP) contributions were calculated. Of 20,019 test encounters, 2,289 (11.4%) had a recorded readmission in less than 30 days. CatBoost achieved an area under the receiver operating characteristic curve of 0.662 (95% confidence interval: 0.649–0.674) and an average precision of 0.215 (95% confidence interval: 0.197–0.233). At a development-selected threshold, precision was 0.183, recall was 0.543, and the F1-score was 0.274. The calibrated Brier score was 0.097. Prior inpatient utilization had the largest mean absolute SHAP contribution in a sample of 500 test encounters. CatBoost showed moderate discrimination, with a substantial trade-off between identifying readmissions and generating false-positive predictions. The results are exploratory because the internal test set had been previously viewed. Independent evaluation in contemporary populations is needed before clinical use.

1. Introduction

[a] Corresponding author email address: Delaram.alizadeh@gmail.com (Delaram Alizadeh), arman_k69@yahoo.com (Arman Kavoosi Ghafi).
DOI: https://doi.org/10.22034/ijsase.v3i2.239
Available online 2026/11/01
Licensee System Analytics. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).
3060-6349/BGSA Ltd.

Diabetes mellitus is a chronic condition that often requires continuous treatment, regular monitoring, and coordination across different healthcare settings. Hospitalized patients with diabetes may have multiple comorbidities, complex medication regimens, and care needs that persist after discharge. Consequently, early hospital readmission is an important concern for patients and healthcare systems because it may be associated with additional treatment burden, disrupted recovery, and increased use of healthcare resources. Although not every readmission is preventable, identifying patients at elevated risk may help clinicians consider appropriate follow-up and discharge planning ([5]; [9], [19-25]).

The growing availability of electronic health records (EHRs) has created opportunities to investigate readmission risk using routinely collected clinical and administrative information. Such records can include demographic characteristics, diagnoses, medication use, previous healthcare encounters, and details of the current hospitalization. The Diabetes 130-US Hospitals dataset is a widely used resource for this purpose. It contains hospital encounters recorded between 1999 and 2008 across 130 US hospitals and integrated delivery networks, with each row representing an encounter rather than necessarily a unique patient [2]. Nevertheless, routinely collected data may be incomplete or heterogeneous, and their usefulness depends on how variables are recorded and processed [7].

Machine learning methods have increasingly been applied to clinical prediction because they can simultaneously incorporate multiple variables and, depending on the model, capture nonlinear relationships among predictors [1]. In diabetes readmission research, Emi-Johnson and Nkrumah [4] compared logistic regression, random forest, XGBoost, and a deep neural network using the Diabetes 130-US Hospitals dataset. In their analysis, XGBoost achieved the highest reported receiver operating characteristic area under the curve (ROC-AUC) among the evaluated models. These findings illustrate the potential of ensemble methods, while also showing that predicting early readmission remains a challenging task. Reported performance should be interpreted in relation to each study's cohort, preprocessing procedures, and evaluation design.

In another health-related application, Ghafi et al. [10] investigated graph-based epidemic forecasting using noisy and incomplete surveillance data, illustrating the importance of data quality in predictive modeling.

Research in other patient populations also demonstrates the importance of defining the clinical setting and prediction horizon. For example, Jiang et al. [6] investigated 90-day readmissions

among critically ill patients with heart failure and used explainability methods to identify influential predictors. Such studies provide useful methodological insights; however, differences in patient populations, outcomes, and follow-up periods limit direct numerical comparisons with models developed for 30-day readmission in patients with diabetes.

Recent work on diabetes readmission has moved beyond a sole emphasis on classification accuracy. Guo et al. [5] examined class imbalance and used several explainability approaches, including SHapley Additive exPlanations (SHAP), to investigate factors associated with model predictions. Similarly, Salim and Ibrahim [9] evaluated several machine learning models on the Diabetes 130-US Hospitals dataset and incorporated nested cross-validation, probability calibration, bootstrap analysis, and SHAP. These studies show that both predictive performance and methodological transparency are relevant when evaluating readmission models; they also provide an important context for further comparative analyses.

Several characteristics of hospital data make such evaluation particularly important. First, early readmission is less common than the combined category of later readmission or no recorded readmission. In an imbalanced dataset, overall accuracy can therefore appear high even when a model identifies relatively few early readmissions. Measures such as precision, recall, F1-score, and average precision provide complementary information about the identification of the minority class. Second, missing values may arise from differences in documentation or clinical practice. The choice between removing variables, imputing missing values, and treating missingness as a separate category can affect the information available to a model and should be examined explicitly ([7]; [9]; [25-30]).

The unit used to divide data is another important consideration. When a dataset contains multiple encounters for the same patient, a record-level split may place those encounters in both the training and evaluation sets. This overlap can make it difficult to assess how the model performs on patients not represented in the training data. A patient-grouped split reduces this particular source of overlap, although it does not by itself establish generalizability across hospitals or time periods. In addition, the availability of a predictor at the intended time of prediction should be considered. Some diagnostic or administrative codes may not be finalized at discharge, creating a potential timing-related concern ([3]; [30-35]).

Similar concerns about validation and generalizability have been observed in other healthcare applications of artificial intelligence. In a systematic review of COVID-19-related studies,

Jodayree et al. [11] found that most included studies relied primarily on internal validation, highlighting the importance of evaluating models across independent datasets and clinical settings. Model assessment should also distinguish discrimination from probability calibration. A model may rank patients by predicted risk reasonably well while producing probability estimates that do not closely match observed event frequencies. Calibration curves and the Brier score can therefore complement discrimination measures when examining predicted probabilities. Explainability methods can provide an additional view of how features contribute to model outputs, although feature contributions should not be interpreted as evidence of causal effects. Together, transparent reporting of data preparation, model evaluation, and limitations is important for assessing the relevance of machine learning findings in healthcare ([5]; [8]; [9], [35-38]).

Against this background, the present study investigates the prediction of 30-day hospital readmission among patients with diabetes using the Diabetes 130-US Hospitals dataset. Logistic regression, XGBoost, and CatBoost are compared under different strategies for handling selected missing variables. Model configurations are assessed using three-fold patient-grouped cross-validation in the development data, with average precision used for model selection. The analysis also examines probability calibration, threshold-based classification performance, patient-level bootstrap confidence intervals, SHAP feature contributions, and sensitivity to selected coded predictors.

The study is intended as a comparative evaluation of these modeling and assessment choices, rather than introducing a new learning algorithm. The patient-disjoint internal test set had been examined in an earlier analysis; its results are therefore reported as a previously viewed internal evaluation, not as untouched or external validation. This distinction is important when interpreting the findings and considering their potential application beyond the present dataset.

2. Related Work

Hospital readmission prediction has been investigated using statistical models and machine learning methods across different patient populations. In an early analysis of diabetes-related hospital records, Strack et al. [15] used multivariable logistic regression to investigate the relationship between HbA1c measurement and early readmission. Their analysis included approximately 70,000 selected inpatient encounters and showed that the association between HbA1c measurement and readmission varied by primary diagnosis. The authors emphasized that the observational design did not establish a causal relationship.

Beyond diabetes-specific studies, Huang et al. [13] reviewed 43 US studies that applied machine learning to hospital readmission prediction. Tree-based methods, neural networks, regularized logistic regression, and support vector machines were among the commonly evaluated approaches. The review reported considerable variation in model performance across studies, with a median area under the receiver operating characteristic curve (ROC-AUC) of 0.68. These findings indicate that readmission prediction has been studied using a range of methods and clinical populations, making the evaluation context essential when interpreting reported results.

Other studies have examined readmission in specific clinical settings. Slezak et al. [14] compared several methods for predicting readmission after total joint replacement and reported an ROC-AUC of 0.67 for a Light Gradient Boosting Machine model using their selected predictors. Although this study is informative for readmission modeling, its surgical population and predictors differ from those of diabetes-specific readmission studies. Therefore, its results should be interpreted in the context of the methodology rather than as a direct benchmark for the present research.

Using the Diabetes 130-US Hospitals dataset, Emi-Johnson and Nkrumah [4] compared logistic regression, random forest, XGBoost, and a deep neural network for predicting early readmission. Their reported ROC-AUC for XGBoost was 0.667. The authors also used SHAP to examine the contribution of variables such as previous inpatient visits, medication-related characteristics, and hospitalization duration. Their results illustrate the value of reporting both discrimination and class-specific performance: a model can achieve a moderate ROC-AUC while still identifying only a limited proportion of positive cases at a particular decision threshold.

Guo et al. [5] also investigated diabetes-related readmission using the same data source. They evaluated multiple classifiers and approaches for addressing class imbalance and applied permutation importance, partial dependence plots, and SHAP to examine model behavior. The authors reported that previous inpatient admissions and emergency visits were influential in their analysis. Their study illustrates how explainability methods can supplement predictive metrics, although the resulting feature associations should be interpreted in the context of the selected model and preprocessing strategy.

Salim and Ibrahim [9] compared several models on the Diabetes 130-US Hospitals dataset and incorporated nested cross-validation, probability calibration, bootstrap analysis, and SHAP. In the results section of their article, the authors report that the XGBoost model achieved a nested cross-

validation ROC-AUC of 0.664. The authors acknowledged that their cross-validation procedure did not enforce patient-level separation, meaning that encounters belonging to the same patient could appear in different folds. They also noted that their decision curve analysis used full-dataset predictions and that external validation was still required. These limitations provide a specific motivation for examining patient-grouped evaluation, while recognizing that patient grouping alone does not eliminate all sources of bias.

More recently, Afridi et al. [12] developed a diabetes readmission framework that uses patient-level partitioning, class imbalance handling, feature selection, and explainability methods. Their hybrid CNN–LSTM model achieved a reported ROC-AUC of 0.861 and a precision–recall area under the curve (PR-AUC) of 0.420 on their held-out test set. However, their study excluded a substantial proportion of the original encounters and used a different feature set that included discharge disposition. These design differences, together with different model selection and evaluation procedures, mean that its numerical results cannot be interpreted as a direct comparison with those of the present study. The authors also reported that external validation had not been performed.

Table 1 summarizes these representative studies and places the present work in their methodological context.

Table 1. Summary of representative studies on hospital readmission prediction

Study	Population or data	Methods and evaluation	Main finding or consideration
Strack et al. [15]	Approximately 70,000 selected diabetes-related inpatient encounters	Multivariable logistic regression	Examined the association between HbA1c measurement and early readmission; the observational analysis did not establish causality.
Huang et al. [13]	Scoping review of 43 US readmission studies	Review of machine learning methods and reported ROC-AUC	Reported a median ROC-AUC of 0.68 across studies with different populations and designs.
Slezak et al. [14]	Patients undergoing total joint replacement	LightGBM and comparator models; train–test split and cross-validation	Reported ROC-AUC of 0.67 for the LightGBM model; the surgical population differs from the present cohort.

Study	Population or data	Methods and evaluation	Main finding or consideration
Emi-Johnson and Nkrumah [4]	Diabetes 130-US Hospitals dataset	Logistic regression, random forest, XGBoost, and a neural network; SHAP analysis	Reported XGBoost ROC-AUC of 0.667; class-specific results illustrate the difficulty of identifying early readmissions.
Guo et al. [5]	Diabetes 130-US Hospitals dataset	Multiple classifiers, class-imbalance strategies, SHAP, and partial dependence plots	Identified prior inpatient admissions and emergency visits as influential features in their analysis.
Salim and Ibrahim [9]	Diabetes 130-US Hospitals dataset	Multiple models, nested cross-validation, calibration, bootstrap analysis, and SHAP	Reported XGBoost nested cross-validation ROC-AUC of 0.664; patient-level separation was not enforced.
Afridi et al. [12]	Diabetes 130-US Hospitals dataset after study-specific exclusions	Patient-level partitioning, multiple models, class-imbalance handling, SHAP, and LIME	Reported CNN–LSTM ROC-AUC of 0.861 and PR-AUC of 0.420; cohort and feature selection differed from the present study.
Present study	Diabetes 130-US Hospitals dataset after eligibility exclusions	Patient-grouped development cross-validation; logistic regression, XGBoost, and CatBoost; calibration and SHAP	CatBoost achieved an ROC-AUC of 0.6616 and an AP of 0.2149 on a previously seen internal test set.

Note. The results are presented as reported by the respective studies. They are not a direct ranking of the methods because eligibility criteria, predictors, class handling, data splits, and evaluation procedures differ. AP and PR-AUC may also be calculated differently and should not automatically be treated as identical measures.

Taken together, previous studies demonstrate that model comparison, calibration, explainability, and patient-level separation have all been considered in the readmission literature. The present study builds on this work by examining a specific combination of three classifiers and missing-data strategies under a patient-grouped development procedure. Its purpose is to characterize performance and methodological trade-offs on the selected historical cohort. Because the internal

test set was previously inspected and no external validation was performed, the findings should be interpreted as exploratory internal evidence rather than confirmation of clinical generalizability.

3. Materials and Methods

3.1. Study Design and Dataset

This retrospective study used the Diabetes 130-US Hospitals for Years 1999–2008 dataset, available from the University of California, Irvine Machine Learning Repository. The original dataset contains 101,766 hospital encounters from 130 US hospitals and integrated delivery networks. Each row represents a hospital encounter, and an individual patient may contribute more than one encounter [2].

The source dataset includes encounters in which diabetes was recorded as a diagnosis, the hospital stay lasted between 1 and 14 days, and laboratory tests and medications were recorded. The present study focused on predicting whether a readmission would occur within 30 days following an eligible encounter.

Encounters with discharge disposition codes indicating death or hospice care were excluded. Specifically, codes 11, 13, 14, 19, 20, and 21 were removed. Following these exclusions, 99,343 encounters from 69,990 unique patients remained for analysis.

The binary target variable was defined as follows:

$$y_i = \begin{cases} 1, & \text{if encounter } i \text{ was followed by readmission in less than 30 days,} \\ 0, & \text{if readmission was recorded as more than 30 days or no readmission was recorded.} \end{cases} \quad (1)$$

Here, y_i denotes the observed outcome for encounter i. The negative class combines the original >30 and NO categories. The NO category indicates that no readmission was recorded in the dataset. The overall workflow is presented in Figure 1.

Table 2. Dataset characteristics and patient-grouped partitions

Characteristic	Encounters	Unique patients	Early readmissions, *n* (%)
Original dataset	101,766	—	—
Excluded death or hospice encounters	2,423	—	—
Eligible dataset	99,343	69,990	11,314 (11.39%)
Development set	79,324	55,987	9,025 (11.38%)
Previously viewed internal test set	20,019	14,003	2,289 (11.43%)

Note. The unit of outcome measurement was the hospital encounter. Patients were assigned to only one of the development or test partitions.

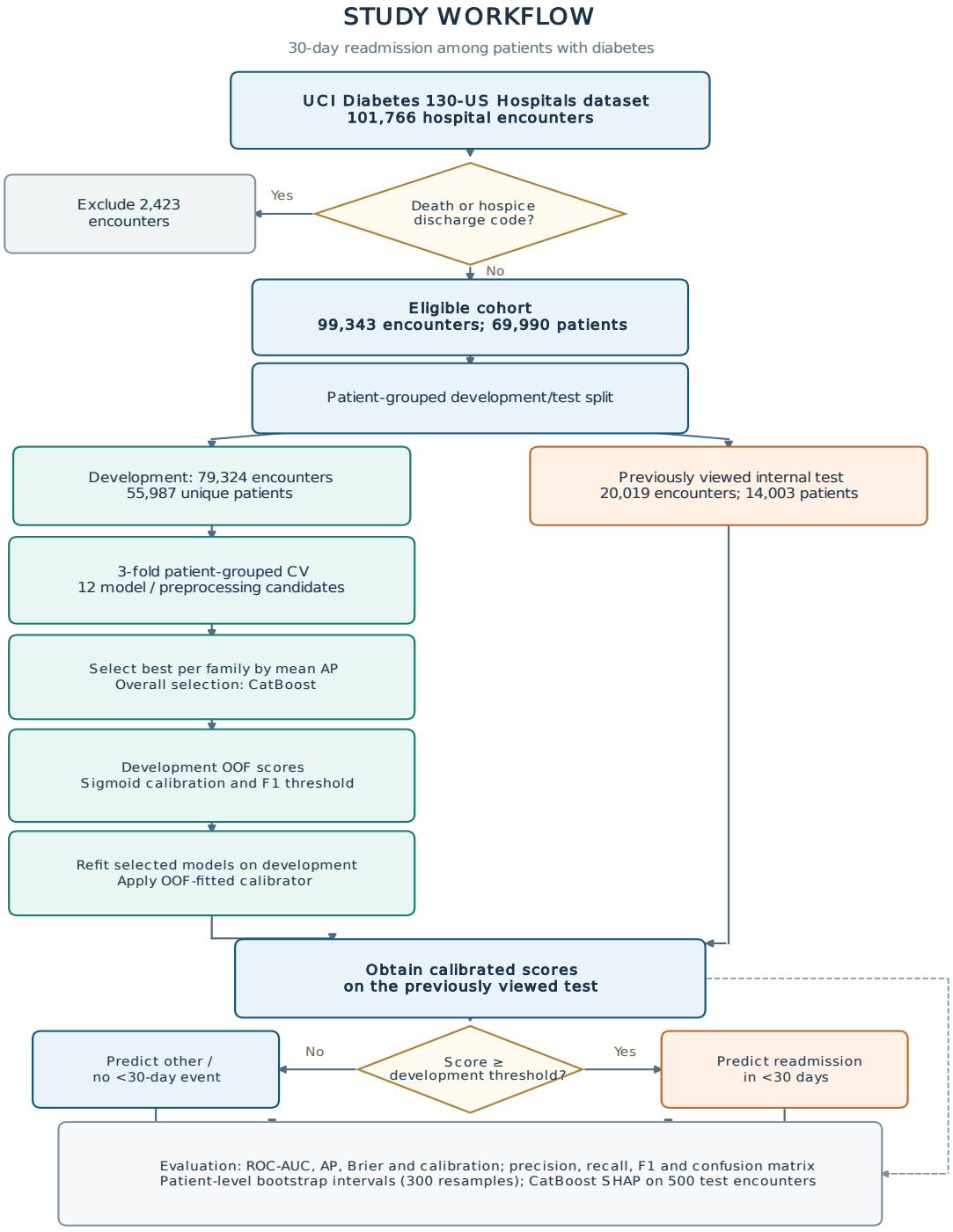

Figure 1. Conditional workflow of the patient-grouped diabetes readmission prediction study. OOF: out-of-fold; AP: average precision; SHAP: SHapley Additive exPlanations. The internal test set had been examined in an earlier analysis.

3.2. Data Preprocessing and Missing-Value Strategies

The variables encounter_id, patient_nbr, and readmitted were excluded from the predictors. The patient identifier was retained only for patient-grouped partitioning and resampling. The variable discharge_disposition_id was used for eligibility exclusions but was not included as a model predictor.

The weight variable was removed from all model configurations because 96.86% of its entries in the original dataset were coded as missing. The proportions of explicitly missing entries were also relatively high for medical_specialty and payer_code, at 49.08% and 39.56%, respectively.

Before applying the missing-data strategies, the full predictor set contained 45 variables: 8 numerical and 37 categorical. Entries coded as blank were treated as missing. For the variables A1Cresult and max_glu_serum, the original value None was retained as a separate category indicating that a test result was not recorded; it was not automatically merged with the category.

Three strategies were evaluated for medical_specialty and payer_code:

1. Drop high-missing variables: both variables were removed.
2. Most-frequent imputation: missing values in these two variables were replaced by the most frequent observed category in the corresponding training fold.
3. Missing category: missing values were retained as an explicit categorical level.

For logistic regression and XGBoost, numerical missing values were handled using median imputation fitted on the relevant training data. Categorical variables were one-hot encoded using only the training data. Numerical variables were additionally scaled for logistic regression. CatBoost accepted categorical variables directly and handled missing numerical values using its native processing.

For logistic regression and XGBoost, the one-hot encoder was fitted on each training fold with a minimum category frequency of 20. Less frequent categories were grouped, and categories not observed during training were ignored during transformation.

Table 3. Candidate configurations examined during model development

Configuration	Description	Number of candidates
Drop high-missing variables	Remove medical_specialty and payer_code; evaluate the three standard models	3

Configuration	Description	Number of candidates
Most-frequent imputation	Replace missing values in the two variables with the training-fold mode; evaluate the three standard models	3
Missing category	Preserve missing values as a separate category; evaluate the three standard models	3
Shallow models	Evaluate shallower CatBoost and XGBoost models with the missing-category strategy	2
Reduced coded features	Evaluate CatBoost after excluding selected diagnosis and payer variables	1
	Total	**12**

Note. For the reduced-coded configuration, diag_1, diag_2, diag_3, number_diagnoses, and payer_code was excluded. This configuration was used for sensitivity analysis and was not treated as proof that all timing-related leakage had been removed.

3.3. Patient-Grouped Partitioning and Cross-Validation

The eligible dataset was divided into development and internal test sets using a stratified patient-grouped procedure with a fixed random seed of 42. The split ensured that encounters from the same patient did not appear in both partitions.

The internal test set was the first fold of a shuffled, five-fold, stratified, patient-grouped split using a random seed of 42; the remaining folds formed the development set.

The development set contained 79,324 encounters from 55,987 patients, while the internal test set contained 20,019 encounters from 14,003 patients. The test set was examined in an earlier analysis and is therefore described as a previously viewed internal test set.

Within the development set, three-fold patient-grouped cross-validation was used to evaluate the 12 candidate configurations. For each fold, models were fitted to the training set of patients and evaluated on a separate validation set. Data-dependent preprocessing, including imputation and, where applicable, encoding, was fitted to the corresponding training fold.

For candidate configuration c, the mean validation average precision was calculated as:

$$\overline{AP}_c = \frac{1}{K}\sum_{k=1}^{K} AP_{c,k}, \qquad K = 3 \tag{2}$$

where$AP_{c,k}$ is the average precision for candidate c on validation fold k. The highest mean fold AP was used to select the best configuration within each model family and the overall model.

Average precision was chosen as the primary selection measure because early readmission was the minority outcome.

Patient-grouped partitioning reduces patient overlap between model development and evaluation. However, it does not establish temporal or external validity, and the availability of all coded predictors at the intended prediction time cannot be verified from this dataset alone [3].

3.4. Machine Learning Models

Three model families were evaluated: **logistic regression, XGBoost, and CatBoost**. Logistic regression served as a linear baseline, whereas XGBoost and CatBoost were gradient-boosted tree approaches. CatBoost was included because it can process categorical predictors directly; XGBoost was evaluated after categorical encoding ([16]; [18]).

Class imbalance was addressed through model-specific class weighting. Logistic regression used balanced class weights, CatBoost used automatic balanced class weights, and XGBoost used a positive-class weight calculated from the class distribution in the relevant training data. No synthetic oversampling was used.

The main model settings were as follows:

- **Logistic regression:** liblinear solver, balanced class weights, and a maximum of 600 iterations.
- **XGBoost:** 300 trees, learning rate of 0.06, and maximum tree depth of 5 for the standard configuration or 3 for the shallow configuration.
- **CatBoost:** 250 iterations, learning rate of 0.06, and tree depth of 6 for the standard configuration or 4 for the shallow configuration.

The standard CatBoost configuration used an L2 leaf regularization value of 3, while the shallow configuration used 6. The standard and shallow XGBoost configurations used L2 regularization values of 1 and 3, respectively. The same subsampling and column-sampling proportions of 0.85 were used for XGBoost.

3.5. Probability Calibration and Threshold Selection

For each selected model family, out-of-fold predictions were generated from the development data. A sigmoid calibration model was fitted using the logit of the model's uncalibrated score:

$$p_i^{cal} = \sigma\left(a \log \frac{s_i}{1-s_i} + b\right), \qquad \sigma(z) = \frac{1}{1+e^{-z}} \tag{3}$$

where s_i is the original model score for encounter i, a and b are parameters estimated by logistic regression, and p_i^{cal} is the calibrated score.

A separate patient-grouped three-fold procedure was applied to the development out-of-fold scores to obtain cross-calibrated scores for threshold selection. For each selected model, the classification threshold was chosen to maximize the F1-score on these development scores:

$$t^* = \arg\max_t F1\left(y_{\text{dev}}, \mathbb{1}\left[p_{\text{dev}}^{cal} \geq t\right]\right) \quad (4)$$

where t^* is the selected threshold and $\mathbb{1}[\cdot]$ is an indicator function. For the selected CatBoost configuration, the development-derived threshold was approximately 0.1202.

After model and threshold selection, each selected model was fitted on the full development set. A final sigmoid calibrator, fitted to the development out-of-fold scores, was then applied to the development-refitted model's scores on the internal test set.

The cross-calibration step was applied to previously generated out-of-fold model scores; the base models were not refitted within each calibration fold. Thus, this procedure should not be described as a fully nested calibration-and-threshold-selection design.

3.6. Performance Evaluation and Uncertainty

Model performance on the previously viewed internal test set was assessed using ROC-AUC, average precision, accuracy, precision, recall, specificity, F1-score, and Brier score. Precision, recall, specificity, and F1-score were calculated from the selected development thresholds:

$$\text{Precision} = \frac{TP}{TP+FP}, \qquad \text{Recall} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN+FP}, \qquad F1 = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \quad (6)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Average precision summarizes the precision–recall relationship over decision thresholds:

$$AP = \sum_{j=1}^{M}\left(R_j - R_{j-1}\right) P_j \quad (7)$$

where P_j and R_j are precision and recall at the$j$-th point, with recall points ordered in increasing order. The Brier score was used to assess the overall accuracy of predicted probabilities:

$$\text{Brier score} = \frac{1}{N}\sum_{i=1}^{N}\left(p_i^{cal} - y_i\right)^2 \quad (8)$$

where N is the number of evaluated encounters. Lower Brier scores indicate better overall probability accuracy.

Uncertainty was estimated using 300 patient-level bootstrap resamples of the internal test set. Patients were sampled with replacement, and all encounters for a given patient were retained

together. The 2.5th and 97.5th percentiles of each bootstrap distribution were used as the limits of a 95% confidence interval.

These intervals describe uncertainty conditional on the fitted models and fixed thresholds; the models were not retrained for each bootstrap sample. Paired bootstrap differences between models were also calculated and interpreted as exploratory, since the test set had already been viewed.

3.7. Model Explainability and Sensitivity Analysis

SHAP values were calculated for the selected CatBoost model on a random sample of **500 test encounters**. Global feature importance was summarized as the mean absolute SHAP value:

$$I_j = \frac{1}{n_s}\sum_{i=1}^{n_s}\left|\phi_{ij}\right| \quad (9)$$

where I_j is the global importance of feature j, $n_s = 500$ is the number of sampled encounters, and ϕ_{ij} is the SHAP contribution of feature j for encounter i. The reported contributions describe the uncalibrated CatBoost model output on the log-odds scale. SHAP values describe model contributions and do not establish causal effects (Lundberg & Lee, 2017).

A development-set sensitivity analysis compared the standard CatBoost configuration with a reduced-coded-feature configuration that excluded three diagnostic variables: the number of diagnoses and payer code. The comparison used mean AP from the patient-grouped development folds.

All analyses were implemented in Python. The recorded environment included Python 3.13, scikit-learn 1.6.1, XGBoost 3.1.3, and CatBoost 1.2.10.

4. Results

4.1. Dataset Characteristics and Missing Values

After the eligibility exclusions, the final dataset contained 99,343 hospital encounters from 69,990 unique patients. Among these encounters, 11,314 cases (11.39%) were followed by a recorded readmission within 30 days, whereas 88,029 encounters (88.61%) were in the other outcome category. The outcome distribution is shown in Figure 2.

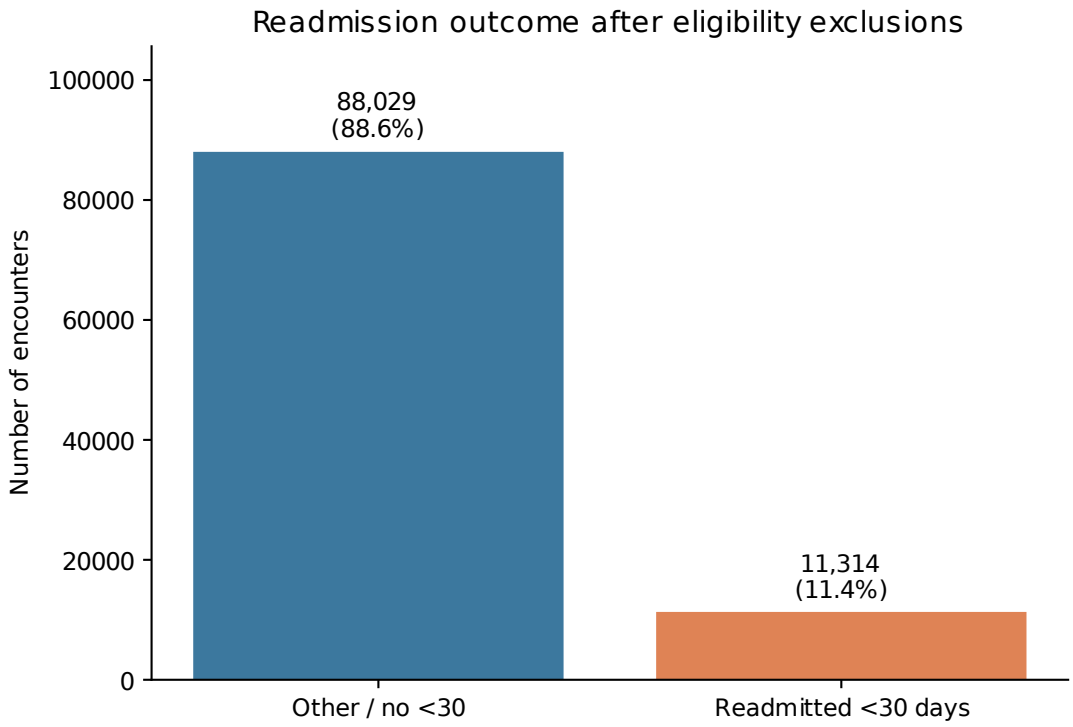

Figure 2. Distribution of early readmission outcomes after eligibility exclusions.

The audit of explicitly missing entries in the original dataset showed that weight had the highest proportion of missing values (96.86%). The corresponding proportions for medical_specialty and payer_code was 49.08% and 39.56%, respectively. Other explicitly missing entries were substantially less frequent. These proportions were calculated using the original dataset and are shown in Figure 3.

4.2. Development Cross-Validation and Model Selection

Twelve candidate configurations were evaluated using three-fold patient-grouped cross-validation in the development set. Table 4 presents the mean validation AP and standard deviation across the three folds.

The standard CatBoost configuration, with missing values retained as a separate category, achieved the highest mean validation AP, 0.2084 ± 0.0060, and was selected as the overall model. The best configurations from the XGBoost and logistic regression families were also retained for comparison on the internal test set.

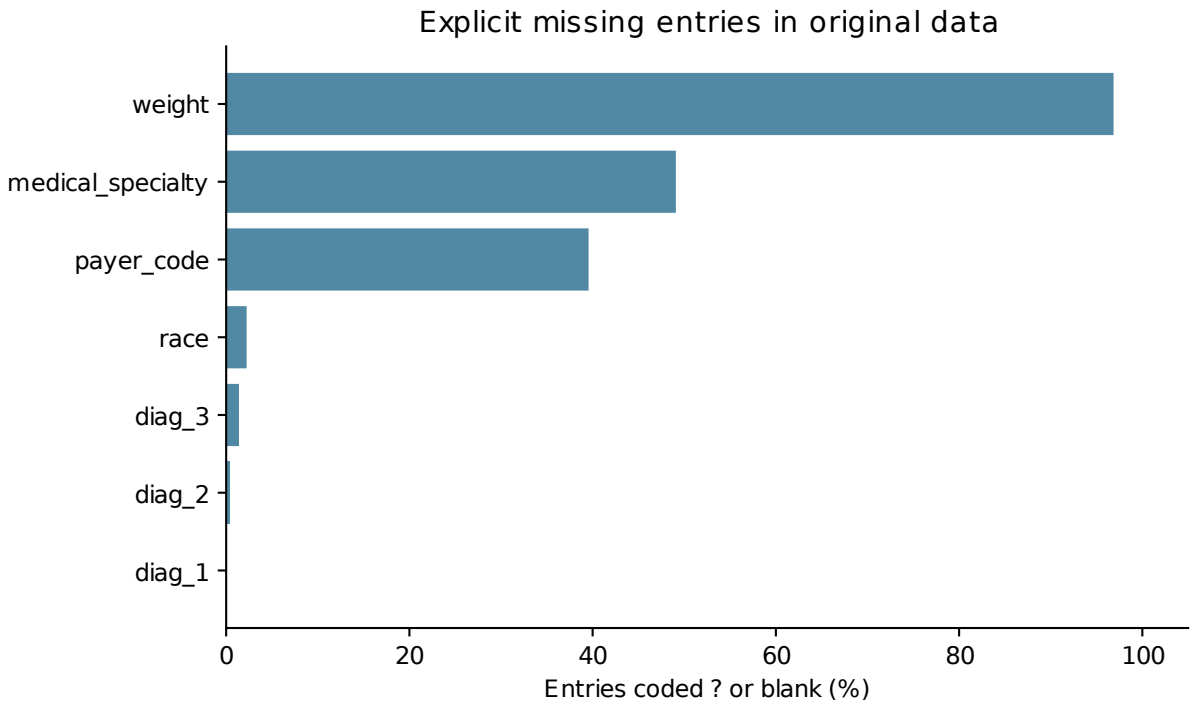

Figure 3. Percentage of explicitly missing entries coded as? or blank in the original dataset.

Table 4. Development cross-validation performance of the candidate configurations

Model	Missing-data strategy	Configuration	Mean validation AP ± SD
Logistic regression	Drop high-missing variables	Standard	0.1954 ± 0.0041
Logistic regression	Most frequent	Standard	0.1950 ± 0.0038
Logistic regression	Missing category	Standard	0.1955 ± 0.0034
XGBoost	Drop high-missing variables	Standard	0.2022 ± 0.0071
XGBoost	Most frequent	Standard	0.2021 ± 0.0061
XGBoost	Missing category	Standard	0.2033 ± 0.0087
XGBoost	Missing category	Shallow	0.2069 ± 0.0074
CatBoost	Drop high-missing variables	Standard	0.2064 ± 0.0055

Model	Missing-data strategy	Configuration	Mean validation AP ± SD
CatBoost	Most frequent	Standard	0.2063 ± 0.0046
CatBoost	**Missing category**	**Standard**	**0.2084 ± 0.0060**
CatBoost	Missing category	Shallow	0.2073 ± 0.0052
CatBoost	Missing category	Reduced coded features	0.2010 ± 0.0094

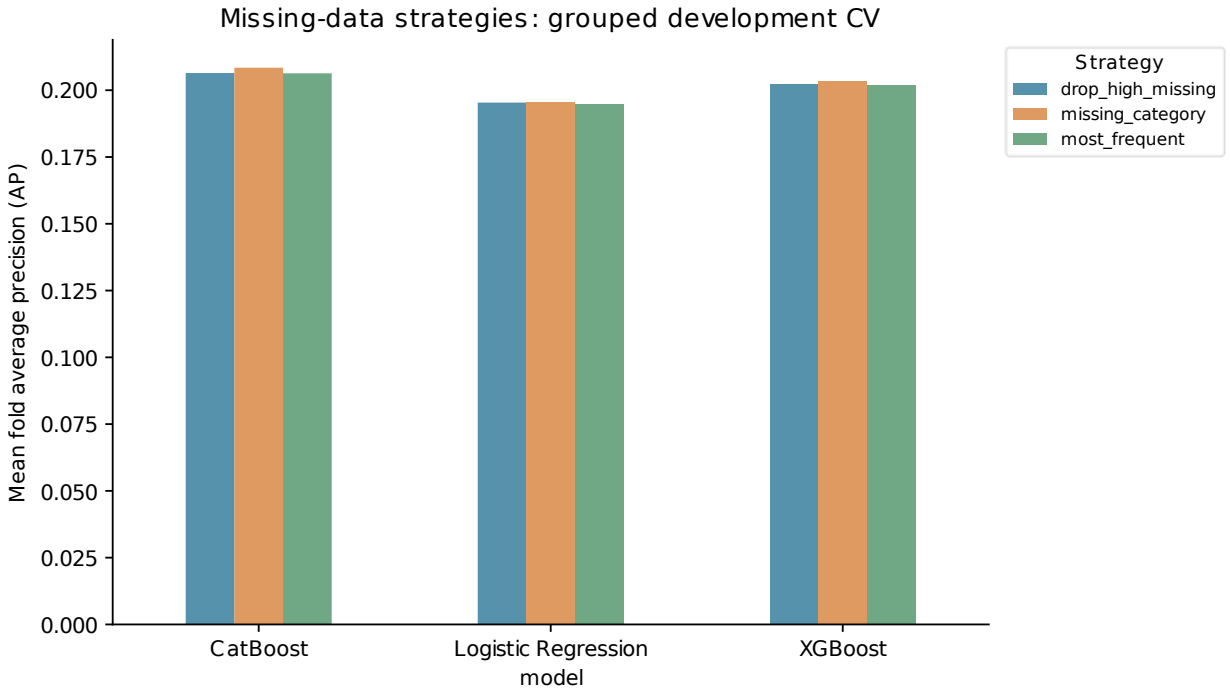

Figure 4. Mean development-fold average precision by model and missing-data strategy for the standard full-feature configurations.

Note. AP values are the means of the three validation-fold scores; SD is the standard deviation across those folds. The bold row indicates the overall configuration selected using development data.

Figure 4 compares the three missing-data strategies for the standard full-feature configurations. The shallower configurations and reduced-coded-feature configuration are included in Table 4 but are not shown in Figure 4.

4.3. Performance on the Previously Viewed Internal Test Set

The test set contained 20,019 encounters, including 2,289 early readmissions and 17,730 other encounters. Table 5 summarizes the performance of the configuration selected from each model

family. The scores were calibrated, and each model used its own threshold selected from the development data.

CatBoost achieved an ROC-AUC of 0.6616 and an AP of 0.2149. XGBoost achieved an ROC-AUC of 0.6575 and an AP of 0.2130, while logistic regression achieved an ROC-AUC of 0.6459 and an AP of 0.2053.

Table 5. Model performance on the previously viewed patient-grouped internal test set

Model	Threshold	ROC-AUC	AP	Accuracy	Precision	Recall	Specificity	F1	Brier
CatBoost	0.1202	0.6616	0.2149	0.6708	0.1832	0.5435	0.6873	0.2741	0.0972
XGBoost	0.1306	0.6575	0.2130	0.7117	0.1922	0.4749	0.7422	0.2736	0.0973
Logistic regression	0.1240	0.6459	0.2053	0.6814	0.1797	0.5011	0.7046	0.2645	0.0978

Note. CatBoost was selected as the overall model based on the mean development-fold AP. Threshold-dependent metrics were calculated using thresholds selected in development data. The test set had been inspected in an earlier analysis.

The ROC and precision–recall curves are presented in Figure 5. The positive-class prevalence in the test set was approximately 0.1143, providing a reference for the precision–recall analysis.

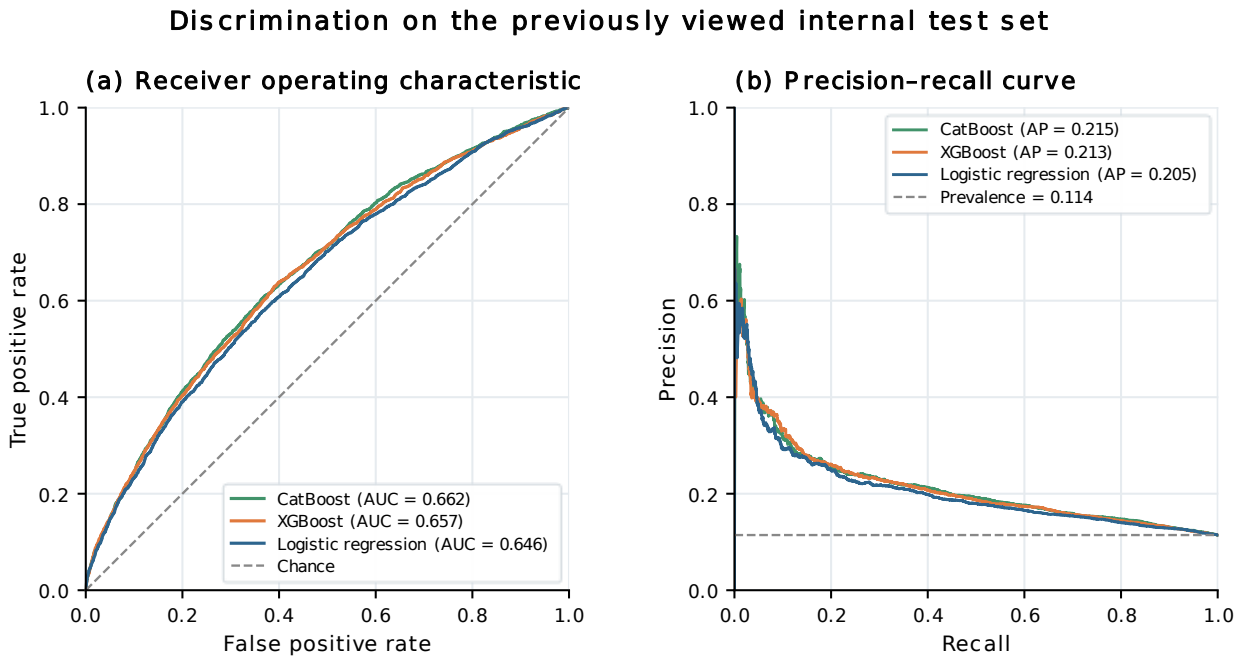

Figure 5. Discrimination performance on the previously viewed internal test set: **(a)** ROC curves and **(b)** precision–recall curves.

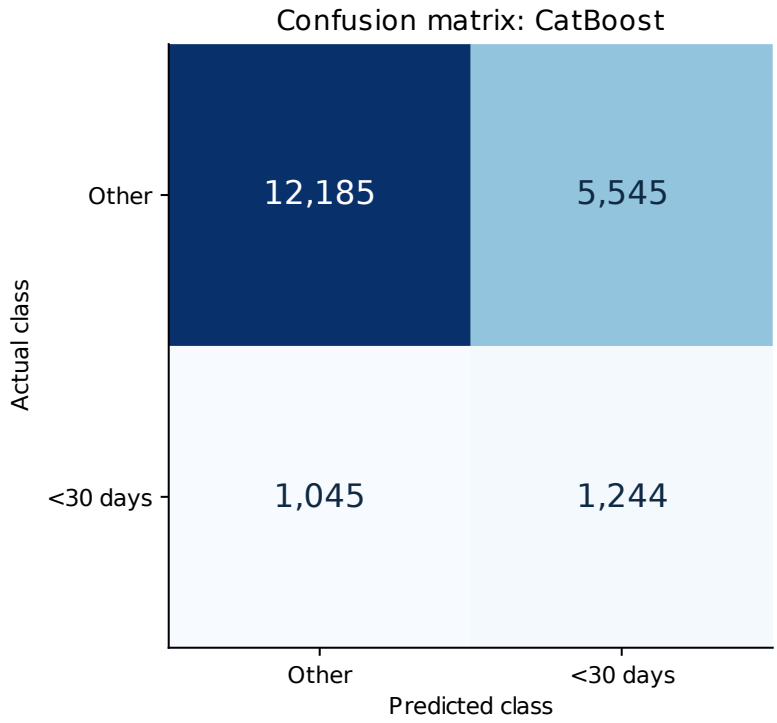

Figure 6. Confusion matrix of the selected CatBoost model on the previously viewed internal test set.

At the selected CatBoost threshold of approximately 0.1202, the model correctly identified 1,244 of the 2,289 early readmissions and missed 1,045. It correctly classified 12,185 other encounters and produced 5,545 false-positive predictions. Thus, among the 6,789 encounters classified as positive, precision was 0.1832 and recall was 0.5435. The corresponding confusion matrix is shown in Figure 6.

For context, a classifier that predicted every test encounter as belonging to the other category would attain an accuracy of approximately 0.8857, but its recall for early readmission would be zero. This illustrates why accuracy alone does not describe minority-class detection.

4.4. Calibration Results

For the selected CatBoost model, the Brier score was 0.2256 for the original class-weighted model scores and 0.0972 after sigmoid calibration. The calibration curve in Figure 7 shows the relationship between mean scores and observed early-readmission proportions before and after calibration.

These results describe the evaluation on the internal test set. The Brier score reflects overall probability accuracy and should not be interpreted as a measure of calibration alone.

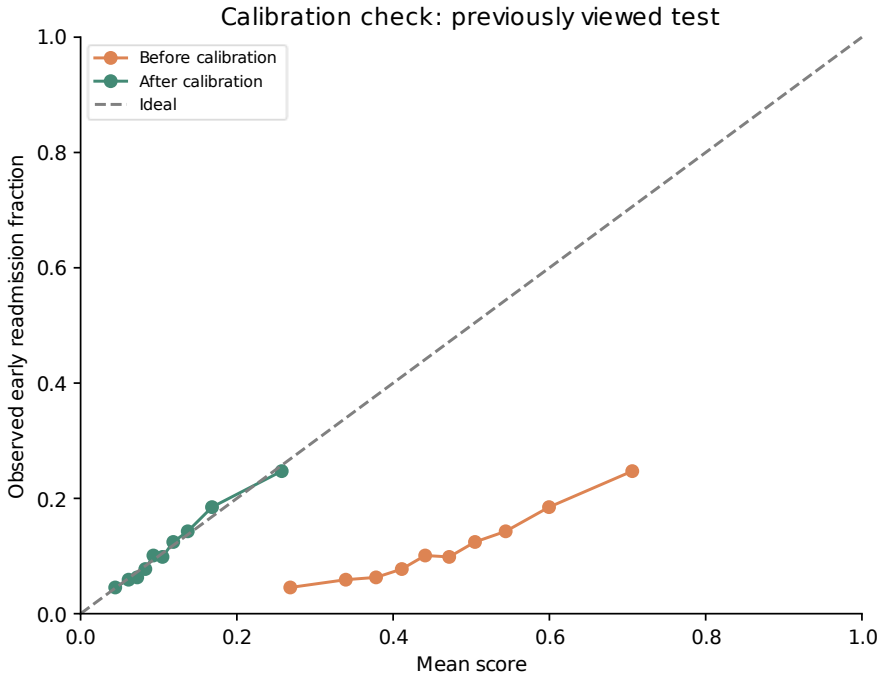

Figure 7. Calibration of the selected CatBoost model before and after sigmoid calibration on the previously viewed internal test set.

4.5. Patient-Bootstrap Confidence Intervals

Table 6 presents the 95% patient-bootstrap confidence intervals for the selected CatBoost model. The ROC-AUC estimate was 0.6616, with a 95% confidence interval of 0.6487–0.6739. The AP estimate was 0.2149, with a 95% confidence interval of 0.1971–0.2334.

Table 6. Patient-bootstrap confidence intervals for the selected CatBoost model

Metric	Estimate	95% CI lower	95% CI upper
AP	0.2149	0.1971	0.2334
ROC-AUC	0.6616	0.6487	0.6739
Precision	0.1832	0.1740	0.1938
Recall	0.5435	0.5212	0.5637
F1	0.2741	0.2617	0.2881
Brier score	0.0972	0.0938	0.1010

Note. Intervals were calculated using 300 patient-level bootstrap resamples. They are conditional on the fitted model and fixed threshold.

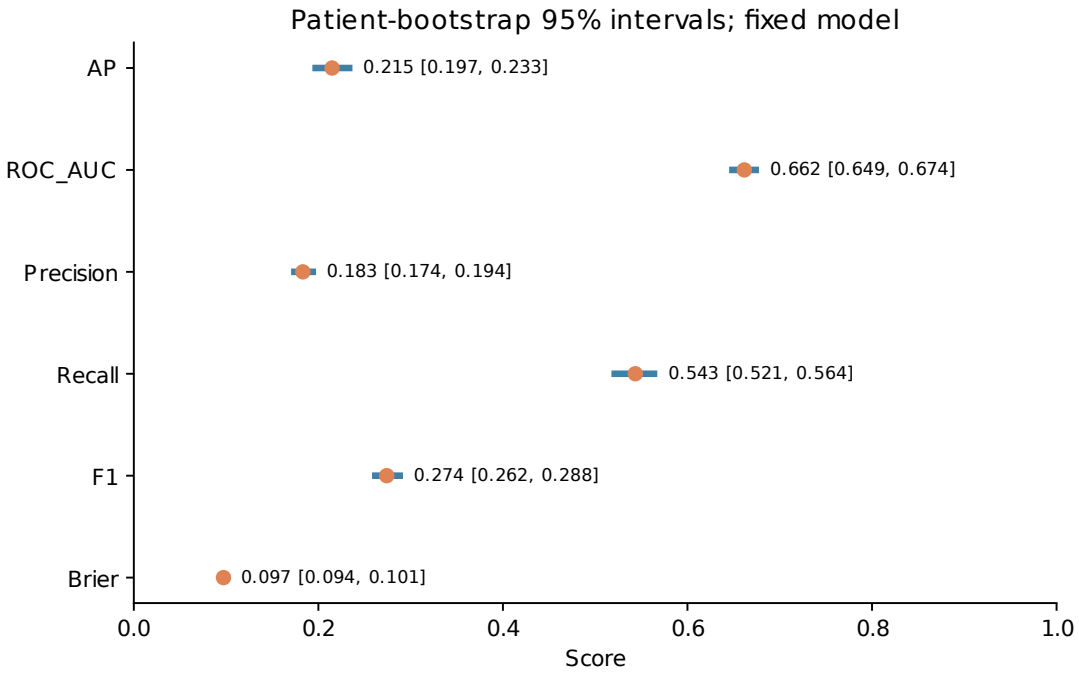

Figure 8. Patient-bootstrap 95% confidence intervals for the selected CatBoost model.

The selected CatBoost configuration had a mean training-fold AP of 0.2552 and a mean validation-fold AP of 0.2084. The corresponding training and validation AP values for the selected XGBoost configuration were 0.2464 and 0.2069. Figure 9 presents these comparisons for the selected

configuration of each model family. The differences describe the training–validation gaps observed in the development folds.

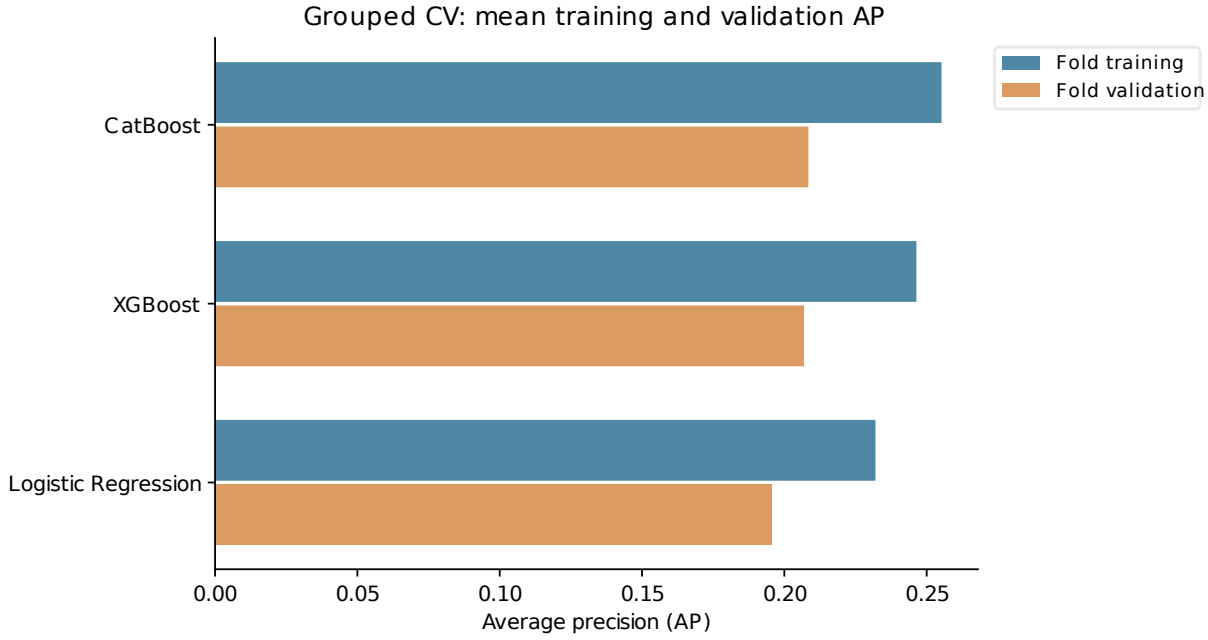

Figure 9. Mean training-fold and validation-fold average precision for the selected configuration of each model family.

4.6. SHAP Feature Contributions

SHAP analysis was performed on 500 test encounters for the selected CatBoost model. The highest mean absolute SHAP contribution was observed for number_inpatient, followed by diag_1, diag_3, and time_in_hospital. The next highly ranked feature was payer_code.

Figure 10 presents the 15 highest-ranked features. These values summarize contributions to the uncalibrated CatBoost model output on the log-odds scale and do not indicate causal effects.

4.7. Sensitivity to Selected Coded Features

The CatBoost sensitivity analysis compared the full predictor set with a reduced set that excluded diag_1, diag_2, diag_3, number_diagnoses, and payer_code. Mean development-fold AP decreased from 0.2084 with the full predictor set to 0.2010 with the reduced set.

The pooled development out-of-fold ROC-AUC was 0.6544 for the full configuration and 0.6388 for the reduced configuration. This analysis was performed on the development data and does not establish that all potential timing-related leakage was removed.

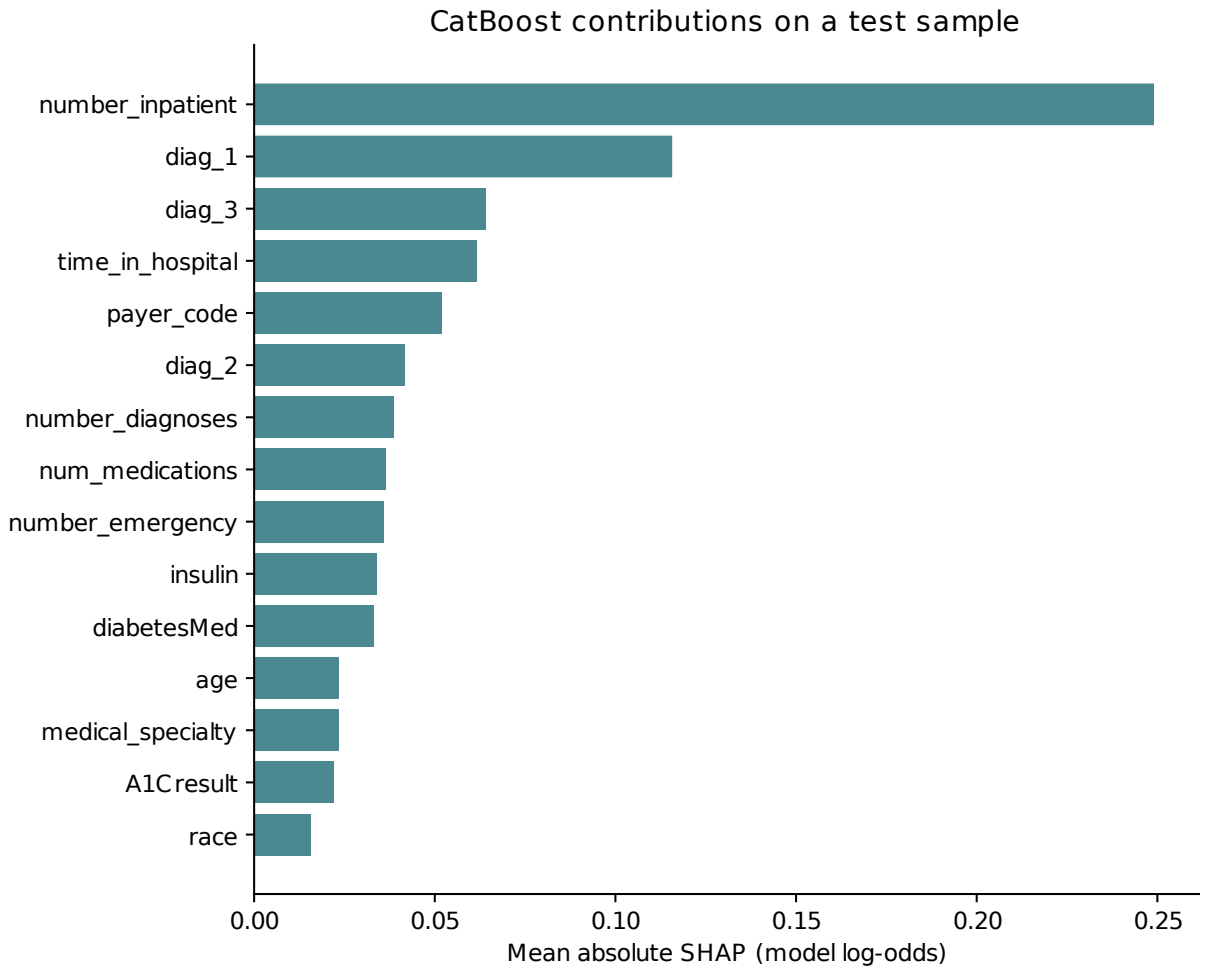

Figure 10. Mean absolute SHAP contributions of the 15 highest-ranked features for the selected CatBoost model, calculated on 500 test encounters.

Table 7. Development-set sensitivity analysis for selected coded features

CatBoost predictor set	Mean fold AP	SD of fold AP	Pooled out-of-fold ROC-AUC
Full predictors	0.2084	0.0060	0.6544
Reduced coded predictors	0.2010	0.0094	0.6388

Note. The reduced configuration excluded three diagnostic variables: number of diagnoses and payer code. The ROC-AUC values were calculated from pooled development out-of-fold scores.

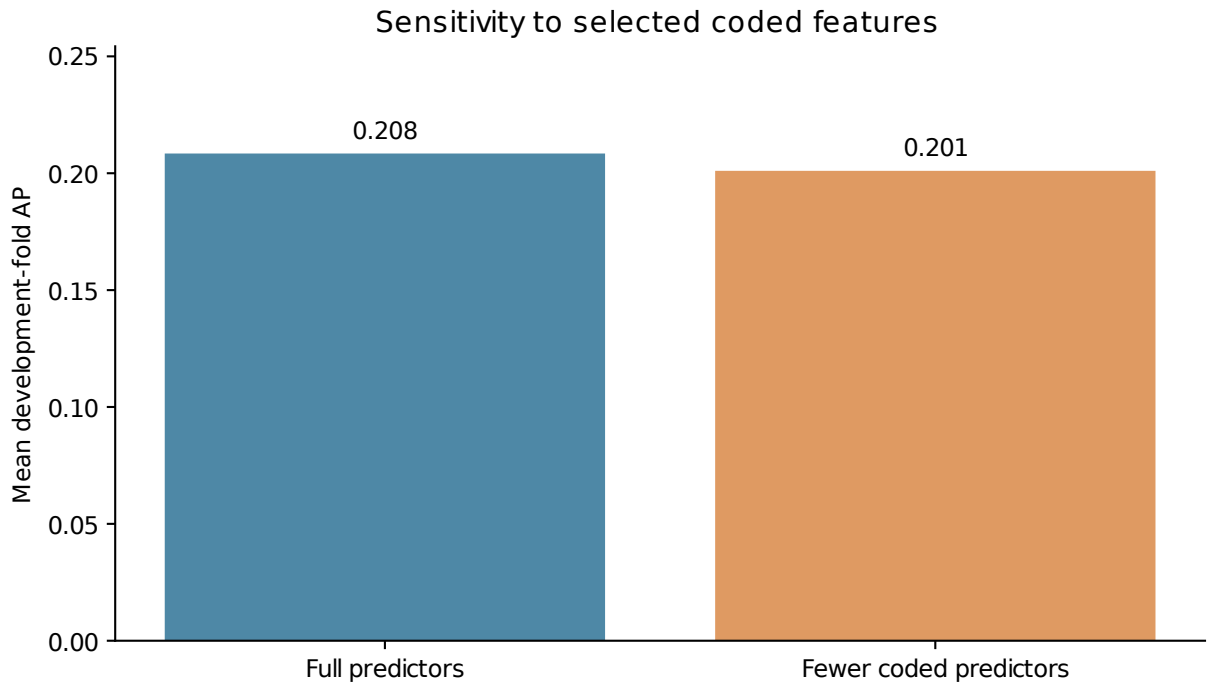

Figure 11. Mean development-fold average precision for the full and reduced-coded-feature CatBoost configurations.

4.8. Exploratory Paired Model Comparisons

Paired patient-bootstrap comparisons were used to examine differences between the selected CatBoost model and the two comparator models. The difference in AP between CatBoost and XGBoost was 0.0018, with a 95% confidence interval of −0.0023 to 0.0058. The corresponding difference in ROC-AUC was 0.0042, with a confidence interval from −0.0002 to 0.0086.

Both CatBoost–XGBoost intervals included zero. The comparisons with logistic regression produced positive point estimates and intervals above zero. All comparisons are exploratory because the test set had previously been examined.

Table 8. Exploratory paired patient-bootstrap differences between models

Comparison	Metric	Difference	95% CI lower	95% CI upper
CatBoost − XGBoost	AP	0.0018	−0.0023	0.0058
CatBoost − XGBoost	ROC-AUC	0.0042	−0.0002	0.0086
CatBoost − logistic regression	AP	0.0095	0.0026	0.0165
CatBoost − logistic regression	ROC-AUC	0.0157	0.0075	0.0220

Note. Differences were calculated using the same patient-bootstrap samples for each pair of models. Intervals are conditional on the fitted models and are interpreted as exploratory.

5. Discussion

5.1. Interpretation of the main findings

This study evaluated machine-learning models for predicting 30-day hospital readmission among patients with diabetes using a historical multicenter dataset. CatBoost was selected based on average precision in patient-grouped development cross-validation. On the previously viewed, patient-disjoint internal test set, it achieved a ROC–AUC of 0.662 and an average precision of 0.215. The patient-bootstrap 95% confidence intervals were 0.649–0.674 for ROC–AUC and 0.197–0.233 for average precision. These results indicate moderate discrimination and demonstrate that evaluation of the positive class remains important in this imbalanced prediction task.

Although CatBoost had the highest point estimates among the evaluated models, its advantage over XGBoost was small. The exploratory patient-bootstrap confidence interval for the difference in average precision between CatBoost and XGBoost was −0.0023 to 0.0058. As this interval included zero, the present analysis does not establish a clear performance difference between the two models. The comparison should also be interpreted with caution because the internal test set had been examined at an earlier stage of the project.

5.2. Comparison with previous studies

The observed ROC–AUC is numerically close to values reported in some studies using the same source dataset. Emi-Johnson and Nkrumah [4] reported a ROC–AUC of 0.667 for XGBoost using a train–test evaluation. Salim and Ibrahim [9] reported a ROC–AUC of 0.664 for XGBoost under nested cross-validation. Although these values provide useful context, differences in cohort construction, predictors, data partitioning, and evaluation procedures prevent a direct comparison of model superiority.

Higher discrimination has also been reported. Afridi et al. [12] reported a ROC–AUC of 0.861 for a hybrid CNN–LSTM model on a held-out test set derived from the same source dataset. Their analysis used a different cohort, predictor set, preprocessing strategy, and modeling procedure. Consequently, the difference between their result and ours cannot be attributed to model architecture without an evaluation under a common protocol. Taken together, these studies

illustrate the importance of reporting the complete evaluation design when comparing readmission models.

5.3. Class imbalance and prediction trade-offs

Because early readmission occurred in approximately 11.4% of the test encounters, average precision provides useful information alongside ROC–AUC. The selected CatBoost model achieved an average precision of 0.215, compared with the test-set prevalence reference of approximately 0.114. Nevertheless, performance at the selected threshold shows a substantial trade-off between identifying readmissions and generating false-positive predictions.

At the F1-based threshold selected using development data, precision was 0.183 and recall was 0.543. Of 6,789 encounters classified as high risk, 1,244 had a recorded readmission in less than 30 days, while 5,545 did not. The model also missed 1,045 of the 2,289 recorded early readmissions, corresponding to approximately 45.7% of positive cases. These findings do not establish whether the model would be useful in practice: a suitable threshold would depend on the available intervention, its cost, the resources required for follow-up, and the consequences of missed cases.

5.4. Calibration of predicted scores

Calibration is relevant when model scores are interpreted as estimated probabilities. On the previously viewed internal test set, the Brier score of the selected model decreased from 0.2256 for the uncalibrated class-weighted scores to 0.0972 after sigmoid calibration. The calibration plot also showed closer agreement between predicted scores and observed proportions after calibration. For context, a constant prediction equal to the observed test-set prevalence yields a Brier score of approximately 0.1013. This is a descriptive, post hoc reference based on the test outcomes, rather than an independently specified clinical benchmark. The lower Brier score after calibration supports improved probability accuracy in this internal evaluation, but the Brier score reflects both calibration and discrimination. Neither this result nor the calibration plot establishes clinical benefit or reliable calibration in a new population.

5.5. Feature contributions and missing data

In the SHAP analysis of 500 test encounters, the number of prior inpatient visits had the largest mean absolute contribution to the uncalibrated CatBoost model output. Primary and secondary diagnosis variables, length of stay, and payer information also contributed to predictions. The prominence of prior inpatient utilization is consistent with findings reported by Guo et al. (2025)

and Emi-Johnson and Nkrumah [4]. However, the SHAP values describe contributions to this model's predictions on the log-odds scale; they do not establish causal effects or, from the mean absolute values alone, the direction of an association.

Missing data also affected the design of the models. In the original dataset, weight was missing for approximately 96.9% of encounters, medical specialty for 49.1%, and payer code for 39.6%. Weight was excluded, whereas alternative strategies were evaluated for the other variables. Differences among the evaluated missing-data strategies were relatively small in development cross-validation. Such comparisons are important because missing entries in clinical records can reflect care and documentation processes rather than purely random omissions [7].

In the sensitivity analysis, the mean development-fold average precision decreased from 0.2084 with the full predictor set to 0.2010 after selected coded variables were removed. This suggests that those variables contributed predictive information in this development sample. It does not, however, prove that all retained variables would be available at the intended time of prediction or that the model is free from timing-related leakage. In particular, clinical diagnosis codes may be finalized after the proposed prediction time [3].

5.6. Strengths and limitations

The analysis combined patient-grouped data partitioning, comparison of multiple model families and missing-data strategies, calibration, patient-level bootstrap intervals, and model explanation. Patient grouping reduced the risk of the same patient appearing in both development and test sets. The use of average precision and threshold-specific measures also helped characterize performance for the minority class.

Several limitations affect interpretation. First, the internal test set had been viewed during a previous iteration, and subsequent analytical changes mean that its results should be regarded as exploratory rather than as an unbiased final validation. Second, the best configuration was selected from several candidates using the same development cross-validation results, which may make its selected development estimate optimistic. Third, threshold selection used cross-calibration of out-of-fold scores, but the underlying models were not refit within a fully nested calibration procedure. Finally, the bootstrap intervals were based on 300 patient-level resamples with the fitted models and thresholds held fixed; they do not capture the full uncertainty associated with model development and selection.

The dataset covers hospital encounters from 1999 to 2008, and no temporal, geographic, or external validation was performed. Changes in clinical practice, patient populations, and recording systems may therefore limit applicability to contemporary settings. Encounters with selected death or hospice discharge codes were excluded, which also affects the population to which the findings apply. In addition, the absence of a recorded early readmission does not necessarily establish that no readmission occurred outside the source data. The analysis did not evaluate subgroup fairness or demonstrate the clinical effect of using the model.

5.7. Implications for future research

Future studies should validate the approach in a contemporary, independent cohort and verify that every predictor is available at the intended time of prediction. Evaluation should include calibration and performance across relevant patient subgroups, as well as assessment of decision thresholds in relation to intervention costs and available resources. Prospective studies would be needed to determine whether using predictions improves care or patient outcomes.

This emphasis on independent evaluation is consistent with the broader concerns about validation maturity and generalizability identified in the systematic review by Jodayree et al. [11]. That review addressed COVID-19 applications, so it provides methodological context rather than direct evidence about diabetes readmission prediction.

6. Conclusion

This study evaluated logistic regression, XGBoost, and CatBoost for predicting 30-day hospital readmission among patients with diabetes. CatBoost was selected using patient-grouped development cross-validation and achieved a ROC–AUC of 0.662 and an average precision of 0.215 on the previously viewed patient-disjoint internal test set. At the threshold selected from development data, the model achieved a precision of 0.183 and a recall of 0.543. These findings indicate moderate discrimination and a substantial trade-off between identifying early readmissions and generating false-positive predictions.

Calibration improved the Brier score of the selected model on the internal test set, while SHAP analysis identified prior inpatient utilization as the largest contributor to the model's predictions in the analyzed sample. However, the exploratory comparison did not establish a clear performance advantage for CatBoost over XGBoost. The results should be interpreted in light of the previously viewed test set, the historical nature of the data, and the absence of external validation.

Future research should evaluate the model in independent, contemporary patient populations, verify the availability of predictors at the intended time of use, and assess performance across relevant subgroups. Prospective evaluation is also needed to determine whether model-guided interventions provide clinical benefit.

Data availability: The dataset analyzed in this study is publicly available from the UCI Machine Learning Repository at https://doi.org/10.24432/C5230J

References

[1] Chakraborty, C., Bhattacharya, M., Pal, S., & Lee, S.-S. (2024). From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. Current Research in Biotechnology, 7, 100164. https://doi.org/10.1016/j.crbiot.2023.100164 1-s2.0-S2590262823000461-main.pdf

[2] Clore, J., Cios, K., DeShazo, J., & Strack, B. (2014). Diabetes 130-US hospitals for years 1999–2008 [Data set]. UCI Machine Learning Repository. https://doi.org/10.24432/C5230J

[3] Davis, S. E., Matheny, M. E., Balu, S., & Sendak, M. P. (2024). A framework for understanding label leakage in machine learning for health care. Journal of the American Medical Informatics Association, 31(1), 274–280. https://doi.org/10.1093/jamia/ocad178

[4] Emi-Johnson, O. G., & Nkrumah, K. J. (2025). Predicting 30-day hospital readmission in patients with diabetes using machine learning on electronic health record data. Cureus, 17(4), e82437. https://doi.org/10.7759/cureus.82437 cureus-0017-00000082437.pdf

[5] Guo, C., Zhou, H., Pires, I. M., Coelho, P. J., Tong, R., & Farid, F. (2025). Towards an explainable machine learning model to reduce readmission risks for diabetes patients. Informatics in Medicine Unlocked, 58, 101686. https://doi.org/10.1016/j.imu.2025.101686 1-s2.0-S2352914825000759-main.pdf

[6] Jiang, M.-H., Yu, F., Yang, H.-Y., Yin, S.-J., Yang, L.-J., Chen, Y., Li, D.-M., Guo, Y., Zhu, J.-D., Cai, W.-K., & He, G.-H. (2025). Predicting readmission rates in critically ill heart failure patients during a 90-day vulnerable phase using interpretable machine learning models. Clinics, 80, 100775. https://doi.org/10.1016/j.clinsp.2025.100775 1-s2.0-S1807593225001930-main.pdf

[7] Mollaie, F., Shakoor, M. H., Sepahvand, M., & Boostani, R. (2026). Medical data types in machine learning and their challenges: A comprehensive review. Intelligence-Based Medicine, 15, 100427. https://doi.org/10.1016/j.ibmed.2026.100427 1-s2.0-S2666521226000852-main.pdf

[8] Padula, W. V., Kreif, N., Vanness, D. J., Adamson, B., Rueda, J.-D., Felizzi, F., Jonsson, P., IJzerman, M. J., Butte, A., & Crown, W. (2022). Machine learning methods in health economics and outcomes research—The PALISADE checklist: A good practices report of an ISPOR task force. Value in Health, 25(7), 1063–1080. https://doi.org/10.1016/j.jval.2022.03.022 1-s2.0-S1098301522001917-main.pdf

[9] Salim, S. S., & Ibrahim, A. A. (2026). A machine learning approach for predicting 30-day hospital readmission in patients with diabetes. Healthcare, 14(9), 1185. https://doi.org/10.3390/healthcare14091185 healthcare-14-01185-v2.pdf

[10] Ghafi, A. K., Pirkhedri, A., Akhbarifar, S., & Shafiabadi, M. H. (2026). Epidemic Forecasting via Hybrid Deep Learning With Unified Visibility and Temporal Graphs Under Stochastic Noise Modelling. Healthcare Technology Letters, 13(1), e70073.

[11] Jodayree, M., Ghafi, A. K., Khodadadi, I., Jahanshahi, J. A., Tofangdarzade, A., & Pirkhedri, A. (2026). A systematic review of machine learning and deep learning methods for COVID-19 diagnosis, prognosis, and outbreak forecasting. Applied Network Science. Advance online publication. https://doi.org/10.1007/s41109-026-00836-w s41109-026-00836-w_reference.pdf

[12] Afridi, A. A., Usman, M., Rubin, D. J., & Rana, M. R. R. (2026). A robust and explainable framework for predicting hospital readmission in patients with diabetes. Diabetology, 7(10), 197. https://doi.org/10.3390/diabetology7100197

[13] Huang, Y., Talwar, A., Chatterjee, S., & Aparasu, R. R. (2021). Application of machine learning in predicting hospital readmissions: A scoping review of the literature. BMC Medical Research Methodology, 21, 96. https://doi.org/10.1186/s12874-021-01284-z

[14] Slezak, J., Butler, L., & Akbilgic, O. (2021). The role of frailty index in predicting readmission risk following total joint replacement using light gradient boosting machines. Informatics in Medicine Unlocked, 25, 100657. https://doi.org/10.1016/j.imu.2021.100657 1-s2.0-S2352914821001453-main.pdf
[15] Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., & Clore, J. N. (2014). Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. BioMed Research International, 2014, 781670. https://doi.org/10.1155/2014/781670
[16] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785
[17] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30.
https://proceedings.neurips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html
[18] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31. https://proceedings.neurips.cc/paper_files/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html
[19] Mohammadi, M. (2025). A Transjective-Holobiontic Rhetoric (THR). Somatechnics, 15(1), 63-75.
[20] Bahador, A., Moghaddam, S. A. M., Mojtahedi, H., Hakim, L. K., & Tebyaniyan, H. (2026). Artificial Intelligence and Digital Workflow in Craniofacial Bone Tissue Engineering: From Cone-Beam Computed Tomography (CBCT) to Personalized Bioceramic Implants. Dentistry Journal, 14(9), 558.
[21] Mohammadi, M. (2024). Embodiment in Posthuman Rhetoric. Available at SSRN 5267892.
[22] Huang, C., Esfani Sarafraz, P., Enayati, P., Mortazavi Mamaghani, E., Babakhanzadeh, E., & Nazari, M. (2023). Circular RNAs in renal cell carcinoma: from mechanistic to clinical perspective. Cancer Cell International, 23(1), 288.
[23] Mohammadi, M. (2024). Rhetoric and Posthumanism. It was published on criticalposthumanism. net.
[24] Hoseini, S. H., Enayati, P., Nazari, M., Babakhanzadeh, E., Rastgoo, M., & Sohrabi, N. B. (2024). Biomarker profile of colorectal cancer: current findings and future perspective. Journal of Gastrointestinal Cancer, 55(2), 497-510.
[25] Mohammadi, M. (2026). Brave new democracy: algorithmic transcontextuality and fluid dynamics of democratic deliberation. International Journal of Studies in Humanities and Social Science, 2(1), 1-17.
[26] Mohammadi, M. (2026). Toward a Post-Temporal Time-consciousness in GenAI-mediated Writing. Available at SSRN 7404941.
[27] Mohammadi, M. (2025). A Postphenomenological Turn in Rhetorical Studies. Available at SSRN 5276597.
[28] Kouser, S., Dheep, G. R., & Bansal, R. C. (2025). Artificial Rabbit Optimized Neural Network Control of Battery and PV-Powered UPQC for Micro-Grid Application. Smart Grids and Sustainable Energy, 10(3), 79.
[29] Mohammadi, M. (2025). Philosophy of Technology in Rhetoric and Writing Studies. Available at SSRN 5267904.
[30] Mazloum, V., Akbari, H., & Gholampour, A. (2023). The comparison of the effects of neuromuscular electrical stimulation and Kinesio Taping on ankle swelling in athletes with lateral ankle sprain. Journal of experimental orthopaedics, 10(1), 63.
[31] Kouser, S., Dheep, R., & Bansal, R. C. (2022, November). Horse herd optimization MPPT for grid connected PV system under partial shading conditions. In 2022 IEEE 10th Power India International Conference (PIICON) (pp. 1-6). IEEE.
[32] Jayachitra, G., & Kouser, S. (2026, May). Modeling and Control of Hybrid Renewable Energy based EV Charging Station with Grid Integration. In 2026 8th International Conference on Inventive Material Science and Applications (ICIMA) (pp. 119-124). IEEE.
[33] Kouser, S., Dheep, R., & Bansal, R. C. (2024, June). Adaptive neuro-fuzzy inference system for optimizing energy flow in grid-linked solar and battery storage systems. In 2024 Advances in Science and Engineering Technology International Conferences (ASET) (pp. 1-7). IEEE.
[34] Jayachitra, G., & Kouser, S. (2026, May). Modeling and Control of Hybrid Renewable Energy based EV Charging Station with Grid Integration. In 2026 8th International Conference on Inventive Material Science and Applications (ICIMA) (pp. 119-124). IEEE.
[35] Kouser, S., Dheep, R., & Bansal, R. C. (2024, June). Adaptive neuro-fuzzy inference system for optimizing energy flow in grid-linked solar and battery storage systems. In 2024 Advances in Science and Engineering Technology International Conferences (ASET) (pp. 1-7). IEEE.
[36] Zamiri, A., Sohrabi, N., Raasch, K., Sumita, M., & Shavezipur, M. (2024, August). Monitoring a pseudouridine 5'-monophosphate synthesis process using an interdigitated MEMS sensor and electrochemical impedance

spectroscopy (EIS). In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 88421, p. V008T08A008). American Society of Mechanical Engineers.
[37] Abbasi, S., Hafezniya, H., Abadi, M. F. L., Ghoroqi, M., & Borchalouei, M. K. (2026). A customer satisfaction–focused green supply chain network under carbon tax and carbon cap across heterogeneous regulatory regions. Cleaner Logistics and Supply Chain, 100352.
[38] Zamiri, A., & Shavezipur, M. (2023, October). A three-electrode three-dimensional impedance-based biochemical sensor for food safety applications. In ASME International Mechanical Engineering Congress and Exposition (Vol. 87691, p. V012T13A028). American Society of Mechanical Engineers.