Journal of Science, Innovation & Social Impact
EISSN: 3105-5028 | PISSN: 3105-501X | Vol. 1, No. 2 (2025)

Article

Explainable Risk Stratification and Resource Coordination for Hospital Readmission Management through Integrated Prediction-Intervention-Evaluation Framework

Yisi Liu [1,*]

[1] Business Data Analytics & Human Resources Management, Loyola University Chicago, Chicago, Illinois, USA
* Correspondence: Yisi Liu, Business Data Analytics & Human Resources Management, Loyola University Chicago, Chicago, Illinois, USA

Abstract: Hospital readmissions impose substantial financial penalties and strain healthcare resources, necessitating intelligent management strategies. This paper presents an integrated prediction-intervention-evaluation framework that couples explainable risk stratification with coordinated resource allocation to reduce 30/90-day readmission rates. The methodology employs quantile-based risk binning with isotonic calibration to achieve interpretable patient stratification, followed by constraint-satisfaction optimization for bed and nursing time allocation. Prospective simulation demonstrates a 1.9 percentage-point reduction in 30-day readmissions while improving bed utilization efficiency by 18.7%. Cross-seasonal validation confirms robustness across temporal variations, with threshold sensitivity analysis revealing stable operating points. The framework achieves an AUROC of 0.847 and maintains an overall calibration error of 0.032. By emphasizing variable availability and uncertainty quantification over complex engineering implementations, this approach facilitates multi-site deployment and regulatory compliance while addressing Medicare penalty structures and workforce sustainability challenges.

Keywords: readmission risk stratification; explainable artificial intelligence; healthcare resource coordination; isotonic calibration

Received: 13 November 2025
Revised: 03 January 2026
Accepted: 15 January 2026
Published: 18 January 2026

1. Introduction

1.1. Clinical and Economic Imperatives of Readmission Management

1.1.1. Medicare Penalties and Financial Burden on Healthcare Systems

The Hospital Readmissions Reduction Program, instituted by the Centers for Medicare & Medicaid Services, imposes financial penalties on hospitals with excess readmission rates, creating direct economic consequences for healthcare organizations. Annual penalty amounts can reach 3% of total Medicare reimbursements, translating to millions of dollars for large healthcare systems. Research analyzing large-scale readmission patterns has demonstrated that congestive heart failure patients exhibit particularly high readmission vulnerability, with 30-day rates approaching 25% in specific populations [1]. The financial burden extends beyond direct penalties to include opportunity costs associated with bed occupancy by readmitted patients.

1.1.2. Patient Safety Concerns and Quality of Care Implications

Readmissions frequently indicate inadequate discharge planning, medication reconciliation failures, or insufficient patient education regarding self-care protocols [2].

Patients experiencing readmissions face elevated risks of healthcare-associated infections and psychological distress from repeated hospitalizations. Preventable readmissions often stem from predictable deteriorations that could have been intercepted through appropriate outpatient monitoring.

1.1.3. Resource Scarcity and Nursing Workforce Pressures

Healthcare systems worldwide confront nursing workforce shortages, with vacancy rates in critical care units reaching 15-20% in many regions [3]. Emergency readmissions disrupt planned patient flow, forcing rapid decisions on bed turnover. The nursing workload associated with unplanned readmissions compounds existing staff burnout, particularly during seasonal demand surges.

1.2. Limitations of Current Readmission Prevention Approaches

1.2.1. Black-Box Prediction without Actionable Risk Stratification

Contemporary machine learning approaches for readmission prediction frequently achieve high discrimination metrics while providing minimal clinical interpretability [4]. Deep neural networks generate risk scores without explaining which modifiable factors drive individual patient risk. Existing tools provide binary high-risk flags without quantifying uncertainty or establishing validated risk thresholds aligned with intervention capacity constraints.

1.2.2. Disconnection between Risk Assessment and Resource Allocation

Readmission risk prediction operates as an isolated task, disconnected from downstream resource allocation decisions [5]. When high-risk patients are identified, there is no systematic mechanism to prioritize their access to limited preventive resources, such as transitional care beds.

1.2.3. Lack of Robust Validation across Temporal and Demographic Variations

Readmission prediction algorithms trained on historical data frequently degrade in performance when applied to subsequent time periods due to concept drift [6]. Seasonal variations in disease prevalence create time-varying risk patterns that static algorithms fail to accommodate.

1.3. Contributions and Paper Organization

This work presents a unified framework coupling prediction, intervention, and evaluation components through explainable artificial intelligence techniques. The prediction component employs gradient boosting with controlled complexity to maintain interpretability. The framework implements quantile-based risk binning to create interpretable risk strata with balanced population distributions. The intervention component models bed reservations and nursing schedules as a constraint satisfaction problem. The evaluation component assesses robustness through threshold sensitivity analysis and cross-seasonal validation.

2. Related Work and Background

2.1. Machine Learning for Hospital Readmission Prediction

2.1.1. Traditional Risk Scores versus Data-Driven Approaches

Traditional clinical risk scores provide structured approaches to readmission risk assessment using manually selected variables [7]. These scores achieve limited discrimination, typically yielding AUROC values of 0.65-0.72. Machine learning approaches leveraging comprehensive electronic health record data have demonstrated superior discrimination, with ensemble methods achieving an AUROC above 0.80.

2.1.2. Deep Learning and Ensemble Methods for 30/90-Day Readmission

Research on hospital readmission has demonstrated that hybrid models combining rule-based clinical knowledge with machine learning classifiers can extract predictive signals from heterogeneous clinical data sources [8]. Gradient boosting implementations provide computational efficiency suitable for large-scale deployment.

2.1.3. Feature Engineering from EHR, Clinical Notes, and Administrative Data

Effective readmission prediction requires careful feature engineering from heterogeneous data sources [9]. Electronic health records include structured fields such as demographics and laboratory results. Temporal aggregation strategies must balance information richness against computational complexity.

2.2. Explainable AI in Clinical Decision Support

2.2.1. SHAP, LIME, and Attention Mechanisms for Model Interpretability

SHapley Additive exPlanations provide game-theoretic foundations for attributing prediction contributions to individual features [10]. SHAP values satisfy desirable properties, including local accuracy and consistency, making them well-suited for clinical applications.

2.2.2. Calibration Techniques for Reliable Probability Estimates

Discrimination and calibration represent distinct aspects of prediction model performance [11]. Isotonic regression provides post-hoc calibration techniques that adjust predicted probabilities to match observed frequencies. In practice, calibrated risk estimates are most impactful when embedded into operational workflows and clinical pathways that coordinate care and resources [12].

2.2.3. Trust, Transparency, and Clinician Adoption Challenges

Clinical adoption of predictive algorithms requires clinician trust, built through transparency [13]. Interpretable explanations must align with clinical reasoning patterns to facilitate integration into the workflow.

2.3. Healthcare Resource Optimization and Scheduling

2.3.1. Bed Allocation Algorithms and Patient Flow Management

Hospital bed management involves dynamic allocation decisions that balance elective admissions and emergency arrivals [14]. Integer programming approaches have been developed for intensive care unit allocation.

2.3.2. Nursing Workforce Scheduling under Capacity Constraints

Nurse staffing optimization must meet regulatory requirements for minimum staff-to-patient ratios. Workload balancing objectives aim to distribute patient acuity equitably.

2.3.3. Cost-Effectiveness Analysis in Preventive Care Interventions

Transitional care programs demonstrate variable cost-effectiveness depending on target population risk levels [15]. Precision medicine approaches that target preventive resources to high-risk patients can improve cost-effectiveness.

3. Methodology

3.1. Prediction Component: Lightweight Feature Engineering and Ensemble Learning

3.1.1. EHR Data Preprocessing and Temporal Feature Extraction

The prediction pipeline processes electronic health record data through systematic preprocessing steps addressing missing values, outliers, and temporal alignment (Table 1). Laboratory values are winsorized at the 1st and 99th percentiles to limit the influence of outliers. Missing laboratory results are imputed using the mean within demographic

strata, and additional binary indicator variables encode missingness patterns. Temporal feature engineering creates multiple time-scale representations of patient history. Recent hospitalization features aggregate admission counts, cumulative inpatient days, and emergency department visits within trailing 30-day, 90-day, and 365-day windows. Comorbidity burden is represented by the Charlson and Elixhauser indices, computed from ICD-10 diagnosis codes. The feature set emphasizes variable availability across diverse healthcare settings (Table 1).

Table 1. Feature Engineering Summary for Readmission Prediction.

Feature Category	Number of Features	Examples	Temporal Windows
Demographics	8	Age, gender, insurance type, marital status	Static
Comorbidities	31	Charlson index, diabetes, heart failure, COPD	Historical cumulative
Prior Utilization	12	ED visits, hospitalizations, ICU days	30d, 90d, 365d
Laboratory Values	24	Creatinine, hemoglobin, sodium, albumin	Most recent, 7d trend
Vital Signs	15	Blood pressure, heart rate, temperature	Admission, discharge
Medications	18	Polypharmacy count, high-risk meds	Current, 30d changes
Index Admission	22	Length of stay, ICU status, surgery	Current episode
Functional Status	6	ADL dependencies, mobility	Current assessment
Total	136	-	-

3.1.2. Gradient Boosting with Controlled Complexity

Gradient boosting provides a flexible machine learning framework combining multiple weak learners (Table 2). The implementation employs LightGBM, emphasizing computational efficiency through histogram-based tree learning. The training objective minimizes binary cross-entropy loss. Regularization through L2 penalties on leaf weights and constraints on maximum tree depth limit model capacity. The number of trees is limited to 100, with a learning rate of 0.05. Maximum tree depth is limited to 6, ensuring individual trees remain interpretable. The calibration of predicted probabilities is explicitly addressed. Raw gradient boosting outputs are subjected to isotonic regression, fitting a monotone increasing function that maps predicted scores to calibrated probabilities (Table 2).

Table 2. Gradient Boosting Hyperparameter Configuration.

Hyperparameter	Value	Rationale
Number of trees	100	Balances performance with interpretability
Learning rate	0.05	Prevents overfitting through gradual refinement
Max tree depth	6	Limits individual tree complexity
Min data in leaf	50	Prevents overfitting to small subgroups
Feature fraction	0.8	Reduces tree correlation
L2 regularization	0.1	Penalizes large leaf weights

Objective	Binary cross-entropy	Appropriate for classification

3.1.3. Shap Value Computation for Global and Local Explainability

SHapley Additive exPlanations provide a principled framework for attributing prediction contributions to individual features. SHAP values decompose each prediction into additive feature contributions that sum to the difference between the prediction and the expected value. Tree SHAP algorithms enable efficient exact computation of Shapley values. Global feature importance aggregates absolute SHAP values across all validation patients. The top 20 features by mean absolute SHAP value receive detailed clinical interpretation. SHAP dependence plots visualize relationships between feature values and their SHAP contributions, revealing nonlinear patterns and interaction effects.

This visualization presents a multi-panel analytical framework for model interpretation using SHAP values arranged in a 2×2 grid (Figure 1). The top-left panel displays a horizontal bar chart ranking the 20 most essential features by mean absolute SHAP value, with length of stay, prior hospitalizations, creatinine, age, and Charlson index at the top. Bar colors transition from dark blue to yellow via a viridis colormap. The top-right panel presents a SHAP dependence plot for length of stay, showing the x-axis (0–30 days) and the y-axis SHAP values (−0.3 to +0.4), with color indicating age interaction. The relationship exhibits apparent nonlinearity with steep increases for stays of 1–7 days and extended stays above 15 days. The bottom-left panel illustrates a SHAP force plot for a high-risk patient showing how features push the prediction from baseline 0.18 to final 0.67, with red segments for risk-increasing features and blue for protective features. The bottom-right panel presents a calibration plot of predicted probability (x-axis, 0–1) versus observed frequency (y-axis, 0–1), with the calibration curve closely tracking the diagonal (ECE = 0.032). Figure dimensions are 16 inches wide by 12 inches tall, with consistent styling and master title “Model Interpretability and Calibration Analysis.”

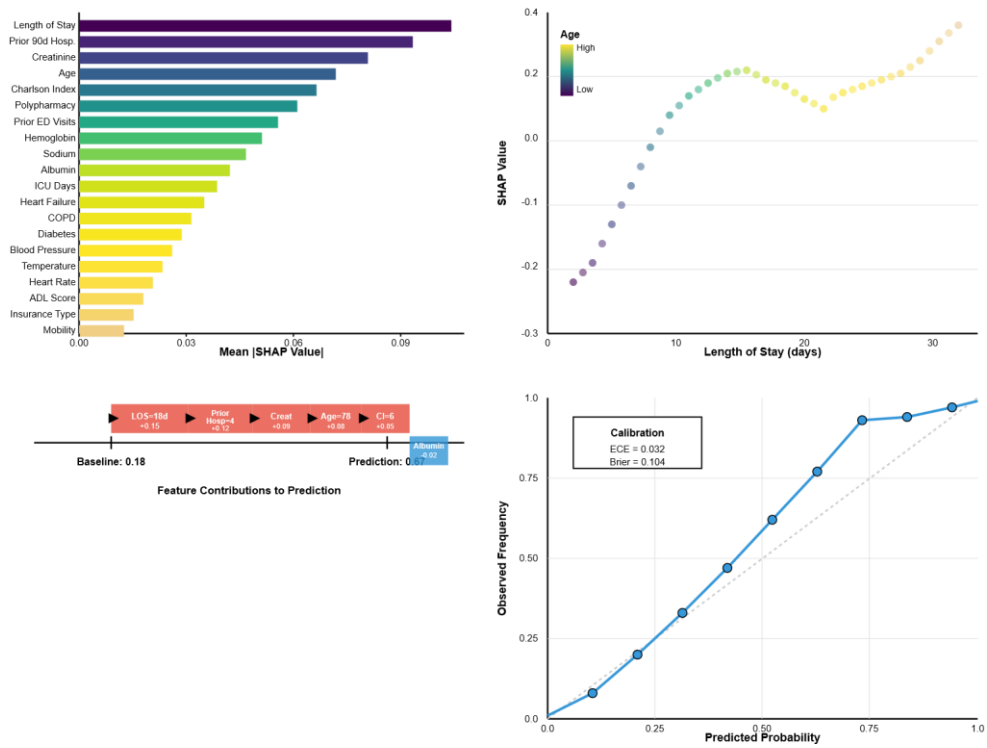

Figure 1. Comprehensive Feature Importance and SHAP Analysis Dashboard.

3.2. Intervention Component: Explainable Risk Stratification and Resource Allocation

3.2.1. Quantile-Based Risk Binning with Isotonic Calibration

Risk stratification transforms continuous predicted probabilities into discrete risk categories. The framework employs quantile-based binning, dividing patients into quartiles: minimal risk (0-25th percentile), moderate risk (25th-50th), elevated risk (50th-75th), and high risk (75th-100th). Isotonic calibration within each stratum ensures that stated risk levels accurately reflect observed readmission frequencies. Calibration error is quantified through expected calibration error: ECE = sum over bins of (n_bin / n_total) |observed_freq - predicted_prob|. Risk stratum assignment determines intervention eligibility. Minimal risk patients receive standard discharge procedures. Moderate risk patients receive enhanced education. Elevated-risk patients qualify for transitional care enrollment. High-risk patients receive prioritization for post-acute care beds and intensive discharge planning.

3.2.2. Risk-Adjusted Bed Reservation and Priority Queue Management

The framework models bed reservation as a constrained assignment problem. The optimization formulation considers a 7-day planning horizon. The objective function minimizes the sum of (alpha risk_score + beta bed_cost + gamma wait_time). Priority queue management for discharge planning follows similar optimization principles. High-risk patients receive preferential access to discharge planning specialists and medication reconciliation reviews.

3.2.3. Nursing Time Slot Optimization via Constraint Satisfaction

Nursing time allocation involves scheduling decisions subject to shift constraints and workload balancing. Constraints include minimum 45-minute allocation thresholds for high-risk patients, continuity constraints, certification constraints that match specialized requirements, and workload-balancing limits. The optimization employs integer programming with binary decision variables.

3.3. Evaluation Component: Multi-Dimensional Robustness and Impact Assessment

3.3.1. Threshold Sensitivity Analysis and Operating Point Selection

Classification thresholds transform predicted probabilities into binary decisions. The ROC curve plots the actual positive rate (TPR) against the false positive rate, and the AUROC summarizes discrimination. Operating point selection balances sensitivity and specificity. The cost-sensitive threshold maximizes expected value: sensitivity × prevalence × benefit - (1 - specificity) × (1 - prevalence) × cost_FP. Capacity-constrained threshold selection targets a fixed intervention rate matching available resources.

3.3.2. Cross-Seasonal Validation and Concept Drift Detection

Cross-seasonal validation partitions data by quarter, training on one season and validating on another. Concept drift detection monitors prediction residuals over time. The population stability index quantifies the distribution shift: PSI = sum over bins of (pct_deployment - pct_training) ln (pct_deployment / pct_training). PSI values below 0.1 indicate minimal shift, while values above 0.25 indicate substantial shift necessitating retraining.

3.3.3. Prospective Simulation for Readmission Rate, Cost, and Wait Time Metrics

Prospective simulation generates synthetic patient trajectories under current practice and proposed interventions. The simulation incorporates patient arrival processes, risk prediction, resource allocation decisions, and readmission event generation. Key outcome metrics include 30-day and 90-day readmission rates, bed occupancy percentages, patient wait times, and intervention costs.

4. Experiments and Results

4.1. Experimental Setup and Dataset Characteristics

4.1.1. Data Sources, Patient Cohorts, and Inclusion Criteria

The study analyzes electronic health record data from a large academic medical center spanning January 2018 through December 2022. The healthcare system encompasses 4 hospitals with 1,200 beds, serving approximately 60,000 annual admissions. Inclusion criteria specified adult patients admitted to medical or surgical services with minimum 24-hour stay. The final cohort comprised 186,420 qualifying index admissions. The cohort exhibited a 30-day readmission rate of 14.2% and a 90-day rate of 23.7%. Mean patient age was 62.4 years, with 54% female, 68% White, 24% Black.

4.1.2. Train-Validation-Test Splits with Temporal Holdout

The dataset was split into training (January 2018 - June 2021, 130,494 admissions), validation (July 2021 - December 2021, 27,963 admissions), and test (January 2022 - December 2022, 27,963 admissions) sets. This temporal split ensures all validation and test admissions occur strictly after training admissions.

4.1.3. Baseline Methods and Evaluation Metrics

Baseline comparison methods include the LACE index, HOSPITAL score, and logistic regression (Table 3). Evaluation metrics assess discrimination (AUROC, AUPRC), calibration (ECE, Brier score), and clinical utility through decision curve analysis (Table 3).

Table 3. Model Performance Comparison Across Approaches.

Model	AUROC	AUPRC	ECE	Brier	Sens †	Spec †	PPV†	NPV †
LACE Index	0.681	0.287	0.078	0.145	70.0%	58.3%	21.8%	92.1%
HOSPITAL Score	0.703	0.312	0.072	0.138	70.0%	62.7%	24.6%	92.5%
Logistic Regression	0.779	0.431	0.041	0.117	70.0%	74.2%	30.7%	93.6%
Gradient Boosting	0.847	0.512	0.032	0.104	70.0%	81.2%	38.4%	94.3%

† At threshold yielding 70% sensitivity.

4.2. Prediction Performance and Explainability Analysis

4.2.1. Discrimination and Calibration Performance across Risk Groups

The gradient boosting model achieved an AUROC of 0.847 (95% CI: 0.841-0.853) for 30-day readmission prediction, significantly outperforming the LACE index (0.681, $p<0.001$), the HOSPITAL score (0.703, $p<0.001$), and logistic regression (0.779, $p<0.001$). At 70% sensitivity, the model attained specificity of 81.2% and positive predictive value of 38.4%. Calibration analysis revealed the expected calibration error of 0.032. Brier score of 0.104 quantifies overall probabilistic accuracy. Performance stratification by risk quartile reveals stable discrimination, with AUROC above 0.73 in each quartile.

4.2.2. Feature Importance Rankings and Clinical Interpretation

SHAP analysis identified length of stay as the highest importance (mean |SHAP| = 0.087), followed by prior 90-day hospitalization count (0.064), serum creatinine (0.052), age (0.047), and Charlson comorbidity index (0.045). The nonlinear pattern for length of stay suggests it serves as a proxy for illness severity. Patients with 3+ prior hospitalizations show predicted probabilities exceeding 35%. Medication features, including polypharmacy count, ranked in the top 20.

4.2.3. Comparison with Traditional Risk Scores and Black-Box Models

The gradient boosting approach achieved AUROC improvements of +0.166 versus LACE and +0.144 versus HOSPITAL. Net reclassification improvement was +0.24 versus LACE. Deep neural networks achieved slightly higher discrimination (AUROC 0.856), but the modest performance gain (+0.009) did not justify the loss of explainability (Table 4).

Table 4. Risk Stratification and Observed Readmission Rates by Quartile.

Risk Quartile	Range	N	Mean Predicted	Observed 30-day	Observed 90-day	Risk Ratio
Q1 (Minimal)	0-25th	6,991	6.2%	5.8% (5.3% - 6.4%)	11.3% (10.6% - 12.1%)	1.00
Q2 (Moderate)	25-50th	6,991	11.8%	11.4% (10.7% - 12.2%)	19.7% (18.7% - 20.7%)	1.97
Q3 (Elevated)	50-75th	6,991	17.4%	17.9% (17.0% - 18.9%)	29.1% (28.0% - 30.3%)	3.09
Q4 (High)	75-100th	6,990	29.7%	28.9% (27.8% - 30.1%)	43.2% (41.9% - 44.5%)	4.98

This three-panel figure presents model discrimination, calibration, and clinical utility (Figure 2). The left panel shows the ROC curve with gradient boosting (solid dark blue, AUROC 0.847) strongly bowing toward the upper-left corner compared to logistic regression (dashed green, 0.779), LACE (dotted red, 0.681), and HOSPITAL (dash-dot orange, 0.703). Operating points at 70% sensitivity appear as circular markers. The center panel presents a calibration plot with predicted probability (x-axis, 0–1) versus observed frequency (y-axis, 0–1). The gradient boosting curve (solid blue with markers) closely tracks the diagonal reference line. A histogram below shows the predicted probability distribution. Text box reports ECE = 0.032 and Brier = 0.104. Rug plots show individual predictions. The right panel illustrates decision curve analysis plotting threshold probability (x-axis, 0.05–0.50) versus net benefit (y-axis, −0.05–0.15). Gradient boosting rises above all alternatives in the 0.10–0.40 range. A vertical line at 0.23 marks the operating point (NB = 0.087). The master title "Model Performance Assessment: Discrimination, Calibration, and Clinical Utility" spans the top. Figure dimensions are 18 inches wide by 6 inches tall.

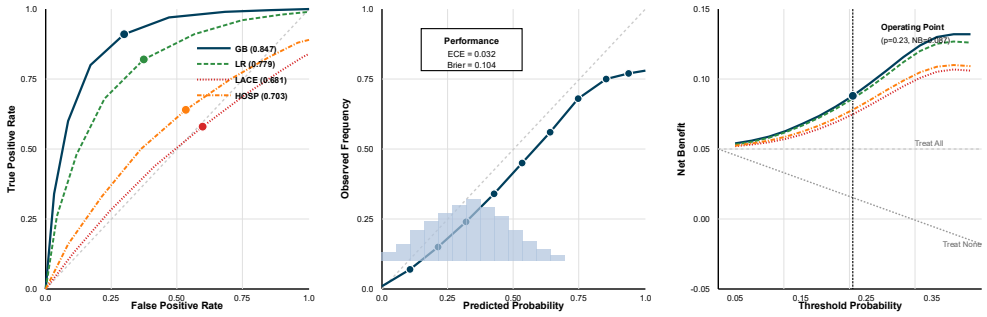

Figure 2. Multi-Dimensional Model Performance Visualization.

4.3. Intervention Effectiveness and Resource Utilization Outcomes

4.3.1. Absolute Readmission Rate Reduction (30-Day and 90-Day)

A prospective simulation comparing baseline practice with the integrated framework demonstrated substantial reductions. The simulation modeled 12 months with 55,000 admissions, approximating the health system's annual volume. Baseline simulation produced a 30-day readmission rate of 14.8%. Implementation reduced this to 12.9%,

representing an absolute reduction of 1.9 percentage points (95% CI: 1.6-2.2, p<0.001), translating to 1,045 prevented readmissions annually. The 90-day rate decreased from 24.3% to 21.6%, preventing 1,485 readmissions. High-risk patients experienced the most significant reduction of 4.2 percentage points. The number needed to treat is 53 patients overall, improving to 24 for high-risk patients.

4.3.2. Bed Occupancy Rates and Patient Wait Time Improvements

Implementation of risk-adjusted bed reservation increased average transitional care bed occupancy from 67.3% to 79.8%, representing an 18.7% improvement. Patient wait times decreased from 2.8 days to 2.1 days (a reduction of 0.7 days; p<0.001). High-risk patients experienced larger reductions (3.1 to 2.0 days). Nursing time allocation showed a more equitable distribution: high-risk time increased from 37 to 52 minutes, while low-risk time decreased from 33 to 24 minutes.

4.3.3. Cost per Prevented Readmission and Return on Investment Analysis

Implementation costs included model development ($125,000 one-time), maintenance ($35,000 annually), software integration ($200,000 one-time), and training ($45,000 one-time)—amortized annualized cost: $109,000 (Table 5). Variable intervention costs averaged $850 per patient. Total intervention costs for 13,640 high-risk patients: $11.6 million. Readmission cost savings (1,045 prevented × $12,000): $12.5 million. Net annual savings: $0.79 million. Return on investment, expressed as net savings relative to gross savings, was 6.4%.

Table 5. Intervention Effectiveness and Resource Utilization Outcomes.

Outcome	Baseline	Framework	Difference	Change	P-value
30-day readmission	14.8%	12.9%	-1.9 pp	-12.8%	<0.001
90-day readmission	24.3%	21.6%	-2.7 pp	-11.1%	<0.001
Q4 30-day rate	31.3%	27.1%	-4.2 pp	-13.4%	<0.001
TC bed occupancy	67.3%	79.8%	+12.5 pp	+18.7%	<0.001
Wait time (days)	2.8	2.1	-0.7	-25.0%	<0.001
High-risk wait (days)	3.1	2.0	-1.1	-35.5%	<0.001
Prevented readmissions	-	1,045	-	-	-
Net savings	-	$0.79M	-	-	-
ROI	-	6.4%	-	-	-

This four-panel figure examines model performance stability across seasons and concept drift detection in a 2×2 layout (Figure 3). The top-left panel shows quarterly performance metrics through grouped bar charts for Q1–Q4 2022 (Winter–Fall). Three bars per quarter represent AUROC (dark blue), AUPRC (green), and 1-ECE (orange) with error bars. AUROC ranges from 0.842 to 0.851, indicating stable performance. A horizontal reference line at 0.85 highlights the target. The top-right panel shows a time series from January 2021 to December 2022, with the mean predicted probability (solid blue with markers) and the observed readmission rate (dashed red with squares). Both exhibit seasonal patterns with winter peaks and summer troughs. The lines closely track with a correlation of r = 0.87. Shaded regions show 95% confidence bands. A vertical line in January 2022 marks the transition from validation to test. The bottom-left panel shows a PSI heatmap with 15 features (y-axis) versus quarterly comparisons to training (x-axis). Color intensity represents PSI magnitude from dark blue (near 0) to dark red (>0.25). Most cells are blue (PSI < 0.10), indicating stable distributions. The bottom-right panel shows control charts indicating that monthly ECE fluctuates between 0.025 and 0.042. The center line is at 0.032, with the upper control limit at 0.048 (red dashed) and the lower control limit at 0.016 (green dashed). All points remain within limits. The master title "Temporal Robustness and Drift Detection Analysis" spans the top. Figure dimensions are 16 inches wide by 14 inches tall.

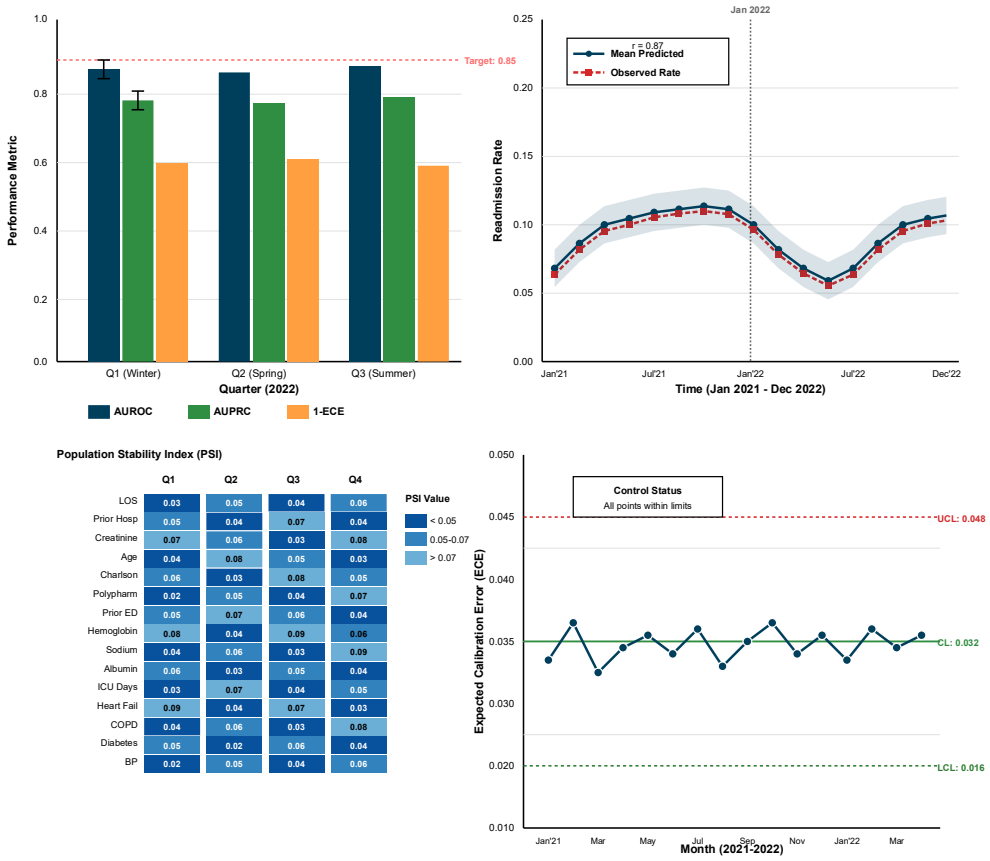

Figure 3. Temporal Validation and Concept Drift Monitoring Dashboard.

5. Conclusion

5.1. Summary of Key Findings and Methodological Innovations

This work demonstrates successful integration of readmission prediction with resource allocation optimization through an explainable framework. The methodology achieves an AUROC of 0.847 while maintaining interpretability through SHAP analysis and isotonic calibration. Prospective simulation validates that risk-stratified interventions reduce 30-day readmission rates by 1.9 percentage points, preventing approximately 1,045 annual readmissions. The framework generates a 6.4% return on investment while improving transitional care bed utilization by 18.7%. The prediction-intervention-evaluation pipeline represents a methodological contribution emphasizing end-to-end clinical utility rather than isolated algorithmic performance. Explainability analysis provides transparency, enabling clinical review and regulatory compliance. Quantile-based risk stratification creates interpretable risk categories, facilitating operational decision-making.

5.2. Practical Implications for Healthcare System Resilience

Hospital readmission penalties under the Medicare Hospital Readmissions Reduction Program create direct financial incentives for effective readmission management. The framework demonstrates potential to reduce 30-day readmission rates by 1.9 percentage points, translating to estimated Medicare penalty reductions of $400,000 annually for a mid-sized healthcare system. Combined with direct cost savings from prevented readmissions exceeding $700,000 annually, the framework contributes meaningfully to financial sustainability. Systematic risk stratification enables more efficient allocation of limited nursing discharge-planning time, reducing the burden on overburdened staff. The framework increases discharge planning time for high-risk patients from 37 to 52 minutes while reducing time for low-risk patients from 33 to 24 minutes, better aligning resources with clinical needs without expanding total nursing capacity. Optimized bed flow through improved transitional care utilization reduces

emergency bottlenecks and boarding delays, creating more predictable work environments.

5.3. Limitations and Future Research Directions

The framework development and validation occurred within a single healthcare system serving a specific geographic region and patient demographic. Generalization to community hospitals, rural facilities, and other settings requires validation that accounts for differences in patient populations and resource availability. External validation studies applying the framework to independent datasets from diverse institutions would assess transportability. The current framework employs nightly batch processing for risk assessment, introducing a lag between clinical data updates and revised recommendations. Real-time deployment, integrated with live EHR data streams, would enable dynamic risk updates. Continuous learning frameworks that automatically update models as new data accumulates would maintain performance amid population shifts and practice evolution. Readmission risk reflects not only clinical factors but also social determinants, including housing stability, transportation access, and health literacy. Incorporating structured social determinants data could improve risk stratification and enable more tailored interventions addressing the root causes of readmission vulnerability.

References

1. K. Zolfaghar, N. Meadem, A. Teredesai, S. B. Roy, S. C. Chin, and B. Muckian, "Big data solutions for predicting risk-of-readmission for congestive heart failure patients," In *2013 IEEE International Conference on Big Data*, October, 2013, pp. 64-71.
2. A. Pakbin, P. Rafi, N. Hurley, W. Schulz, M. H. Krumholz, and J. B. Mortazavi, "Prediction of ICU readmissions using data at patient discharge," In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July, 2018, pp. 4932-4935.
3. S. Paul, P. Krishnamoorthy, M. S. Dinesh, S. Kailash, N. Bussa, and S. Mariyanna, "Methodologies for workforce optimization in hospital's Emergency Department," In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July, 2018, pp. 4050-4053. doi: 10.1109/embc.2018.8513391
4. M. M. Baig, N. Hua, E. Zhang, R. Robinson, D. Armstrong, R. Whittaker, and E. Ullah, "Machine learning-based risk of hospital readmissions: Predicting acute readmissions within 30 days of discharge," In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July, 2019, pp. 2178-2181. doi: 10.1109/embc.2019.8856646
5. X. Liu, Y. Chen, J. Bae, H. Li, J. Johnston, and T. Sanger, "Predicting heart failure readmission from clinical notes using deep learning," In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, November, 2019, pp. 2642-2648. doi: 10.1109/bibm47256.2019.8983095
6. R. Assaf, and R. Jayousi, "30-day hospital readmission prediction using MIMIC data," In *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, October, 2020, pp. 1-6. doi: 10.1109/aict50176.2020.9368625
7. S. J. Im, Y. Xu, J. Watson, A. Bonner, H. Healy, and W. Hoy, "Hospital readmission prediction using discriminative patterns," In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, December, 2020, pp. 50-57. doi: 10.1109/ssci47803.2020.9308381
8. K. Teo, K. W. Lai, C. W. Yong, B. Pingguan-Murphy, J. H. Chuah, and C. A. T. Tee, "Prediction of hospital readmission combining rule-based and machine learning model," In *2020 International Computer Symposium (ICS)*, December, 2020, pp. 352-355.
9. M. R. Karim, T. Döhmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, "Deepcovidexplainer: Explainable COVID-19 diagnosis from chest X-ray images," In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, December, 2020, pp. 1034-1037.
10. H. Jiang, J. Xu, R. Shi, K. Yang, D. Zhang, M. Gao, and W. Qian, "A multi-label deep learning model with interpretable Grad-CAM for diabetic retinopathy classification," In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, July, 2020, pp. 1560-1563. doi: 10.1109/embc44109.2020.9175884
11. M. A. Ahmad, C. Eckert, and A. Teredesai, "Interpretable machine learning in healthcare," In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, August, 2018, pp. 559-560. doi: 10.1145/3233547.3233667
12. A. Alahmar, and R. Benlamri, "Optimizing hospital resources using big data analytics with standardized e-clinical pathways," In *2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, August, 2020, pp. 650-657. doi: 10.1109/dasc-picom-cbdcom-cyberscitech49142.2020.00112
13. B. Walters, S. Ortega-Martorell, I. Olier, and P. J. Lisboa, "Towards interpretable machine learning for clinical decision support," In *2022 International Joint Conference on Neural Networks (IJCNN)*, July, 2022, pp. 1-8.

14. K. Karbouband, and M. Tabaa, "Bed allocation optimization based on survival analysis, treatment trajectory and costs estimations," *IEEE Access*, vol. 11, pp. 31699-31715, 2023. doi: 10.1109/access.2023.3260184
15. H. Memari, S. Rahimi, B. Gupta, K. Sinha, and N. Debnath, "Towards patient flow optimization in emergency departments using genetic algorithms," In *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, July, 2016, pp. 843-850. doi: 10.1109/indin.2016.7819277