Technical Report

Explainable Risk-Stratification and Clinical Decision-Support Framework for Remote Patient Monitoring

An Explainable AI Framework Integrating FHIR R4 Data Normalization, Physiological Feature Engineering, and SHAP-Based Clinical Explanation

Li, Xinyang

Health Informatics Tools Corporation (HITC)

HITC-RD-2026-002

Document Version: 1.0 | July 2026

Classification: Public — Methodology Report

Project Code: HITC-RD-2026-002

Domain: Remote Patient Monitoring, Clinical Decision Support, XAI

Repository: https://xinyangli.tech/line2/

Abstract

Remote patient monitoring (RPM) has expanded rapidly, yet most deployments remain passive: devices collect biometric data and rely on manual clinical review. This approach is unsustainable at scale. We present the RPM-CDS Framework, an explainable artificial intelligence framework designed to transform continuous patient-generated health data into structured clinical risk intelligence. The framework integrates HL7 FHIR R4-based data normalization, a structured 80-feature physiological dictionary spanning 17 clinical domains, XGBoost-based risk stratification with probability calibration, and SHAP-based explainable AI that generates clinical explanation artifacts as FHIR DiagnosticReport resources.

The framework was validated through prospective pilot deployments encompassing 340 heart failure telemonitoring patients and 275 post-surgical recovery patients. Results demonstrate an AUROC of 0.89 for 72-hour readmission prediction (95% CI: 0.84-0.93), median alert-to-clinician latency of 74 seconds, and a 23% improvement in patient engagement adherence compared to baseline RPM usage. SHAP-based explanations identified the top five clinical predictors, providing interpretable decision support that clinicians rated 4.2/5.0 on satisfaction surveys.

Keywords: Remote Patient Monitoring - Clinical Decision Support - FHIR R4 - Explainable AI - SHAP - Risk Stratification - XGBoost - Healthcare Interoperability

1. RPM Challenges

The global shift toward value-based care and the rapid expansion of telehealth infrastructure have created an urgent need for intelligent RPM platforms that do more than passively collect biometric data. Healthcare organizations increasingly require systems that can analyze streaming patient-generated health data in near-real time, predict clinical deterioration before it occurs, and deliver actionable decision support directly to care teams.

A 2025 survey by the American Telemedicine Association found that 68% of RPM programs report that their clinical teams are overwhelmed by data volume, and 42% report delayed detection of clinical deterioration due to alert fatigue from non-discriminating thresholds. Three specific deficiencies characterize current RPM systems:

- Static threshold-based alerting fails to account for patient-specific baselines, temporal trends, or multi-variable interactions, producing high false-positive rates that erode clinician trust.
- Few platforms integrate streaming ML inference capable of updating risk predictions as new data arrives; most rely on batch-processed models with update cycles measured in hours or days.
- Patient engagement features are typically minimal, consisting of generic push notifications rather than personalized, context-aware interactions adapted to individual behavior patterns.

The RPM-CDS Framework was designed to address these gaps through a purpose-built architecture that treats continuous physiological data as a first-class streaming computation problem, applies ensemble ML models optimized for temporal clinical data, and incorporates a flexible CDS engine enabling multi-factorial alerting logic.

2. Framework Overview

The RPM-CDS Framework is structured as a six-layer pipeline spanning data acquisition through clinical action. Each layer provides well-defined interfaces enabling independent deployment and evolution.

Layer	Component	Function
Layer 1	Data Acquisition	Wearable device ingestion, BLE/Wi-Fi/cellular connectivity, ISO 11073 PHD protocol
Layer 2	FHIR R4 Data Normalization	Observation-based standardization, LOINC terminology, UCUM units

Layer 3	Feature Engineering	80-Feature Dictionary, 17 clinical domains, derived metrics
Layer 4	ML Risk Stratification	XGBoost ensemble, probability calibration, risk score generation
Layer 5	Explainable AI Layer	SHAP feature attribution, explanation generation, clinical mapping
Layer 6	CDS Workflow Integration	Alerting, risk review, SMART-on-FHIR integration

3. FHIR R4 Data Normalization

All incoming RPM observations are normalized into FHIR R4 Observation and QuestionnaireResponse resources, adhering to the HITC Physiological Data Extension (PDE) profiles. This model supports continuous vital signs (heart rate, respiratory rate, SpO2, blood pressure), continuous glucose monitoring traces, actigraphy (step count, sleep stages), patient-reported symptom surveys (PHQ-9, PROMIS), and medication adherence events.

Each observation is coded using LOINC terminology with UCUM units enforced through invariant constraints. Observations are timestamped with UTC time and linked to the originating device, patient, and encounter context through FHIR reference elements. Normalized data is persisted in a time-series-optimized FHIR store supporting efficient range-based queries for historical trend retrieval and ML feature construction.

4. 80-Feature Physiological Model

The framework defines a structured 80-feature physiological dictionary spanning 17 clinical domains. Features are organized into five categories: cardiovascular (heart rate, blood pressure, heart rate variability metrics), respiratory (respiratory rate, SpO2, oxygen saturation trends), metabolic (glucose, weight, BMI, temperature), activity (step count, sleep stages, activity classification), and patient-reported indicators (pain, dyspnea, fatigue, medication adherence).

Each feature is specified with a unique identifier, clinical domain classification, source data stream reference, LOINC code, UCUM unit, temporal window definition (e.g., 7-day rolling average), derivation logic (e.g., SDNN for HRV), and missing-data policy. The feature dictionary is published as a structured CSV artifact at https://xinyangli.tech/line2/.

Category	Features	Clinical Domain Examples
Cardiovascular	18	HR, BP, HRV (SDNN, RMSSD), pulse pressure, MAP
Respiratory	10	RR, SpO2, O2 therapy status, SpO2 variability
Metabolic	12	Glucose, weight, BMI, temperature, HbA1c trend
Activity	15	Step count, sleep stages, activity classification, sit-stand transitions
Patient-Reported	15	Pain, dyspnea, fatigue, PHQ-9, medication adherence
Data Quality	10	Completeness score, signal quality, dropout indicators

5. Risk Stratification Model

The risk stratification engine hosts two primary model types. The ensemble risk stratification model uses gradient-boosted tree ensembles (XGBoost/LightGBM) trained on historical EHR data combined with RPM time-series features, producing 24-hour, 72-hour, and 30-day risk scores for targeted outcomes including readmission, decompensation, and infection. The anomaly detection model uses deep learning-based autoencoders that learn patient-specific physiological baselines and emit anomaly scores when streaming observations deviate significantly from expected patterns.

Model training used a de-identified dataset of 12,800 telemonitoring patient records spanning four clinical conditions (heart failure, type 2 diabetes, COPD, post-surgical recovery) from two academic medical centers. Data was split

70/15/15 for training, validation, and held-out testing. The feature engineering team developed 385 candidate features from ECG-derived heart rate variability metrics, actigraphy patterns, patient-reported symptom trajectories, and medication adherence sequences.

Inference is executed via an ONNX Runtime serving layer with GPU acceleration, achieving sub-50-millisecond per-inference latency at batch sizes of 1, enabling real-time risk scoring with each new observation stream.

6. Explainable AI Architecture

The explainable AI component is a distinguishing architectural contribution. Rather than treating model interpretability as a post-hoc reporting feature, the framework generates structured explanation artifacts that are clinically meaningful and interoperable.

6.1 SHAP Feature Attribution

SHAP (SHapley Additive exPlanations) values are computed for each risk prediction using a KernelSHAP approximation optimized for tree-based models. Feature contributions are computed at the patient level for each prediction window, producing a ranked list of contributing features with their direction (increasing or decreasing risk) and magnitude. SHAP values are calculated on the same feature space used for model inference, ensuring explanation fidelity.

6.2 Clinical Explanation as FHIR DiagnosticReport

The framework introduces a novel approach: mapping SHAP explanation output to a structured FHIR DiagnosticReport resource. The report includes: the risk score with confidence interval, the top-K contributing features with LOINC-coded labels and direction indicators, textual narrative summarizing the clinical rationale, and references to the underlying Observation resources used for prediction. This enables explanations to be stored, queried, and shared alongside clinical data through standard FHIR infrastructure.

6.3 Top Predictors Identified

Model interpretability analysis using SHAP values identified the top five predictors of 72-hour readmission across both clinical cohorts: (1) heart rate variability metrics (SDNN < 50 ms sustained over 6+ hours); (2) deviation from patient-specific step count baseline exceeding 40%; (3) increase in daytime resting heart rate > 8 bpm above 7-day rolling average; (4) non-completion of scheduled symptom survey for 48+ consecutive hours; and (5) self-reported dyspnea score trending upward over 3+ consecutive assessments.

7. CDS Workflow

The CDS rules engine evaluates both ML inference outputs and clinician-defined rules against each patient's current state. Rules are expressed using a YAML-based domain-specific language supporting conditions on vital sign values, risk score thresholds, trend slope and volatility metrics, medication adherence gaps, and patient-reported outcome escalation. The engine triggers configurable actions: CDS alerts pushed to the clinician's EHR via SMART-on-FHIR, automated escalation through the HITC-Interop platform, or personalized patient-facing messages.

All CDS actions are logged to the audit subsystem with full provenance, enabling retrospective analysis, quality improvement, and regulatory compliance documentation. The CDS workflow supports multiple alert severity levels, escalation paths, and configurable suppression rules to manage alert burden.

8. Validation Protocol

Validation was conducted in two phases. Phase 1 (retrospective) used the 12,800-patient training dataset with cross-validation and held-out testing. Phase 2 (prospective) comprised pilot deployments at two partner sites: a heart failure telemonitoring program (n=340, mean follow-up 112 days) and a post-surgical joint replacement recovery program (n=275, mean follow-up 42 days).

Metric	Heart Failure	Post-Surgical
72-hour readmission AUROC	0.89 (95% CI: 0.84-0.93)	0.86 (95% CI: 0.80-0.91)
Median alert latency	74 seconds	88 seconds
Alert false-positive rate	21.3%	18.7%
Patient engagement (baseline)	64%	71%
Patient engagement (post-deployment)	87%	85%
Clinician satisfaction (1-5)	4.2	4.4

9. Limitations

Several limitations should be acknowledged. The prospective pilot cohorts were moderate in size and limited to two clinical conditions; generalizability to broader patient populations requires further validation. The SHAP explanation approach, while clinically interpretable, does not capture feature interaction effects in the explanation narrative. The 80-feature dictionary was derived from a specific set of wearable devices; device heterogeneity may affect feature availability. Alert false-positive rates of 18-21%, while substantially improved over static thresholds, still represent an area requiring ongoing refinement.

Planned future work includes: (1) multi-site randomized controlled trial to formally evaluate clinical outcomes; (2) expansion of the ML model registry to additional conditions including hypertension, pediatric asthma, and perinatal monitoring; (3) federated learning for multi-institutional model training without centralizing identifiable patient data; and (4) integration of social determinants of health data streams as additional model features.

References

1. Chen, Y., Hoffman, M. A., & Williams, J. P. (2025). Machine Learning for Continuous Risk Stratification in Remote Patient Monitoring. Nature Digital Medicine, 8, 112.
2. Health Informatics Tools Corporation. (2025). HITC Physiological Data Extension (PDE): FHIR R4 Profiles for Wearable Device and Patient-Generated Health Data. HITC TR-2025-007.
3. Johnson, A. E. W., Pollard, T. J., & Mark, R. G. (2024). Benchmarking Temporal Deep Learning Models for Clinical Deterioration Prediction. Proceedings of MLHC, PMLR 225, 487-503.
4. Martinez, D. L., & O'Brien, K. (2025). Alert Fatigue in Remote Patient Monitoring Programs. Telemedicine and e-Health, 31(4), 298-310.
5. Patel, S., & Thompson, R. (2026). Patient Engagement Patterns in Digital Health Monitoring. Journal of Medical Internet Research, 28(2), e38912.
6. Li, X. (2026). Explainable RPM-CDS Framework. https://xinyangli.tech/line2/