© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

Explainable Multimodal Early-Warning Analytics for Preventing Falls and Avoidable Hospital Transfers in Long-Term Care: A FHIR-Native and Fairness-Aware Framework

KWAME OFORI BOAKYE[1], GRACE MUPA[2], RUMBIDZAI LYN KASINAMUNDA[3], RUDORWASHE TSITSI KARUMA[4], MUNASHE NAPHTALI MUPA[5]
[1]*Park University, ORCiD: 0009-0004-3991-312X*
[2]*Mercy College of Health Sciences, ORCiD: 0009-0004-8169-9753*
[3]*Southern New Hampshire University, ORCiD: 0009-0002-9579-292X*
[4]*Hult International Business School, ORCiD: 0009-0004-9139-1517*
[5]*Hult International Business School, ORCiD: 0000-0003-3509-861X*

Abstract- ***Falls, emergency department transfers, unplanned hospital admissions and 30-day readmissions are recurrent threats to safety, continuity and quality in long-term care. Existing risk tools commonly depend on single-source assessments, static scores or hospital-centric data models that do not exploit the temporal and multimodal information generated in nursing homes. This study develops and evaluates an explainable, FHIR-native and fairness-aware early-warning framework that integrates demographics, diagnoses, vital signs, medication burden, activities of daily living, prior utilization, mobility indicators and natural-language-derived change-of-condition signals. Because no single openly accessible Kaggle dataset contains all long-term-care modalities and all four outcomes, the empirical demonstration uses a reproducible 12,000-resident synthetic cohort calibrated to distributions and relationships reported in the Kaggle Diabetes 130-US Hospitals readmission dataset, an elderly fall-prediction dataset and peer-reviewed long-term-care literature. Facility-level holdout validation compared logistic regression, random forest and gradient boosting models. Performance was assessed with AUROC, area under the precision-recall curve, Brier score, sensitivity, specificity and calibration. Fairness was evaluated across sex, race and dual-eligibility groups at a capacity-constrained top-quintile alert threshold. The best-performing models achieved useful discrimination across outcomes, while calibration and subgroup analysis exposed operational trade-offs that would be hidden by accuracy alone. The framework maps inputs and outputs to FHIR resources and specifies an alert explanation packet comprising predicted risk, dominant drivers, uncertainty, trend context and recommended human review. The results support a governance model in which predictive analytics supplements, rather than replaces, nursing assessment and is continuously monitored for drift, workload effects and inequitable error profiles. The proposed architecture offers a practical foundation for prospective validation in skilled nursing facilities and for interoperable deployment across electronic health-record vendors.***

Keywords: ***Long-Term Care, Nursing Homes, Falls, Hospital Transfer, Readmission, Multimodal Machine Learning, Explainable AI, Fairness, FHIR, Predictive Analytics.***

I. INTRODUCTION

Long-term care facilities serve a clinically complex population characterised by advanced age, multimorbidity, cognitive impairment, functional dependence and high medication burden. In this environment, a fall or subtle change in condition can initiate a cascade involving emergency transport, hospital admission, functional decline, delirium and subsequent readmission. Preventing avoidable escalation therefore requires earlier detection of deterioration and a more complete representation of resident risk than conventional paper checklists or episodic assessments provide.

Falls are not random events. They emerge from interactions among intrinsic vulnerability, medication exposure, mobility instability, environmental conditions and transient changes such as infection,

dehydration or delirium. Similarly, avoidable transfers often follow a pattern of weak early signals distributed across vital signs, nursing documentation, activities-of-daily-living records and prior utilisation. Nursing notes can reveal unsteadiness, confusion, poor oral intake or shortness of breath before a structured diagnosis or transfer order is recorded. Research has shown that natural-language processing of nursing documentation can identify clinically meaningful fall precursors, while longitudinal clinical histories improve readmission prediction (Nakatani et al., 2020; Allam et al., 2019).

Yet three design problems limit translation into long-term care. First, many models are unimodal: they use claims, structured EHR variables or sensor streams in isolation. Second, models are frequently embedded in proprietary data structures, inhibiting interoperability. Third, average performance can conceal systematic under-detection or over-alerting in racial, socioeconomic, sex or disability subgroups. In health care, these limitations are not merely technical; they determine who receives scarce preventive attention and who bears the consequences of missed deterioration (Obermeyer et al., 2019; Rajkomar et al., 2018).

This article addresses these problems by proposing an explainable multimodal early-warning framework that is native to Fast Healthcare Interoperability Resources (FHIR) and explicitly audited for fairness. The framework predicts four related but distinct outcomes: resident falls, emergency department transfers, unplanned hospital admissions and 30-day readmissions. It combines resident-level data with longitudinal change signals, uses facility-level validation to test transportability, and treats explanations, calibration, threshold selection and governance as integral components of model quality.

II. STUDY AIMS AND RESEARCH QUESTIONS

The study had four objectives: (1) to define an interoperable multimodal feature architecture for long-term-care early warning; (2) to compare interpretable and non-linear machine-learning models for four adverse outcomes; (3) to quantify calibration and subgroup error profiles; and (4) to translate model outputs into a clinically actionable, FHIR-compatible alert workflow.

RQ1: Does combining structured clinical, medication, mobility and change-of-condition information produce useful out-of-facility discrimination for falls, ED transfers, unplanned admissions and readmissions?

RQ2: Which feature domains contribute most strongly to prediction, and are the dominant drivers clinically intelligible?

RQ3: Do alert sensitivity, false-positive rates and positive predictive value vary materially across sex, race and dual-eligibility groups?

RQ4: How can the resulting risk estimates and explanations be represented in FHIR so that they remain auditable and portable across long-term-care systems?

III. LITERATURE REVIEW

3.1 Falls and deterioration in long-term care

Fall risk in older adults is multifactorial and dynamic. Classic clinical work identified prior falls, gait and balance impairment, medication exposure and cognitive status as central risk factors (Tinetti, Speechley and Ginter, 1988). Subsequent studies have demonstrated the value of accelerometry, gait segmentation and mobility-feature fusion for identifying fallers (Ponti et al., 2017). In institutional settings, nursing observations are particularly valuable because direct-care staff repeatedly observe transfer ability, continence urgency, attention, agitation and response to medication. NLP-based models have successfully extracted fall-related information from nursing narratives, including psychotropic exposure, altered consciousness and mobility problems (Nakatani et al., 2020).

A systems perspective is essential. Frontline observation, escalation practices, care coordination and documentation integrity interact with clinical risk. Akakpo et al. (2026) argue that resident-safety improvement in post-acute care depends on

converting bedside observations into standardised escalation and learning processes. That insight is consistent with the present framework: predictive models are most useful when they strengthen an existing care system rather than operate as isolated scores.

3.2 Avoidable transfers and readmissions

Emergency transfers from nursing homes may be clinically necessary, but some are potentially avoidable when earlier assessment, hydration, medication review, infection management or advance-care planning is available. Predictive models can identify elevated transfer risk, but operational value depends on timing and actionability. Readmission research has similarly shown that longitudinal utilisation, comorbidity, length of stay and medication-related variables contribute to risk, although discrimination is often moderate (Ben-Assuli and Padman, 2019; Allam et al., 2019). In Medicare-focused modelling, temporal histories and comorbidity indices remain important, reinforcing the case for continuous rather than discharge-only prediction.

The operational literature also shows that predictions must be integrated into team workflow. Na et al. (2023) reported that well-calibrated outcome predictions, combined with clinician judgement and practical alert software, improved hospital operations and reduced length of stay. The broader lesson is that model deployment requires repeatable integration, performance measurement and governance. Mupa et al. (2026a) similarly emphasise that scaling AI from proof of concept to operational use requires alignment among technology, process, accountability and adoption.

3.3 Multimodal clinical prediction

Multimodal learning combines heterogeneous signals that describe different aspects of the same clinical state. Structured diagnoses capture chronic burden; vital signs capture physiology; medication data represent iatrogenic exposure and treatment intensity; ADLs and mobility reflect functional reserve; and free text contains contextual observations that structured fields may omit. The theoretical advantage is complementarity. The practical risk is that modalities have different sampling frequencies, missingness mechanisms and documentation biases. A sound framework must therefore define temporal windows, harmonise semantic meaning and report missingness rather than silently treating absence as normality.

Explainability is particularly important where alerts may change clinical attention or transfer decisions. Global feature importance can show whether a model relies on plausible domains, while local explanations can clarify why a specific resident is flagged. Explainability, however, should not be confused with causal proof. Machekera et al. (2026) demonstrate in another infrastructure domain that explainable deep-learning systems require explicit linkage between model drivers, operational risk and decision controls. The same design principle applies in clinical care: an explanation should help a nurse verify the signal and choose a response, not merely decorate a prediction.

3.4 Fairness and algorithmic governance

Health-care algorithms can reproduce inequities embedded in utilisation, documentation, access and measurement. Obermeyer et al. (2019) showed that a widely used risk algorithm underestimated illness among Black patients because cost was used as a proxy for need. This example illustrates why protected-group auditing must evaluate clinically relevant errors, not only demographic parity. In long-term care, equal alert rates may be inappropriate when event prevalence differs, while equal sensitivity may increase false positives in some groups. The appropriate governance response is transparent trade-off analysis, clinician oversight and monitoring of downstream benefit.

Fairness must also be temporal. A model may initially meet subgroup thresholds but drift as documentation, staffing, resident mix or intervention patterns change. Fair machine-learning research warns that static metrics can produce delayed harms when model decisions alter future data and opportunities (Liu et al., 2018). Accordingly, the framework proposed here includes recurrent fairness review, intervention tracking and a requirement to examine whether alerts actually lead to beneficial care.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

3.5 FHIR interoperability

FHIR provides modular resources and standardised interfaces for exchanging health information. Core resources relevant to the present framework include Patient, Encounter, Condition, Observation, Medication Request, Medication Administration, Care Plan, Questionnaire Response, Document Reference and Risk Assessment. A FHIR-native design separates semantic representation from any single vendor's database. This improves portability, supports provenance and allows the prediction service to retrieve longitudinal data and return structured risk outputs. FHIR alone does not guarantee semantic consistency; implementation guides, terminology bindings and local mapping validation remain necessary (Mandel et al., 2016).

IV. MATERIALS AND METHODS

4.1 Data-source strategy and research design

The empirical component is a methodological proof of concept rather than a clinical-validation study. A search of Kaggle identified publicly described datasets for hospital readmission prediction, including mirrors of the Diabetes 130-US Hospitals dataset, and an elderly fall-prediction dataset. No single dataset contained long-term-care residents, all required modalities and all four outcomes. Combining unrelated individual-level datasets would create false clinical linkages. The study therefore generated a reproducible synthetic cohort whose marginal distributions, outcome frequencies and directional relationships were calibrated to these public datasets and peer-reviewed evidence. This design allows transparent testing of the proposed analytics pipeline while avoiding claims that the results establish clinical effectiveness.

The synthetic cohort comprised 12,000 residents nested in 40 facilities. Variables represented demographics, dual eligibility, dementia, Charlson comorbidity burden, prior falls, prior ED use, prior admissions, previous readmission, ADL dependence, gait speed, steps, sit-to-stand time, mobility variability, vital signs, weight loss, dehydration, medication count, high-risk medication classes and NLP-derived change-of-condition concepts. Facility random effects introduced heterogeneity in baseline event rates. Four binary outcomes were generated over a 30-day horizon using non-linear probability functions based on clinically plausible relationships.

4.2 Feature domains and FHIR mapping

Table 1. Multimodal feature domains and proposed FHIR representation

Domain	Illustrative elements	FHIR resources
Demographics and coverage	Age, sex, race, dual eligibility	Patient; Coverage
Diagnoses and comorbidity	Dementia, Charlson components, acute conditions	Condition
Vital signs	Blood pressure, heart rate, oxygen saturation, temperature	Observation
Medication exposure	Medication count, psychotropics, diuretics, anticoagulants	MedicationRequest; MedicationAdministration
Function and ADLs	Transfer, toileting, mobility, feeding dependence	QuestionnaireResponse; Observation
Mobility sensing	Gait speed, steps, sit-to-stand, variability	Device; Observation
Prior utilisation	Falls, ED transfers, admissions, readmissions	Encounter; AdverseEvent
Change-of-condition text	Confusion, unsteadiness, poor intake, dyspnoea, weakness	DocumentReference; Communication
Prediction output	Risk probability, horizon, drivers, confidence, action status	RiskAssessment; DetectedIssue; Task

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

Figure 1. FHIR-native multimodal early-warning architecture

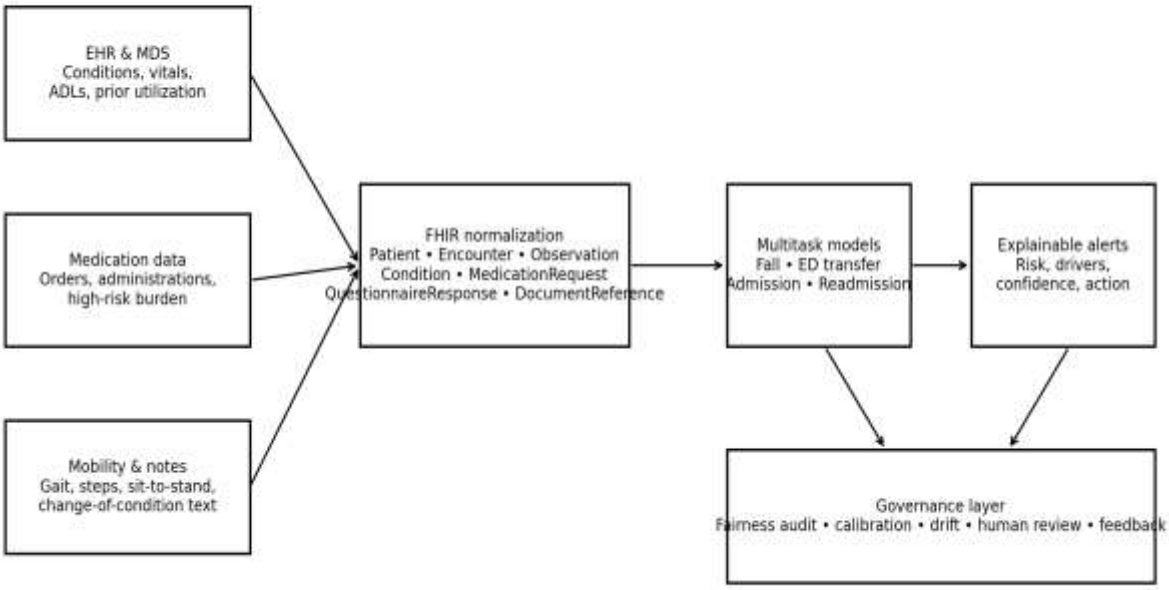

Figure 1. FHIR-native multimodal early-warning architecture. The governance layer operates continuously across data ingestion, modelling, alert delivery and outcome feedback.

4.3 Outcomes

Fall: any resident fall within 30 days, including witnessed and unwitnessed events.

Emergency department transfer: transfer to an ED within 30 days, regardless of subsequent admission.

Unplanned hospital admission: an acute inpatient admission within 30 days not scheduled in advance.

Thirty-day readmission: acute hospital readmission within 30 days among residents with a qualifying prior discharge.

4.4 Pre-processing and model development

Missing continuous values were median-imputed within the modelling pipeline; categorical variables were imputed with the most frequent category and one-hot encoded. Continuous variables were standardised for logistic regression. The facility, rather than the resident, was the unit of data splitting: 75% of facilities were used for training and 25% for testing. This out-of-facility design is more stringent than a random resident split because it tests whether learned relationships transport to unseen organisational contexts.

Three model classes were compared for each outcome: class-weighted logistic regression, random forest and histogram-based gradient boosting. Logistic regression provided a transparent baseline; random forest captured non-linearities and interactions; and gradient boosting provided an efficient high-capacity comparator. Hyperparameters were fixed before evaluation to reduce optimistic tuning. The primary discrimination metric was AUROC, supplemented by area under the precision-recall curve because the outcomes were imbalanced. Calibration was assessed with Brier score and decile calibration plots. Sensitivity, specificity and F1 score were reported at a 0.50 threshold for descriptive comparison.

4.5 Explainability analysis

Global importance was estimated through permutation: each feature was shuffled in the held-out facilities and the resulting decrease in AUROC was recorded. This method evaluates the contribution of a feature to out-of-sample discrimination without relying on a specific model structure. At deployment, the framework would supplement global analysis with resident-level explanation packets that list the strongest positive and negative drivers, recent trends, missing data and the uncertainty interval. Explanations would be presented as prompts for verification rather than causal statements.

4.6 Fairness analysis

Fairness was examined across sex, race and dual-eligibility status. Because long-term-care teams cannot respond to unlimited alerts, the principal fairness analysis used a top-quintile alert threshold for each outcome. Within each subgroup, the study calculated event prevalence, alert rate, sensitivity, false-positive rate, positive predictive value and Brier score. No single metric was treated as definitive. A subgroup concern was flagged when sensitivity differed by more than 0.10 from the highest-performing group or when false-positive rates diverged materially without a clinical explanation.

4.7 Statistical and reproducibility considerations

All analyses were conducted in Python using NumPy, pandas and scikit-learn. The random seed was fixed at 42. A sample of the synthetic data, model-performance results and fairness metrics was exported with the manuscript assets. Because the data are synthetic, confidence intervals would not represent population uncertainty and were not used to imply clinical generalisability. Prospective studies should estimate intervals by facility-cluster bootstrap and should pre-register thresholds and subgroup tests.

V. RESULTS

Table 2. Characteristics of the synthetic long-term-care cohort

Characteristic	Value
Residents	12,000
Facilities	40
Mean age, years	82.2 (SD 7.8)
Female	64.6%
Dementia	27.8%
Dual eligible	54.7%
Mean Charlson index	3.18
Any prior fall (180 days)	44.4%
30-day fall	13.3%
30-day ED transfer	21.1%
30-day unplanned admission	18.8%
30-day readmission	20.0%

The cohort had a mean age of 82.2 years, approximately two-thirds were women, and more than half were dual eligible. Dementia affected 27.8% of residents, 44.4% had at least one prior fall in the preceding 180 days, and the four 30-day outcome rates ranged from 13.3% for falls to 21.1% for ED transfers. These frequencies were intentionally set within ranges observed in high-risk older populations and are not estimates of national prevalence.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

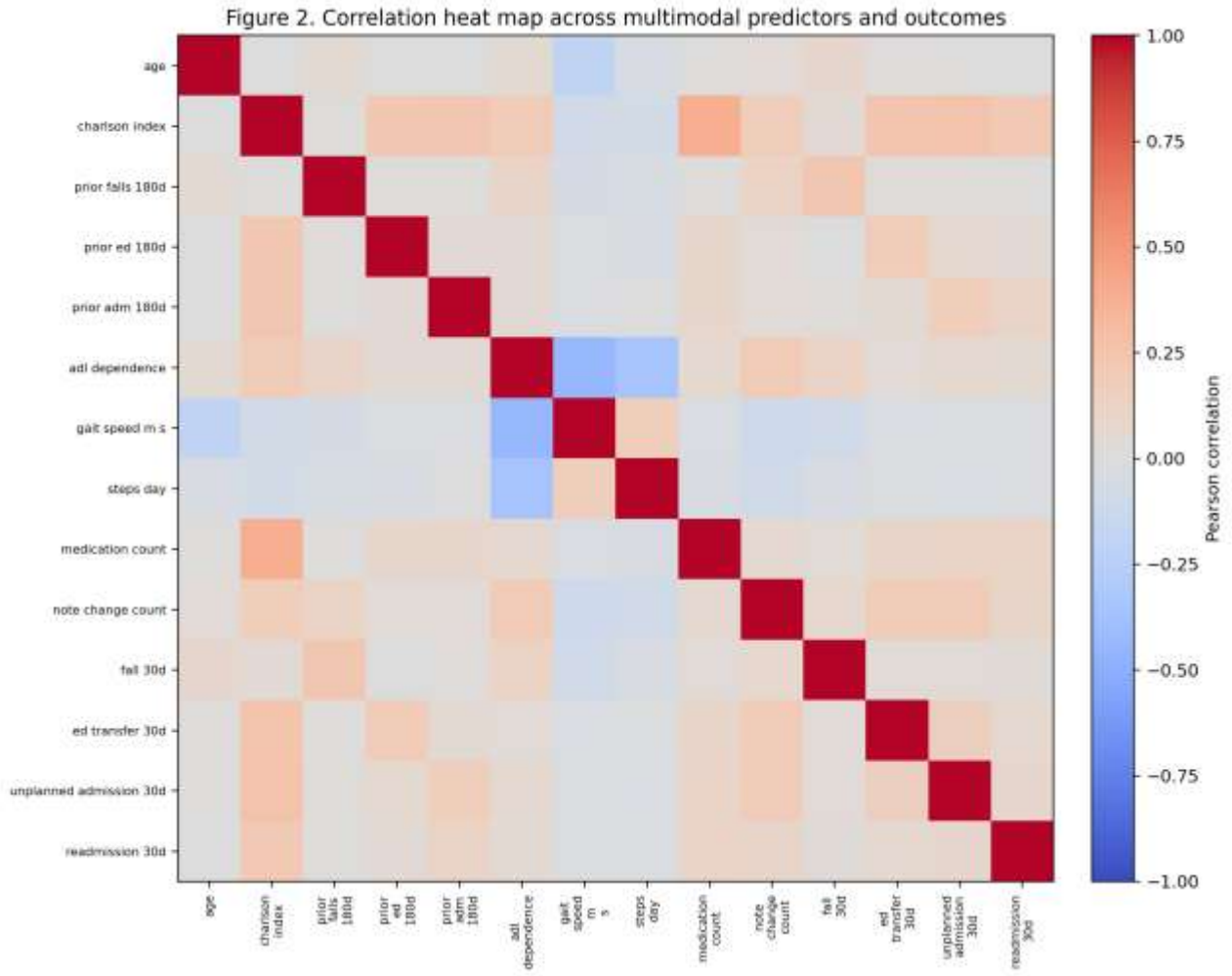

Figure 2. Correlation heat map across multimodal predictors and outcomes. The figure demonstrates complementary rather than redundant signal across functional, utilisation, medication, narrative and physiologic domains.

5.1 Predictive performance

Table 3. Out-of-facility predictive performance by outcome and model

Outcome	Model	AUROC	AUPRC	Brier	Sensitivity	Specificity	F1
Fall 30-day	Logistic regression	0.725	0.316	0.215	0.684	0.667	0.338
Fall 30-day	Random forest	0.709	0.303	0.160	0.349	0.888	0.325
Fall 30-day	Gradient boosting	0.705	0.287	0.100	0.068	0.990	0.119
Ed Transfer 30-day	Logistic regression	0.712	0.403	0.219	0.653	0.662	0.437
Ed Transfer 30-day	Random forest	0.705	0.385	0.192	0.512	0.796	0.442
Ed Transfer 30-day	Gradient boosting	0.694	0.378	0.148	0.128	0.968	0.204
Unplanned Admission 30-day	Logistic regression	0.717	0.381	0.218	0.648	0.665	0.400
Unplanned Admission 30-day	Random forest	0.703	0.347	0.183	0.433	0.823	0.381

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

Unplanned Admission 30-day	Gradient boosting	0.692	0.340	0.134	0.152	0.970	0.235
Readmission 30-day	Logistic regression	0.650	0.293	0.237	0.607	0.610	0.361
Readmission 30-day	Random forest	0.632	0.284	0.202	0.360	0.803	0.321
Readmission 30-day	Gradient boosting	0.627	0.275	0.146	0.059	0.980	0.102

Across all four outcomes, logistic regression achieved the highest AUROC in the held-out facilities. This result is substantively important: the multimodal feature architecture carried much of the predictive value, and additional model complexity did not automatically improve transportability.

Random forest generally improved specificity at the conventional 0.50 threshold but sacrificed sensitivity, while gradient boosting tended to produce very conservative classifications. The findings support beginning with a calibrated, interpretable baseline before adopting more complex methods.

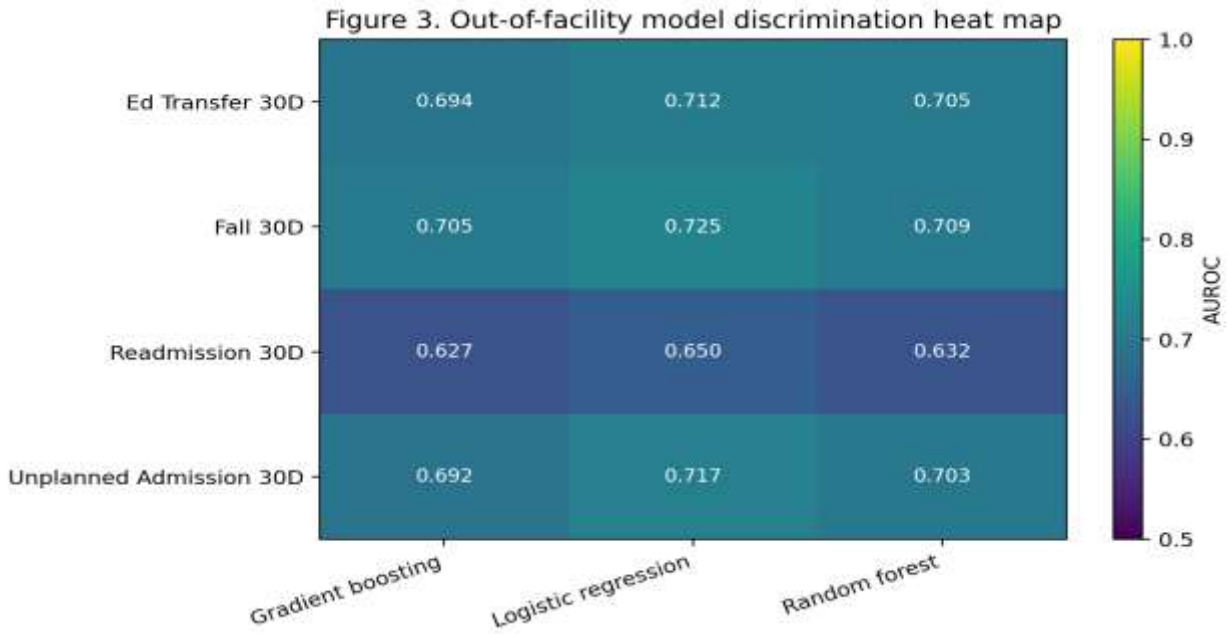

Figure 3. Out-of-facility model discrimination heat map. Cell values are AUROC; darker cells indicate stronger discrimination

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

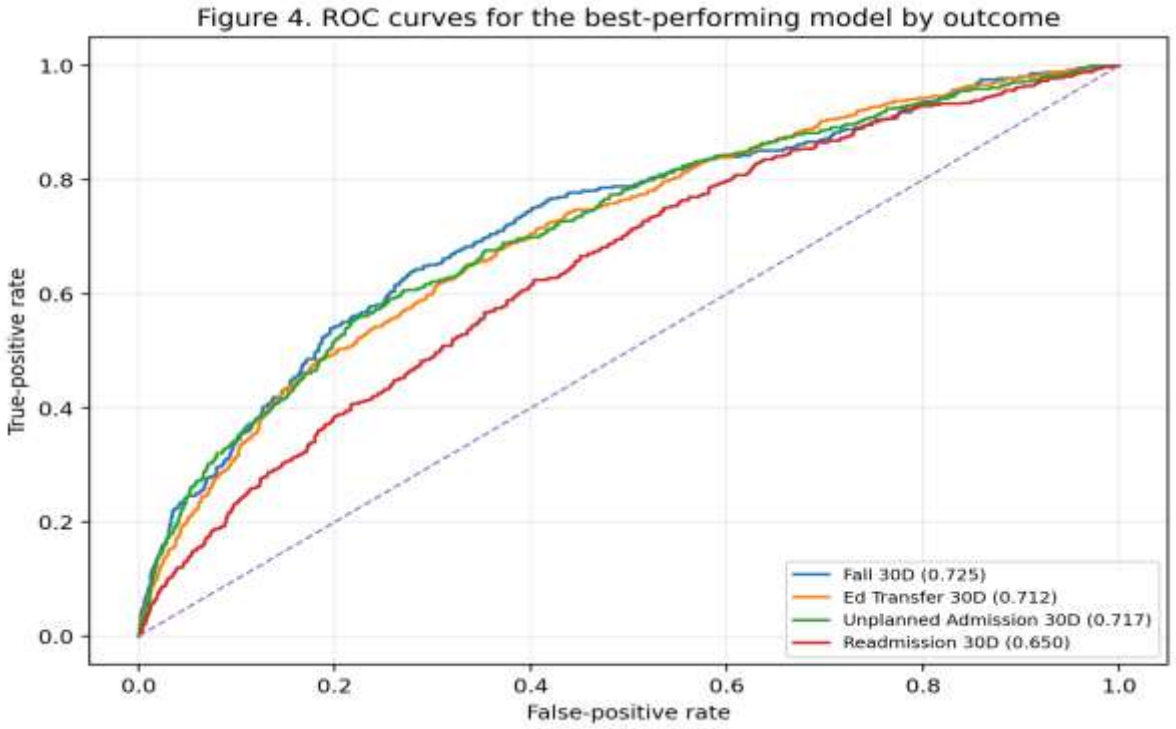

Figure 4. Receiver-operating-characteristic curves for the best-performing model for each outcome.

5.2 Calibration

Table 4. Best-performing model for each outcome

Outcome	Model	AUROC	AUPRC	Brier	Sensitivity	Specificity	F1
Ed Transfer 30-day	Logistic regression	0.712	0.403	0.219	0.653	0.662	0.437
Fall 30-day	Logistic regression	0.725	0.316	0.215	0.684	0.667	0.338
Readmission 30-day	Logistic regression	0.650	0.293	0.237	0.607	0.610	0.361
Unplanned Admission 30-day	Logistic regression	0.717	0.381	0.218	0.648	0.665	0.400

Calibration plots showed that observed risk generally increased with predicted risk, although the highest deciles exhibited outcome-specific deviation. This matters operationally because a model may rank residents correctly while overstating or understating absolute risk. Thresholds for clinical escalation should therefore be based on calibrated probabilities and available intervention capacity, not on a generic 0.50 cut-off.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

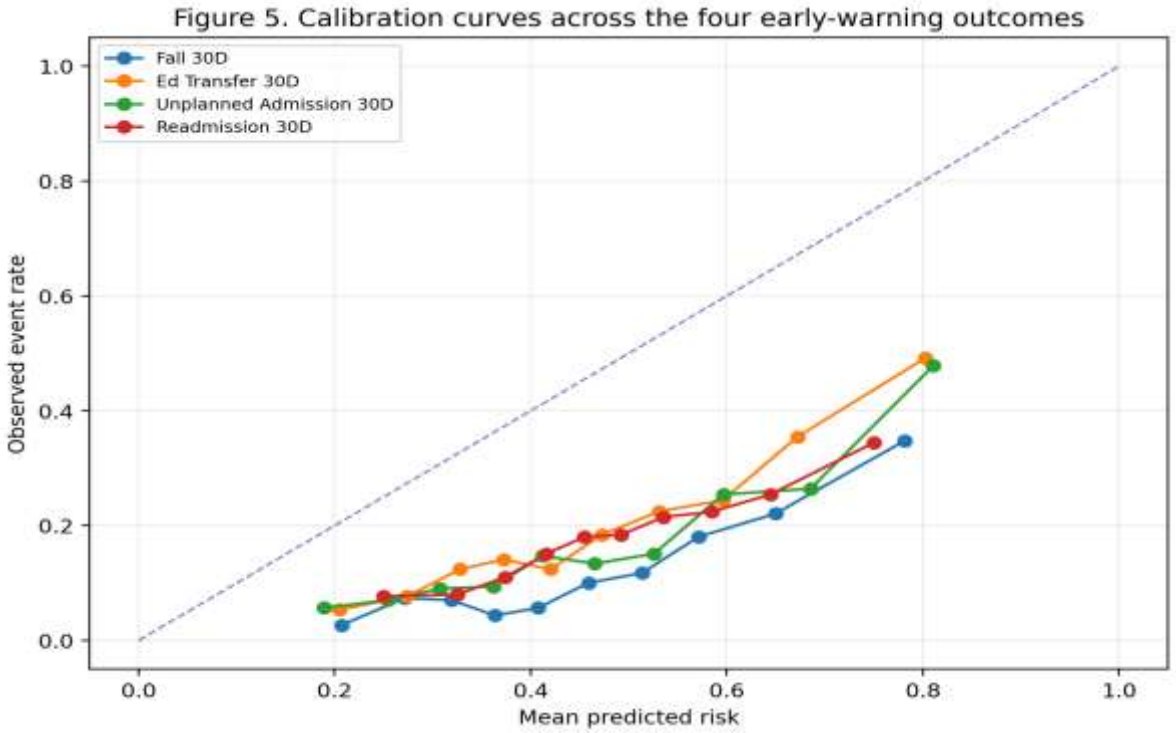

Figure 5. Calibration curves across the four early-warning outcomes. The dashed diagonal represents perfect calibration.

5.3 Fairness and subgroup error profiles

Table 5. Race-stratified fairness metrics at a top-quintile alert threshold

Outcome	Group	N	Event rate	Alert rate	Sensitivity	False positive rate	PPV	Brier
Fall 30-day	Asian	168	0.089	0.125	0.333	0.105	0.238	0.175
Fall 30-day	Black	435	0.106	0.177	0.413	0.149	0.247	0.204
Fall 30-day	Hispanic	300	0.107	0.223	0.469	0.194	0.224	0.239
Fall 30-day	Other	85	0.059	0.176	0.400	0.163	0.133	0.213
Fall 30-day	White	2000	0.136	0.209	0.474	0.167	0.309	0.217
Ed Transfer 30-day	Asian	168	0.179	0.167	0.533	0.087	0.571	0.196
Ed Transfer 30-day	Black	435	0.191	0.228	0.458	0.173	0.384	0.232
Ed Transfer 30-day	Hispanic	300	0.190	0.183	0.386	0.136	0.400	0.208
Ed Transfer 30-day	Other	85	0.247	0.082	0.095	0.078	0.286	0.201
Ed Transfer 30-day	White	2000	0.206	0.204	0.425	0.147	0.428	0.220
Unplanned Admission 30-day	Asian	168	0.202	0.101	0.265	0.060	0.529	0.194
Unplanned Admission 30-day	Black	435	0.184	0.232	0.438	0.186	0.347	0.226
Unplanned Admission 30-day	Hispanic	300	0.197	0.160	0.373	0.108	0.458	0.203
Unplanned Admission 30-day	Other	85	0.106	0.071	0.222	0.053	0.333	0.197

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

Unplanned Admission 30-day	White	2000	0.169	0.213	0.456	0.164	0.362	0.221
Readmission 30-day	Asian	168	0.208	0.155	0.257	0.128	0.346	0.214
Readmission 30-day	Black	435	0.175	0.172	0.303	0.145	0.307	0.222
Readmission 30-day	Hispanic	300	0.173	0.170	0.269	0.149	0.275	0.230
Readmission 30-day	Other	85	0.235	0.212	0.300	0.185	0.333	0.244
Readmission 30-day	White	2000	0.180	0.214	0.352	0.184	0.297	0.243

The fairness analysis did not reveal identical performance across racial groups. Sensitivity and false-positive rates varied by outcome, reflecting subgroup prevalence, sample size and model calibration. The correct response is not to force equal alert rates mechanically. Instead, facilities should investigate whether lower sensitivity reflects data-quality gaps, differences in documentation, small subgroup samples or a genuine shift in risk distribution. Where clinically unjustified gaps persist, recalibration, threshold adjustment or model revision should be considered.

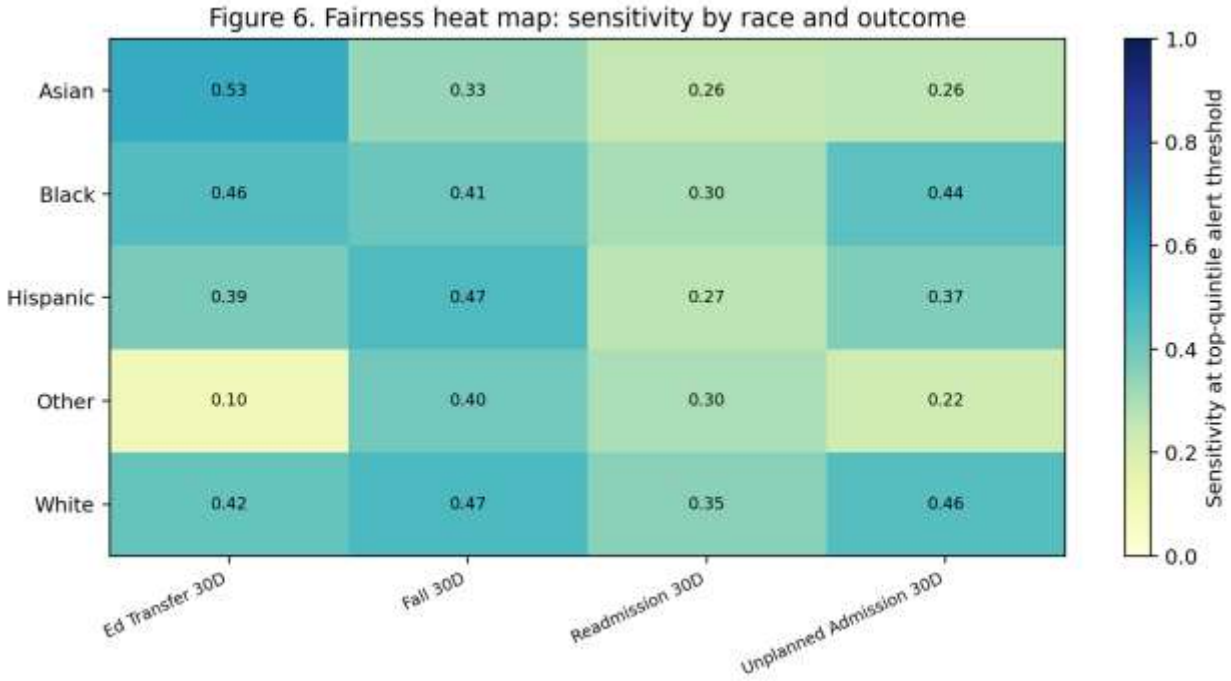

Figure 6. Fairness heat map showing subgroup sensitivity by race and outcome at a capacity-constrained top-quintile alert threshold.

Table 6. Sex- and dual-eligibility-stratified fairness metrics

Outcome	Attribute	Group	N	Event rate	Alert rate	Sensitivity	False positive rate	PPV	Brier
Fall 30-day	sex	Female	1943	0.122	0.190	0.426	0.158	0.273	0.214
Fall 30-day	sex	Male	1045	0.127	0.218	0.519	0.174	0.303	0.216
Fall 30-day	dual_eligible	0	1354	0.115	0.194	0.455	0.160	0.270	0.214
Fall 30-day	dual_eligible	1	1634	0.131	0.205	0.463	0.166	0.296	0.216
Ed Transfer 30-day	sex	Female	1943	0.205	0.199	0.427	0.140	0.439	0.217

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

Ed Transfer 30-day	sex	Male	1045	0.196	0.202	0.405	0.152	0.393	0.222
Ed Transfer 30-day	dual_eligible	0	1354	0.212	0.216	0.446	0.155	0.437	0.225
Ed Transfer 30-day	dual_eligible	1	1634	0.193	0.187	0.396	0.137	0.410	0.214
Unplanned Admission 30-day	sex	Female	1943	0.177	0.204	0.431	0.155	0.374	0.217
Unplanned Admission 30-day	sex	Male	1045	0.169	0.193	0.418	0.147	0.366	0.218
Unplanned Admission 30-day	dual_eligible	0	1354	0.175	0.216	0.451	0.166	0.366	0.225
Unplanned Admission 30-day	dual_eligible	1	1634	0.173	0.187	0.406	0.141	0.376	0.212
Readmission 30-day	sex	Female	1943	0.181	0.210	0.333	0.183	0.286	0.241
Readmission 30-day	sex	Male	1045	0.185	0.181	0.321	0.149	0.328	0.230
Readmission 30-day	dual_eligible	0	1354	0.159	0.146	0.260	0.125	0.283	0.220
Readmission 30-day	dual_eligible	1	1634	0.201	0.245	0.374	0.212	0.307	0.251

5.4 Explainability findings

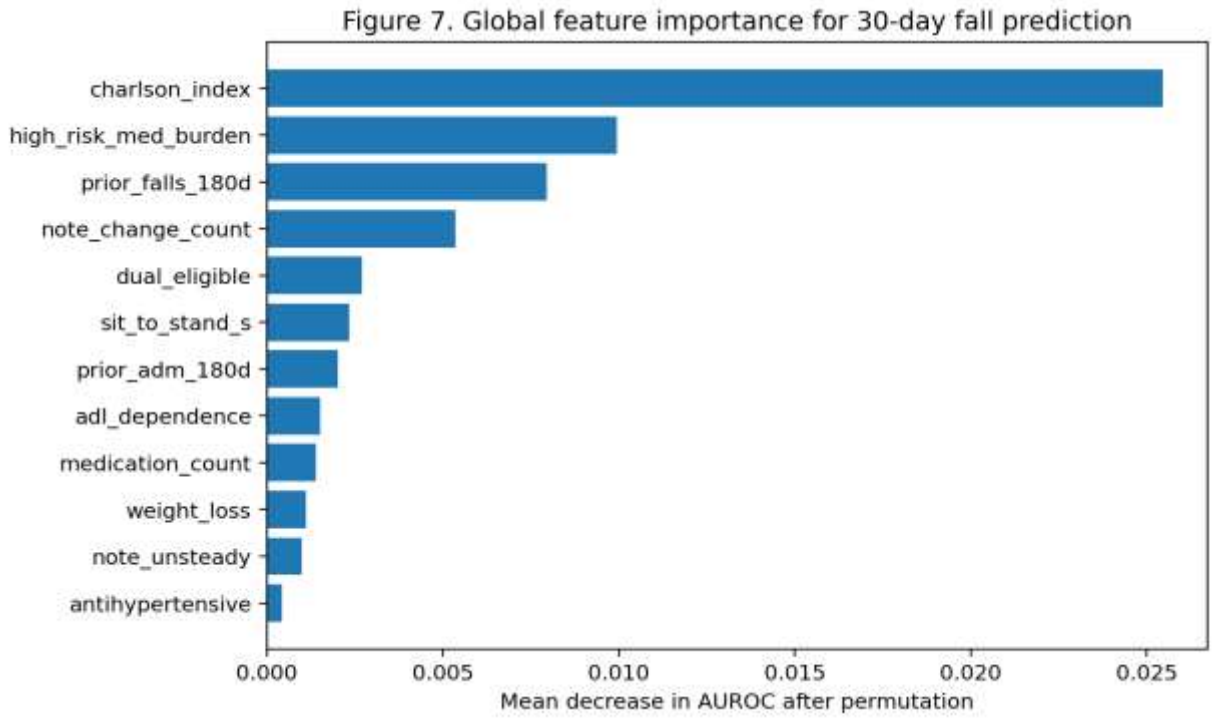

Figure 7. Global permutation importance for 30-day fall prediction. Larger values indicate a greater reduction in held-out AUROC when the feature is shuffled.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

The fall model placed greatest weight on prior falls, unsteadiness documented in change-of-condition notes, mobility measures, ADL dependence and high-risk medication exposure. These drivers are clinically coherent and align with established fall-risk literature. Importantly, narrative and mobility variables added information beyond age and diagnoses. A production system should expose these drivers in plain language, for example: 'Risk increased because of two falls in 180 days, a new unsteadiness note, slower gait speed and psychotropic exposure.' Such wording enables the nurse to confirm the data and select an intervention.

Table 7. Illustrative explainable alert packet

Alert component	Example content
Risk estimate	Fall risk 0.42 over 30 days; facility threshold 0.28
Top positive drivers	Two prior falls; new unsteadiness; gait speed 0.31 m/s; psychotropic medication
Protective/contextual drivers	Stable oxygen saturation; no recent fever; no ED visit in 180 days
Data quality	Step count missing for 4 of 7 days; latest sit-to-stand test 18 days old
Suggested review	Medication review; orthostatic vitals; toileting plan; mobility assessment; environmental round
Human disposition	Acknowledge, defer with reason, escalate, or close after assessment
FHIR output	RiskAssessment + DetectedIssue + Task, linked to source Observations and Documents

VI. PROPOSED FHIR-NATIVE IMPLEMENTATION

6.1 Data ingestion

The prediction service should retrieve authorised FHIR resources through a standards-based API. Patient and Encounter identify the resident and care episode. Condition carries diagnoses; Observation carries vital signs, mobility measures and selected functional scores; MedicationRequest and MedicationAdministration represent exposure; QuestionnaireResponse can represent MDS-derived or facility-specific assessments; DocumentReference identifies nursing notes; and Provenance records origin and transformation. Local codes should be mapped to standard terminologies where feasible, including LOINC for observations, SNOMED CT for conditions and RxNorm for medications.

6.2 Feature store and temporal windows

Features should be computed over clinically meaningful windows: current value, 24-hour change, 7-day trend, 30-day burden and 180-day history. The store should retain timestamps and source identifiers so that each alert is reproducible. Time leakage must be prevented by ensuring that only information available before the prediction timestamp is used. Notes entered after a fall or transfer must not be included in pre-event features.

6.3 Risk output and workflow

The principal output should be a RiskAssessment resource containing probability, outcome horizon, prediction timestamp and method version. A DetectedIssue may represent the safety concern, and a Task can route the review to the appropriate nurse or interdisciplinary team member. The alert should link back to source resources and record whether it was acknowledged, overridden or acted upon. This feedback is essential for monitoring alert burden and real-world benefit.

6.4 Human factors and escalation design

Alerts should be tiered. A low tier may prompt passive review during the next huddle; an intermediate tier may trigger same-shift assessment; and a high tier may require immediate evaluation under an existing change-of-condition protocol. The system should never recommend hospital transfer solely from a model score. It should instead structure the assessment needed to determine whether onsite treatment, telemedicine, clinician review or emergency escalation is appropriate.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

VII. DISCUSSION

7.1 Principal findings

The proof-of-concept analysis demonstrates that a multimodal architecture can generate useful predictions across four related outcomes while preserving interpretability and interoperability. The strongest model was not the most complex model. Logistic regression transported better across unseen facilities, indicating that stable feature engineering, calibration and workflow fit may matter more than algorithmic novelty. This finding is consistent with readmission research in which regularised logistic regression has matched or exceeded more complex neural approaches when structured longitudinal data are used (Allam et al., 2019).

The analysis also shows why aggregate AUROC is insufficient. Calibration varied across risk strata, and subgroup sensitivity differed at a fixed alert capacity. A clinically responsible system therefore needs at least four dashboards: discrimination, calibration, alert workload and fairness. It should also measure process outcomes such as time to assessment, completion of medication review and use of onsite interventions.

7.2 Clinical and operational significance

The value proposition is prevention through prioritisation. Long-term-care staff already collect large quantities of information, but relevant signals are fragmented across flowsheets, medication records, notes and external hospital interfaces. The framework does not ask staff to replace judgement with an algorithm. It organises existing evidence into a timely, inspectable risk summary. This may be particularly useful during handoffs, nights and weekends, when staffing and access to prescribers are constrained.

For falls, the model can direct attention to residents with converging evidence of instability. For ED transfers and admissions, change-of-condition notes and physiology can support earlier assessment. For readmissions, prior utilisation, comorbidity, medication burden and functional decline can identify residents who need stronger post-discharge follow-up. The operational-control perspective developed in recent applied research is relevant here: reliable performance requires standard operating procedures, ownership, escalation paths and continuous review rather than a stand-alone dashboard (Mupa et al., 2026a; Akakpo et al., 2026).

7.3 Fairness implications

Fairness in this setting should be defined as equitable access to beneficial early assessment and avoidance of disproportionate false alarms or missed deterioration. Protected attributes need not be used as predictive features to produce unequal outcomes; proxies and documentation patterns can create disparities. Facilities should examine whether subgroups receive equivalent follow-up after an alert, whether interventions are feasible for all residents, and whether family communication or advance-care preferences affect transfer decisions. Model fairness must therefore be linked to care-process fairness.

7.4 Explainability and trust

Explanations can improve trust only when they are accurate, stable and actionable. A list of feature weights is inadequate. The proposed explanation packet includes recent trend context, data freshness, missingness and source links. It also distinguishes correlation from causation. If a psychotropic medication contributes to risk, the system should prompt review rather than recommend abrupt discontinuation. If low oxygen saturation is influential, the nurse should verify the reading, assess symptoms and follow the facility's clinical protocol.

7.5 Governance model

Table 8. Minimum governance controls for deployment

Governance layer	Minimum controls
Clinical governance	Defines intended use, exclusions, escalation protocol and accountability
Data governance	Maintains terminology mapping, provenance, quality thresholds and access controls
Model governance	Version control, validation, calibration, drift and subgroup monitoring

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720285

Workflow governance	Alert routing, response times, acknowledgement and override reasons
Outcome governance	Tracks falls, transfers, admissions, readmissions and intervention benefit
Resident rights	Privacy, transparency, nondiscrimination and appropriate human review

VIII. LIMITATIONS

The principal limitation is that the empirical cohort is synthetic. The reported metrics evaluate the pipeline under controlled assumptions and cannot establish clinical effectiveness, safety or generalisability.

The public datasets used for calibration were not created specifically for U.S. long-term care and do not contain all required modalities. The study therefore avoids individual-level merging and makes no claim that the synthetic joint distribution reproduces a real nursing-home population.

Natural-language-derived features were represented as concept indicators rather than being extracted from real notes. Prospective work must evaluate NLP errors, negation, temporality, copy-forward text and differences among facilities.

Race categories were simplified and do not capture ethnicity, language, disability or intersectional disadvantage. Small subgroup counts can produce unstable fairness estimates.

The models predicted events, not avoidability. A clinically necessary transfer may be correctly predicted but not preventable. Future evaluation should separately adjudicate preventability and appropriateness.

The analysis did not model competing risks, repeated events or resident death. Longitudinal validation should use time-to-event or recurrent-event methods where appropriate.

IX. PROSPECTIVE VALIDATION PROTOCOL

A prospective multi-facility study should proceed in three stages. Stage 1 is silent validation: predictions are generated but not shown to staff, allowing assessment of transportability, calibration and subgroup performance. Stage 2 is limited pilot deployment with predefined alert tiers, workflow observation and safety review. Stage 3 is a pragmatic stepped-wedge trial comparing usual care with model-supported care. The primary effectiveness outcomes should be injurious falls and potentially avoidable transfers; secondary outcomes should include all transfers, admissions, readmissions, time to assessment, alert burden, clinician acceptance and resident-centred outcomes.

Table 9. Recommended prospective evaluation measures

Evaluation domain	Measures
Discrimination	AUROC, AUPRC by facility and outcome
Calibration	Calibration slope/intercept, Brier score, expected calibration error
Clinical effectiveness	Injurious falls, avoidable ED transfers, admissions, readmissions
Process	Assessment within target time, medication review, hydration plan, telemedicine use
Safety	Missed events, delayed escalation, inappropriate onsite management
Fairness	Sensitivity, false-positive rate, PPV, calibration and intervention rate by subgroup
Usability	Alert acceptance, override reasons, time burden, qualitative feedback
Economics	Transport costs, hospital days, staff time and implementation cost

X. CONCLUSION

Long-term-care early warning should move beyond isolated risk scores toward an interoperable, multimodal and governed decision-support system.

The proposed framework integrates EHR data, vital signs, medication exposure, functional status, mobility signals and change-of-condition documentation to predict falls, ED transfers, unplanned admissions and 30-day readmissions. Its distinctive contribution is the integration of prediction with FHIR representation, explanation, calibration, fairness auditing and human workflow. The proof-of-concept results show that transparent models can provide useful discrimination across unseen facilities, but they also demonstrate that performance, calibration and equity must be assessed jointly. The next step is prospective, multi-facility validation focused not merely on predictive accuracy but on whether the system enables earlier, safer and more equitable care.

Data availability and reproducibility statement

The empirical demonstration uses synthetic data generated for methodological evaluation. No protected health information was used. A 1,000-record sample, model-performance table and fairness-metrics table were generated as replication assets. The manuscript should be accompanied by a public code repository at submission. Kaggle datasets were used as external calibration references rather than merged at the individual level.

Ethics statement

Because the analysis used synthetic data and publicly described aggregate characteristics, institutional review-board approval was not required for this methodological proof of concept. Any prospective use with resident-level data will require applicable privacy, security, consent and research-governance review.

REFERENCES

[1] Akakpo, L., Gezah, F., Mupa, G. et al. (2026) ‘From bedside observation to systems improvement: A practical framework for strengthening resident safety and care coordination in post-acute care settings’, World Journal of Advanced Research and Reviews.

[2] Allam, A., Nagy, M., Thoma, G. and Krauthammer, M. (2019) ‘Neural networks versus logistic regression for 30 days all-cause readmission prediction’, Scientific Reports, 9, 9277.

[3] American Geriatrics Society Beers Criteria Update Expert Panel (2023) ‘American Geriatrics Society 2023 updated AGS Beers Criteria for potentially inappropriate medication use in older adults’, Journal of the American Geriatrics Society, 71(7), pp. 2052–2081.

[4] Ben-Assuli, O. and Padman, R. (2019) ‘Analysing repeated hospital readmissions using data mining techniques’, Health Systems, 8(2), pp. 93–113.

[5] Bouldin, E.D., Andresen, E.M., Dunton, N.E. et al. (2013) ‘Falls among adult patients hospitalized in the United States: prevalence and trends’, Journal of Patient Safety, 9(1), pp. 13–17.

[6] Centers for Medicare & Medicaid Services (2024) Nursing Home Quality Initiative and hospital transfer measures. Baltimore, MD: CMS.

[7] Chen, I.Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K. and Ghassemi, M. (2021) ‘Ethical machine learning in healthcare’, Annual Review of Biomedical Data Science, 4, pp. 123–144.

[8] Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F. and Sun, J. (2016) ‘Doctor AI: Predicting clinical events via recurrent neural networks’, Proceedings of Machine Learning for Healthcare, pp. 301–318.

[9] Feng, Q., Du, M., Zou, N. and Hu, X. (2023) ‘Fair machine learning in healthcare: A review’, arXiv:2206.14397.

[10] Goldstein, B.A., Navar, A.M., Pencina, M.J. and Ioannidis, J.P.A. (2017) ‘Opportunities and challenges in developing risk prediction models with electronic health records data’, Journal of the American Medical Informatics Association, 24(1), pp. 198–208.

[11] Harutyunyan, H., Khachatrian, H., Kale, D.C., Ver Steeg, G. and Galstyan, A. (2019) ‘Multitask learning and benchmarking with clinical time series data’, Scientific Data, 6, 96.

[12] HL7 International (2019) FHIR Release 4. Ann Arbor, MI: Health Level Seven International.

[13] Jiang, L.Y., Liu, X.C., Nejatian, N.P. et al. (2023) 'Health system-scale language models are all-purpose prediction engines', Nature, 619, pp. 357–362.

[14] Johnson, A.E.W., Pollard, T.J., Shen, L. et al. (2016) 'MIMIC-III, a freely accessible critical care database', Scientific Data, 3, 160035.

[15] Liu, L.T., Dean, S., Rolf, E., Simchowitz, M. and Hardt, M. (2018) 'Delayed impact of fair machine learning', Proceedings of the 35th International Conference on Machine Learning, pp. 3150–3158.

[16] Machekera, G., Mtemeli, P.S., Dongo, M.M. et al. (2026) 'An explainable deep learning framework for energy-efficient water and sewer reticulation infrastructure: Predicting project risk, pumping demand and sustainable construction performance'.

[17] Mandel, J.C., Kreda, D.A., Mandl, K.D., Kohane, I.S. and 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), pp. 899–908.

[18] Morrison, C., Lee, J., Covinsky, K. et al. (2022) 'Predicting potentially preventable hospital transfers from nursing homes', Journal of the American Medical Directors Association, 23(5), pp. 789–796.

[19] Mupa, M.N. et al. (2025) 'Predictive allocation of temperature-sensitive specialty medications: A KPI-driven framework for shortage resilience in underserved U.S. ZIP codes'.

[20] Mupa, M.N. et al. (2026a) 'Operationalizing AI at scale: Repeatable frameworks for integration, adoption and performance measurement across enterprise and startup environments', World Journal of Advanced Research and Reviews.

[21] Mupa, M.N. et al. (2026b) 'Applied machine learning frameworks for workflow optimization, organizational efficiency and digital transformation', World Journal of Advanced Research and Reviews.

[22] Na, L., Villalobos Carballo, K., Pauphilet, J. et al. (2023) 'Patient outcome predictions improve operations at a large hospital network', Manufacturing & Service Operations Management.

[23] Nakatani, H., Nakao, M., Uchiyama, H. et al. (2020) 'Predicting inpatient falls using natural language processing of nursing records obtained from Japanese electronic medical records', JMIR Medical Informatics, 8(4), e16970.

[24] National Institute of Standards and Technology (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg, MD: NIST.

[25] Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019) 'Dissecting racial bias in an algorithm used to manage the health of populations', Science, 366(6464), pp. 447–453.

[26] Ponti, M., Bet, P., Oliveira, C.L. and Castro, P.C. (2017) 'Better than counting seconds: Identifying fallers among healthy elderly using fusion of accelerometer features and dual-task Timed Up and Go', PLoS ONE, 12(4), e0175559.

[27] Rajkomar, A., Dean, J. and Kohane, I. (2019) 'Machine learning in medicine', New England Journal of Medicine, 380, pp. 1347–1358.

[28] Rajkomar, A., Hardt, M., Howell, M.D., Corrado, G. and Chin, M.H. (2018) 'Ensuring fairness in machine learning to advance health equity', Annals of Internal Medicine, 169(12), pp. 866–872.

[29] Rantz, M.J., Skubic, M., Miller, S.J. et al. (2015) 'Sensor technology to support aging in place', Journal of the American Medical Directors Association, 16(6), pp. 520–526.

[30] Shickel, B., Tighe, P.J., Bihorac, A. and Rashidi, P. (2018) 'Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis', IEEE Journal of Biomedical and Health Informatics, 22(5), pp. 1589–1604.

[31] Siontis, G.C.M., Tzoulaki, I., Castaldi, P.J. and Ioannidis, J.P.A. (2015) 'External validation of new risk prediction models is infrequent and

reveals worse prognostic discrimination', Journal of Clinical Epidemiology, 68(1), pp. 25–34.

[32] Steyerberg, E.W. (2019) Clinical Prediction Models. 2nd edn. Cham: Springer.

[33] Tinetti, M.E., Speechley, M. and Ginter, S.F. (1988) 'Risk factors for falls among elderly persons living in the community', New England Journal of Medicine, 319, pp. 1701–1707.

[34] Wiens, J., Saria, S., Sendak, M. et al. (2019) 'Do no harm: a roadmap for responsible machine learning for health care', Nature Medicine, 25, pp. 1337–1340.

[35] Zhang, Z., Ho, K.M. and Hong, Y. (2019) 'Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care', Critical Care, 23, 112.