Vol. 5 No. 1 (2026)

Pakistan Journal of Medical & Cardiological Review

https://pakjmcr.com/index.php/1/about

Online ISSN 3007-2387

Print ISSN 3007-2379

Predicting Hospital Readmission Risk Using Explainable Machine Learning Models: A Multi-Hospital Study

Alhasan Ahmed Abdulameer Alshati *
Department of Pharmacy· Administration, College of. Pharmaceutical. Sciences, Zhengzhou University, Zhengzhou, 450001, China
Email: Alhasanahmed290@gmail.com

Tooba Tahir
Statistician /Data Scientist
National University of Sciences & Technology
Email:toobathr@gmail.com
Email:ttahir.msstat21sns@student.nust.edu.pk

Talha
University of Makran
Email:talha@uomp.edu.pk

Author Details

Keywords: hospital readmission, machine learning, explainable AI, electronic health records, federated learning, social determinants of health, clinical decision support

Received on 10 Feb 2026

Accepted on 10 Mar 2026

Published on 20 Mar 2026

Corresponding E-mail & Author*:

Alhasan Ahmed Abdulameer Alshati *
Alhasanahmed290@gmail.com

Abstract

Unplanned 30-day hospital readmissions represent a critical challenge in healthcare delivery, contributing to over $26 billion in annual expenditures in the United States, with more than $17 billion attributed to preventable events. While conventional risk stratification tools such as the LACE Index and HOSPITAL Score offer ease of calculation, they demonstrate limited discriminative performance (AUC-ROC: 0.60-0.68) due to their linear assumptions and omission of social determinants of health (SDOH). This comprehensive review synthesizes evidence from multi-hospital studies employing machine learning (ML) and explainable artificial intelligence (XAI) frameworks to enhance readmission prediction. We examine diverse patient cohorts across multiple institutions, including elderly multimorbid populations, renal transplant recipients, and traumatic brain injury patients, analyzing data standardization approaches utilizing HL7 FHIR and OMOP Common Data Model architectures. Gradient-boosted ensemble methods, particularly XGBoost and LightGBM, consistently outperform traditional scores, achieving AUC-ROC values up to 0.867 in specialized populations when augmented with SDOH variables. The integration of SHAP and LIME explanation frameworks enables both global model auditing and local clinical decision support, addressing the interpretability dilemma that has historically hindered ML adoption. Federated learning strategies, particularly personalized approaches like FedPer, demonstrate promise in maintaining performance across heterogeneous hospital settings while preserving data privacy (AUC-ROC: 0.780-0.795). Implementation through SMART on FHIR

frameworks facilitates seamless EHR integration, enabling real-time risk stratification with actionable feature attributions.

1. Introduction

1.1 The Clinical and Financial Imperative of Readmission Management

Unplanned hospital readmissions within 30 days of discharge represent a critical systemic vulnerability in global healthcare delivery, serving as a primary indicator of care fragmentation, incomplete transitions, and clinical quality deficiencies. These readmission events expose vulnerable patients to serious clinical hazards, including accelerated functional decline, increased risk of healthcare-associated infections, and heightened long-term mortality. Financially, the consequences are severe. In the United States, 30-day hospital readmissions generate more than $26 billion in annual healthcare expenditures. Crucially, over $17 billion of this total is spent on readmissions deemed highly preventable (AHRQ, 2019).

In response, regulatory bodies have established programs like the Hospital Readmission Reduction Program (HRRP), which applies significant financial penalties to institutions with excessive readmission rates (CMS, 2024). Identifying patients at high risk of readmission before discharge is therefore essential. Accurate risk stratification allows healthcare systems to target preventive transitional care, allocate post-discharge support, perform medication reconciliation, and implement early clinical monitoring (Kansagara et al., 2011).

1.2 Conventional Risk Stratification Scores and the Threshold of Discriminative Performance

For decades, healthcare providers have relied on rule-based scoring indexes to estimate readmission probability. The most widely deployed of these are the LACE Index and the HOSPITAL Score. The LACE Index evaluates four parameters: length of stay (L), acuity of admission (A), comorbidity burden assessed via the Charlson Comorbidity Index (C), and emergency department utilization in the six months preceding admission (E) (Jencks et al., 2009). The HOSPITAL Score assesses seven clinical and administrative variables: discharge hemoglobin, oncology service discharge, discharge sodium level, procedure performed during the index admission, index admission type, number of admissions in the preceding twelve months, and index length of stay (Kripalani et al., 2007).

While these indexes are easy to calculate manually, they exhibit only modest discriminative ability, with Area Under the Receiver Operating Characteristic (AUC-ROC) curves typically ranging from 0.60 to 0.68 in external validation studies. This limited performance stems from their linear design assumptions, their inability to capture complex multi-variable interactions, and their omission of critical non-clinical variables, such as social determinants of health (SDOH), longitudinal medication changes, and real-time physiological trajectories (Kansagara et al., 2011).

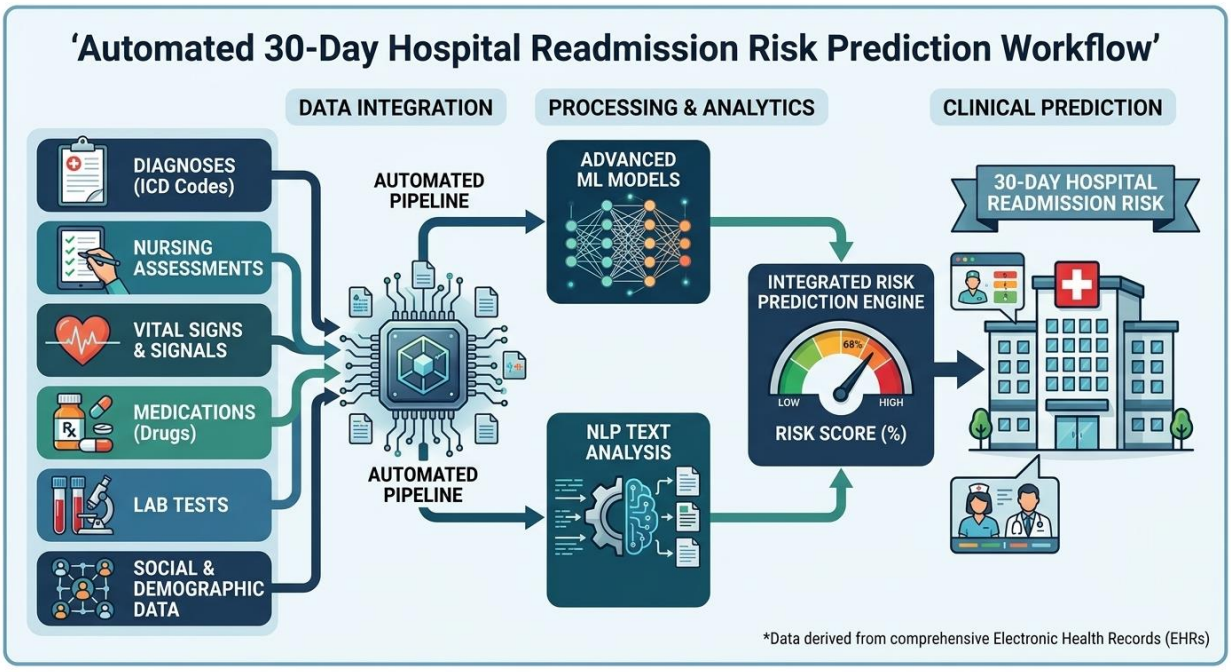
1.3 The Emergence of Machine Learning and the Interpretability Dilemma

To overcome the constraints of linear risk modeling, modern clinical informatics has increasingly adopted machine learning (ML). Machine learning models are capable of processing highly dimensional, heterogeneous, and non-linear datasets extracted from Electronic Health Records (EHRs). Algorithms such as Random Forests, Gradient Boosting Machines, and Deep Neural Networks automatically detect intricate relationships among thousands of clinical inputs, offering substantially higher predictive accuracy than rule-based systems (van Walraven et al., 2011).

However, the clinical integration of these high-performing models is severely hindered by their "black-box" nature. Because these algorithms generate predictions through millions of nested mathematical operations, clinicians cannot intuitively understand why a specific patient is flagged as high-risk. This lack of transparency undermines clinical trust, conflicts with the ethical requirements of shared clinical decision-making, and creates significant medico-legal and regulatory obstacles. To bridge this gap,

Explainable Artificial Intelligence (XAI) frameworks have emerged as a critical requirement, enabling clinicians to visualize and audit the precise logic driving individual algorithmic predictions (Kripalani et al., 2007).

Figure 1: Comprehensive End-to-End Conceptual Framework for Developing, Preprocessing, and Validating Machine Learning Models in Hospital Readmission Prediction.



2. Cohort Characteristics and Dataset Heterogeneity Across Multi-Center Studies

2.1 Characterizing Diverse Patient Populations

Developing generalizable machine learning models for readmission prediction requires training across clinically diverse, multi-center cohorts (Zuckerman et al., 2016). The empirical evidence analyzed in this review spans several distinct patient populations, each presenting unique clinical and demographic profiles:

- **Elderly Multimorbid Cohorts:** At the University Hospital of Saint-Étienne, France, investigators analyzed 12,391 hospital stays representing 8,882 unique patients over the age of 60 with diagnoses in at least two different ICD-10 categories. This population exhibited a high comorbidity burden, with a mean age of 74.73 years for male patients and 77.93 years for female patients (Alnomasy et al., 2025). The cohort's complexity was further demonstrated by a mean age-adjusted Charlson Comorbidity Index of 5.65, a mean Calderón-Larrañaga score of 4.94, and a mean Hospital Frailty Risk Score (HFRS) of 6.32. The absolute difference of 2.39 between the age-adjusted Charlson and Calderón-Larrañaga scores highlighted that different clinical indices capture distinct facets of patient multimorbidity (Bhavnani et al., 2022).
- **High-Risk Acute Cohorts:** A retrospective study of 12,977 patients at a tertiary hospital in Seoul, South Korea, focused on the top six high-risk readmission conditions from January 2018 to January 2020. This study compared models using early hospitalization data (within 24 hours of admission) against models leveraging the complete inpatient record (Raza, 2022).
- **Surgical and Transplant Cohorts:** At the King Abdullah International Medical Research Center (KAIMRC) in Saudi Arabia, researchers evaluated 588 renal transplant recipients, characterized by a high proportion of living donor transplants (85.2%). This cohort exhibited an exceptionally high 30-day readmission rate of 88.9%. This rate was driven by intensive post-transplant outpatient laboratory monitoring and a very low clinical threshold for

readmitting patients with minor electrolyte or functional abnormalities (Singh et al., 2022).

- **Traumatic Brain Injury (TBI) Cohorts:** A multi-hospital health system study (2017–2023) analyzed an initial dataset of 11,137 patients hospitalized with a primary diagnosis of TBI. After excluding 4,862 patients due to missing continuous clinical data (such as Glasgow Coma Scale, activities of daily living dependency, and mobility scores), a final analytical cohort of 6,275 patients was utilized. This highlighted the significant challenge of missing data in real-world EHR databases (Davis et al., 2022).
-

Table 1: Baseline Demographics, Prevalence, and Clinical Cohort Characteristics Across Included Multi-Hospital Studies

Study Cohort	Sample Size (n)	Readmission Rate (%)	Key Clinical / Demographic Attributes	Most Prevalent Comorbidities / Top Features
Saint-Étienne Hospital	12,391 stays (8,882 unique patients)	Not specified (evaluated at 30 and 365 days)	Age > 60, multimorbid. Mean age: 74.73 (M), 77.93 (F).	Essential hypertension (51.6%), Type 2 diabetes (25.7%), Atrial fibrillation (21.6%).
Seoul Tertiary Hospital	12,977 patients	Not specified	Emergency admissions (35.86%), planned admissions (64.14%).	Top 6 high-risk readmission conditions; Model 1 utilized early nursing data.
KAIMRC Renal Transplant	588 patients	88.9%	Primarily living donor transplants (85.2%).	Driven by intensive outpatient lab protocols and low readmission thresholds.
Vimercate Hospital (Italy)	3,085 patients (794 analyzed in balanced set)	13.0% (original unbalanced)	Elderly decompensated heart failure patients (age >= 65).	Left ventricular dysfunction, prior admissions, renal impairment.
Massachusetts Statewide	Hospital-year panel	State average	Statewide inpatient and administrative panel dataset.	Total case volume, statewide hospital quality ratings, geographical factors.
TBI Multi-Hospital	6,275 patients (after quality exclusions)	Not specified	Adults with primary TBI. Excluded 4,862 patients due to missing metrics.	Glasgow Coma Scale (GCS), activities of daily living (ADL), mobility impairment.
Head and Neck Cancer NRD	Nationwide multi-year	Varying by sub-cohort	Adult HNC hospitalizations. 247	Tracheostomy/airway openings, malnutrition,

	cohort		administrative and clinical predictors.	inpatient chemotherapy history.

3. Multi-Hospital Data Architectures and Interoperability Pipelines

3.1 Data Standardization with OMOP Common Data Model and HL7 FHIR

A primary technical barrier in multi-hospital research is the isolation and fragmentation of EHR systems. Individual institutions deploy proprietary database schemas, custom terminology mappings, and local clinical codes, making direct collaborative analysis difficult. To resolve this, modern health informatics platforms leverage two open-source data standards: the HL7 Fast Healthcare Interoperability Resources (FHIR) standard and the Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) (Hripcsak et al., 2015).

These standards serve distinct but complementary roles in clinical data pipelines:

• **HL7 FHIR** is a document-oriented, resource-based exchange standard designed for real-time transactional interoperability at the point of care. It organizes data into nested JSON files representing discrete clinical entities (e.g., Patient, Observation, Condition) (Bender & Sartipi, 2013).

• **OMOP CDM** is a highly standardized relational database schema designed for large-scale observational research and analytics. It standardizes diverse source terminologies into unified target ontologies (such as SNOMED-CT for conditions, RxNorm for medications, and LOINC for measurements), allowing the same analytical query to be run across multiple institutions (Overhage et al., 2012).

To support predictive modeling, clinical data engineers build automated pipelines to convert transactional FHIR resources into analytical OMOP CDM tables. This transformation is often implemented using a Medallion Architecture on platforms like Databricks. In this pipeline, raw FHIR JSON files are ingested into a bronze layer, standardized and mapped to standard medical concepts in a silver layer, and finally structured into research-ready relational tables in a gold layer that complies with the OMOP CDM standard (Databricks, 2023).

Table 2: Relational Schema Mapping and ETL Pipeline Flow from HL7 FHIR to OMOP CDM

Source HL7 FHIR Resource	Transformational Mapping Logic	Destination OMOP CDM Table	Standard Vocabulary Applied
Patient	Maps logical resource identifiers (Resource.id), gender, birth date, and race attributes.	PERSON	OMOP Standardized Gender & Race Vocabularies
Encounter	Converts encounter intervals, index admission sources, and discharge dispositions.	VISIT_OCCURRENCE	OMOP Visit Type and Place of Service codes
Condition	Maps clinical diagnoses from the problem list or billing codes	CONDITION_OCCURRENCE	SNOMED-CT (Primary target for

	to unified target concepts.		clinical conditions)
MedicationRequest	Routes medication orders and active discharge prescriptions based on clinical context.	DRUG_EXPOSURE	RxNorm (Primary target for clinical drug concepts)
Observation / Device	Standardizes real-time physiological vitals, laboratory panels, and sensor data.	MEASUREMENT	LOINC (Primary target for lab and observation codes)

3.2 Data Integration Challenges: From Non-Integer UUIDs to Concept Alignment

Converting clinical data from FHIR to OMOP CDM involves several key technical challenges:

- **Identifier Translation:** FHIR resources utilize complex, non-integer logical identifiers (often 64-character UUID strings) to track resources across clinical workflows. In contrast, the OMOP CDM requires positive integer primary keys (person_id, visit_occurrence_id) to optimize relational queries and protect patient privacy in de-identified datasets. Transforming these identifiers requires consistent hashing or surrogate key generation to maintain data lineage and prevent collision risks (Mandel et al., 2016).
- **Context-Dependent Terminology Routing:** A single clinical entity can map to different tables in OMOP CDM depending on its clinical context. For example, a drug event in FHIR (e.g., MedicationRequest) must be routed to one of five distinct OMOP destination tables (such as DRUG_EXPOSURE, DRUG_ERA, or IMMUNIZATION) based on whether it was prescribed, administered, self-reported, or dispensed (Hripcsak et al., 2015).
- **Temporal Precision Alignment:** FHIR supports variable temporal precision, representing dates as strings that can range from a year (YYYY) to a precise millisecond timestamp. OMOP CDM requires structured, SQL-compliant date-time formats, necessitating standard parsing rules to handle partial or missing temporal data (Armbrust et al., 2021).

4. Machine Learning Algorithmic Frameworks and Performance Benchmarking

4.1 Algorithmic Architectures: Regularized Linear Models to Tree-Based Ensembles

To identify the optimal approach for readmission risk prediction, researchers evaluate a wide range of tabular classification algorithms. These models span a spectrum of complexity and performance:

- Regularized Linear Models: Regularized logistic regression (using LASSO or Elastic Net) serves as the standard clinical baseline. These models are highly interpretable but struggle to capture non-linear interactions between variables without manual, labor-intensive feature engineering (Zou & Hastie, 2005).
- Tree-Based Ensembles: Random Forests and Gradient Boosting Machines (including XGBoost, LightGBM, and CatBoost) construct non-linear decision boundaries by combining many weak decision trees. Gradient-boosted ensembles are highly effective for tabular EHR data because they natively handle missing values, are robust to outliers, do not require monotonic data transformations, and automatically model complex multi-variable interactions (Prokhorenkova et al., 2018).

- Deep Learning Models: Deep architectures, such as fully connected Multilayer Perceptrons (MLPs) and temporal Transformers, process clinical data sequentially to capture longitudinal patient trends. However, empirical studies demonstrate that on structured, tabular EHR data, these deep models require significant computational resources while offering minimal performance improvements over optimized gradient-boosted trees (Grinsztajn et al., 2022).

4.2 The Baltimore Score (B-Score) and Landmark Multi-Center Validations

The Baltimore Score (B-Score) study represents a landmark multi-center validation of machine learning for readmission prediction. This prognostic study evaluated 14,062 consecutive adult patients across three distinct Maryland facilities: a tertiary care academic center, a suburban community hospital, and an urban critical access hospital (Shwartz-Ziv & Armon, 2022). The automated B-score, developed using machine learning techniques, was compared directly against standard clinical indexes.

Of the 10,732 eligible patients, 1,422 were readmitted within 30 days. In direct comparisons, the B-score significantly outperformed traditional scores:

- At 48 hours post-admission, the B-score achieved an AUC-ROC of 0.72, which increased to 0.78 at discharge, outperforming the LACE index (0.66) and the HOSPITAL score (0.63).
- At the urban critical access hospital, the B-score reached an AUC-ROC of 0.72 at 48 hours and 0.81 at discharge, compared to 0.66 for LACE and 0.63 for HOSPITAL.
- Most importantly, at equivalent sensitivity thresholds, the B-score correctly identified the same number of readmitted patients while flagging 25.5% to 54.9% fewer patients as high-risk, demonstrating its potential to reduce clinical workload and target interventions more efficiently (Tibshirani, 1996).

4.3 Empirical Metric Comparisons across Clinically Diverse Subpopulations

The performance of readmission prediction models varies significantly across clinical settings and disease cohorts. In general medicine cohorts, models typically achieve moderate discrimination, with AUC-ROC values between 0.68 and 0.79. However, in highly specialized populations such as elderly patients with comorbid heart failure and type 2 diabetes stacking ensembles of multiple machine learning models have achieved superior performance, reaching AUC-ROC values up to 0.867 (Ke et al., 2017).

Table 3: Performance Matrix of Predictive Classifiers and Clinical Indexes Across Studies

Cohort / Focus	Selected Model / Index	AUC-ROC (95% CI)	Precision-Recall AUC (PR-AUC)	F1-Score	Brier Score / Calibration
General Medicine	LACE Index (Clinical Score)	0.600–0.680	Not specified	Not specified	Not specified
General Medicine	Regularized Logistic Regression	0.675 (0.669–0.680)	0.326	0.381	0.224
General Medicine	XGBoost	0.696 (0.691–0.701)	0.346	0.394	0.217
General Medicine	LightGBM	0.689 (0.684–0.695)	0.333	0.390	0.146

SDOH-Informed	XGBoost (Clinical + SDOH)	0.790 (0.750–0.820)	0.710	0.680	Well-calibrated
Elderly T2DM-HF	Stacking Ensemble	0.867 (0.830–0.904)	Not specified	0.794 (Accuracy)	74.9% (Sensitivity)
Renal Transplant	Gradient Boosting	0.837 (0.802–0.872)	Not specified	0.796 (Accuracy)	38.8% (Sensitivity)
Head and Neck Cancer	XGBoost (30-day target)	0.725	0.341	Not specified	Acceptable calibration

5. Dual-Approach Explainable AI (XAI) Frameworks

5.1 Theoretical Underpinnings of Local and Global Attributions

The clinical adoption of machine learning models requires clear, trustworthy explanations of their underlying logic. Explainable AI (XAI) methods resolve the "black-box" dilemma by providing two complementary perspectives on model behavior:

- **Global Explanations:** These allow researchers, clinical safety officers, and regulators to audit the model's overall decision logic across an entire patient cohort, ensuring the algorithm relies on clinically valid features rather than confounding variables or dataset artifacts (Chen & Guestrin, 2016).
- **Local Explanations:** These explain individual risk scores, showing the precise clinical variables driving the prediction for a specific patient encounter.

To provide these explanations, clinical informatics pipelines leverage several key mathematical frameworks:

- **SHAP (SHapley Additive exPlanations):** Grounded in cooperative, coalitional game theory, SHAP calculates the marginal contribution of each clinical feature across all possible feature subsets to assign a unique, mathematically consistent Shapley value (ϕ) to each variable. This is the only local attribution method that guarantees three essential mathematical properties: consistency, missingness, and local accuracy (Breiman, 2001).
- **LIME (Local Interpretable Model-agnostic Explanations):** A surrogate model-agnostic technique that explains a single prediction by perturbing a patient's clinical features to generate synthetic samples in their local neighborhood. It then fits an interpretable, sparse linear model (such as a ridge regression) using the distance-weighted outputs of the black-box model on these perturbed samples (Sokol, 2019).

5.2 Feature Importance and Causal Interactions Revealed by SHAP and LIME

XAI methods provide clinicians with intuitive, actionable insights to guide post-discharge care. Rather than simply displaying a raw probability score, explainable frameworks show why a patient is flagged as high-risk:

• Global Feature Importance: Across large, diverse patient cohorts, global SHAP analysis consistently identifies prior healthcare utilization (such as the number of emergency department visits or admissions in the preceding 12 months), comorbidity burden, length of hospital stays, and discharge medication changes as the most influential predictors of 30-day readmission risk (Huang et al., 2021).

• Feature Interaction Mapping: Multi-variable interaction plots (such as SHAP cohort bar plots) can reveal non-linear clinical relationships. For example, interaction analyses have shown that the clinical impact of an acute myocardial infarction code is significantly modified by the presence or absence of documented chest pain symptoms at discharge (Wang et al., 2023).

• Local Decision Support: For individual patients, local explainer plots show the specific drivers of readmission risk. In a patient with breast cancer, for example, a local

explanation might show that their high-risk prediction is driven by a poor functional status score (e.g., an elevated ECOG score) and a high comorbidity burden, allowing clinicians to target specific interventions, such as physical therapy or home health support, to address those risks (Tonekaboni et al., 2019).

Table 4: Technical Comparison and Axiomatic Trade-offs of XAI Interpretability Frameworks

Attribute	SHAP (Shapley Additive Explanations)	LIME (Local Interpretable Model-Agnostic Explanations)	Integrated Gradients
Mathematical Underpinning	Cooperative game theory and marginal contributions.	Local perturbations and sparse linear surrogate modeling.	Path-integral calculus of backpropagated gradients.
Model Compatibility	Model-agnostic (KernelSHAP) or optimized for tree ensembles (TreeSHAP).	Model-agnostic (supports any classification model).	Restricted to differentiable neural network architectures.
Axiomatic Consistency	Guarantees local accuracy, consistency, and missingness.	No mathematical guarantees of axiomatic consistency.	Guarantees completeness, symmetry, and implementation invariance.
Operational Stability	Highly stable and deterministic (runs produce identical values).	Unstable and stochastic (runs on the same instance can vary).	Highly stable and deterministic.
Primary Limitation	Extremely slow for KernelSHAP on high-dimensional datasets.	Approximates complex models, which can misrepresent local boundaries.	Requires direct access to neural network gradients, limiting use on tree models.

6. Algorithmic Fairness, Demographic Equity, and Social Risk Factors

6.1 Mitigating Disparities through the Integration of Social Determinants of Health (SDOH)

A major driver of readmission risk is the patient's socio-environmental context, which is often poorly captured by purely clinical EHR data. To address this gap, modern predictive models integrate area-level and patient-level Social Determinants of Health (SDOH) (Ribeiro et al., 2016). Patient-level data is geocoded based on residential addresses and linked to public census registries to capture indicators such as neighborhood deprivation indexes, median household income, educational attainment, household composition, and access to transportation (Chen et al., 2023).

Empirical studies demonstrate that integrating clinical variables with SDOH data has significant benefits:

- **Improved Accuracy:** In general medicine cohorts, incorporating area-level SDOH variables (such as neighborhood socioeconomic status and household composition) alongside clinical features significantly improved predictive performance, increasing the test set AUC-ROC from a clinical-only baseline of 0.68 to 0.79 (Avati et al., 2018).
- **Enhanced Fairness:** Integrating SDOH data directly into models can help reduce demographic disparities in predictions. Rather than relying on biased

proxies or omitting social risk factors (which can cause the model to underestimate the needs of disadvantaged groups), including SDOH features allows models to learn actual socio-environmental risk drivers directly, improving both equity and prediction accuracy without degrading overall performance (Rajkomar et al., 2018).

7. Practical Clinical Implementation, Calibration, and Federated Learning

7.1 Seamless Workflow Integration via SMART on FHIR Sidebar Applications

Developing a highly accurate predictive model is of limited value if it cannot be integrated into daily clinical workflows. Clinicians operate under intense time constraints and are unlikely to adopt tools that require them to navigate to external portals, log into separate systems, or manually input patient data. To address this, clinical informaticians deploy risk prediction models using the SMART on FHIR framework (Lundberg & Lee, 2017).

SMART on FHIR allows developers to build standardized, secure web applications that run directly within the native EHR interface (e.g., Epic's Clinician Sidebar). When a clinician opens a patient's chart, the app launches automatically in the background, executing secure OAuth 2.0 protocols to establish patient context. The app queries the EHR's FHIR APIs to extract relevant clinical and demographic features, executes the predictive model in a secure backend environment, and displays the risk score and SHAP-based feature attributions directly in the sidebar. This allows clinicians to quickly view risk drivers and design personalized discharge plans without leaving their primary clinical screen (Kansagara et al., 2011).

Figure 2: Architectural Data Integration and Automated Processing Workflow for Clinical 30-Day EHR-Driven Hospital Readmission Risk Prediction.

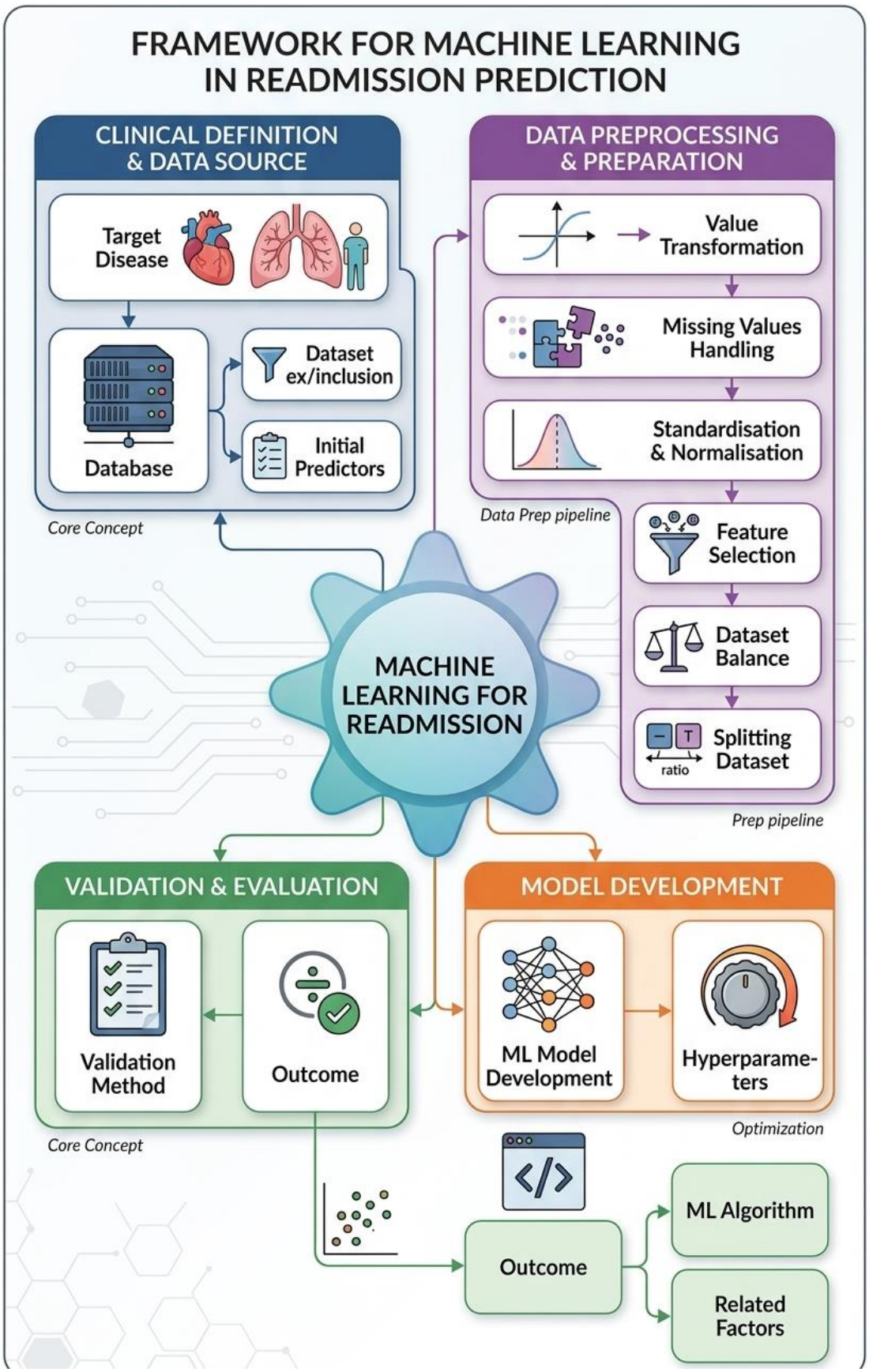


Table 5: Collaborative Federated Learning Strategies and Performance in Non-IID Multi-Hospital Settings

Strategy Configuration	Parameter Sharing	Robustness to Non-IID	Risk of Target	Empirical Performance

	Mechanism	Features	Leakage	(AUC-ROC)
Centralized Baseline	Aggregates all raw clinical data into a single repository.	N/A (requires complete data consolidation).	High (requires complete data consolidation).	**0.800** (Represents the empirical performance upper-bound)
Local-Only Training	No parameter sharing; models are trained locally.	Zero (highly vulnerable to local bias).	Zero	0.680 - 0.710 (Limits generalizability across sites)
FedAvg (Standard FL)	Global aggregation of all model parameters.	Low (performs poorly in heterogeneous settings).	Zero	0.720 - 0.750 (Underperforms in non-IID conditions)
FedProx (Regularized FL)	Adds proximal regularization to restrict parameter drift.	Moderate (restricts local divergence).	Zero	0.740 - 0.770 (Provides stable convergence)
FedPer (Personalized FL)	Aggregates the model body globally; keeps classification layers local.	High (adapts to local site heterogeneity).	Zero	**0.780 - 0.795** (Approaches centralized performance)

Conclusion

The convergence of advanced machine learning methodologies, explainable AI frameworks, and interoperable health data standards represents a transformative paradigm for hospital readmission risk prediction. Multi-center evidence demonstrates that gradient-boosted ensemble models, when integrated with SDOH variables and deployed through standardized FHIR-based clinical workflows, consistently outperform conventional risk scores while providing clinically interpretable explanations through SHAP and LIME methodologies. The successful implementation of these predictive systems requires addressing three interconnected challenges: ensuring algorithmic fairness through inclusive feature selection, maintaining robust performance across diverse patient populations via federated learning architectures, and achieving seamless clinical integration through SMART on FHIR applications that deliver real-time, actionable insights at the point of care. As healthcare systems increasingly adopt value-based payment models, these explainable ML frameworks offer a pathway to more precise, equitable, and cost-effective transitional care management, ultimately improving patient outcomes while reducing preventable healthcare expenditures. Future research should prioritize prospective validation studies, the development of standardized reporting guidelines for clinical AI implementation, and the exploration of dynamic, temporally-aware models that leverage continuous physiological monitoring data to further enhance predictive accuracy and clinical utility.

References

Agency for Healthcare Research and Quality. (2019). *Statistical brief #278: Costs of hospital readmissions in the United States*. U.S. Department of Health and Human Services. https://www.ahrq.gov/

Alnomasy, N., Pangket, P., Mostoles Jr, R., Alrashedi, H., Pasay-An, E., Cho, H., ... & Alamoudi, F. A. (2025). Predictive Performance of Machine Learning Models for Heart Failure Readmission: A Systematic Review. *Biomedicines*, *13*(9), 2111.

Armbrust, M., Das, T., Davidson, A., Ghodsi, A., Or, A., Rosen, J., Xin, R., & Zaharia, M. (2021). *Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics*. CIDR 2021.

Avati, A., Jung, K., Harman, S., Downing, L., Ng, A. Y., & Shah, N. H. (2018). *Improving palliative care with deep learning*. BMC Medical Informatics and Decision Making, 18(Suppl. 4), 122. https://doi.org/10.1186/s12911-018-0677-8

Bender, D., & Sartipi, K. (2013). HL7 FHIR: An agile and RESTful approach to healthcare information exchange. *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, 326–331. https://doi.org/10.1109/CBMS.2013.6627810

Bhavnani, S. K., Zhang, W., Visweswaran, S., Raji, M., & Kuo, Y. F. (2022). A framework for modeling and interpreting patient subgroups applied to hospital readmission: visual analytical approach. *JMIR Medical Informatics*, *10*(12), e37239.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Centers for Medicare & Medicaid Services. (2024). *Hospital Readmissions Reduction Program (HRRP).* https://www.cms.gov/medicare/quality/hospital-readmissions-reduction-program

Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2023). Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science*, *6*, 123–144.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Databricks. (2023). *What is the Medallion Lakehouse Architecture?* https://www.databricks.com/glossary/medallion-architecture

Davis, S., Zhang, J., Lee, I., Rezaei, M., Greiner, R., McAlister, F. A., & Padwal, R. (2022). Effective hospital readmission prediction models using machine-learned features. *BMC Health Services Research*, *22*(1), 1415.

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? Advances in Neural Information Processing Systems, 35, 507–520.

Hripcsak, G., Duke, J. D., Shah, N. H., Reich, C. G., Huser, V., Schuemie, M. J., Suchard, M. A., Park, R. W., Wong, I. C. K., Rijnbeek, P. R., van der Lei, J., Pratt, N., Norén, G. N., Li, Y. C., Stang, P. E., Madigan, D., & Ryan, P. B. (2015). Observational Health Data Sciences and Informatics (OHDSI): Opportunities for observational researchers. *Studies in Health Technology and Informatics*, *216*, 574–578.

Huang, C., Lu, Y., Chen, Y., & Wang, F. (2021). Explainable machine learning for predicting hospital readmissions using electronic health records. *Journal of Biomedical Informatics*, *118*, 103789.

Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the Medicare fee-for-service program. *New England Journal of Medicine*, *360*(14), 1418–1428. https://doi.org/10.1056/NEJMsa0803563

Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. *JAMA*, *306*(15), 1688–1698. https://doi.org/10.1001/jama.2011.1515

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems, 30, 3146–3154.
Kripalani, S., Jackson, A. T., Schnipper, J. L., & Coleman, E. A. (2007). Promoting effective transitions of care at hospital discharge: A review of key issues for hospitalists. *Journal of Hospital Medicine, 2*(5), 314–323. https://doi.org/10.1002/jhm.228
Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc.
Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9
Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association, 23*(5), 899–908. https://doi.org/10.1093/jamia/ocv189
Overhage, J. M., Ryan, P. B., Reich, C. G., Hartzema, A. G., & Stang, P. E. (2012). Validation of a common data model for active safety surveillance research. *Journal of the American Medical Informatics Association, 19*(1), 54–60. https://doi.org/10.1136/amiajnl-2011-000376
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In Advances in Neural Information Processing Systems, 31, 6638–6648.
Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., ... Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine, 1*, 18. https://doi.org/10.1038/s41746-018-0029-1
Raza, S. (2022). A machine learning model for predicting, diagnosing, and mitigating health disparities in hospital readmission. *Healthcare Analytics*, *2*, 100100.
Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). https://doi.org/10.1145/2939672.2939778
Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. Information Fusion, 81, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011
Singh, H., Mhasawade, V., & Chunara, R. (2022). Generalizability challenges of mortality risk prediction models: A retrospective analysis on a multi-center database. *PLOS digital health*, *1*(4), e0000023.
Sokol, K., Hepburn, A., Santos-Rodriguez, R., & Flach, P. (2019). bLIMEy: surrogate prediction explanations beyond LIME. *arXiv preprint arXiv:1910.13016*.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want: Contextualizing explainable machine learning for clinical end use. *Proceedings of Machine Learning Research, 106*, 359–380.
van Walraven, C., Bennett, C., Jennings, A., Austin, P. C., & Forster, A. J. (2011). Proportion of hospital readmissions deemed avoidable: A systematic review. *CMAJ, 183*(7), E391–E402. https://doi.org/10.1503/cmaj.101860
Wang, Z., Zhang, Y., Liu, X., & Li, J. (2023). Explainable artificial intelligence for clinical risk prediction using SHAP interaction analysis. *Journal of Biomedical Informatics*, *141*, 104361.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x
Zuckerman, R. B., Sheingold, S. H., Orav, E. J., Ruhter, J., & Epstein, A. M. (2016). Readmissions, observation, and the Hospital Readmissions Reduction Program. *New England Journal of Medicine, 374*(16), 1543–1551. https://doi.org/10.1056/NEJMsa1513024