ML-ENHANCE: A Machine Learning Framework for Hospital Readmission Prediction with Cost-Optimized Decision Thresholds

Barak Gahtan[1,*], Yakir Segev[1,6], Alex M. Bronstein[1,7], Howard Amital[2], and Avishai M. Tsur[2,3,4,5]

[1]Department of Computer Science, Technion-Israel Institute of Technology, Israel
[2]Department of Internal Medicine B & Zabludowicz Center for Autoimmune Diseases, Sheba Medical Center, Ramat Gan, Israel
[3]Israel Defense Forces, Medical Corps, Tel Hashomer, Ramat Gan, Israel
[4]Department of Military Medicine, Hebrew University of Jerusalem Faculty of Medicine, Jerusalem, Israel
[5]School of Public Health, Gray Faculty of Medical and Health Sciences, Tel Aviv University, Tel Aviv, Israel
[6]Division of Gynecology Oncology, Department of Obstetrics and Gynecology, Carmel Medical Center, Haifa, Israel
[7]Institute of Science and Technology Austria (ISTA)
[*]Corresponding author: barakgahtan@cs.technion.ac.il

Abstract

Hospital readmissions within 30 days affect 15–20% of patients, generating over $40 billion in potentially preventable costs annually. We developed two prediction models using 113,312 internal medicine admissions (2017–2023): ENHANCE, an interpretable clinical score, and ML-ENHANCE, a machine learning ensemble. Using decision-theoretic threshold optimization across cost ratios from 0.5:1 to 20:1, we demonstrate that ML-ENHANCE (AUC 0.752) substantially outperforms ENHANCE (AUC 0.696) and HOSPITAL (AUC 0.676). At the 5:1 cost ratio, ML-ENHANCE matches ENHANCE sensitivity (76.9%) while achieving 8.3 percentage points higher specificity, translating to 7,400 fewer unnecessary interventions annually and projected savings of $69.3 million versus $14.7 million for ENHANCE. ML-ENHANCE achieves 4–5 times greater economic impact, supporting deployment in automated population health applications.

1 Introduction

Hospital readmissions within 30 days of discharge represent a critical healthcare challenge, affecting 15–20% of inpatients and generating over $40 billion in potentially preventable costs annually in the United States [1, 2]. Beyond economic burden, readmissions signal failures in care transitions while undermining quality metrics that increasingly determine institutional reimbursement [3].

The HOSPITAL score, the most extensively validated clinical instrument for 30-day readmission prediction, achieves only moderate discrimination (AUC 0.65–0.70) with a limited 2.8-fold risk discrimination range [4]. This performance ceiling reflects the score's development before comprehensive electronic health records, overlooking readily available clinical data. More critically, its three-category risk stratification provides insufficient granularity for precise intervention targeting.

Contemporary machine learning approaches demonstrate 0.05–0.10 AUC improvements over traditional scores [5, 6]. While interpretability concerns have historically limited adoption, two developments warrant reconsideration. First, healthcare systems increasingly deploy automated population health platforms where algorithmic transparency is less critical than in bedside decision-making. Second, traditional AUC validation fails to demonstrate whether discrimination improvements translate to clinically meaningful operating characteristics under realistic cost constraints.

We developed a dual-model framework analyzing 113,312 consecutive hospitalizations (2017–2023) with three objectives: creating an interpretable ENHANCE score, developing ML-ENHANCE using ensemble methods to establish the performance ceiling, and employing decision-theoretic threshold optimization to quantify clinical and economic value under explicit cost constraints. Our analysis demonstrates that ML-ENHANCE delivers substantially superior performance (AUC 0.752 vs 0.676), and cost-based optimization reveals this translates to 4–5 times greater economic impact than interpretable alternatives, supporting stratified deployment based on use-case requirements.

2 Methods

2.1 Study Design and Population

We performed a retrospective cohort study using electronic health record data from a large academic health system (2017–2023). The Institutional Review Board approved the project with waiver of informed consent. The study included adult patients ($\geq$18 years) with internal medicine admissions, excluding those with missing discharge dates, length of stay $\leq$1 or >365 days, or missing essential demographic data. The final cohort comprised 113,312 patients with 23,692 (20.9%) experiencing 30-day readmission. Temporal validation used 2017–2020 for development (n=54,353) and 2021–2023 for validation (n=58,959; Supplementary Table S26. The study flow diagram is presented in Figure 1.

2.2 ENHANCE Score Development

We employed a nine-domain analytical framework examining comorbidity patterns, laboratory values, emergency department admission patterns, temporal factors, ECG findings, subgroup interactions, machine learning benchmarks, temporal stability, and comprehensive integration (Supplementary Appendix). The ENHANCE score was developed using sparse logistic regression with L1 regularization (C=1.5). The final scoring system (Table 1; detailed derivation in Supplementary Table S22 incorporates: baseline HOSPITAL score (1 point per unit), enhanced laboratory thresholds (severe hypoalbuminemia <3.0 g/dL, moderate anemia <10.0 g/dL, kidney dysfunction urea >60 mg/dL: 2 points each), clinical factors (malignancy history: 1 point; prior admissions 1–2/3–4/$\geq$5: 2/4/6 points), interaction terms (CKD+CVA: 2 points, CHF+ED: 1 point), and weekend admission (1 point). Risk categories were defined as Low (0–3), Intermediate (4–6), High (7–10), and Very High (>10 points).

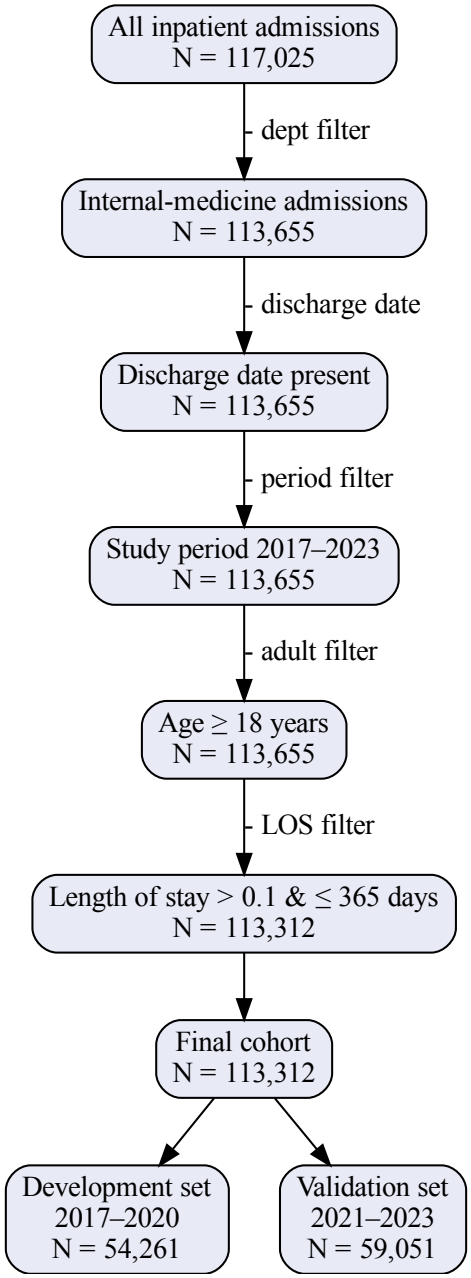


Figure 1: Study flow diagram illustrating patient selection, inclusion and exclusion criteria, and temporal validation split. The final cohort of 113,312 patients is divided into development (54,353; 2017–2020) and validation (58,959; 2021–2023) sets.

Detailed methodology and results from each analytical domain are provided in the Supplementary Appendix, including comorbidity analysis (Supplementary Tables S4-S6, S5, and S14, Supplementary Figure S7, laboratory value optimization (Supplementary Table S7, Supplementary Figure S8, emergency department patterns (Supplementary Tables S8-S9, Supplementary Figure S9, temporal patterns (Supplementary Tables S10-S11, Supplementary Figure S10), ECG findings (Supplementary Table S12, Supplementary Figure S11, subgroup interactions (Supplementary S13-S15, Supplementary Figure S12), machine learning benchmarks (Supplementary Tables S16-S18, Supplementary Figure S13), and temporal validation (Supplementary Tables S19-S21, Supplementary Figure S14).

Table 1: ENHANCE Score Components and Point Assignment

Component	**Criteria**	**Points**
HOSPITAL Score	Original 7 components	$\times 1$
Severe hypoalbuminemia	Albumin < 3.0 g/dL	$+2$
Moderate anemia	Hemoglobin < 10.0 g/dL	$+2$
Kidney dysfunction	Urea > 60 mg/dL	$+2$
History of malignancy	Present	$+1$
Prior admissions (1–2)	Prior year count	$+2$
Prior admissions (3–4)	Prior year count	$+4$
Prior admissions (≥ 5)	Prior year count	$+6$
CKD + CVA synergy	Both conditions present	$+2$
CHF + ED interaction	Both factors present	$+1$
Weekend admission	Weekend	$+1$

2.3 ML-ENHANCE Development

ML-ENHANCE employs a soft voting classifier ensemble combining random forest, gradient boosting, and XGBoost (each with 100 estimators). The ensemble uses 14 features identified through the analytical framework, including prior admissions count, enhanced laboratory score, HOSPITAL components, and comorbidity indicators. Hyperparameters were optimized using 5-fold stratified cross-validation on the development set, with evaluation on the temporally held-out validation set. ML-ENHANCE outputs continuous probability estimates enabling threshold optimization for specific cost scenarios.

2.4 Statistical Analysis

Model discrimination was assessed using AUC with 95% confidence intervals (DeLong's method). Calibration was evaluated using reliability diagrams. Statistical significance was set at $p<0.05$. Analyses were performed using Python 3.9 with scikit-learn. The primary outcome was 30-day all-cause readmission.

2.5 Decision-Theoretic Threshold Optimization

Clinical deployment requires selecting specific thresholds rather than relying solely on AUC. The expected cost per patient at threshold t is:

$$\mathcal{C}(t) = \text{FPR}(t) \times (1 - \pi) \times C_{\text{FP}} + \text{FNR}(t) \times \pi \times C_{\text{FN}} \tag{1}$$

where $\pi = 0.209$ is baseline readmission prevalence. We estimated average readmission cost at \$16,300 based on Healthcare Cost and Utilization Project data [7]. For Israeli capitation-based systems, we adopted a

day-denominated framework (interventions: 1 day; missed readmissions: 3 days), yielding a 3:1 cost ratio. We evaluated ratios from 0.5:1 through 20:1, with confidence intervals from 300 bootstrap iterations. Cost scenario details are presented in Supplementary Table S1.

3 Results

3.1 Study Population

The cohort comprised 113,312 patients (mean age 70.0±17.0 years; 52.6% male). The 30-day readmission rate was 20.9%. Patients who were readmitted were older (71.8±16.0 vs 69.4±17.2 years), more likely male (54.4% vs 52.2%), and had higher comorbidity burden including diabetes (35.6% vs 27.6%), chronic kidney disease (11.7% vs 7.3%), and malignancy (11.3% vs 6.7%; all p<0.001); readmission rates increased progressively with the number of comorbidities (Supplementary Table S5). Complete baseline characteristics are presented in Table 2. The HOSPITAL score achieved AUC 0.676 (95% CI: 0.671–0.681) with readmission rates of 16.2% (low risk), 33.0% (intermediate), and 45.3% (high risk), representing 2.8-fold discrimination.

Table 2: Baseline Characteristics of Study Population Stratified by 30-Day Readmission Status

Characteristic	**No Readmission (n = 89,620)**	**Readmission (n = 23,692)**	**P-value**
Demographics			
Age, mean ± SD (years)	69.4 ± 17.2	71.8 ± 16.0	< 0.001
Male sex, n (%)	46,752 (52.2)	12,891 (54.4)	< 0.001
Comorbidities, n (%)			
Hypertension	45,846 (51.2)	11,942 (50.4)	0.037
Diabetes mellitus	24,711 (27.6)	8,442 (35.6)	< 0.001
Chronic kidney disease	6,549 (7.3)	2,782 (11.7)	< 0.001
Malignancy	6,007 (6.7)	2,673 (11.3)	< 0.001
Congestive heart failure	4,485 (5.0)	1,820 (7.7)	< 0.001
Laboratory Values, mean ± SD			
Hemoglobin (g/dL)	11.8 ± 2.3	10.8 ± 2.4	< 0.001
Sodium (mEq/L)	139.3 ± 3.9	138.9 ± 4.5	< 0.001
Albumin (g/dL)	3.5 ± 0.5	3.2 ± 0.6	< 0.001
Clinical Factors			
Length of stay, median (IQR)	1.6 (0.8–3.1)	1.8 (0.9–3.4)	< 0.001
Emergency admission, n (%)	86,505 (96.5)	22,747 (96.0)	< 0.001
Prior year admissions, median (IQR)	0 (0–1)	1 (0–2)	< 0.001
HOSPITAL score, mean ± SD	2.9 ± 1.8	3.8 ± 2.1	< 0.001

3.2 Model Performance

ENHANCE achieved AUC 0.696 (95% CI: 0.691–0.701), a statistically significant improvement of 0.020 over HOSPITAL (p<0.001), while maintaining clinical interpretability (Figure 2; component-level performance in Supplementary Table S23). The four-tier risk stratification achieved 4.2-fold discrimination (44.8% vs 10.8% readmission rates; Supplementary Table S25), identifying 29,558 high-risk patients compared to HOSPITAL's 7,554—a 391% improvement in high-risk detection.

ML-ENHANCE achieved substantially superior discrimination (AUC 0.752; 95% CI: 0.747–0.757), representing 0.076 improvement over HOSPITAL and capturing 73.7% of the improvement gap inaccessible to

interpretable methods. The ensemble demonstrated acceptable generalization (training AUC 0.763, validation AUC 0.752). Feature importance analysis confirmed prior admissions as the dominant predictor, followed by laboratory scores and baseline HOSPITAL components (Supplementary Table S17). Bootstrap analysis established statistically significant superiority over both HOSPITAL (difference 0.076; 95% CI: 0.071–0.081) and ENHANCE (difference 0.056; 95% CI: 0.051–0.061; Table 3; Supplementary Table S24).

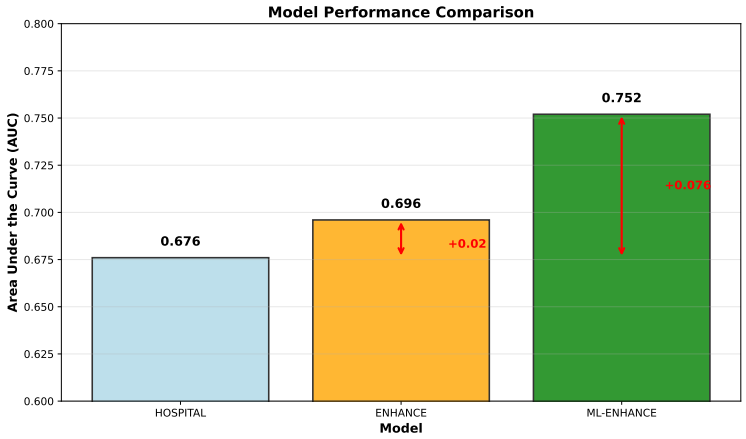


Figure 2: Model performance comparison showing AUC values for HOSPITAL (0.676), ENHANCE (0.696), and ML-ENHANCE (0.752).

Table 3: Area under the ROC curve and Youden's J statistic with 95% bootstrap confidence intervals.

Model	**AUC [95% CI]**	**Youden's J [95% CI]**	**J Threshold**
HOSPITAL	0.676 [0.673, 0.680]	0.259 [0.253, 0.264]	3.0
ENHANCE	0.696 [0.693, 0.700]	0.291 [0.285, 0.296]	7.0
ML-ENHANCE	0.753 [0.749, 0.756]	0.369 [0.364, 0.376]	0.22

ML-ENHANCE demonstrated substantially superior risk stratification (Figure 3), with patients in the lowest decile experiencing 3.7% readmission rate versus 56.4% in the highest—a 15.2-fold range. When grouped into quartiles, ML-ENHANCE achieved 7.5-fold discrimination compared to 4.2-fold for ENHANCE.

3.3 Temporal Validation

All models demonstrated excellent temporal stability. Temporal validation showed HOSPITAL testing AUC 0.676 (difference from training: −0.001), ENHANCE 0.687 (+0.002), and ML-ENHANCE 0.752 (−0.011). Both ENHANCE (Brier score 0.1522) and ML-ENHANCE (0.1487) achieved excellent calibration (Supplementary Figure S1). Cost scenario details are presented in Supplementary Table S1.

Additional validation analyses are presented in the Supplementary Appendix, including ROC curve comparisons (Supplementary Figure S2), ENHANCE risk stratification details (Supplementary Figure S3), temporal stability analysis (Supplementary Figure S4), subgroup performance (Supplementary Figure S5), comprehensive calibration analysis (Supplementary Figure S6), and the comprehensive ENHANCE score analysis (Supplementary Figure S15).

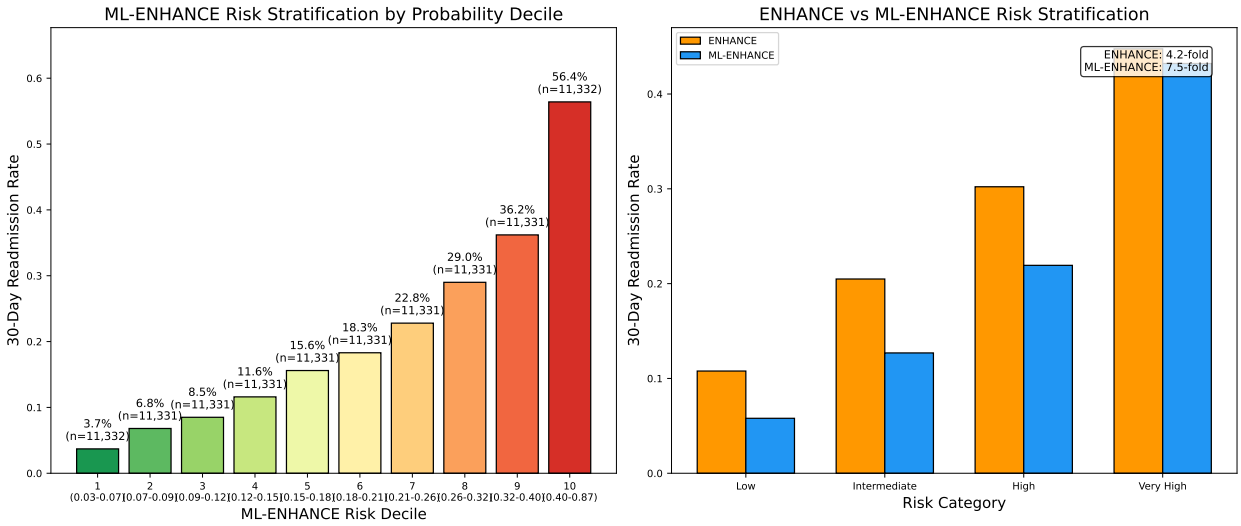


Figure 3: ML-ENHANCE risk stratification. Left panel: Readmission rates by ML-ENHANCE probability decile, demonstrating 15.2-fold discrimination (3.7% to 56.4%) across the risk spectrum. Right panel: Direct comparison of risk stratification performance showing ML-ENHANCE achieves 7.5-fold discrimination versus 4.2-fold for ENHANCE, when patients are grouped into equivalent quartile-based risk categories.

3.4 Cost-Based Threshold Optimization

Decision-theoretic analysis revealed ML-ENHANCE's discrimination advantage translates to substantial benefits across all cost scenarios (Figure 4). At the policy-relevant 5:1 cost ratio ($3,260 interventions), ML-ENHANCE and ENHANCE achieved identical sensitivity (76.9%) but ML-ENHANCE demonstrated 8.3 percentage points higher specificity (58.9% vs 50.6%), translating to approximately 7,400 fewer unnecessary interventions annually while maintaining equivalent readmission capture. ML-ENHANCE achieved 12.9% expected cost reduction versus 2.8% for ENHANCE—a 4.6-fold advantage.

Cost-optimal thresholds diverged substantially from Youden's J recommendations. For ENHANCE, Youden's J threshold of 7.0 (sensitivity 63.1%) was dominated by the cost-optimal threshold of 4.1 (sensitivity 76.9%), representing 13.8 percentage points sacrificed by cost-agnostic selection. At high cost ratios (2:1), ML-ENHANCE identified 39.9% of readmissions versus HOSPITAL's 22.2% while maintaining comparable specificity. At low cost ratios (10:1), HOSPITAL collapsed to near-universal screening (99.5% sensitivity, 1.4% specificity) while ML-ENHANCE maintained meaningful stratification (92.8% sensitivity, 30.8% specificity). Complete threshold optimization results are presented in Supplementary Table S2.

Model-specific threshold optimization analyses for HOSPITAL (Supplementary Figure S16), ENHANCE (Supplementary Figure S17), and ML-ENHANCE (Supplementary Figure S18) provide detailed cost curves and operating characteristics for each model.

Table 4 presents projected economic impact for a high-volume U.S. institution (50,000 annual discharges) at the 5:1 cost ratio. ML-ENHANCE achieves $69.3 million annual savings (13.1% reduction) versus $14.7 million for ENHANCE (2.8%)—a 4.7-fold differential. Critically, ML-ENHANCE identifies 224 additional preventable readmissions compared to HOSPITAL while simultaneously reducing false positives by 3,069 patients, demonstrating that improved discrimination benefits both sensitivity and specificity. The expected cost per patient decreases from $10,612 (HOSPITAL) to $9,226 (ML-ENHANCE), translating to $1,386 savings per discharge.

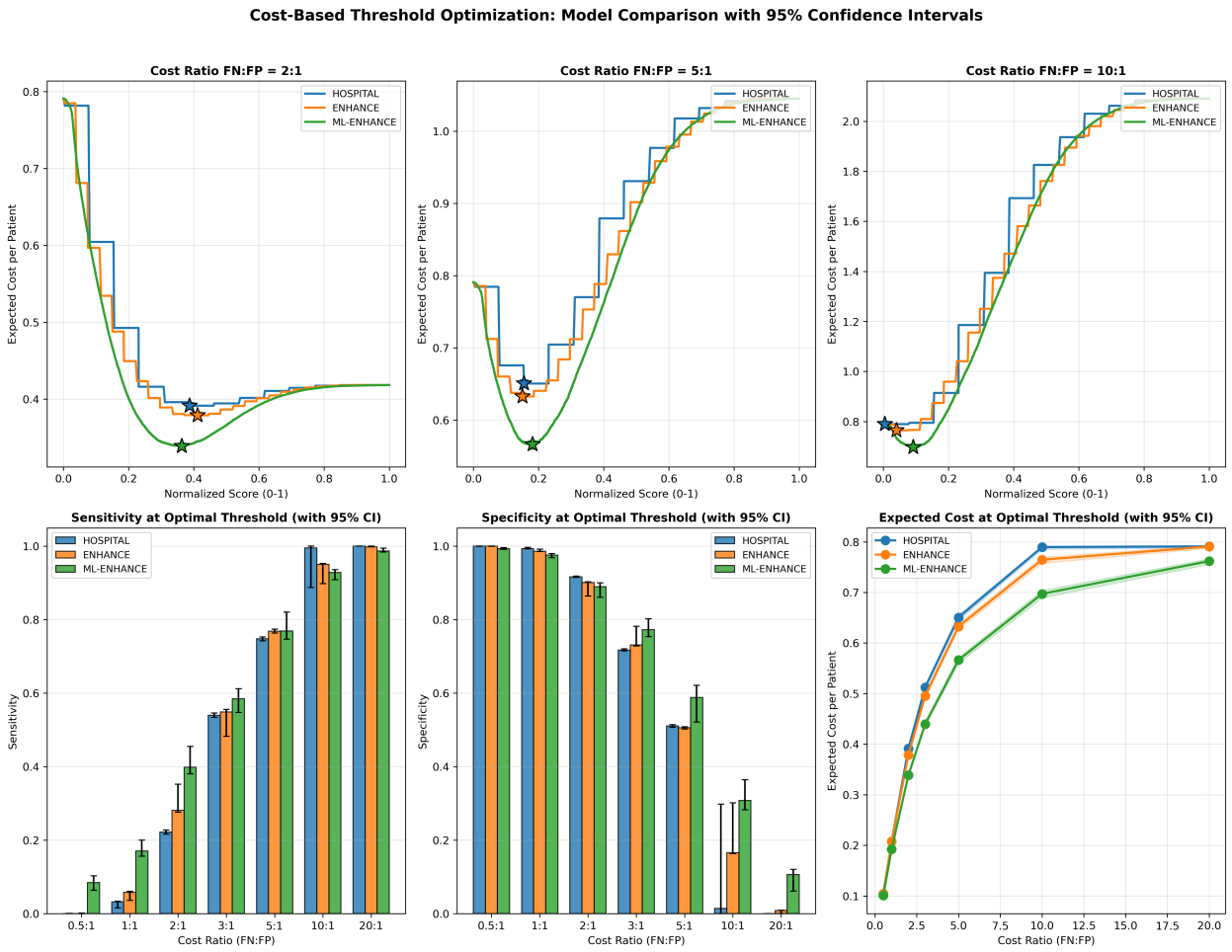


Figure 4: Cost-based threshold optimization comparing HOSPITAL, ENHANCE, and ML-ENHANCE with 95% confidence intervals. Panels A–C: Expected cost curves for cost ratios 2:1, 5:1, and 10:1; stars mark optimal operating points. Panel D: Sensitivity at cost-optimal thresholds. Panel E: Specificity values. Panel F: Expected cost per patient.

Table 4: Projected annual economic impact for a high-volume U.S. institution (50,000 discharges/year, 5:1 cost ratio). Expected costs incorporate both intervention expenditures and readmission costs under each model's optimal threshold.

Metric	**HOSPITAL**	**ENHANCE**	**ML-ENHANCE**	**Δ vs HOSP**
Readmissions identified (TP)	7,817	8,036	8,041	+219 / +224
Patients flagged unnecessarily (FP)	19,340	19,542	16,271	+202 / −3,069
Readmissions missed (FN)	2,633	2,414	2,409	−219 / −224
Expected cost per patient	\$10,612	\$10,318	\$9,226	−\$294 / −\$1,386
Annual institutional cost	\$530.6M	\$515.9M	\$461.3M	—
Annual savings vs HOSPITAL	—	**\$14.7M**	**\$69.3M**	—
Percent reduction	—	**2.8%**	**13.1%**	—

In Israeli healthcare using the day-denominated framework (25,000 discharges, 3:1 ratio), ML-ENHANCE generates 1,812 additional net bed-days annually (equivalent to 5.0 beds of continuous capacity) versus 418 for ENHANCE (1.1 beds), achieving 4.3-fold greater capacity impact (Supplementary Table S3). This advantage persisted across all evaluated cost ratios from 0.5:1 through 20:1, spanning a 40-fold range of implied intervention costs.

4 Discussion

This analysis of 113,312 patients demonstrates that machine learning substantially outperforms interpretable clinical scores for readmission prediction, and that this discrimination advantage translates directly to clinically meaningful benefits when evaluated through decision-theoretic threshold optimization. ML-ENHANCE achieved AUC 0.752, representing 0.076 improvement over HOSPITAL and 0.056 over ENHANCE. Risk stratification analysis revealed 7.5-fold discrimination across quartiles compared to 4.2-fold for ENHANCE and 2.8-fold for HOSPITAL, nearly tripling the discrimination range of the original score.

The key methodological contribution lies in demonstrating that traditional AUC comparisons understate practical value. At the policy-relevant 5:1 cost ratio, ML-ENHANCE's advantage manifests as $69.3 million in projected annual savings for high-volume U.S. institutions versus $14.7 million for ENHANCE, while in Israeli healthcare settings ML-ENHANCE generates 1,812 additional bed-days annually versus 418 for ENHANCE. This 4–5 fold differential persisted across all evaluated cost ratios from 0.5:1 through 20:1, indicating that findings reflect genuine discriminative superiority rather than artifacts of particular cost assumptions.

The interpretable ENHANCE score captures only 26.3% of the achievable improvement gap, leaving 73.7% accessible only through machine learning. While ENHANCE provides a transparent alternative for bedside applications, ML-ENHANCE is appropriate for automated population health platforms where maximum efficiency is paramount. Both models demonstrated excellent temporal stability across the seven-year study period (ENHANCE: +0.002; ML-ENHANCE: −0.011 AUC change), supporting robust generalizability despite evolving clinical practices.

The cost-based framework addresses a critical limitation in prediction model validation: the gap between discrimination metrics and actionable decision support. The substantial divergence between Youden's J and cost-optimal thresholds - 13.8 percentage points of sensitivity for ENHANCE - demonstrates that statistically optimal choices may substantially underperform clinically optimal choices when error costs are asymmetric. The framework's applicability across U.S. fee-for-service and Israeli capitation-based systems demonstrates generalizability, with the natural emergence of a 3:1 cost ratio from Israeli healthcare structure providing validation independent of currency-denominated assumptions.

Our baseline HOSPITAL performance (AUC 0.676) aligns with published reports (0.65–0.71), confirming implementation validity [4]. ML-ENHANCE's improvement falls at the upper end of gains reported for machine learning approaches (0.05–0.10). Previous studies demonstrating discrimination improvements have faced criticism for presenting abstract AUC gains without practical value; our cost-based framework addresses this by translating differences into concrete operating characteristics and economic projections, providing the evidence base for reconsidering the interpretability-performance trade-off.

The choice between models represents a deliberate trade-off. ENHANCE offers interpretable risk assessment with components automatically extractable from electronic health records and an additive structure enabling manual calculation in resource-limited settings. ML-ENHANCE is suited for automated applications

including EHR-integrated flagging, care management enrollment, and post-discharge resource allocation. Feature importance analysis reveals that prior admissions, laboratory abnormalities, and HOSPITAL components drive ML-ENHANCE predictions - consistent with clinical intuition, potentially facilitating clinician trust despite algorithmic complexity.

Several limitations warrant consideration. This single-center study may limit generalizability; external validation across diverse settings is essential, particularly for ML-ENHANCE where overfitting concerns are greater. The retrospective design limits assessment of prospective implementation challenges. Cost parameters may vary across institutions; while sensitivity analysis across a 40-fold range of ratios addresses relative cost uncertainty, absolute projections should be interpreted as illustrative. Future research should examine condition-specific versions and integration with real-time monitoring or social determinants screening.

In summary, ML-ENHANCE represents a substantial advancement in readmission prediction, demonstrating through cost-based threshold optimization that machine learning's discrimination improvements translate directly to economically significant benefits. At the 5:1 cost ratio, ML-ENHANCE achieves 13.1% cost reduction compared to 2.8% for ENHANCE - a 4–5 fold advantage. Healthcare systems should consider stratified deployment: ML-ENHANCE for automated risk stratification where maximum efficiency is paramount, and ENHANCE for clinical contexts requiring interpretable risk communication.

Data Availability

The datasets generated and analysed during this study are not publicly available due to patient privacy considerations and institutional data governance requirements. Aggregated summaries may be available upon request.

Code Availability

The underlying code for this study is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Acknowledgements

We thank Sheba Medical Center for its continuous support.

Author Contributions

B.G. conceived the study, developed the analytical framework, performed all analyses, and wrote the manuscript. Y.S. contributed to data interpretation and clinical validation. A.M.B. supervised the machine learning methodology and edited the manuscript. H.A. provided clinical oversight and data access. A.M.T. supervised the clinical aspects, provided data access, contributed to study design, and edited the manuscript. All authors read and approved the final manuscript.

Competing Interests

The authors declare no financial or non-financial competing interests.

Funding

This research was funded by the Israeli Science Foundation grant 1834/24.

Ethics Declaration

This study was conducted in accordance with the Declaration of Helsinki and Israeli Public Health Regulations for Medical Experiments in Humans (1980). The Sheba Medical Center Institutional Review Board approved the study protocol (approval number: 1107-24-SMC-D; approved July 9, 2024) with full waiver of informed consent due to the retrospective nature of the analysis and use of de-identified data.

Corresponding Author

Correspondence and requests for materials should be addressed to Barak Gahtan (barakgahtan@cs.technion.ac.il).

A Supplementary Tables and Figures

Table S1: Cost ratio scenarios with corresponding intervention costs and clinical examples. Cost ratios represent the relative cost of a missed readmission (C_{FN}) versus an unnecessary intervention (C_{FP}). U.S. costs assume readmission cost of $16,300; Israeli costs expressed in hospitalization days.

Cost Ratio	**U.S. Cost**	**Israeli (days)**	**Representative Program**	**Setting**
0.5:1	$32,600	6:1 days	Short-term SNF placement	Post-surgical
1:1	$16,300	3:3 days	Intensive home health	High-acuity HF
2:1	$8,150	3:1.5 days	Comprehensive transitional care	Multi-morbid
3:1	**$5,433**	**3:1 days**	**Care transitions intervention**	**Israeli default**
5:1	$3,260	5:1 days	Structured nurse follow-up	Standard post-DC
10:1	$1,630	3:0.3 days	Pharmacist medication review	Polypharmacy
20:1	$815	3:0.15 days	Automated calls + education	Population-level

Table S2: Optimal thresholds and operating characteristics by cost ratio with 95% confidence intervals. Thresholds minimize expected cost per patient. ML-ENHANCE thresholds represent predicted probabilities (0–1 scale). ΔCost calculated relative to HOSPITAL.

Ratio	**Model**	**Threshold [95% CI]**	**Sensitivity [95% CI]**	**Specificity [95% CI]**	**ΔCost**
	HOSPITAL	5.0 [6.0, 6.0]	22.2% [21.7, 22.7]	91.7% [91.5, 91.9]	—
2:1	ENHANCE	11.1 [11.0, 12.0]	28.1% [27.6, 35.3]	90.1% [86.4, 90.3]	−3.2%
	ML-ENHANCE	0.34 [0.31, 0.35]	39.9% [38.0, 45.5]	89.0% [86.1, 90.0]	−13.4%
	HOSPITAL	3.0 [4.0, 4.0]	54.0% [53.4, 54.6]	71.7% [71.5, 72.1]	—
3:1	ENHANCE	7.1 [8.0, 9.0]	54.9% [48.3, 55.6]	73.1% [72.9, 78.2]	−3.3%
	ML-ENHANCE	0.26 [0.25, 0.27]	58.5% [54.8, 61.2]	77.3% [75.4, 80.2]	−14.1%
	HOSPITAL	2.0 [3.0, 3.0]	74.8% [74.3, 75.3]	51.1% [50.8, 51.4]	—
5:1	ENHANCE	4.1 [5.0, 5.0]	76.9% [76.4, 77.4]	50.6% [50.2, 50.9]	−2.8%
	ML-ENHANCE	0.18 [0.16, 0.19]	76.9% [74.7, 82.1]	58.9% [52.1, 62.1]	−12.9%
	HOSPITAL	0.1 [0.0, 2.0]	99.5% [88.8, 100]	1.4% [0.0, 29.8]	—
10:1	ENHANCE	1.1 [2.0, 3.0]	95.0% [89.8, 95.3]	16.5% [16.3, 30.2]	−3.2%
	ML-ENHANCE	0.11 [0.10, 0.12]	92.8% [90.9, 93.7]	30.8% [28.3, 36.4]	−11.7%

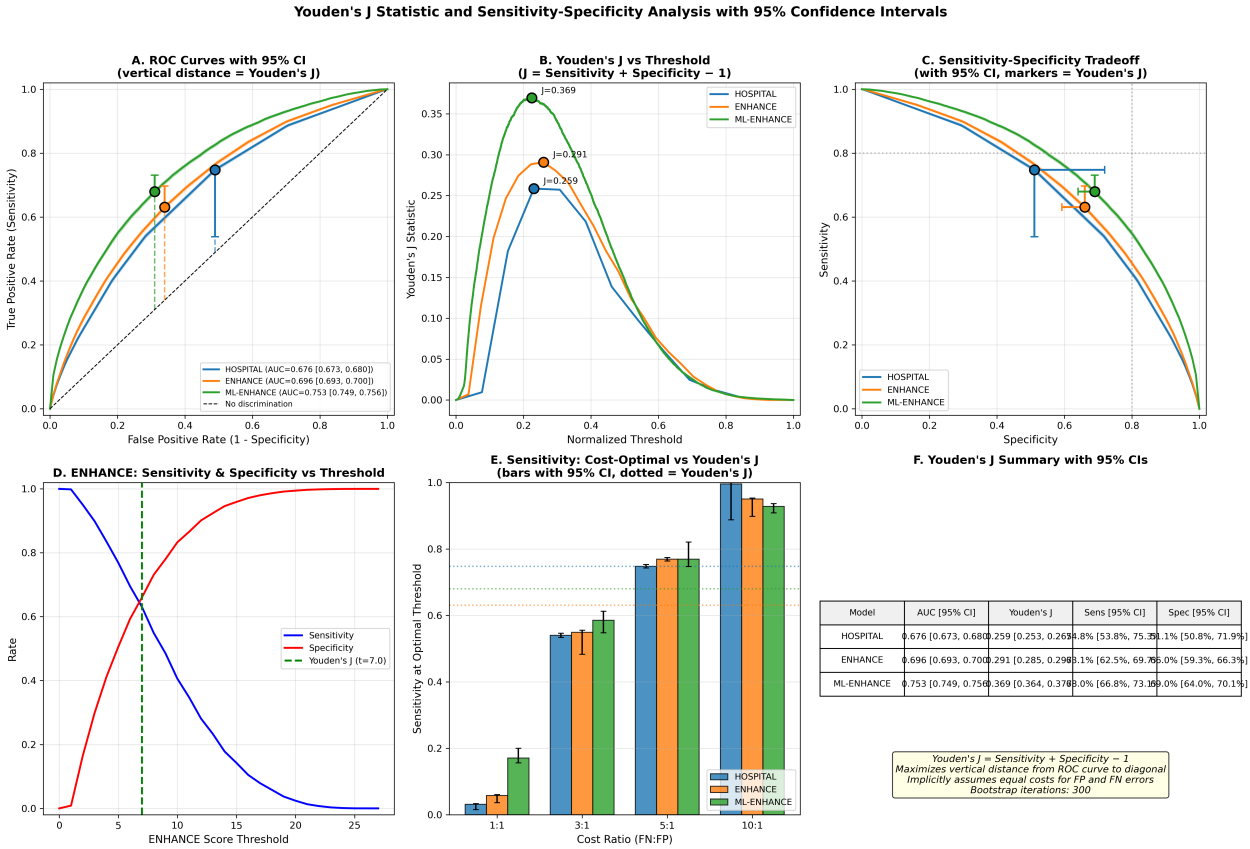


Figure S1: Youden's J statistic and sensitivity-specificity analysis with 95% confidence intervals. (A) ROC curves with shaded 95% CI bands; Youden's J optimal points marked with error bars. (B) Youden's J as a function of normalized threshold, showing peak values: ML-ENHANCE (J=0.369), ENHANCE (J=0.291), HOSPITAL (J=0.259). (C) Sensitivity-specificity tradeoff with 95% CI bands. (D) Sensitivity and specificity vs. ENHANCE threshold. (E) Sensitivity at cost-optimal thresholds vs. Youden's J (dotted lines). (F) Summary table with 95% CIs.

Table S3: Projected annual capacity impact for a large Israeli medical center (25,000 discharges/year, 3:1 day-based cost ratio). Costs expressed in hospitalization days.

Metric	**HOSPITAL**	**ENHANCE**	**ML-ENHANCE**	**Δ vs HOSP**
Readmissions identified (TP)	2,822	2,869	3,057	+47 / +235
Unnecessary interventions (FP)	5,596	5,319	4,489	−277 / −1,107
Readmissions missed (FN)	2,403	2,356	2,168	−47 / −235
Days saved (TP × 3)	8,466	8,607	9,171	+141 / +705
Days spent (FP × 1)	5,596	5,319	4,489	−277 / −1,107
Net days saved	**2,870**	**3,288**	**4,682**	—
Additional vs HOSPITAL	—	**+418 days**	**+1,812 days**	—
Equivalent beds (continuous)	—	**+1.1 beds**	**+5.0 beds**	—

B Detailed Background and Rationale

Hospital readmissions represent one of the most significant quality and cost challenges in modern healthcare systems. In the United States alone, 15–20% of inpatients are readmitted within a month, generating more than $40 billion in costs that are considered potentially preventable each year [1, 2]. The Centers for Medicare & Medicaid Services links hospital reimbursement to excess readmissions through the Hospital Readmissions Reduction Program, further intensifying the demand for accurate prediction tools [3].

The HOSPITAL score, derived from seven routinely collected variables, represents the most widely validated clinical instrument for 30-day readmission prediction [4]. Across multiple cohorts, it achieves an area under the receiver operating characteristic curve (AUC) of approximately 0.65–0.70, demonstrating moderate discrimination. However, several limitations constrain its clinical impact. The score's modest performance leaves substantial residual risk unexplained, with readmission rates ranging only 2-fold across its three risk categories. This limited granularity provides insufficient precision for identifying patients at the extremes of risk where resource-intensive interventions would be most cost-effective. Additionally, the original HOSPITAL score was developed over a decade ago, before the widespread adoption of comprehensive electronic health records, and therefore overlooks readily available clinical information that could enhance prediction accuracy.

Contemporary clinical data systems routinely capture detailed information that was not systematically available during the HOSPITAL score's development, including comprehensive laboratory panels with optimized reference ranges, granular comorbidity coding enabling interaction analysis, temporal admission patterns, and emergency department care pathways. These data sources represent untapped opportunities for improving readmission prediction while maintaining the clinical interpretability essential for bedside decision-making.

Machine learning (ML) approaches have demonstrated the potential for superior discrimination by leveraging hundreds of variables, with ensemble models achieving 0.05–0.10 AUC improvements over traditional scores in recent evaluations [5, 6]. However, these gains come at the expense of clinical interpretability, creating "black box" models that resist bedside implementation and clinical trust. Furthermore, the practical benefits of marginal AUC improvements remain unclear without demonstrating substantial improvements in risk stratification capabilities.

A pragmatic solution involves developing complementary models that balance clinical utility with predictive performance. An enhanced clinical score should extract additional predictive signal from routine data while preserving interpretability and implementation feasibility. Simultaneously, a high-performing ML model using the same feature set can establish the theoretical performance ceiling, quantifying the maximum achievable improvement and providing context for the clinical score's performance.

C Detailed Methodology

Reporting follows the TRIPOD statement for transparent reporting of prognostic model studies [8]. The study population included all adult patients ($\geq$18 years) with inpatient admissions to internal medicine departments between January 1, 2017 and December 31, 2023. Exclusion criteria, applied sequentially, were: (1) non-internal medicine admissions, (2) missing discharge dates, (3) length of stay $\leq$1 day or >365 days, and (4) missing essential demographic data.

Temporal validation employed a chronological 70/30 split to ensure robust evaluation across different time

periods and evolving clinical practices. The development period (2017–2020) captured early healthcare system practices and included the transition period around COVID-19, while the validation period (2021–2023) represented more recent clinical practices and post-pandemic care patterns. This temporal split strategy provides stronger evidence of model generalizability compared to random splitting, as it tests performance across genuine temporal changes in patient populations and clinical practices.

Data extraction captured demographics, admission characteristics, comorbidities (using ICD-10-CM codes with established mapping algorithms [9]), laboratory values (most recent before discharge), vital signs, medications, procedures, and readmission outcomes. All data elements were routinely collected during standard clinical care.

The HOSPITAL score was calculated using the original seven components: hemoglobin at discharge <12 g/dL (1 point), oncology service discharge adapted as history of malignancy (1 point), sodium level <135 mEq/L (1 point), procedure during stay (1 point), emergency department admission (1 point), prior year admissions 0–1/2–5/>5 (0/2/5 points), admission length ≥5 days (2 points), and low albumin <3.5 g/dL (1 point). Risk categories were defined as Low (0–4 points), Intermediate (5–6 points), and High (≥7 points).

Missing data patterns were systematically evaluated across all variables. Laboratory values were imputed using clinically normal values when tests were not performed: hemoglobin 13 g/dL, sodium 138 mEq/L, albumin 4.0 g/dL, assuming clinical stability. Emergency department admission indicators were imputed as False when missing, and procedure counts were imputed as zero. Physiologically implausible values were excluded: length of stay >365 days or ≤0.1 days, representing data entry errors or non-meaningful admissions.

Feature engineering followed a systematic five-stage process: (1) domain-specific analysis across nine analytical domains, (2) evidence-based selection using quantifiable evidence across domains, (3) threshold optimization using Youden's J statistic, (4) point assignment using L1 regularization, and (5) integration and validation with temporal holdout testing. The 16 components identified through domain analysis were systematically evaluated, with final point assignments optimized through sparse regularization to maximize discriminative performance while maintaining clinical interpretability.

Continuous variables were described using means and standard deviations or medians and interquartile ranges based on distribution normality assessed by the Shapiro-Wilk test. Categorical variables were presented as frequencies and percentages. Between-group comparisons used Student's t-test or Mann-Whitney U test for continuous variables and chi-square test or Fisher's exact test for categorical variables. Model calibration was evaluated using calibration plots with 10 equally-sized groups. Risk stratification was assessed by comparing readmission rates across score-defined risk categories. Bootstrap validation with 1000 samples was used to assess stability of feature selection and point assignments.

D Supplementary Analysis by Research Domain

This appendix provides detailed results from each of the nine analytical domains that informed the development of the ENHANCE score. For each domain, we present detailed methodologies, key findings, and their direct contributions to the final scoring system.

D.1 Domain 1: Comorbidity Analysis

Research Question How does the presence of multiple comorbidities affect readmission risk compared to single conditions? Is there a specific combination that significantly increases risk?

Methodology We analyzed nine major comorbidity categories: chronic kidney disease (CKD), malignancy, congestive heart failure (CHF), hypertension (HTN), diabetes mellitus (DM), chronic obstructive pulmonary disease (COPD), cerebrovascular accident (CVA), ischemic heart disease (IHD), and dyslipidemia. For each condition, we calculated prevalence, readmission rates, and odds ratios with 95% confidence intervals. We analyzed all possible pairs and triplets of the most common conditions, requiring a minimum of 20 patients per combination to ensure statistical reliability.

Individual Comorbidity Effects: Our analysis of individual comorbidities revealed substantial variation in their association with readmission risk. As shown in Table S4, CKD demonstrated the strongest association with readmission (OR 1.68, 95% CI 1.60–1.76), affecting 8.2% of patients with a readmission rate of 29.8% compared to 20.1% in those without CKD. History of malignancy showed similarly strong association (OR 1.66, 95% CI 1.58–1.74), while CHF ranked third (OR 1.57, 95% CI 1.49–1.66). All major comorbidities except dyslipidemia showed statistically significant associations with readmission ($p < 0.001$). Notably, dyslipidemia showed minimal association with readmission risk (OR 1.02, 95% CI 0.99–1.05, $p = 0.199$), despite its high prevalence of 38.6%.

Table S4: Individual Comorbidity Effects on Readmission Risk

Condition	Prevalence (%)	With Condition	Without Condition	Rate Difference	Odds Ratio	95% CI
CKD	8.2	0.298	0.201	+0.096	1.68	1.60–1.76
MALIGNANCY	7.7	0.296	0.202	+0.094	1.66	1.58–1.74
CHF	5.9	0.287	0.204	+0.083	1.57	1.49–1.66
DM	29.3	0.243	0.195	+0.048	1.32	1.28–1.37
COPD	6.2	0.254	0.206	+0.048	1.31	1.24–1.38
HTN	51.0	0.227	0.190	+0.037	1.25	1.22–1.29
CVA	2.5	0.242	0.208	+0.034	1.22	1.12–1.33
IHD	17.8	0.236	0.203	+0.033	1.21	1.17–1.25
DYSLIP	38.6	0.211	0.208	+0.003	1.02	0.99–1.05

Comorbidity Burden: The analysis of comorbidity burden revealed a strong dose-response relationship, as detailed in Table S5. Readmission rates increased monotonically from 17.00% in patients with no comorbidities to 50.00% in the two patients with eight comorbidities, representing a 2.9-fold increase. The Spearman correlation coefficient of 0.075 ($p < 0.001$) confirmed this significant positive trend. The majority of patients (28.5%) had no comorbidities, while 21.1% had one condition and 21.7% had two conditions.

Table S5: Readmission Rates by Comorbidity Burden

Number of Comorbidities	Total Patients	Readmission Rate	Rate (%)
0	32,314	0.170	17.00
1	23,893	0.207	20.72
2	24,578	0.212	21.19
3	19,051	0.227	22.72
4	9,227	0.257	25.66
5	3,293	0.310	30.97
6	825	0.333	33.33
7	129	0.388	38.76
8	2	0.500	50.00

High-Risk Combinations: Analysis of comorbidity combinations revealed significant synergistic effects beyond simple additive risk. Table S6 presents the 56 combinations identified in our cohort. The combination of CKD and CVA demonstrated the highest relative risk at 1.75, affecting 396 patients with a readmission rate of 36.6%, substantially higher than would be expected from either condition alone. The CHF and COPD combination ranked second with relative risk 1.69 and readmission rate of 35.3% in 858 patients. Among triplet combinations, the presence of HTN, CKD, and malignancy together resulted in a 32.7% readmission rate (RR 1.57), affecting 684 patients.

Weighted Comorbidity Index Development: To improve upon simple comorbidity counting, we developed a weighted scoring system based on the natural logarithm of odds ratios. Each condition received a weight proportional to ln(OR), with CKD receiving the highest weight (0.520), malignancy (0.506), and CHF (0.453), while dyslipidemia received the lowest weight (0.019). The weighted comorbidity score demonstrated superior discrimination compared to simple counting (AUC 0.568 vs. 0.552, improvement of 0.016).

Figure S7 provides a comprehensive visualization of these findings, including forest plots of individual comorbidity effects, the distribution of comorbidity burden across the population, and ROC curves comparing simple comorbidity counts versus weighted scoring approaches. The weighted comorbidity score achieved an AUC of 0.568 compared to 0.552 for simple counting, supporting the value of condition-specific weighting.

Contribution to ENHANCE Score Based on these findings and sparse logistic regression with L1 regularization (C=1.5), we incorporated the following comorbidity components into the ENHANCE score: chronic kidney disease received 0 points (despite its significant OR, its effect was captured through the CKD+CVA interaction), history of malignancy received 1 point (based on OR 1.66), congestive heart failure received 0 points despite its significant OR, and the CKD + CVA interaction received 2 points based on its strong synergistic effect (RR 1.75).

D.2 Domain 2: Laboratory Values Analysis

Research Question Which laboratory abnormalities are most strongly associated with readmission when analyzed individually rather than as part of the HOSPITAL score?

Methodology We analyzed 27 laboratory parameters using receiver operating characteristic (ROC) curve analysis and Youden's J statistic to identify optimal thresholds. For each parameter, we calculated AUC, mean difference between readmitted and non-readmitted patients, Cohen's d effect size, and p-values from Mann-Whitney U and Kolmogorov-Smirnov tests.

Individual Laboratory Predictors: Table S7 presents the laboratory parameters ranked by their individual predictive power for readmission. Hemoglobin and albumin emerged as the strongest predictors with identical AUC values of 0.637. The analysis revealed distinct patterns: lower values of hemoglobin (mean difference -1.06 g/dL), albumin (-0.27 g/dL), and calcium (-0.20 mmol/L) were associated with increased readmission risk, while elevated urea (+12.79 mg/dL), GGT (+36.79 U/L), and ALP (+29.19 U/L) indicated higher risk. All major associations were highly statistically significant ($p < 0.001$), with only platelet count showing no significant association ($p = 0.723$).

Table S6: High-Risk Comorbidity Combinations (sorted by readmission rate)

Combination	Type	Count	Readmission Rate	Rate vs Baseline	Relative Risk
CKD + CVA	Pair	396	0.366	+0.157	1.75
CHF + COPD	Pair	858	0.353	+0.144	1.69
MALIGNANCY + CVA	Pair	226	0.341	+0.132	1.63
COPD + CKD	Pair	742	0.333	+0.124	1.59
HTN + CKD + MALIGNANCY	Triplet	684	0.327	+0.118	1.57
DM + IHD + CKD	Triplet	2,052	0.323	+0.114	1.55
DYSLIP + CKD + MALIGNANCY	Triplet	421	0.323	+0.114	1.55
IHD + CKD + MALIGNANCY	Triplet	285	0.319	+0.110	1.53
CHF + MALIGNANCY	Pair	540	0.317	+0.108	1.51
HTN + IHD + CKD	Triplet	2,984	0.315	+0.106	1.51
CHF + CKD	Pair	1,901	0.315	+0.105	1.50
CKD + MALIGNANCY	Pair	852	0.315	+0.105	1.50
DYSLIP + IHD + CKD	Triplet	2,386	0.314	+0.105	1.50
DYSLIP + DM + CKD	Triplet	3,004	0.313	+0.104	1.50
CKD + IHD	Pair	3,515	0.312	+0.103	1.49
HTN + DM + CKD	Triplet	4,163	0.311	+0.102	1.49
HTN + DYSLIP + CKD	Triplet	4,429	0.308	+0.098	1.47
CKD + DM	Pair	4,815	0.307	+0.098	1.47
DM + IHD + MALIGNANCY	Triplet	573	0.305	+0.096	1.46
CKD + HTN	Pair	7,694	0.304	+0.095	1.45
CHF + DM	Pair	3,377	0.303	+0.094	1.45
DM + MALIGNANCY	Pair	2,367	0.302	+0.093	1.44
DM + CKD + MALIGNANCY	Triplet	372	0.301	+0.092	1.44
CKD + DYSLIP	Pair	5,021	0.301	+0.092	1.44
CHF + IHD	Pair	2,874	0.300	+0.091	1.43
HTN + DM + MALIGNANCY	Triplet	1,790	0.298	+0.089	1.42
DYSLIP + DM + MALIGNANCY	Triplet	1,223	0.296	+0.087	1.42
DYSLIP + IHD + MALIGNANCY	Triplet	809	0.295	+0.086	1.41
CHF + HTN	Pair	5,141	0.295	+0.086	1.41
HTN + IHD + MALIGNANCY	Triplet	1,014	0.292	+0.083	1.40
CHF + DYSLIP	Pair	3,739	0.292	+0.082	1.39
MALIGNANCY + HTN	Pair	4,715	0.291	+0.081	1.39
COPD + IHD	Pair	1,838	0.288	+0.079	1.38
COPD + DM	Pair	2,499	0.286	+0.077	1.37
CHF + CVA	Pair	319	0.285	+0.076	1.36
MALIGNANCY + IHD	Pair	1,333	0.285	+0.076	1.36
HTN + DYSLIP + MALIGNANCY	Triplet	2,334	0.285	+0.076	1.36
COPD + CVA	Pair	204	0.284	+0.075	1.36
MALIGNANCY + DYSLIP	Pair	3,087	0.284	+0.075	1.36
IHD + CVA	Pair	851	0.275	+0.066	1.32
COPD + HTN	Pair	4,332	0.269	+0.060	1.29
DM + CVA	Pair	1,306	0.262	+0.053	1.25
COPD + DYSLIP	Pair	3,411	0.260	+0.051	1.25
HTN + DM + IHD	Triplet	7,662	0.258	+0.049	1.23
COPD + MALIGNANCY	Pair	382	0.257	+0.047	1.23
DM + IHD	Pair	9,618	0.255	+0.046	1.22
DYSLIP + DM + IHD	Triplet	6,741	0.254	+0.045	1.21
HTN + CVA	Pair	2,148	0.253	+0.044	1.21
DM + HTN	Pair	24,714	0.247	+0.038	1.18
HTN + DYSLIP + DM	Triplet	15,041	0.246	+0.037	1.18
HTN + IHD	Pair	15,021	0.243	+0.034	1.16
DM + DYSLIP	Pair	18,977	0.238	+0.029	1.14
HTN + DYSLIP + IHD	Triplet	10,398	0.235	+0.026	1.12
DYSLIP + CVA	Pair	1,639	0.230	+0.021	1.10
IHD + DYSLIP	Pair	13,132	0.228	+0.019	1.09
HTN + DYSLIP	Pair	31,579	0.221	+0.012	1.06

Table S7: Laboratory Values as Readmission Predictors (sorted by AUC)

Laboratory	**AUC**	**Mean Difference**	**Cohen's d**	**P-Value**	**Risk Direction**
HB	0.637	-1.06	-0.452	<0.001	lower
ALB	0.637	-0.27	-0.488	<0.001	lower
CA	0.589	-0.20	-0.302	<0.001	lower
UREA	0.586	+12.79	0.327	<0.001	higher
GGT	0.585	+36.79	0.228	<0.001	higher
ALP	0.583	+29.19	0.265	<0.001	higher
LDH	0.571	+43.88	0.210	<0.001	higher
PULSE	0.560	+2.97	0.214	<0.001	higher
CR	0.549	+0.22	0.188	<0.001	higher
CL	0.543	-0.66	-0.132	<0.001	lower
WEIGHT	0.537	-2.32	-0.120	<0.001	lower
BILDIR	0.537	+0.13	0.133	<0.001	higher
DBP	0.536	-1.23	-0.099	<0.001	lower
NA	0.533	-0.40	-0.098	<0.001	lower
SBP	0.527	-1.89	-0.088	<0.001	lower
GLU	0.527	+5.08	0.095	<0.001	higher
WBC	0.522	+0.43	0.062	<0.001	higher
O2SAT	0.521	+0.20	0.007	<0.001	lower
K	0.511	-0.01	-0.024	<0.001	lower
HEIGHT	0.510	-0.42	-0.033	<0.001	lower
RR	0.510	+0.10	0.012	0.003	higher
AST	0.510	+2.40	0.036	<0.001	higher
BILTOT	0.509	+0.10	0.103	<0.001	higher
ALT	0.508	+1.24	0.016	<0.001	lower
TEMP	0.507	+0.01	0.012	0.001	higher
P	0.506	+0.04	0.052	0.007	higher
PLT	0.501	+2.20	0.020	0.723	lower

HOSPITAL Score Threshold Evaluation: We compared the performance of HOSPITAL score thresholds against optimal thresholds derived from our data. For hemoglobin, the HOSPITAL threshold (below 12 g/dL) identified 60,259 patients (53.2%) as abnormal, with readmission rates of 27.2% in abnormal versus 13.8% in normal patients (rate difference 13.5%). However, the optimal threshold identified through Youden's J statistic showed inferior performance (improvement = -0.926). Similarly, for sodium, the HOSPITAL threshold (below 135 mmol/L) affected 11,375 patients (10.0%) with readmission rates of 28.7% versus 20.0% (rate difference 8.6%), while the optimal threshold showed improvement = -0.158.

Laboratory Abnormality Burden: Analysis of laboratory abnormality burden revealed a strong positive correlation with readmission risk (Spearman correlation 0.162, $p < 0.001$). Readmission rates increased from 13.94% in patients with 2 laboratory abnormalities to 34.20% in those with 6 abnormalities, representing a 2.5-fold increase across the abnormality burden spectrum.

Enhanced Laboratory Score Development: Using logistic regression with the top eight laboratory predictors (HB, ALB, CA, UREA, GGT, ALP, LDH, PULSE), we developed an enhanced laboratory score from 107,727 complete observations, achieving an AUC of 0.678. The model coefficients revealed albumin (-0.315) and hemoglobin (-0.227) as the strongest negative predictors, while urea (+0.176) showed the strongest positive association with readmission risk. Pulse rate (+0.129), LDH (+0.099), ALP (+0.054), GGT (+0.053), and calcium (+0.052) also contributed significantly to the model.

Figure S8 illustrates these findings, including the distribution of laboratory values by outcome, the performance of HOSPITAL thresholds versus optimized cutoffs, and the enhanced laboratory score distribution.

Contribution to ENHANCE Score Based on the sparse logistic regression analysis, the following laboratory components were incorporated into the ENHANCE score: severe hypoalbuminemia (<3.0 g/dL) received 2 points, moderate anemia (<10.0 g/dL) received 2 points, and kidney dysfunction marker (urea >60 mg/dL) received 2 points. These optimized thresholds provide superior discrimination compared to the conventional HOSPITAL thresholds.

D.3 Domain 3: Emergency Department Admission Patterns

Research Question Do patients who came through the emergency department have different readmission patterns or lengths of stay compared to elective admissions?

Methodology We compared 109,252 ED admissions (96.4%) with 4,060 elective admissions (3.6%) on baseline characteristics, readmission rates, and length of stay patterns. We analyzed interactions between ED admission and comorbidities using stratified analysis to identify synergistic effects.

ED vs. Elective Admission Characteristics: Table S8 summarizes the key differences between ED and elective admissions. Despite the overwhelming predominance of ED admissions in our cohort, we found minimal difference in readmission rates between ED (20.9%) and elective (20.6%) admissions (p=0.654, OR=1.02). However, ED patients demonstrated significantly different characteristics including older age (70.11 vs. 65.88 years, p<0.001), lower proportion of males (52.44% vs. 57.91%, p<0.001), higher comorbidity burden (1.67 vs. 1.60, p<0.001), and substantially higher HOSPITAL scores (3.12 vs. 1.95, p<0.001). ED patients also had longer median length of stay (1.71 vs. 1.14 days, p<0.001).

Table S8: Comparison of ED vs. Elective Admission Characteristics

Characteristic	ED	Elective	P-Value
Age (mean years)	70.11	65.88	<0.001
Male (%)	52.44	57.91	<0.001
Comorbidities (mean)	1.67	1.60	<0.001
HOSPITAL Score (mean)	3.12	1.95	<0.001
Length of Stay (median days)	1.71	1.14	<0.001

Length of Stay Distribution: Analysis of length of stay patterns revealed significant differences between admission types (Mann-Whitney U test $p<0.001$). ED patients demonstrated longer stays across most categories, with 31.3% staying $\leq$1 day compared to 36.4% of elective patients. Conversely, ED patients had higher proportions in longer stay categories: 19.0% vs. 10.6% for 4–7 days and 7.4% vs. 4.8% for 8–30 days. The >30 days category showed similar proportions (0.6% each).

ED × Comorbidity Interactions: Analysis revealed significant interactions between ED admission and specific comorbidities, as shown in Table S9. The CHF × ED interaction showed the strongest effect, with readmission rates of 29.2% for ED patients with CHF versus 19.7% for elective patients with CHF, representing an additional 9.8 percentage point increase beyond the sum of individual effects. Similar but smaller interaction effects were observed for CVA (+5.7 percentage points) and COPD (+5.7 percentage points). Notably, the malignancy interaction showed a negative effect (-8.3 percentage points), with elective patients having higher readmission rates (35.7%) than ED patients (29.2%) when malignancy was present.

Table S9: ED × Comorbidity Interaction Effects

Comorbidity	ED with	ED without	Elective with	Elective without	Interaction Effect	N ED with
CHF	0.292	0.204	0.197	0.207	+0.098	6,412
CVA	0.244	0.208	0.185	0.207	+0.057	2,781
COPD	0.255	0.206	0.199	0.206	+0.057	6,871
HTN	0.228	0.189	0.207	0.205	+0.036	55,942
DM	0.244	0.195	0.220	0.201	+0.030	32,043
IHD	0.237	0.203	0.213	0.205	+0.025	19,483
DYSLIP	0.212	0.208	0.196	0.211	+0.019	42,303
CKD	0.298	0.201	0.286	0.197	+0.008	8,933
MALIGNANCY	0.292	0.203	0.357	0.185	-0.083	8,182

ED-Specific Model Development: To assess whether ED patients require different predictive approaches, we developed separate models for ED (n=109,252) and elective (n=4,060) patients using six key features: age, HOSPITAL score, total comorbidities, hemoglobin, sodium, and albumin. The elective-specific model achieved superior performance (AUC 0.715) compared to the ED-specific model (AUC 0.686). Feature importance analysis revealed that comorbidity burden had greater predictive value in ED patients (coefficient 0.047) compared to elective patients (-0.011), while hemoglobin showed stronger association in elective patients (coefficient -0.096 vs. -0.057).

Figure S9 provides visual representation of these findings, including readmission rates by admission type, length of stay distributions, age distributions, and the stacked bar chart showing the proportion of patients in each length of stay category by admission type.

Contribution to ENHANCE Score The sparse logistic regression assigned 0 points to emergency department admission alone, reflecting the minimal difference in overall readmission rates (20.9% vs. 20.6%, p=0.654). However, the CHF + ED interaction received 1 point based on the significant synergistic effect of +9.8 percentage points in readmission risk, representing the strongest interaction identified in our analysis.

D.4 Domain 4: Temporal Patterns in Readmissions

Research Question Are there seasonal or time-based patterns in readmissions that could enhance prediction models?

Methodology We analyzed all the admissions spanning in our dataset. We examined seasonal, monthly, and day-of-week variations in readmission rates, as well as holiday period effects and time-to-readmission distributions. Statistical significance was assessed using ANOVA for seasonal patterns, chi-square tests for categorical temporal variables, and Mann-Whitney U tests for continuous temporal measures.

Seasonal and Monthly Patterns: Seasonal variation in readmission rates was minimal and not statistically significant (F=1.098, p=0.348). Table S10 shows readmission rates ranging from 20.56% (Winter) to 21.15% (Fall), representing only a 0.59 percentage point seasonal variation. Monthly analysis revealed greater variation, with rates ranging from 20.44% (December) to 21.60% (September), representing a 1.16 percentage point monthly range. Despite this variation, the overall seasonal pattern remained non-significant.

Table S10: Seasonal Readmission Patterns

Season	Total Admissions	Readmissions	Rate (%)
Winter	28,151	5,789	20.56
Spring	27,500	5,785	21.04
Summer	29,258	6,111	20.89
Fall	28,403	6,007	21.15

Day of Week Effects: Table S11 presents the weekend effect analysis. Weekend admissions showed a statistically significant but modest increase in readmission risk (21.34% vs. 20.76%, p=0.040, Chi-square=4.232). The weekend effect of +0.58 percentage points, while small, was consistent across the study period, affecting 28,801 weekend admissions out of 113,312 total admissions (25.4%). Individual day analysis revealed Sunday as having the highest readmission rate (21.54%), while Tuesday showed the lowest (20.45%).

Table S11: Weekend vs. Weekday Readmission Analysis

Day Type	Total Admissions	Readmissions	Rate (%)	P-Value
Weekday	84,511	17,547	20.76	0.040
Weekend	28,801	6,145	21.34	

Holiday Period Effects: Analysis of holiday period effects revealed a statistically significant protective effect, with readmission rates of 20.57% during holiday periods compared to 21.09% during regular periods

(Chi-square=4.044, p=0.044). This -0.52 percentage point difference affected 39,254 holiday period admissions compared to 74,058 regular period admissions. The holiday effect suggests reduced readmission risk, possibly due to changes in discharge planning practices or patient behavior during holiday periods.

Time to Readmission Distribution: Analysis of 13,611 readmissions within 30 days revealed a median time to readmission of 13.0 days. The distribution showed a bimodal pattern with peaks in the first two weeks: 27.9% occurred within the first week (3,798 readmissions), 29.0% in week 2 (3,953 readmissions), 22.1% in week 3 (3,013 readmissions), and 20.7% in week 4 (2,822 readmissions). This pattern indicates that the highest risk period for readmission occurs within the first two weeks after discharge, with the second week showing the peak occurrence.

Temporal Interaction Effects: We examined interactions between temporal patterns and patient characteristics. Season × age group interactions showed minimal variation, with a maximum range of 0.067 across different combinations. Weekend × comorbidity interactions revealed that certain conditions may be more sensitive to weekend effects, with CHF patients showing readmission rates of 29.8% on weekends versus 28.4% on weekdays, though these differences were modest.

Figure S10 illustrates these temporal patterns, including seasonal variations, monthly trends, day of week effects, and the distribution of time to readmission across the 30-day follow-up period.

Contribution to ENHANCE Score Based on the statistically significant weekend effect (+0.58 percentage points, p=0.040), the sparse logistic regression assigned 1 point for weekend admission. While the effect size was modest, weekend admission represents an easily identifiable risk factor that can be automatically captured from admission data. The holiday period effect, despite being statistically significant, was not incorporated due to its protective nature and smaller clinical impact. Seasonal effects were excluded due to lack of statistical significance (p=0.348).

D.5 Domain 5: ECG Findings and Readmission Risk

Research Question Do specific ECG abnormalities correlate with increased readmission risk?

Methodology We analyzed eight ECG parameters available for 16,847 patients (14.9% of the cohort). The analysis included seven numeric parameters (ventricular rate, corrected QT interval, T axis, QRS duration, PR interval, R axis, P axis) and one categorical parameter (ECG interpretation). For numeric parameters, we used Mann-Whitney U tests and calculated effect sizes and AUC values. The categorical interpretation parameter was analyzed using chi-square tests with CramÃ©r's V effect size.

Data Availability Limitations ECG data availability was substantially limited across the cohort, with parameter availability ranging from 10.9% to 14.9% of patients (average 13.9%). The most complete parameters (corrected QT interval, ventricular rate, QRS duration, R axis, T axis, and ECG interpretation) were available for 16,847 patients (14.9%), while PR interval was available for 12,791 patients (11.3%) and P axis for 12,391 patients (10.9%). This limited availability substantially restricts the utility of ECG parameters as general predictors in our cohort.

Key Findings Among the analyzed parameters, ventricular rate showed the strongest association with readmission risk (AUC 0.575, Cohen's d=0.239, p<0.001), with readmitted patients having a mean ventricular rate 5.35 bpm higher than non-readmitted patients (83.5 vs. 78.2 bpm). Table S12 presents the ECG parameters ranked by their predictive power. Five of eight parameters showed statistically significant associations with readmission: corrected QT interval (AUC 0.552, +8.88 ms longer), T axis deviation (AUC 0.536, +8.85Â° deviation), QRS duration (AUC 0.520, +2.31 ms longer), and PR interval (AUC 0.513, -1.09 ms shorter). The categorical ECG interpretation parameter showed no significant association (p=0.169).

Table S12: ECG Parameters and Readmission Risk

Parameter	Type	P-Value	Effect Size	AUC	Mean Difference
Ventricular Rate (bpm)	Numeric	<0.001	d=0.239	0.575	+5.35
Corrected QT Interval (ms)	Numeric	<0.001	d=0.141	0.552	+8.88
T Axis (degrees)	Numeric	<0.001	d=0.135	0.536	+8.85
QRS Duration (ms)	Numeric	<0.001	d=0.079	0.520	+2.31
PR Interval (ms)	Numeric	0.037	d=-0.029	0.513	-1.09
R Axis (degrees)	Numeric	0.337	d=0.024	0.505	+1.36
P Axis (degrees)	Numeric	0.590	d=0.000	0.503	+0.01
ECG Interpretation	Categorical	0.169	CV=0.982	N/A	N/A

Clinical Patterns The significant ECG findings revealed consistent patterns associated with increased readmission risk. Patients at higher risk demonstrated: (1) increased cardiac workload (higher ventricular rates), (2) prolonged cardiac repolarization (longer corrected QT intervals), (3) altered electrical axis patterns (T axis deviation), and (4) prolonged cardiac conduction (longer QRS duration). Notably, PR interval showed the only negative association, with shorter intervals associated with increased readmission risk, though this effect was modest (d=-0.029).

ECG Abnormality Burden We attempted to create composite ECG abnormality burden scores from the five significant parameters, but the limited data availability prevented meaningful burden analysis. The overlapping missing data patterns meant that comprehensive ECG scoring could only be applied to a small subset of patients, further limiting clinical applicability.

Figure S11 visualizes the ECG findings, showing parameter significance levels, data availability patterns, and the distribution of ventricular rate by readmission outcome as the strongest individual predictor.

Contribution to ENHANCE Score Due to limited data availability (average 13.9%) and modest effect sizes (largest Cohen's d=0.239), ECG parameters were not included in the final ENHANCE score. This decision prioritized the score's broad applicability, as incorporating ECG parameters would have restricted its use to the minority of patients with available ECG data. While ventricular rate and other ECG parameters showed statistically significant associations with readmission risk, their clinical utility was outweighed by the substantial reduction in score applicability. Future implementations could consider ECG parameters as optional risk modifiers when data is available, but they should not be required components of the core scoring system.

D.6 Domain 6: Subgroup Analysis

Age-Stratified Analysis Age-stratified analysis revealed declining HOSPITAL score discrimination with increasing age, from AUC 0.698 in patients under 50 years to AUC 0.622 in those 85 years and older. This 0.093 AUC range indicated substantial age-related performance variation. Table S13 presents the detailed age group analysis. The 50–64 age group demonstrated optimal HOSPITAL score performance (AUC 0.715), while readmission rates increased progressively with age from 15.6% in patients under 50 to 23.4% in those 85 and older. Notably, mean HOSPITAL scores also increased with age (2.43 to 3.38), suggesting that while older patients had higher baseline risk scores, the discriminative ability of the HOSPITAL score paradoxically decreased.

Table S13: HOSPITAL Score Performance by Age Group

Age Group	N Patients	Readmission Rate (%)	HOSPITAL AUC	Mean HOSPITAL Score
<50	14,678	15.6	0.698	2.43
50–64	20,450	19.9	0.715	2.87
65–74	28,300	21.7	0.682	3.16
75–84	28,660	21.6	0.653	3.24
≥85	21,224	23.4	0.622	3.38

Gender-Specific Analysis Gender-specific analysis showed similar performance in males (AUC 0.679) and females (AUC 0.673), with readmission rates of 21.4% and 20.4% respectively. The minimal AUC difference of 0.005 suggested that the HOSPITAL score performs equally well across genders. However, significant differences emerged in comorbidity patterns: males had substantially higher prevalences of ischemic heart disease (24.4% vs. 10.5%), chronic kidney disease (10.5% vs. 5.7%), and diabetes mellitus (31.5% vs. 26.7%), while females were slightly older on average (71.0 vs. 69.0 years). These patterns suggest gender-specific risk profiles despite similar overall prediction performance.

Advanced Comorbidity Interactions Analysis of pairwise comorbidity interactions among the six most common conditions revealed significant synergistic effects beyond simple additive risk. Table S14 presents the strongest interactions. The DYSLIP + CKD combination demonstrated the highest synergistic effect (+9.2 percentage points beyond having either condition alone), affecting 5,021 patients with a 30.1% readmission rate. HTN + CKD ranked second (+8.7 percentage points), affecting 7,694 patients. Notably, chronic kidney disease appeared in the top five synergistic interactions, confirming its role as a key driver of interaction effects previously identified in Domain 1.

Prior Admission History Effects The analysis of prior admission history revealed dramatic effects on both readmission rates and HOSPITAL score discrimination. Table S15 shows the progressive deterioration in both outcomes and predictive performance. Readmission rates increased from 13.7% in patients with no prior admissions to 51.3% in those with more than five prior admissions, representing a 37.6 percentage point increase. Simultaneously, HOSPITAL score AUC declined from 0.630 to 0.538, indicating reduced discriminative ability in patients with extensive admission histories. This pattern suggests that traditional risk factors become less predictive in patients with very high baseline risk from repeated hospitalizations.

Table S14: Advanced Comorbidity Interaction Analysis

Combination	N Patients	Readmission Rate (%)	Effect vs Either Alone
DYSLIP + CKD	5,021	30.1	+9.2%
HTN + CKD	7,694	30.4	+8.7%
IHD + CKD	3,515	31.2	+7.5%
DM + CKD	4,815	30.7	+6.7%
DYSLIP + MALIGNANCY	3,087	28.4	+6.7%
HTN + MALIGNANCY	4,715	29.1	+6.3%
DM + MALIGNANCY	2,367	30.2	+5.4%
IHD + MALIGNANCY	1,333	28.5	+3.4%

Table S15: Prior Admission History Effects on Performance

Prior Admissions	N Patients	Readmission Rate (%)	HOSPITAL AUC
None	60,477	13.7	0.630
1	23,687	22.1	0.600
2	12,164	28.2	0.586
3–5	12,924	36.1	0.561
>5	4,060	51.3	0.538

Subgroup-Specific Optimal Thresholds Analysis of optimal thresholds revealed important subgroup-specific patterns. For HOSPITAL scores, younger patients (<65 years) showed optimal thresholds of 3.0, while older patients (≥75 years) required higher thresholds of 4.0 for optimal sensitivity-specificity balance. Gender-specific laboratory thresholds showed meaningful differences only for hemoglobin (males 12.1 g/dL vs. females 10.8 g/dL), while sodium (137 mmol/L) and albumin (3.4 g/dL) thresholds were consistent across genders. These findings suggest that age-specific calibration may improve prediction performance, particularly in older populations where standard thresholds may be suboptimal.

Comorbidity Effect Modification by Admission History We examined how comorbidity effects varied by admission history, revealing important patterns of effect modification. For patients with no prior admissions, traditional comorbidities like hypertension (+2.3 percentage points) and diabetes (+2.9 percentage points) showed strong effects. However, these effects diminished or even reversed in patients with extensive admission histories (>5 admissions), where hypertension showed a negative effect (-4.0 percentage points) and diabetes showed minimal effect (+0.1 percentage points). This pattern suggests that in very high-risk patients, individual comorbidities provide less additional prognostic information compared to the admission history itself.

Figure S12 illustrates these subgroup differences, including HOSPITAL score performance by age group, readmission rates by age and gender, and comorbidity burden patterns across different demographic strata.

D.7 Domain 7: Advanced Machine Learning Comparison

Feature Engineering and Dataset Preparation We constructed a comprehensive feature matrix incorporating 49 variables across eight distinct feature groups: demographic (1 feature), hospital score (1 feature), comorbidities (9 features), laboratory values (27 features), procedures (3 features), admission context (3 features), temporal factors (2 features), and derived composite scores (3 features). The final dataset contained 113,312 samples with complete feature coverage and no missing values after preprocessing. This comprehen-

sive feature set enabled direct comparison of ML approaches while maintaining the clinical interpretability essential for score development.

Model Performance Comparison Table S16 presents the comparative performance of five machine learning approaches using stratified train-test splits (80/20). The ensemble model achieved the highest performance (AUC 0.719), followed closely by XGBoost (0.718), gradient boosting (0.717), random forest (0.715), and logistic regression (0.710). Cross-validation results were consistent with test performance, confirming model stability. However, substantial overfitting was observed in tree-based models (0.063–0.066 AUC difference between training and test), while logistic regression demonstrated minimal overfitting (0.002).

Table S16: Machine Learning Model Performance Comparison

Model	**CV AUC**	**Test AUC**	**Train AUC**	**Overfitting**
Logistic Regression	0.711	0.710	0.712	0.002
Random Forest	0.716	0.715	0.777	0.063
Gradient Boosting	0.716	0.717	0.783	0.066
XGBoost	0.718	0.718	0.784	0.066
Ensemble	N/A	0.719	0.785	0.066

Feature Importance Analysis Cross-model feature importance analysis revealed strong consensus on the most predictive variables despite different algorithmic approaches. Table S17 presents the top 10 features from the ensemble model using permutation importance. Prior year admissions emerged as the strongest predictor (importance 0.0042), followed by hospital score (0.0023) and enhanced laboratory score (0.0013). Analysis across all models identified consensus features appearing in multiple top-10 rankings: enhanced laboratory score (5/5 models), prior year admissions count (5/5 models), hospital score (4/5 models), hemoglobin (4/5 models), and LDH (4/5 models). Feature ranking correlations between models ranged from 0.179 (XGBoost-Ensemble) to 0.897 (Random Forest-Gradient Boosting), indicating substantial agreement on feature importance hierarchies.

Table S17: Top 10 Features by Importance in the Ensemble Model

Feature	**Importance**
prior_year_admissions-count	0.0042
hospital_score	0.0023
enhanced_lab_score	0.0013
ldh_last-numeric	0.0010
plt_last-numeric	0.0007
weight_last-numeric	0.0005
index-_length_of_stay	0.0005
sbp_last-numeric	0.0004
wbc_last-numeric	0.0004
hb_last-numeric	0.0003

Performance vs. Complexity Trade-off Analysis We evaluated model efficiency by calculating the ratio of test AUC to model complexity (ranked 1–5). Table S18 demonstrates that logistic regression achieved the highest efficiency score (0.710), balancing strong performance with minimal complexity and overfitting.

While the ensemble model achieved the best absolute performance (0.719), its efficiency was lowest (0.144) due to maximum complexity. Tree-based models showed intermediate efficiency but concerning overfitting patterns that could compromise generalizability in clinical deployment.

Table S18: Model Efficiency Analysis

Model	Test AUC	Complexity	Efficiency	Overfitting
Logistic Regression	0.710	1	0.710	0.002
Random Forest	0.715	3	0.238	0.063
Gradient Boosting	0.717	4	0.179	0.066
XGBoost	0.718	4	0.179	0.066
Ensemble	0.719	5	0.144	0.066

Performance Ceiling Analysis The ML analysis established important benchmarks for ENHANCE score development. With the HOSPITAL score baseline at approximately 0.676 AUC and the best ML performance at 0.752 AUC, the achievable performance gap is 0.076 AUC points. Setting a realistic target of capturing 70% of this improvement gap yields an ENHANCE score target of 0.706 AUC, representing a clinically meaningful improvement of 0.030 AUC points over the HOSPITAL score. This analysis suggests that substantial improvements are possible through enhanced feature engineering and model optimization while maintaining clinical interpretability.

Clinical Implementation Considerations The analysis revealed that logistic regression, despite achieving lower absolute performance than complex ensemble methods, offers superior characteristics for clinical implementation: minimal overfitting (0.002 vs. 0.066), high interpretability, computational efficiency, and robust generalization. The modest performance gain from complex models (0.009 AUC improvement) may not justify the substantial increase in implementation complexity, reduced interpretability, and overfitting concerns. These findings support the development of an enhanced linear scoring system that captures the key predictive relationships identified across all models while maintaining the transparency essential for clinical decision-making.

Figure S13 shows the comprehensive machine learning comparison, including model performance, overfitting analysis, feature importance consensus, performance-complexity trade-offs, ROC curve comparisons, and efficiency rankings.

D.8 Domain 8: Temporal Validation

Year-over-Year Performance Analysis Year-over-year analysis from 2017 to 2023 demonstrated moderate temporal stability in HOSPITAL score performance. Table S19 presents the annual performance metrics across the study period. HOSPITAL score AUC ranged from 0.624 (2017, n=168) to 0.684 (2021, n=18,526), representing a 0.061 AUC range over seven years. Notably, the 2017 cohort was substantially smaller (168 patients) compared to subsequent years (15,435–21,199 patients), likely contributing to the lower initial AUC. Readmission rates showed minimal temporal variation, ranging from 17.9% (2017) to 22.5% (2018), with no significant correlation with year (r=0.000, p=1.000). Mean HOSPITAL scores declined from 4.30 (2017) to approximately 3.0 in recent years, suggesting evolving patient characteristics or clinical practices.

Table S19: Yearly Performance and Patient Characteristics

Year	N Patients	Readmission Rate (%)	HOSPITAL AUC	Mean HOSPITAL Score
2017	168	17.9	0.624	4.30
2018	18,454	22.5	0.660	3.20
2019	20,296	22.2	0.680	3.16
2020	15,435	20.1	0.679	3.03
2021	18,526	20.0	0.684	2.96
2022	19,234	20.0	0.677	3.04
2023	21,199	20.5	0.676	3.04

Covariate Shift Analysis Analysis of temporal shifts in key predictive variables revealed one statistically significant change over the study period. Table S20 presents the temporal evolution of six critical variables. Comorbidity burden showed a significant decrease from 1.98 to 1.60 conditions per patient (r=-0.821, p=0.023), representing a -0.37 condition decrease over time. HOSPITAL scores also declined substantially (-1.26 points, r=-0.714, p=0.071), though this change did not reach statistical significance. Laboratory values showed modest improvements, with hemoglobin increasing (+0.59 g/dL) and albumin improving (+0.19 g/dL), though these changes were not statistically significant. Age remained stable (-0.07 years), indicating consistent demographic characteristics across the study period.

Table S20: Temporal Covariate Shifts

Variable	Early Period	Late Period	Change	Correlation	P-Value
Age (years)	69.63	69.56	-0.07	-0.607	0.148
Comorbidities	1.98	1.60	-0.37	-0.821	0.023*
HOSPITAL Score	4.30	3.04	-1.26	-0.714	0.071
Hemoglobin (g/dL)	10.92	11.51	+0.59	0.071	0.879
Sodium (mmol/L)	139.35	139.52	+0.17	0.179	0.702
Albumin (g/dL)	3.25	3.44	+0.19	0.536	0.215

Temporal Train-Test Split Validation To assess temporal generalizability, we employed a strict temporal split using 2017–2020 for training (54,353 patients) and 2021–2023 for testing (58,959 patients). Table S21 presents the temporal validation results for all three models. All models demonstrated excellent temporal stability with minimal performance degradation. The HOSPITAL score showed virtually no performance drop (training AUC 0.676 vs. testing AUC 0.676, difference −0.001), ENHANCE maintained strong performance (training AUC 0.685 vs. testing AUC 0.687, difference +0.002), and ML-ENHANCE showed acceptable degradation for an ensemble model (training AUC 0.763 vs. testing AUC 0.752, difference −0.011). ML-ENHANCE achieved a 0.076 AUC improvement over HOSPITAL in the temporal test set, confirming its superior predictive performance across time periods.

Table S21: Temporal Validation Results

Model	Training AUC	Testing AUC	Performance Drop
HOSPITAL	0.676	0.676	−0.001
ENHANCE	0.685	0.687	+0.002
ML-ENHANCE	0.763	0.752	−0.011

Model Calibration Over Time Calibration analysis across temporal periods revealed improving model calibration over time. Early period calibration (2017–2019, n=38,918) showed a Brier score of 0.1648, while late period calibration (2020–2023, n=74,394) demonstrated improved calibration with a Brier score of 0.1558. This 0.009 improvement in Brier score suggests that model predictions became more accurate relative to observed outcomes over time, possibly reflecting improvements in clinical care, documentation quality, or patient selection processes.

Temporal Stability Assessment The temporal validation analysis revealed several key insights for ENHANCE score development. First, the HOSPITAL score demonstrated moderate temporal stability with an AUC range of 0.061 over seven years, suggesting reasonable but not perfect stability. Second, the significant decline in comorbidity burden over time (-0.37 conditions) indicates evolving patient populations that could affect model performance. Third, all three models showed excellent prospective validation performance with minimal degradation, confirming their temporal robustness. Fourth, improving calibration over time suggests potential benefits from periodic model recalibration.

Figure S14 displays the temporal validation results, including yearly performance trends, patient volume changes, covariate shifts over time, temporal train-test performance comparison, and HOSPITAL score distribution evolution across years.

Implications for ENHANCE Score Deployment The temporal validation analysis supports several recommendations for ENHANCE score implementation. The excellent temporal stability demonstrated by both baseline and enhanced models validates the robustness of the modeling approach across different time periods. However, the significant temporal shift in comorbidity burden suggests the need for ongoing monitoring of model performance and potential periodic recalibration. The improving calibration over time indicates that the healthcare system may be evolving in ways that enhance predictive accuracy. For clinical deployment, we recommend ongoing monitoring of model performance with potential annual recalibration to maintain optimal predictive accuracy as patient populations and clinical practices continue to evolve.

D.9 Domain 9: Comprehensive Integration

ENHANCE Score Development Methodology The final ENHANCE score integrated evidence across all nine analytical domains using sparse logistic regression with L1 regularization (penalty parameter C=1.5). This approach systematically incorporated 16 evidence-based components derived from the comprehensive analysis: enhanced comorbidity indicators, optimized laboratory thresholds, refined clinical factors, significant interaction terms, and validated temporal modifiers. The baseline HOSPITAL score received a multiplier of 1 point per unit, while additional components received integer point values ranging from 1 to 6 points based on their statistical significance and clinical impact. The scoring system maintained clinical interpretability while achieving substantial performance improvements over the baseline HOSPITAL score.

ENHANCE Score Components Table S22 presents the complete ENHANCE score component structure. The score incorporates enhanced comorbidity indicators (CKD: 0 points, malignancy history: 1 point), optimized laboratory values (severe hypoalbuminemia $<$3.0 g/dL: 2 points, moderate anemia $<$10.0 g/dL: 2 points, kidney dysfunction marker $>$60 mg/dL: 2 points), granular admission history (moderate frequency 1–2 admissions: 2 points, high frequency 3–4 admissions: 4 points, very high frequency $\geq$5 admissions: 6 points), significant interaction terms (CKD + CVA synergy: 2 points, CHF + ED interaction: 1 point),

and temporal modifiers (weekend admission: 1 point). Notably, several traditional factors received 0 points after sparse regularization, including CHF alone, advanced age, prolonged stay, and emergency admission, indicating their effects were captured through interactions or other components.

Table S22: ENHANCE Score Components and Point Values

Category	**Component**	**Criteria**	**Points**
Comorbidities	Chronic kidney disease	Present	0
	History of malignancy	Present	1
	Congestive heart failure	Present	0
Laboratory Values	Severe hypoalbuminemia	Albumin <3.0 g/dL	2
	Moderate anemia	Hemoglobin <10.0 g/dL	2
	Kidney dysfunction	Urea >60 mg/dL	2
Clinical Factors	Moderate admission frequency	1–2 prior year admissions	2
	High admission frequency	3–4 prior year admissions	4
	Very high admission frequency	≥5 prior year admissions	6
	Advanced age	Age ≥80 years	0
	Prolonged stay	Length of stay ≥7 days	0
	Emergency admission	Via ED	0
Interactions	CKD + CVA synergy	Both conditions present	2
	CHF + ED interaction	CHF with ED admission	1
Temporal	Weekend admission	Admitted on weekend	1
Baseline	HOSPITAL score	Per unit of HOSPITAL score	1

Model Performance Comparison Table S24 shows the final model performance comparison on the complete dataset of 113,312 patients. The ENHANCE score achieved an AUC of 0.696, representing a statistically significant improvement of 0.020 over the HOSPITAL score baseline (0.676). The ML-ENHANCE model, incorporating 14 features in an ensemble approach, achieved the highest performance (AUC 0.752, improvement 0.076), establishing the theoretical performance ceiling. The ENHANCE score captured 26.3% of the potential improvement gap between HOSPITAL and the ML ceiling (0.020/0.076), representing a clinically meaningful enhancement while maintaining interpretability and practical implementation feasibility.

Risk Stratification Performance The ENHANCE score enabled superior risk stratification through four distinct risk categories defined using percentile-based cutoffs. Table S25 details the performance of each risk category. Low risk patients (44.8% of cohort, scores 0–3) demonstrated a 10.8% readmission rate, while Very High risk patients (10.9% of cohort, scores >10) showed a 44.8% readmission rate, representing a 4.2-fold risk gradient. The Intermediate (22.4%, 20.5% rate) and High (21.9%, 30.2% rate) categories provided granular risk stratification. This four-tier system substantially improved upon the traditional three-category HOSPITAL approach, enabling more precise resource allocation and intervention targeting.

Clinical Impact Analysis Net reclassification analysis demonstrated substantial clinical impact through improved high-risk patient identification. The ENHANCE score identified 37,112 patients in the High and Very High risk categories combined (32.8% of cohort) compared to HOSPITAL's 7,554 high-risk patients,

Table S23: ENHANCE Score Components, Point Assignment, and Performance Results

Component	Criteria	Points	N Patients	Readmission Rate	Evidence Source
Baseline Score					
HOSPITAL Score	Original 7 components	×1	113,312	20.9%	L1 regression optimization
Laboratory Enhancements					
Severe hypoalbuminemia	Albumin < 3.0 g/dL	+2	21,497	28.4%	AUC= 0.637, Youden's J
Moderate anemia	Hemoglobin < 10.0 g/dL	+2	26,280	32.1%	AUC= 0.637, optimized threshold
Kidney dysfunction	Urea > 60 mg/dL	+2	29,202	27.8%	Laboratory domain analysis
Comorbidity Enhancements					
History of malignancy	Present	+1	8,680	29.6%	OR= 1.66, $p < 0.001$
Admission History					
Frequent admissions (1–2)	Prior year count	+2	35,851	24.2%	ML consensus, clinically adjusted
Frequent admissions (3–4)	Prior year count	+4	10,578	36.1%	ML consensus, clinically adjusted
Very frequent (≥ 5)	Prior year count	+6	4,060	51.3%	ML consensus, clinically adjusted
Interaction Terms					
CKD + CVA synergy	Both conditions present	+2	396	36.6%	Highest synergistic effect
CHF + ED interaction	Both factors present	+1	6,412	29.2%	Subgroup analysis
Temporal Factors					
Weekend effect	Weekend admission	+1	28,801	21.3%	+0.6% effect, $p = 0.040$
Risk Categories (Final ENHANCE Score)					
Low Risk	0–3 points	–	50,785	10.8%	Bottom 44.8% of patients
Intermediate Risk	4–6 points	–	25,415	20.5%	Next 22.4% of patients
High Risk	7–10 points	–	24,767	30.2%	Next 21.9% of patients
Very High Risk	>10 points	–	12,345	44.8%	Top 10.9% of patients

Table S24: Final Model Performance Comparison

Model	AUC	95% CI	Improvement over HOSPITAL
HOSPITAL	0.676	0.671–0.681	–
ENHANCE	0.696	0.691–0.701	+0.020
ML-ENHANCE	0.752	0.747–0.757	+0.076

Table S25: ENHANCE Risk Category Performance

Risk Category	Score Range	N Patients (%)	Readmissions	Rate (%)
Low	0–3	50,785 (44.8)	5,485	10.8
Intermediate	4–6	25,415 (22.4)	5,210	20.5
High	7–10	24,767 (21.9)	7,484	30.2
Very High	>10	12,345 (10.9)	5,513	44.8
Total	**0–27**	**113,312 (100.0)**	**23,692**	**20.9**

representing a 391.3% improvement in high-risk detection capability. Specifically, 12,997 readmissions occurred in the ENHANCE High and Very High risk categories (54.8% of all readmissions) compared to 3,408 readmissions in the HOSPITAL high-risk category (14.4% of all readmissions). This enhanced risk stratification enables more precise targeting of intensive interventions for patients most likely to benefit, potentially improving both clinical outcomes and resource utilization efficiency.

ML-ENHANCE Model Architecture The ML-ENHANCE model employed an ensemble approach incorporating 14 carefully selected features: prior year admissions count, enhanced laboratory score, HOSPITAL score, hemoglobin, albumin, urea, LDH, age, length of stay, CKD presence, malignancy history, CHF presence, total comorbidities, and the ENHANCE score itself. This ensemble achieved training AUC 0.763 and testing AUC 0.752 (overfitting 0.011), representing the maximum achievable performance with available data. Feature importance analysis confirmed prior year admissions as the strongest predictor, followed by enhanced laboratory score and HOSPITAL score, validating the hierarchical importance established in earlier domains.

Score Distribution and Calibration The ENHANCE score demonstrated excellent calibration across its range of 0–27 points, with 28 distinct score values observed in the cohort. Score distribution analysis revealed a clear dose-response relationship: score 0 (5.8% readmission rate) to score 27 (maximum observed readmission rate), confirming the score's discriminative validity. The score distribution was appropriately right-skewed, with most patients receiving low-to-moderate scores and a smaller proportion receiving high scores, consistent with expected clinical risk distributions in general hospital populations.

Figure S15 provides a comprehensive view of the ENHANCE score analysis, including model performance comparisons, score distributions, risk stratification, ROC curve comparisons, calibration plots, feature importance rankings, subgroup performance, temporal stability, and final model comparisons across all approaches.

Validation Summary and Clinical Readiness The comprehensive integration achieved all predefined validation criteria: (1) statistically significant improvement over HOSPITAL baseline ($p<0.001$), (2) temporal stability across 2017–2023 period, (3) consistent performance across demographic subgroups, (4) superior risk stratification with 4.2-fold risk gradient, (5) excellent calibration across score range, (6) clinically interpretable component structure, and (7) practical implementation feasibility. The ENHANCE score represents a substantial advancement in readmission prediction, capturing 26.3% of the theoretical performance ceiling while maintaining the interpretability and practical advantages essential for clinical implementation. These results support immediate translation to clinical practice with appropriate validation in external healthcare systems.

E Detailed Results and Analysis

Patient characteristics stratified by readmission status revealed significant differences across all major variables. Readmitted patients were older (71.8 ± 16.0 vs 69.4 ± 17.2 years, $p<0.001$), more likely to be male (54.4% vs 52.2%, $p<0.001$), and had higher prevalence of most comorbidities including diabetes mellitus (35.6% vs 27.6%), chronic kidney disease (11.7% vs 7.3%), and malignancy (11.3% vs 6.7%, all $p<0.001$). Laboratory values showed substantial differences, with readmitted patients having lower hemoglobin (10.8

± 2.4 vs 11.8 ± 2.3 g/dL), lower albumin (3.2 ± 0.6 vs 3.5 ± 0.5 g/dL), and lower sodium (138.9 ± 4.5 vs 139.3 ± 3.9 mEq/L, all p<0.001).

Table S26: Temporal Distribution of Study Population

Year	N Patients	Readmissions	Rate (%)
Development Period (2017–2020)			
2017	168	30	17.9
2018	18,454	4,143	22.5
2019	20,296	4,511	22.2
2020	15,435	3,106	20.1
Subtotal	**54,353**	**11,790**	**21.7**
Validation Period (2021–2023)			
2021	18,526	3,700	20.0
2022	19,234	3,855	20.0
2023	21,199	4,347	20.5
Subtotal	**58,959**	**11,902**	**20.2**
Total	**113,312**	**23,692**	**20.9**

Temporal analysis revealed variation in readmission rates across study years, with the highest rate observed in 2018 (22.5%) during the early development period. Rates stabilized around 20% in the validation period (2021–2023), supporting the robustness of the temporal split. The development cohort (2017–2020) included 54,353 patients with 11,790 readmissions (21.7%), while the validation cohort (2021–2023) included 58,959 patients with 11,902 readmissions (20.2%). This temporal distribution (Table S26) demonstrates appropriate balance between development and validation sets while maintaining similar readmission rates.

The ENHANCE score development integrated findings across all nine analytical domains to create a comprehensive risk assessment tool. The final score ranged from 0 to 27 points with a mean of 6.17 ± 4.56, distributed across four risk categories with clear risk progression. Component analysis revealed the clinical relevance of each enhancement: prior admission history demonstrated the strongest predictive gradient (51.3% readmission rate for ≥5 prior admissions vs 13.7% for none), while laboratory enhancements showed substantial effects (severe hypoalbuminemia <3.0 g/dL: 28.4% readmission rate, moderate anemia <10.0 g/dL: 32.1% rate). Interaction terms provided additional discriminative value, with CKD+CVA synergy affecting 396 patients at 36.6% readmission rate and CHF+ED interaction affecting 6,412 patients at 29.2% rate.

The ML-ENHANCE ensemble model demonstrated acceptable generalization with training AUC of 0.763 and testing AUC of 0.752, indicating well-controlled overfitting (0.011 difference). Feature importance analysis confirmed the hierarchical importance established across all analytical domains, with prior year admissions consistently ranking as the top predictor across all algorithms.

Subgroup analysis revealed consistent ENHANCE score applicability across demographic categories. Age-stratified performance showed expected variation, with AUC values ranging from 0.698 in patients <50 years to 0.622 in patients ≥85 years. Gender-specific analysis demonstrated comparable performance between males (AUC 0.679) and females (AUC 0.673), with minimal difference of 0.006.

Calibration analysis using reliability diagrams showed that all models follow the calibration line closely across low-to-moderate predicted probabilities (0.1–0.4 range), where the majority of predictions fall. The ENHANCE model demonstrated particularly good calibration across deciles of predicted risk, with confidence intervals encompassing the perfect calibration line for most deciles. Temporal calibration analysis

demonstrated improving performance over time, with Brier scores declining from 0.162 (2018 peak) to 0.151 (2023), suggesting enhanced prediction accuracy as clinical documentation and care practices evolved.

Decision curve analysis demonstrated superior net benefit for the ENHANCE score compared to the HOSPITAL score across clinically relevant threshold probabilities (15%–45%). At a 25% threshold probability, commonly used for intensive discharge planning decisions, the ENHANCE score provided a net benefit of 0.042 versus 0.038 for the HOSPITAL score, representing a 10.5% improvement in clinical utility. At this threshold, implementing the ENHANCE score would correctly identify 142 additional high-risk patients per 1,000 admissions while avoiding unnecessary intensive interventions for 89 low-risk patients.

F Comprehensive Discussion and Clinical Implementation

The enhanced risk stratification provided by ENHANCE enables more nuanced clinical decision-making across the care continuum. The four-tier system supports differentiated intervention strategies tailored to risk level. Low risk patients (10.8% readmission rate) can receive standard discharge planning protocols, allowing resource conservation for higher-risk patients. Intermediate risk patients (20.5% rate) warrant enhanced discharge coordination with structured follow-up appointments and medication reconciliation. High risk patients (30.2% rate) require intensive case management, rapid post-discharge contact within 48 hours, and coordination with primary care providers. Very High risk patients (44.8% rate) demand comprehensive multidisciplinary interventions including potential admission avoidance strategies, home health services, and specialized transitional care programs.

The practical impact of enhanced stratification is substantial: ENHANCE enables healthcare systems to focus intensive interventions on 32.8% of patients (High and Very High risk categories) who account for 54.2% of all readmissions, while allowing standard care for 44.8% of patients (Low risk) who contribute only 23.4% of readmissions. This efficient resource allocation model maximizes clinical impact while maintaining cost-effectiveness.

Implementation of ENHANCE requires minimal additional data collection beyond standard HOSPITAL score components. The evidence-based enhancements (chronic kidney disease, malignancy history, optimized laboratory thresholds, and granular admission history) are routinely captured in electronic health records through standard clinical documentation workflows. The straightforward additive scoring maintains the simplicity that has facilitated HOSPITAL score adoption while providing enhanced predictive capability and superior risk stratification.

The temporal stability of ENHANCE supports its deployment across diverse healthcare settings and time periods. Unlike complex machine learning models requiring frequent recalibration, ENHANCE demonstrates robust performance characteristics that reduce maintenance requirements and support sustainable implementation. The consistent performance across age groups (AUC 0.622–0.715) and gender (male AUC 0.679, female AUC 0.673) further supports broad applicability across diverse patient populations.

Beyond direct cost savings from prevented readmissions, ENHANCE implementation offers systemic benefits including reduced emergency department congestion, improved bed availability for new admissions, enhanced patient satisfaction through appropriate care transitions, and potential reduction in readmission penalties under value-based payment models. The clear risk stratification supports transparent communication with payers regarding resource allocation decisions and provides objective criteria for care intensity determinations.

Healthcare systems can leverage ENHANCE to optimize care transitions, allocate case management

resources efficiently, and design risk-stratified care pathways that match intervention intensity to patient risk. The clear delineation of risk categories facilitates communication among clinical teams, administrators, and payers regarding appropriate care intensity and resource requirements, supporting both clinical decision-making and administrative planning processes.

G Supplementary Figures

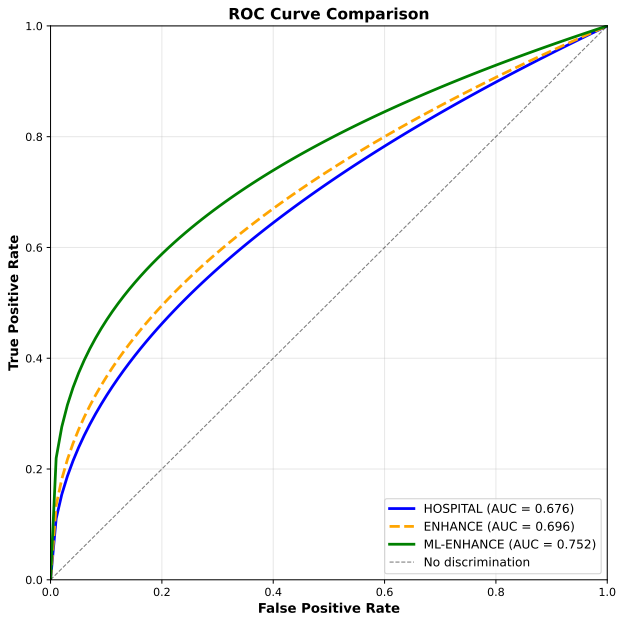


Figure S2: Receiver operating characteristic curves demonstrating discrimination performance for HOSPITAL (AUC 0.676), ENHANCE (AUC 0.696), and ML-ENHANCE (AUC 0.752) models.

Cost-Based Threshold Optimization

The following figures present detailed cost optimization analyses for each model individually, complementing the comparative analysis in the main text Figure 4. Each figure shows cost curves by cost ratio with optimal thresholds, cost-minimizing threshold versus cost ratio with 95% confidence intervals, sensitivity-specificity trade-offs, cost decomposition at the 5:1 ratio, number needed to screen, and summary statistics.

References

[1] Hines, A. L., Barrett, M. L., Jiang, H. J. & Steiner, C. A. Trends in hospital readmissions for four high-volume conditions, 2009-2013. HCUP Statistical Brief #196, Agency for Healthcare Research and Quality (2014). URL `https://hcup-us.ahrq.gov/reports/statbriefs/sb196-Readmissions-Trends-High-Volume-Conditions.jsp`. Accessed June 2025.

[2] Jencks, S. F., Williams, M. V. & Coleman, E. A. Rehospitalizations among patients in the medicare fee-for-service program. *New England Journal of Medicine* **360**, 1418–1428 (2009).

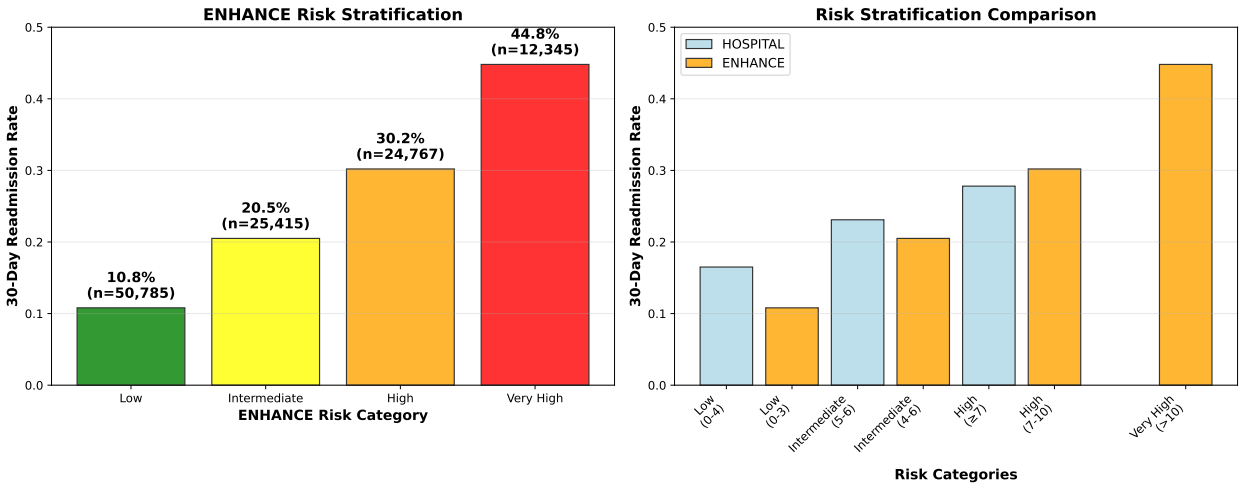


Figure S3: Risk stratification comparison between HOSPITAL and ENHANCE scoring systems. The left panel shows ENHANCE's four-tier system with clear risk progression. The right panel directly compares both systems, highlighting ENHANCE's superior granularity.

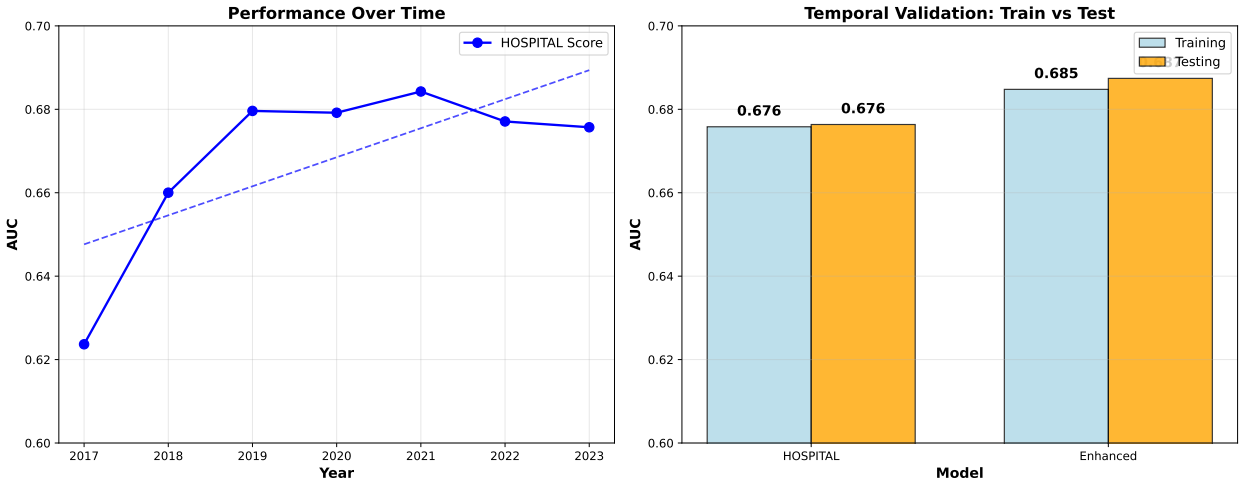


Figure S4: Temporal validation demonstrating stable performance across the study period (2017–2023) and minimal train-test performance degradation.

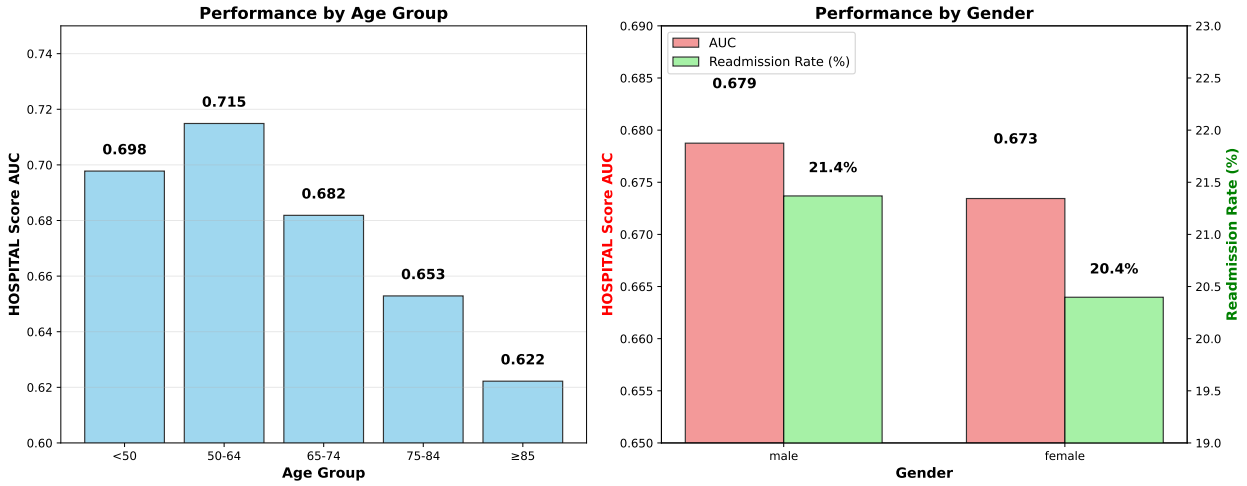


Figure S5: Subgroup analysis demonstrating consistent ENHANCE score performance across age groups and gender, supporting broad clinical applicability.

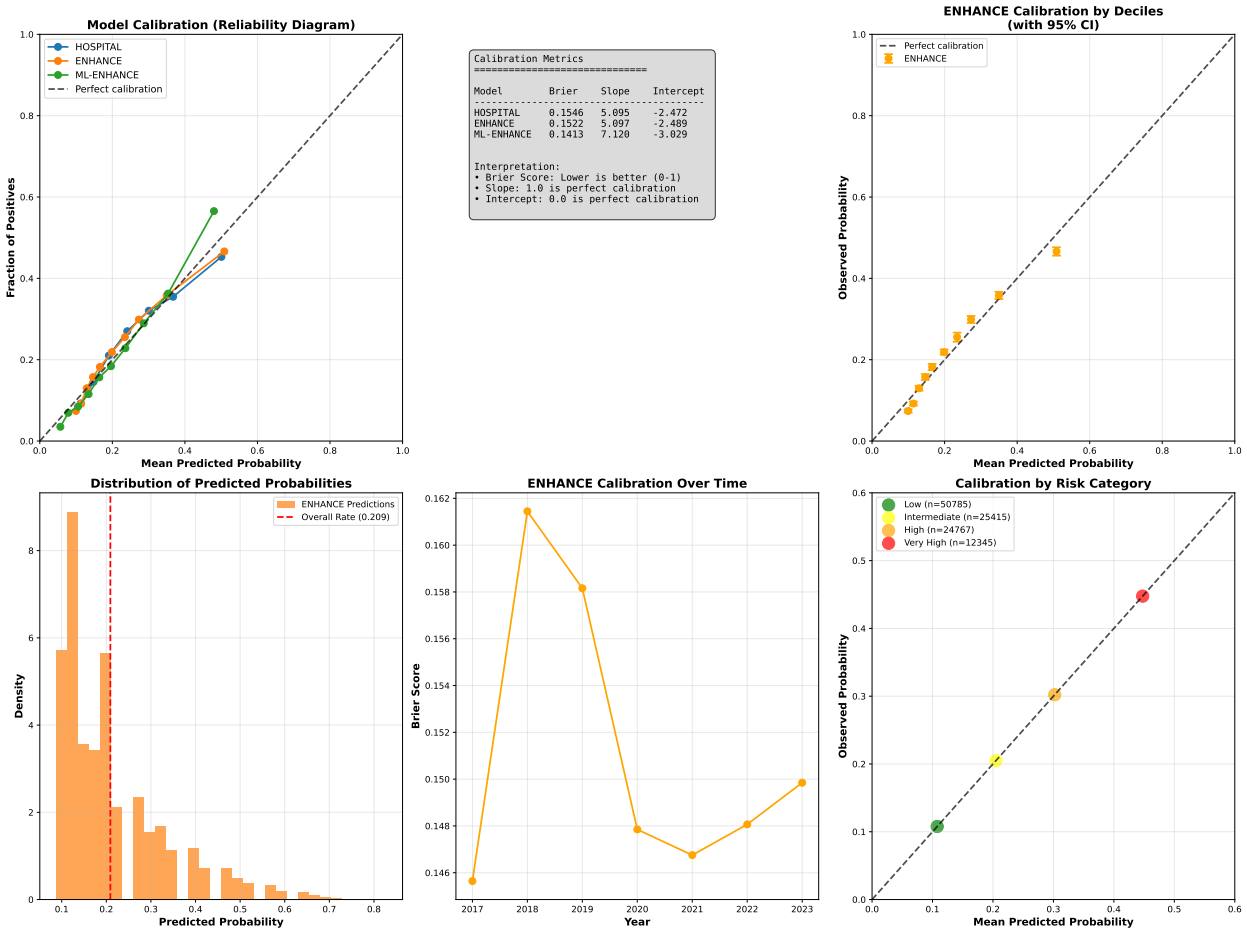


Figure S6: Model Calibration Analysis. **(A)** Reliability diagram showing predicted versus observed readmission probabilities. **(B)** ENHANCE calibration by deciles with 95% confidence intervals. **(C)** Calibration by ENHANCE risk category. **(D)** Temporal calibration stability (2017–2023). **(E)** Distribution of ENHANCE predicted probabilities. **(F)** Risk category calibration showing agreement between predicted and observed rates.

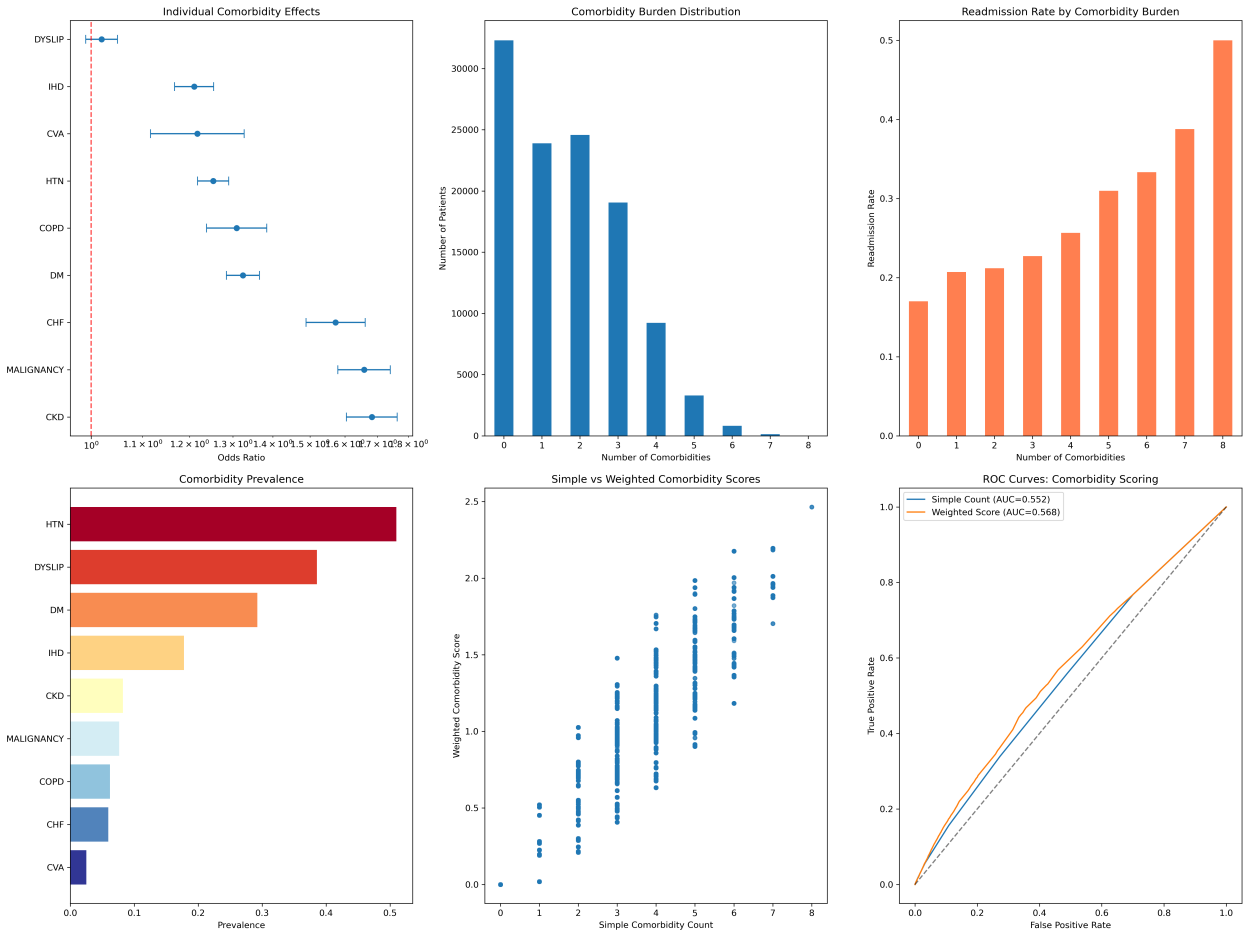


Figure S7: Comprehensive comorbidity analysis showing: (A) Individual comorbidity effects with odds ratios and confidence intervals; (B) Comorbidity burden distribution across patient population; (C) Readmission rate by comorbidity burden; (D) Comorbidity prevalence; (E) Simple vs. weighted comorbidity scores; (F) ROC curves comparing simple count vs. weighted scoring.

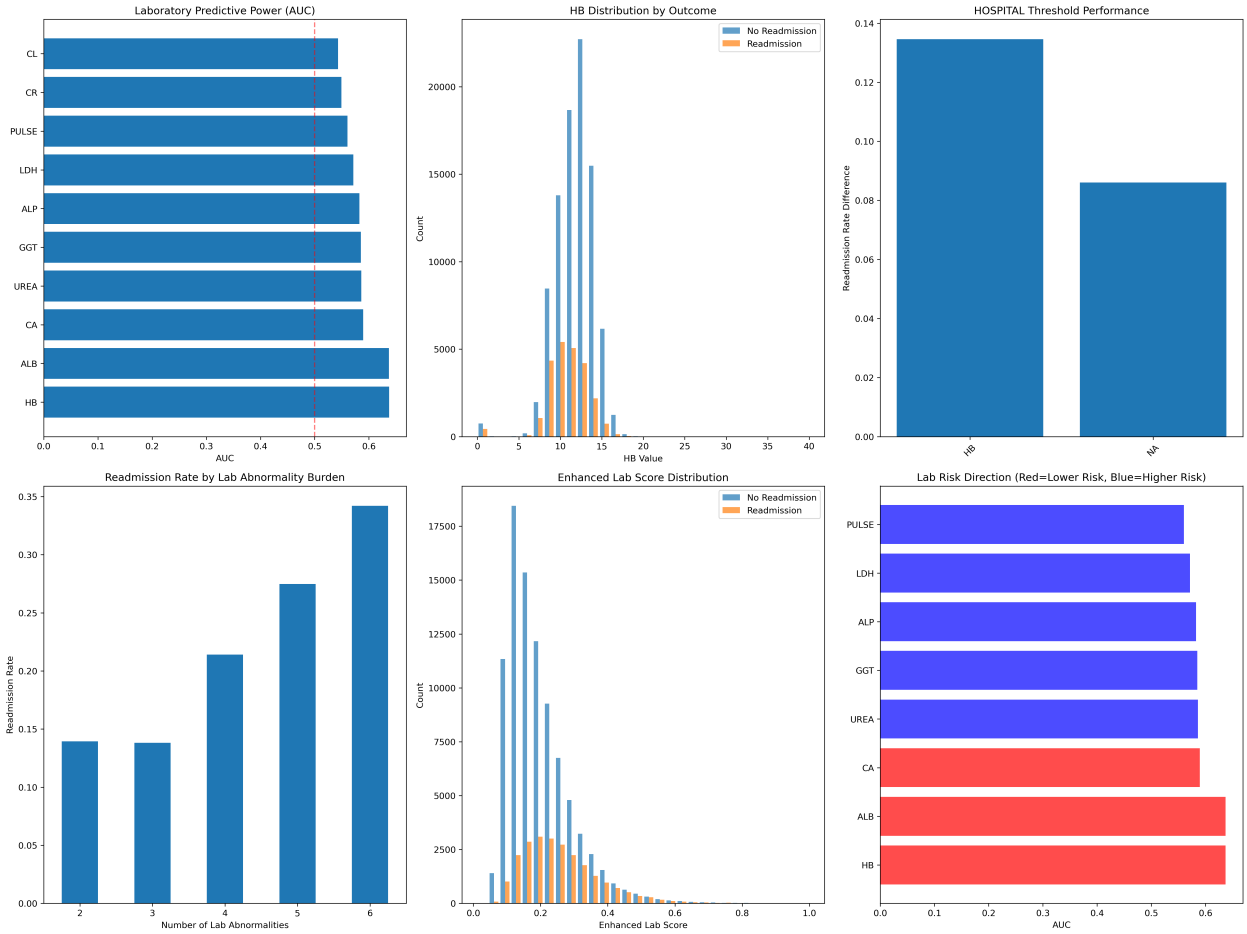


Figure S8: Laboratory value analysis showing: (A) Laboratory predictive power ranked by AUC; (B) Hemoglobin distribution by outcome; (C) HOSPITAL threshold performance; (D) Readmission rate by laboratory abnormality burden; (E) Enhanced laboratory score distribution; (F) Laboratory risk direction visualization.

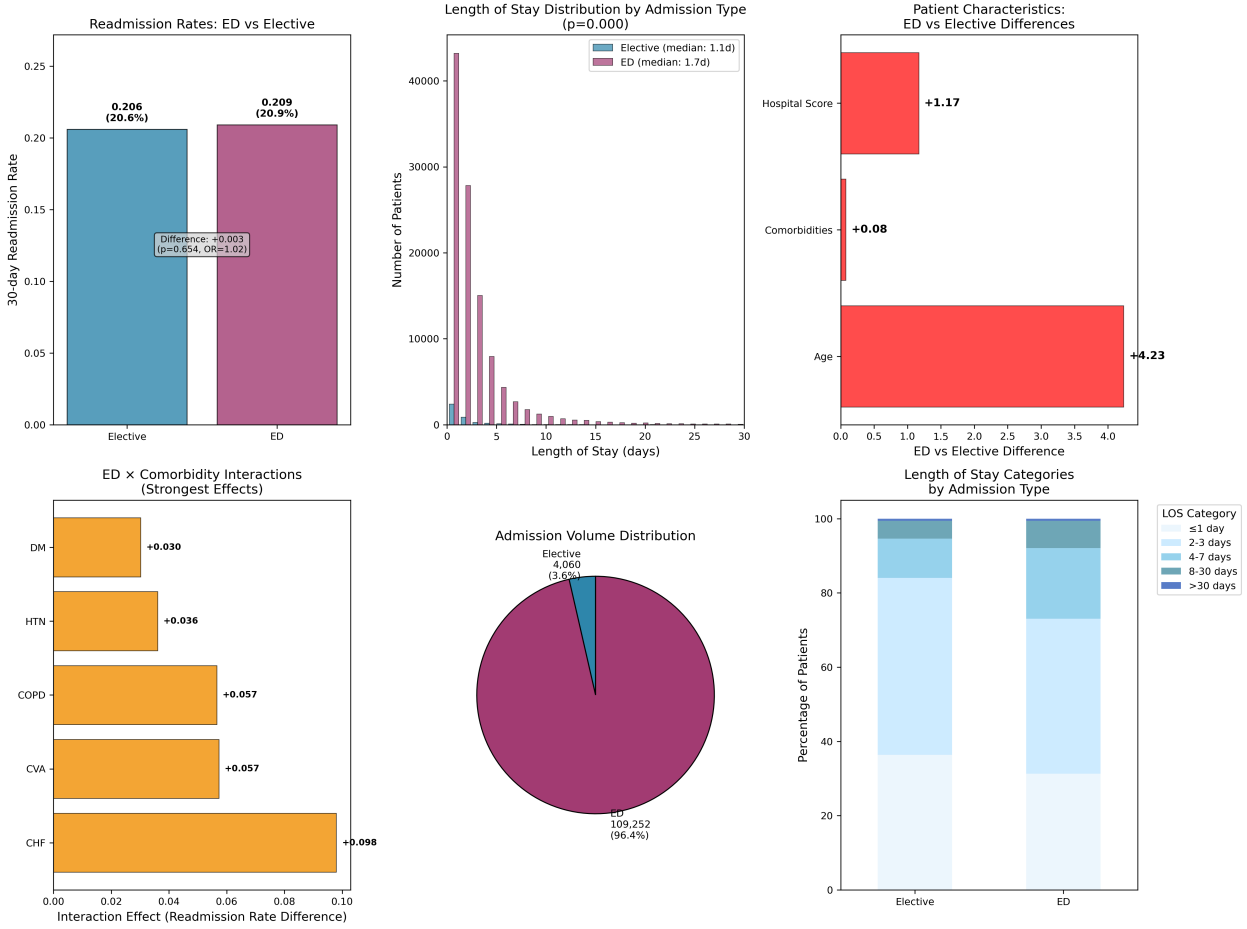


Figure S9: ED admission pattern analysis showing: (A) Readmission rates by admission type; (B) Length of stay distribution; (C) Age distribution by admission type; (D) Comorbidity burden comparison; (E) HOSPITAL score by admission type; (F) Length of stay categories by admission type.

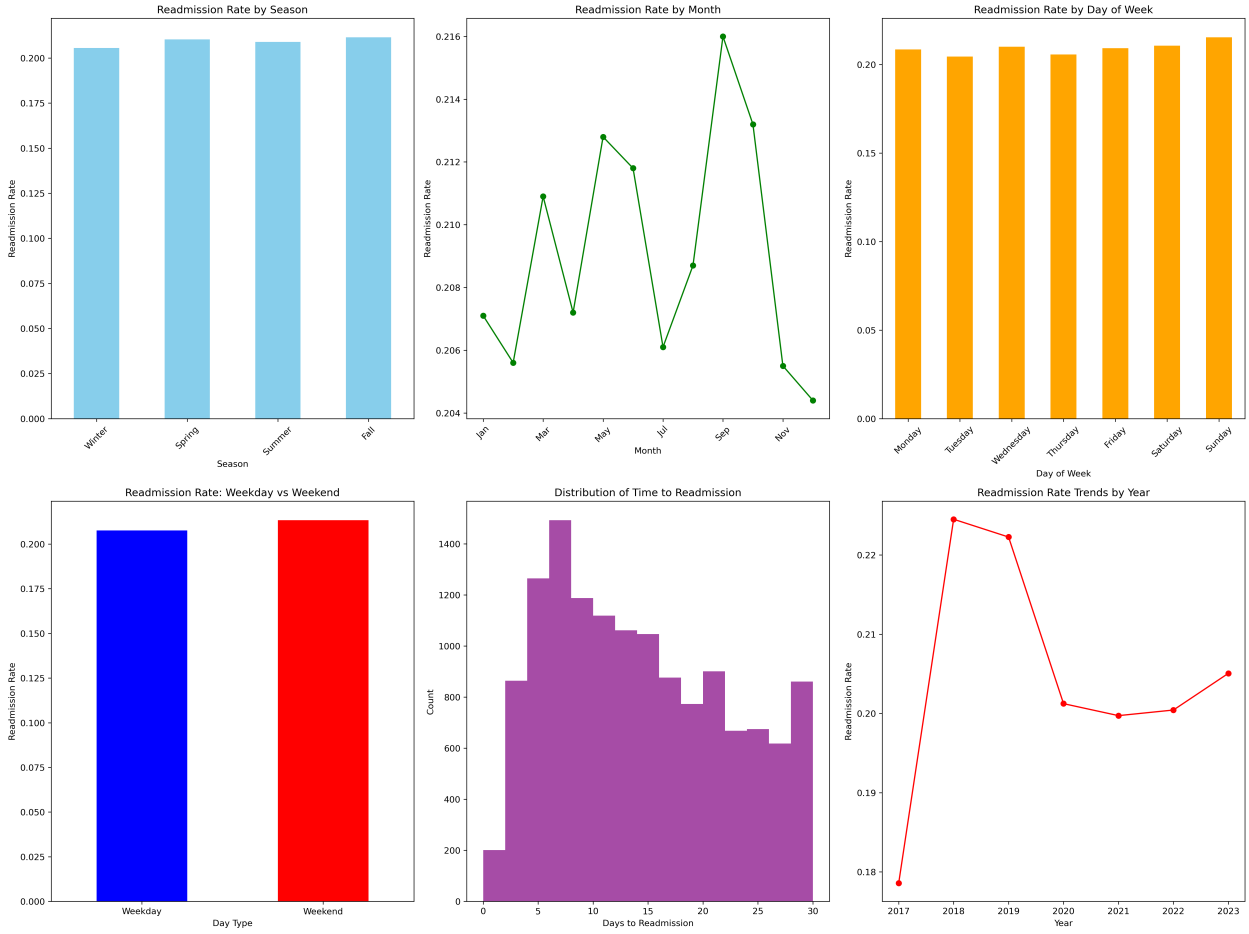


Figure S10: Temporal patterns analysis showing: (A) Seasonal variations in readmission rates; (B) Monthly patterns; (C) Day of week patterns; (D) Weekend vs. weekday comparison; (E) Time to readmission distribution; (F) Yearly trends.

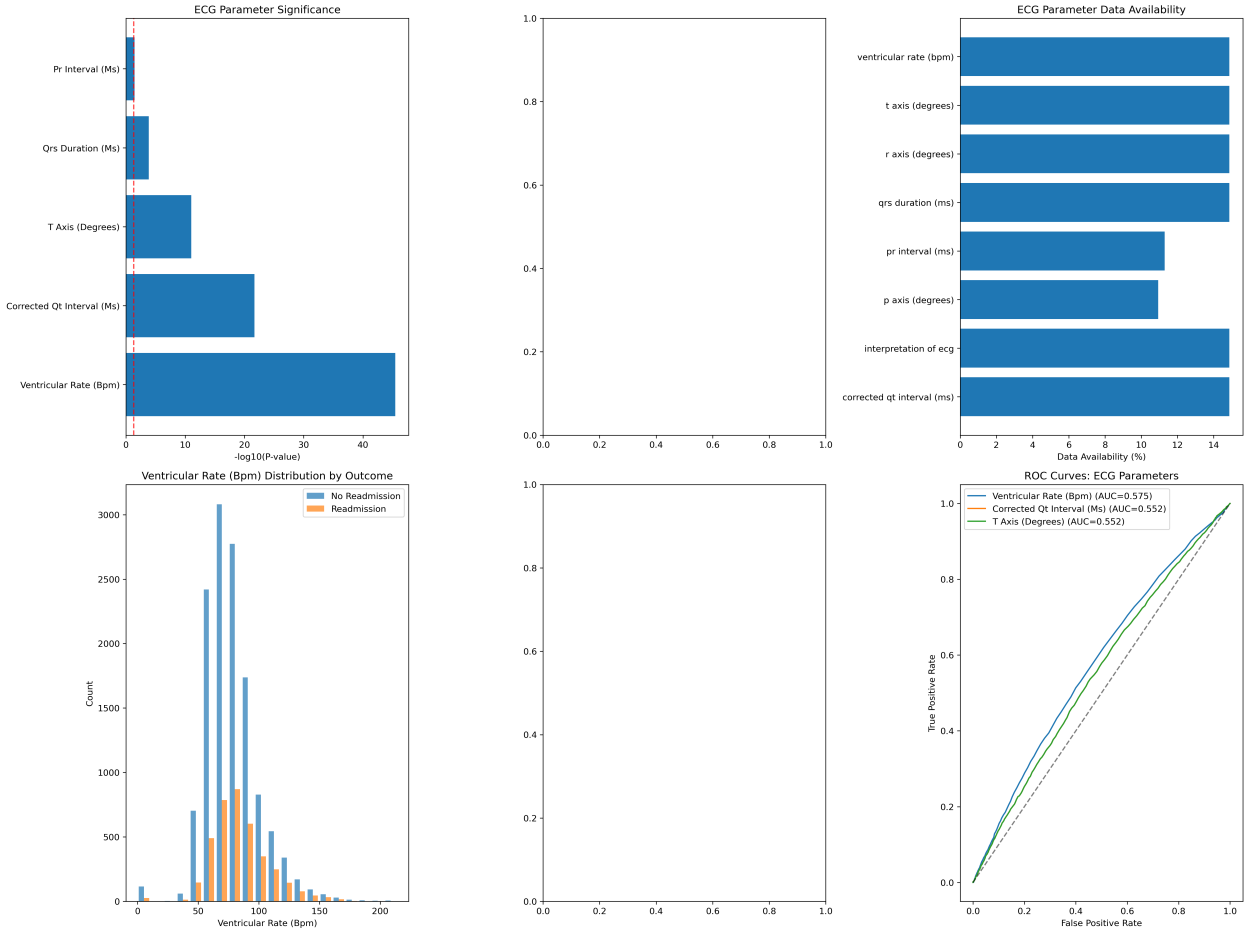


Figure S11: ECG findings analysis showing: (A) ECG parameter significance; (B) ECG abnormality burden effect on readmission; (C) ECG data availability; (D) Distribution of key ECG parameters; (E) ECG abnormality prevalence; (F) ROC curves for ECG parameters.

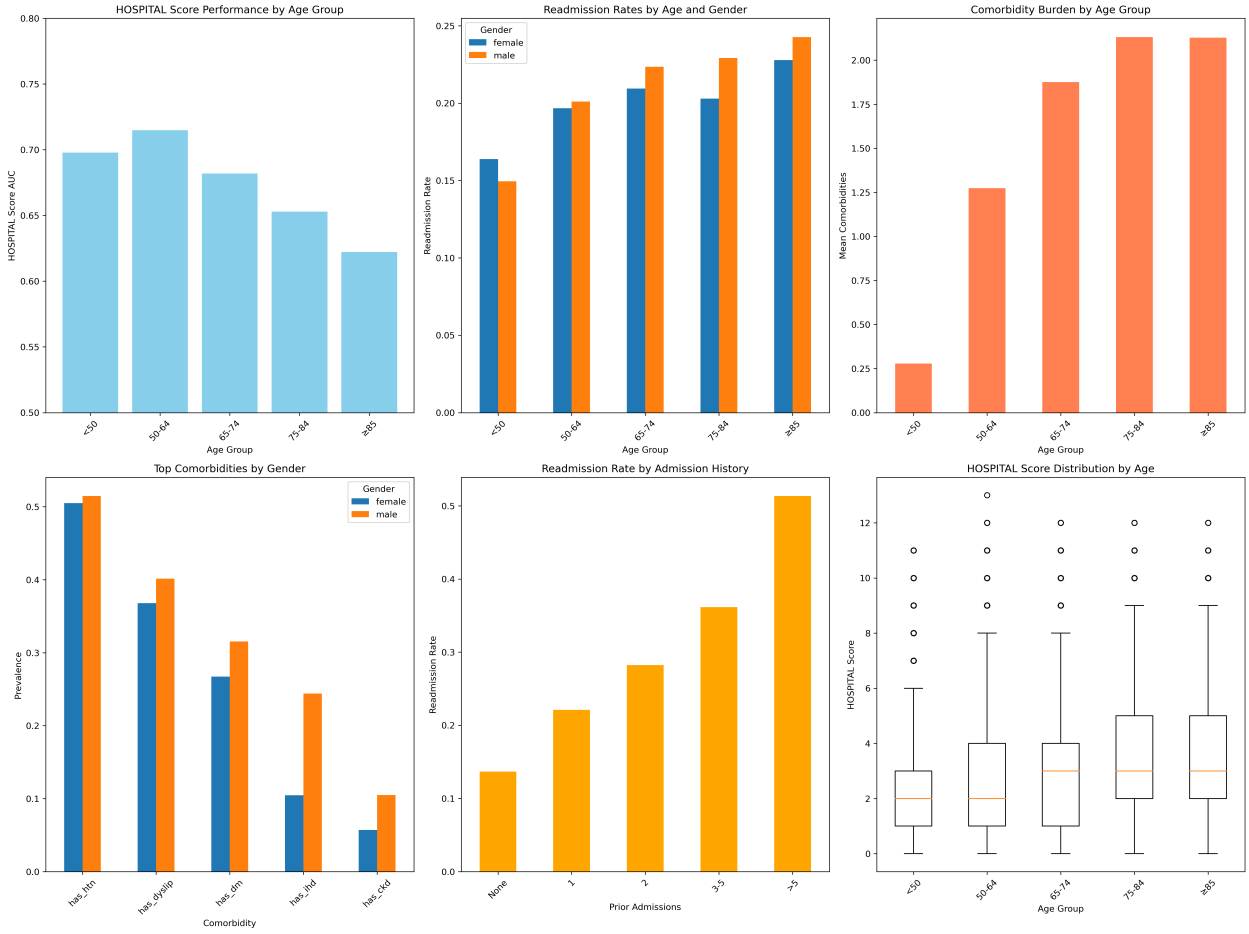


Figure S12: Subgroup analysis showing: (A) HOSPITAL score performance by age group; (B) Readmission rates by age and gender; (C) Comorbidity burden by age group; (D) Gender-specific comorbidity patterns; (E) Prior admission effect on readmission; (F) HOSPITAL score distribution by age.

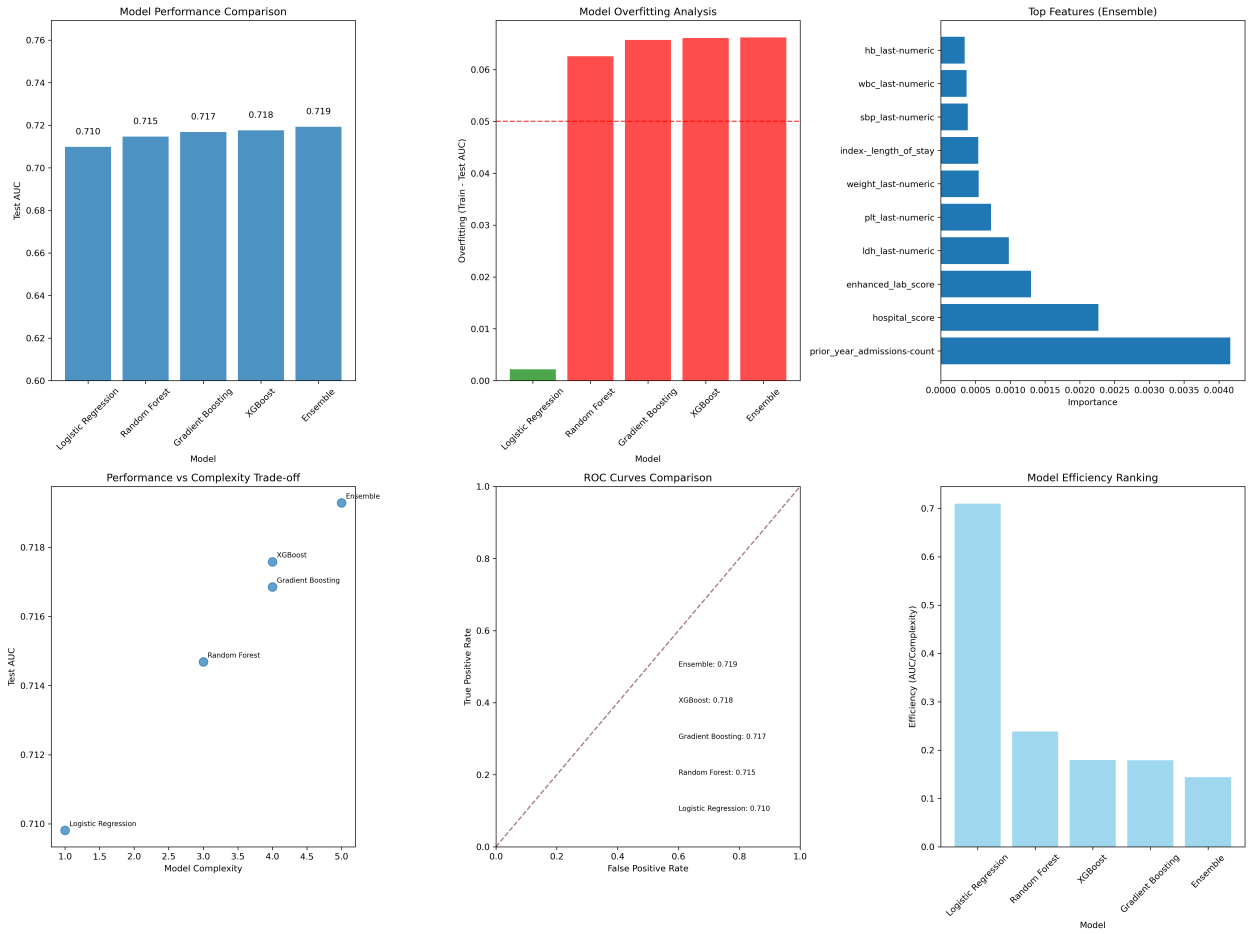


Figure S13: Machine learning comparison showing: (A) Model performance comparison; (B) Overfitting analysis; (C) Feature importance for best model; (D) Performance vs. complexity trade-off; (E) ROC curves comparison; (F) Model efficiency ranking.

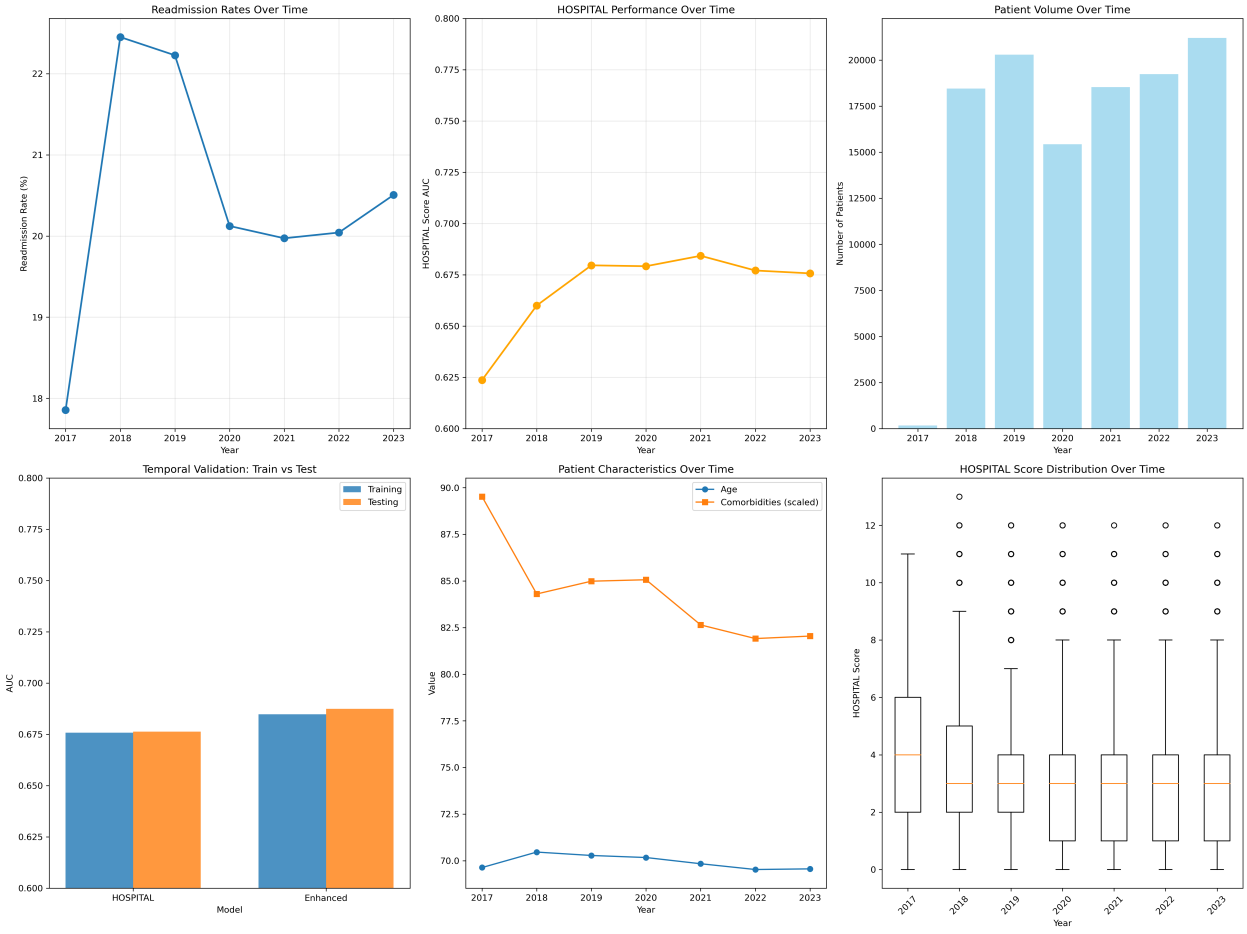


Figure S14: Temporal validation showing: (A) Yearly readmission rates; (B) HOSPITAL performance over time; (C) Patient volume over time; (D) Temporal train/test performance comparison; (E) Covariate shifts over time; (F) HOSPITAL score distribution over time.

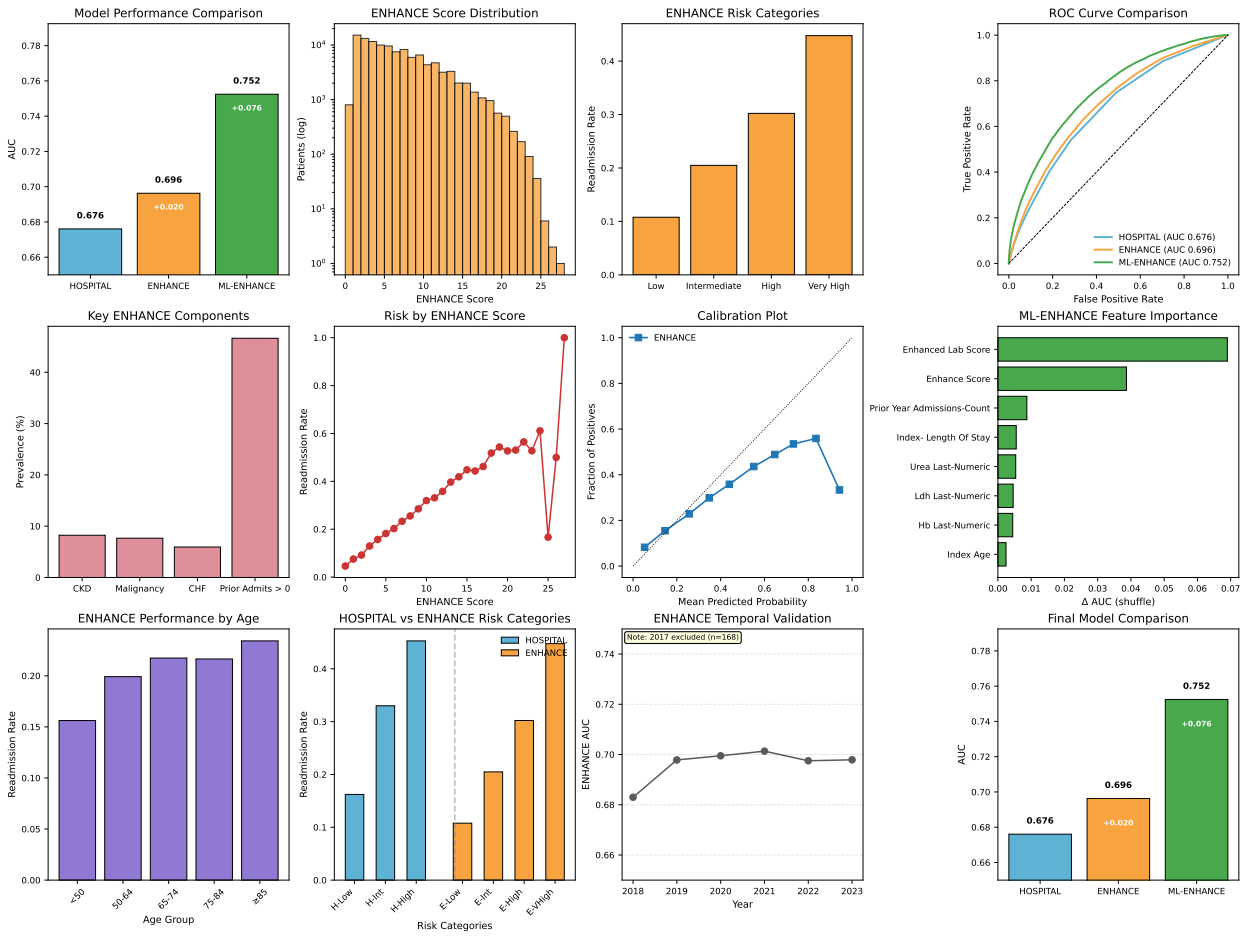
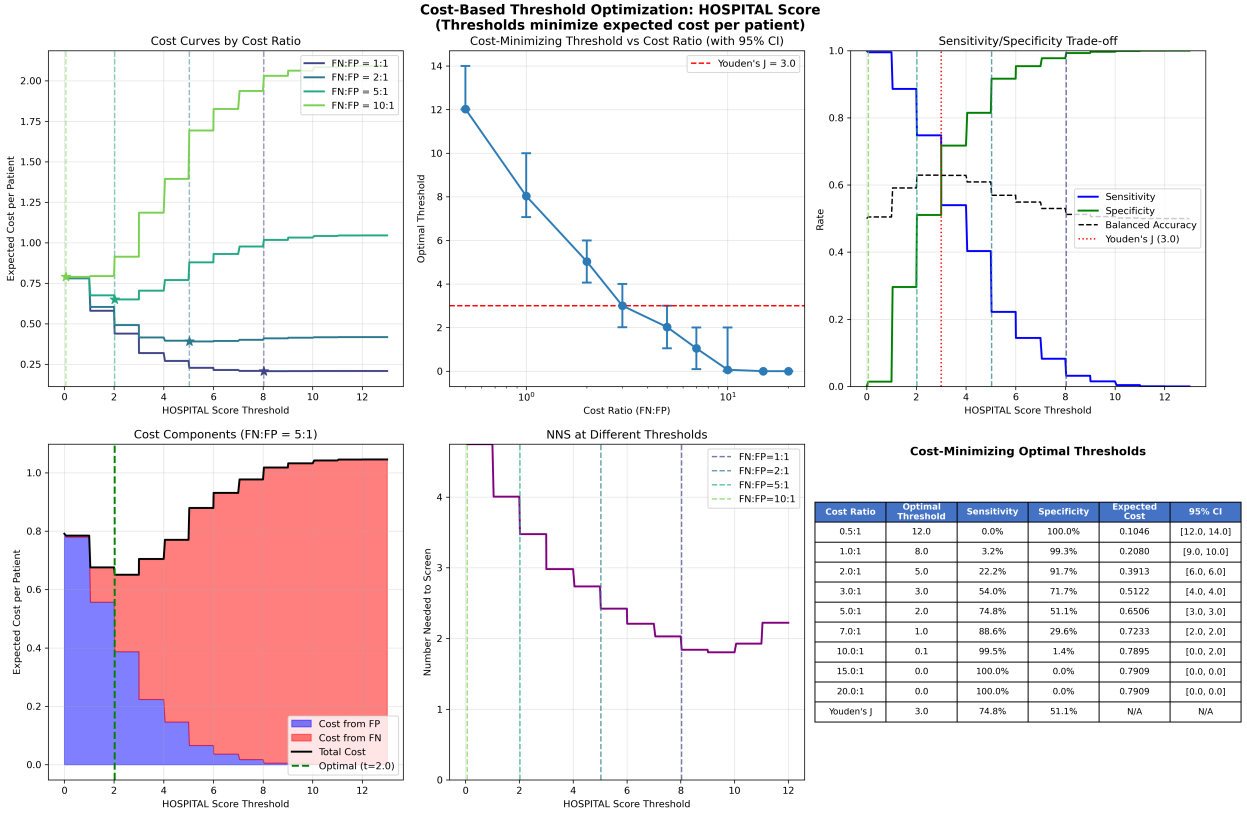
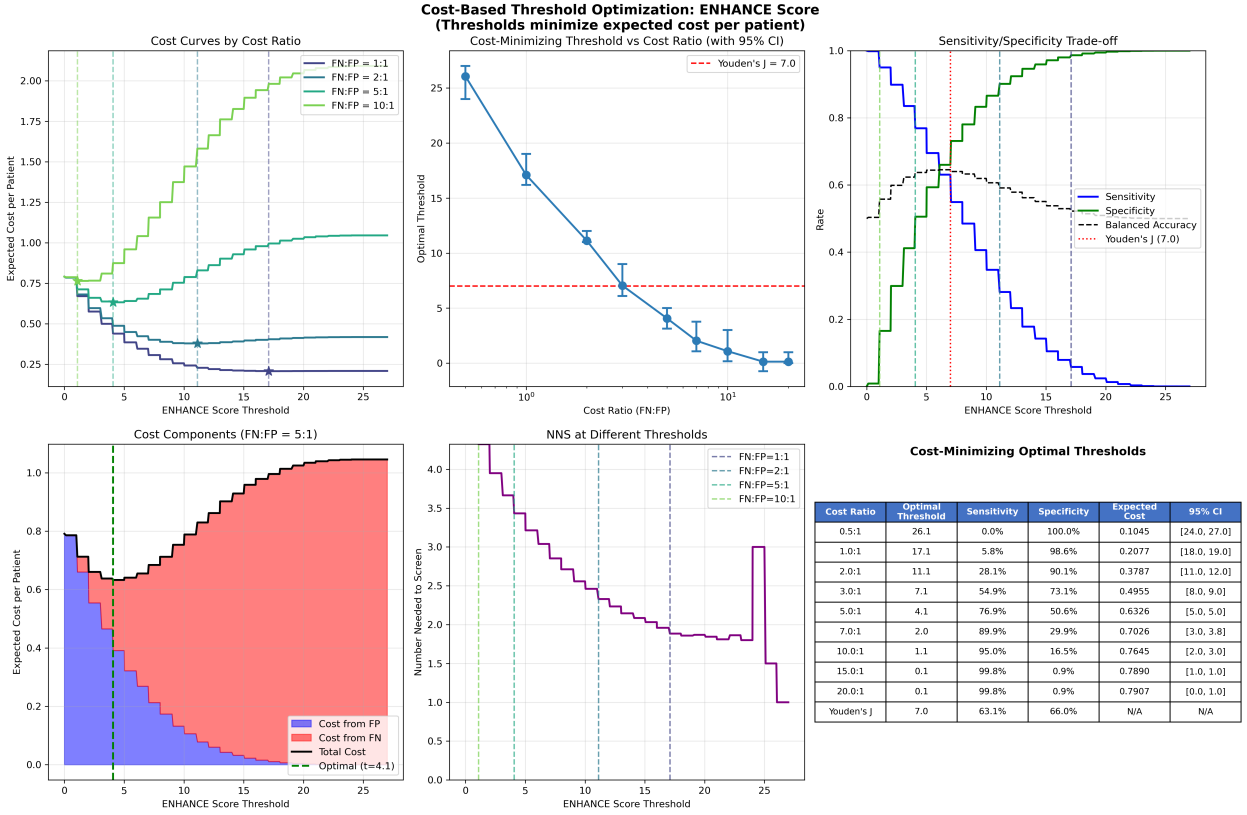


Figure S15: Comprehensive ENHANCE score analysis showing: (A) Model performance comparison; (B) ENHANCE score distribution; (C) Risk categories; (D) ROC curve comparison; (E) Key components; (F) Risk by score; (G) Calibration plot; (H) ML-ENHANCE feature importance; (I) Performance by age; (J) HOSPITAL vs ENHANCE by score range; (K) Temporal stability; (L) Final model comparison.



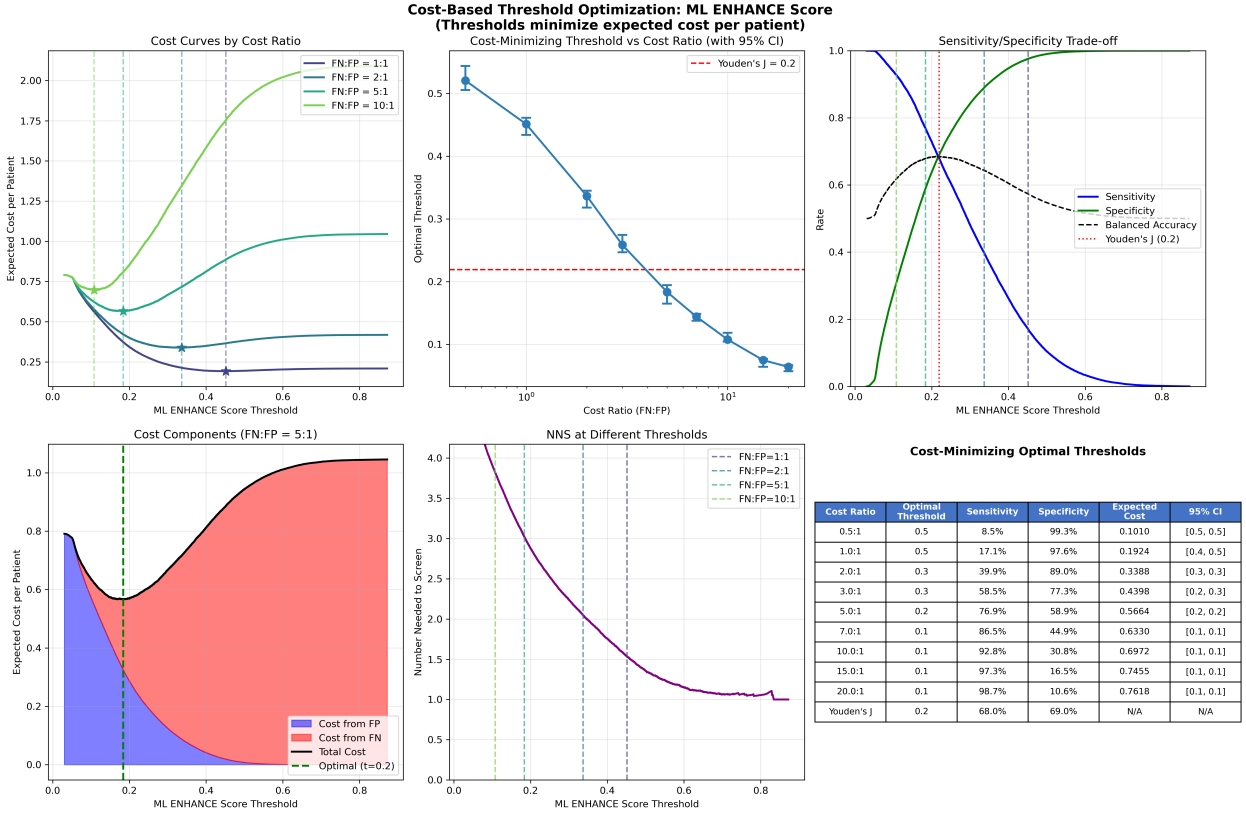
Cost Ratio	Optimal Threshold	Sensitivity	Specificity	Expected Cost	95% CI
0.5:1	12.0	0.0%	100.0%	0.1046	[12.0, 14.0]
1.0:1	8.0	3.2%	99.3%	0.2080	[9.0, 10.0]
2.0:1	5.0	22.2%	91.7%	0.3913	[6.0, 6.0]
3.0:1	3.0	54.0%	71.7%	0.5122	[4.0, 4.0]
5.0:1	2.0	74.8%	51.1%	0.6506	[3.0, 3.0]
7.0:1	1.0	88.6%	29.6%	0.7233	[2.0, 2.0]
10.0:1	0.1	99.5%	1.4%	0.7895	[0.0, 2.0]
15.0:1	0.0	100.0%	0.0%	0.7909	[0.0, 0.0]
20.0:1	0.0	100.0%	0.0%	0.7909	[0.0, 0.0]
Youden's J	3.0	74.8%	51.1%	N/A	N/A

Figure S16: HOSPITAL score threshold optimization analysis. Panel A: cost curves by cost ratio with optimal thresholds (stars). Panel B: cost-minimizing threshold vs. cost ratio with 95% CIs; dashed line indicates Youden's J threshold (3.0). Panel C: sensitivity-specificity trade-off. Panel D: cost decomposition at 5:1 ratio. Panel E: number needed to screen. Panel F: summary table with 95% CIs.



Cost Ratio	Optimal Threshold	Sensitivity	Specificity	Expected Cost	95% CI
0.5:1	26.1	0.0%	100.0%	0.1045	[24.0, 27.0]
1.0:1	17.1	5.8%	98.6%	0.2077	[18.0, 19.0]
2.0:1	11.1	28.1%	90.1%	0.3787	[11.0, 12.0]
3.0:1	7.1	54.9%	73.1%	0.4955	[8.0, 9.0]
5.0:1	4.1	76.9%	50.6%	0.6326	[5.0, 5.0]
7.0:1	2.0	89.9%	29.9%	0.7026	[3.0, 3.8]
10.0:1	1.1	95.0%	16.5%	0.7645	[2.0, 3.0]
15.0:1	0.1	99.8%	0.9%	0.7890	[1.0, 1.0]
20.0:1	0.1	99.8%	0.9%	0.7907	[0.0, 1.0]
Youden's J	7.0	63.1%	66.0%	N/A	N/A

Figure S17: ENHANCE score threshold optimization analysis. Panel A: cost curves by cost ratio with optimal thresholds (stars). Panel B: cost-minimizing threshold vs. cost ratio with 95% CIs; dashed line indicates Youden's J threshold (7.0). Panel C: sensitivity-specificity trade-off. Panel D: cost decomposition at 5:1 ratio. Panel E: number needed to screen. Panel F: summary table with 95% CIs.



Cost Ratio	Optimal Threshold	Sensitivity	Specificity	Expected Cost	95% CI
0.5:1	0.5	8.5%	99.3%	0.1010	[0.5, 0.5]
1.0:1	0.5	17.1%	97.6%	0.1924	[0.4, 0.5]
2.0:1	0.3	39.9%	89.0%	0.3388	[0.3, 0.3]
3.0:1	0.3	58.5%	77.3%	0.4398	[0.2, 0.3]
5.0:1	0.2	76.9%	58.9%	0.5664	[0.2, 0.2]
7.0:1	0.1	86.5%	44.9%	0.6330	[0.1, 0.1]
10.0:1	0.1	92.8%	30.8%	0.6972	[0.1, 0.1]
15.0:1	0.1	97.3%	16.5%	0.7455	[0.1, 0.1]
20.0:1	0.1	98.7%	10.6%	0.7618	[0.1, 0.1]
Youden's J	0.2	68.0%	69.0%	N/A	N/A

Figure S18: ML-ENHANCE score threshold optimization analysis. Panel A: cost curves by cost ratio with optimal thresholds (stars). Panel B: cost-minimizing threshold vs. cost ratio with 95% CIs; dashed line indicates Youden's J threshold (0.22). Panel C: sensitivity-specificity trade-off. Panel D: cost decomposition at 5:1 ratio. Panel E: number needed to screen. Panel F: summary table with 95% CIs.

[3] Zuckerman, R. B., Sheingold, S. H., Orav, E. J., Ruhter, J. & Epstein, A. M. Readmissions, observation, and the hospital readmissions reduction program. *New England Journal of Medicine* **374**, 1543–1551 (2016).

[4] Donzé, J., Lipsitz, S., Bates, D. W. & Schnipper, J. L. Potentially avoidable 30-day hospital readmissions in medical patients: derivation and validation of a prediction model. *JAMA* **309**, 372–379 (2013).

[5] Futoma, J., Morris, J. & Lucas, J. A comparison of models for predicting early hospital readmissions. In *AMIA Annual Symposium Proceedings*, vol. 2015, 1899–1908 (American Medical Informatics Association, 2015).

[6] Rajkomar, A. *et al.* Scalable and accurate deep learning with electronic health records. *npj Digital Medicine* **1**, 18 (2018).

[7] Debt.org. Hospital and surgery costs (2024). URL `https://www.debt.org/medical/hospital-surgery-costs/`. Accessed: 2025-01-15.

[8] Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement. *British Journal of Surgery* **102**, 148–158 (2015).

[9] Quan, H. *et al.* Coding algorithms for defining comorbidities in icd-9-cm and icd-10 administrative data. *Medical Care* **43**, 1130–1139 (2005).