© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

Responsible Artificial Intelligence for Polypharmacy and Medication-Safety Risk in Older Adults: An Explainable FHIR-Based Clinical Decision-Support 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- ***Medication-related harm in older adults is driven by the interaction of multimorbidity, age-related pharmacokinetic change, polypharmacy, potentially inappropriate medications, drug-drug interactions, fall-risk-increasing drugs and impaired renal function. Existing electronic prescribing alerts often treat these risks as isolated rules, generating high alert burden without ranking residents according to the combined probability and clinical significance of harm. This study develops an explainable, FHIR-based and fairness-aware clinical decision-support framework for identifying potentially inappropriate medications, high-risk combinations, excessive medication burden, medication-related fall risk, renal- and age-sensitive prescribing risks, and residents requiring pharmacist or clinical review. Because no single openly accessible Kaggle dataset contains all medication, laboratory, renal, fall and long-term-care variables needed for the full framework, a reproducible synthetic cohort of 10,000 older residents across 30 facilities was calibrated to the Kaggle Drug Adverse Event Detection Dataset, public adverse-drug-event datasets and relationships reported in geriatric pharmacotherapy literature. Logistic regression, random forest and gradient boosting models were compared using facility-level holdout validation. Performance was assessed with AUROC, area under the precision-recall curve, Brier score, sensitivity, specificity, positive predictive value and F1 score. Fairness was evaluated by sex, race and chronic-kidney-disease status at a capacity-constrained top-quintile review threshold. Logistic regression produced the most transportable discrimination across medication harm, medication-related falls and pharmacist-review need, while calibration and subgroup analysis showed that average performance alone was insufficient for safe deployment. The framework maps medication and clinical inputs to FHIR resources and produces an explanation packet that identifies risk level, contributing medicines, interaction mechanisms, renal-dose concerns, data quality and recommended human review. The study offers a transparent foundation for prospective validation in nursing homes, ambulatory geriatric practice and post-acute care.***

Keywords: ***Polypharmacy, Medication Safety, Older Adults, Potentially Inappropriate Medications, Adverse Drug Events, Falls, Renal Dosing, Explainable Artificial Intelligence, FHIR, Clinical Decision Support.***

I. INTRODUCTION

Older adults experience a disproportionate burden of medication-related harm. Multimorbidity often requires several evidence-based therapies, but physiological ageing changes drug absorption, distribution, metabolism and elimination. Reduced renal reserve, altered body composition, cognitive impairment and frailty can transform a standard regimen into a high-risk regimen. Polypharmacy is therefore not intrinsically inappropriate; the central problem is whether the combined regimen remains necessary, tolerable, correctly dosed and aligned with the resident's goals of care.

Potentially inappropriate medications are one component of this problem. The 2023 American Geriatrics Society Beers Criteria identify medicines and combinations for which risks may exceed benefits in many adults aged 65 years and older. The criteria include drugs to avoid generally, drugs to

avoid in selected conditions, drug-drug interactions, medicines requiring renal-dose adjustment and drugs that increase falls or other geriatric syndromes. Explicit criteria are valuable, but they cannot replace individual clinical judgement because indication, dose, duration, prognosis and previous treatment response matter.

Medication safety systems commonly generate one alert per rule. A resident receiving an anticoagulant, non-steroidal anti-inflammatory drug, diuretic, antidepressant and benzodiazepine may therefore generate several disconnected warnings. Clinicians must mentally combine bleeding risk, renal risk, orthostasis, sedation and fall history. Repeated low-value alerts can cause alert fatigue, whereas a resident-level risk summary can prioritise review and explain why the regimen is hazardous.

Artificial intelligence offers a means of combining medication, diagnosis, laboratory, renal-function, fall-history and clinical-observation data. Responsible use requires more than predictive accuracy. The model must be clinically interpretable, interoperable, calibrated, monitored across population groups and embedded in pharmacist or prescriber review. Research on organisational AI adoption stresses that operational value depends on data governance, workflow integration, accountability and performance monitoring, not merely the selection of an algorithm (Mupa et al., 2026a; Mupa et al., 2026b).

This article proposes and tests a medication-specific framework that differs from general adverse-event prediction. It explicitly represents potentially inappropriate medications, high-risk combinations, excessive medication burden, fall-risk-increasing drugs, renal mismatch, anticholinergic burden and sedative burden. It then predicts three actionable outcomes: medication-related harm within 30 days, medication-related fall within 30 days and need for pharmacist or clinical review.

II. RESEARCH OBJECTIVES AND QUESTIONS

The study aimed to: (1) construct an interoperable medication-safety feature model; (2) compare interpretable and non-linear predictive models; (3) evaluate calibration and subgroup error profiles; (4) design explainable review outputs; and (5) specify a FHIR-based implementation and governance pathway.

RQ1: Can a combined representation of medications, diagnoses, renal function, laboratory values, fall history and clinical observations predict medication-related harm and review need?

RQ2: Which medication-safety domains contribute most strongly to risk?

RQ3: Does predictive performance vary by sex, race or chronic-kidney-disease status?

RQ4: How can the model's inputs, outputs and explanations be represented in FHIR for auditable clinical decision support?

III. LITERATURE REVIEW

3.1 Polypharmacy and potentially inappropriate medication use

Polypharmacy is commonly defined as the use of five or more medicines, while excessive polypharmacy is often defined as ten or more. These thresholds are screening concepts rather than clinical verdicts. Appropriate polypharmacy may improve survival and function, but inappropriate polypharmacy increases treatment burden, non-adherence, prescribing cascades and adverse drug reactions. Reviews consistently identify medication count, frailty, multimorbidity and transitions of care as important risk factors (Davies et al., 2020; Zazzara et al., 2021). The Beers Criteria and STOPP/START criteria translate accumulated evidence and expert consensus into explicit review prompts. Their strongest use is to identify cases requiring individual reassessment. Digital systems that incorporate these criteria can improve prescribing practices, but evidence for reductions in hard clinical outcomes remains mixed

and depends heavily on implementation, clinician acceptance and the availability of actionable alternatives (Marasinghe et al., 2015; Ng et al., 2025).

3.2 Drug-drug interactions and cumulative burden

Drug-drug interactions increase combinatorially as medication counts rise. Particularly important geriatric combinations include opioids with benzodiazepines, anticoagulants with NSAIDs, multiple central-nervous-system depressants, medicines that prolong the QT interval and combinations that increase hyperkalaemia or acute kidney injury. Large EHR analyses show that interaction exposure differs by age and sex and cannot be explained solely by medication count (Correia et al., 2019). Network and representation-learning studies further show that domain-specific drug-interaction models can improve detection, although clinical validation and explanation remain essential (Wang and Taylor, 2023).

3.3 Medication-related falls

Falls are a clinically important medication-safety outcome. Polypharmacy has been associated with higher fall rates, and the combination of persistent polypharmacy with fall-risk-increasing drugs appears particularly hazardous (Dhalwani et al., 2017; Xue et al., 2021). Benzodiazepines, antipsychotics, antidepressants, opioids, gabapentinoids and blood-pressure-lowering drugs can contribute through sedation, impaired attention, gait instability or orthostatic hypotension. A medication-specific model should therefore combine drug exposure with prior falls, functional vulnerability and clinical observations such as dizziness or somnolence.

3.4 Renal and age-sensitive prescribing

Renal function commonly declines with age, and serum creatinine may underestimate impairment in residents with low muscle mass. Renally cleared medicines may accumulate, while nephrotoxic combinations can precipitate acute kidney injury. Dosing decisions should consider estimated filtration or creatinine clearance, indication, body size and the specific product label. Studies continue to identify clinically important discordance between prescribed doses and renal recommendations among older adults (Munar and Singh, 2007; AlQashqri et al., 2023; Očovská et al., 2024).

3.5 Explainable and interoperable clinical decision support

Medication review is cognitively demanding. Simulation evidence suggests that visual, intelligent decision support can help pharmacists identify more drug-related problems without increasing review time (Mouazer et al., 2024). Explainability is crucial because the safest action may be dose reduction, monitoring, substitution, deprescribing, continuation with justification or no change. The system must expose the medicines and clinical facts driving the alert.

FHIR provides a standard representation for medication orders, administrations, conditions, observations, encounters and risk assessments. A FHIR-native service can retrieve data across vendors and return a structured safety assessment. The architecture must preserve provenance and terminology mapping. Analogous explainable-risk frameworks developed in other regulated domains emphasise linking model drivers to operational controls and documented review, a principle directly applicable to medication safety (Machekera et al., 2026).

IV. METHODS

4.1 Data strategy

A targeted search identified the Kaggle Drug Adverse Event Detection Dataset, which simulates patients receiving multiple medicines and adverse outcomes, together with other public adverse-drug-reaction resources. No accessible Kaggle dataset contained the full combination of long-term-care medication lists, diagnoses, laboratory values, renal function, fall history, clinical observations and adjudicated medication harm. Individual-level merging of unrelated datasets would be methodologically invalid. The empirical study therefore used a reproducible synthetic cohort calibrated to the Kaggle dataset structure and peer-reviewed geriatric medication-safety literature. The analysis is a proof of concept and does not claim clinical validation.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

4.2 Cohort and variables

The cohort contained 10,000 residents aged 65 years and older nested within 30 facilities. Demographic variables included age, sex and race. Clinical variables included dementia, diabetes, heart failure, atrial fibrillation, chronic kidney disease, Charlson comorbidity index, prior falls, systolic blood pressure and orthostasis. Laboratory variables included eGFR, creatinine, potassium, sodium and haemoglobin. Medication exposures represented benzodiazepines, antipsychotics, antidepressants, opioids, anticholinergics, diuretics, ACE inhibitors or ARBs, anticoagulants, NSAIDs, sulfonylureas, insulin and gabapentinoids.

Table 1. Proposed FHIR representation of medication-safety data

Domain	FHIR resource	Illustrative content
Medication list and dose	MedicationRequest; MedicationAdministration	Active ingredient, dose, route, frequency, status
Diagnoses	Condition	Dementia, diabetes, heart failure, atrial fibrillation, CKD
Laboratory values	Observation	eGFR, creatinine, potassium, sodium, haemoglobin
Falls and encounters	AdverseEvent; Encounter	Prior falls, ED visits and admissions
Clinical observations	Observation; DocumentReference	Dizziness, confusion, somnolence, poor intake, bleeding
Safety assessment	RiskAssessment; DetectedIssue	Probability, risk category, contributing factors
Clinical work queue	Task; CarePlan	Pharmacist review, prescriber response, monitoring plan
Provenance	Provenance	Source, transformation, model version and reviewer

Figure 1. FHIR-based medication-safety decision-support architecture

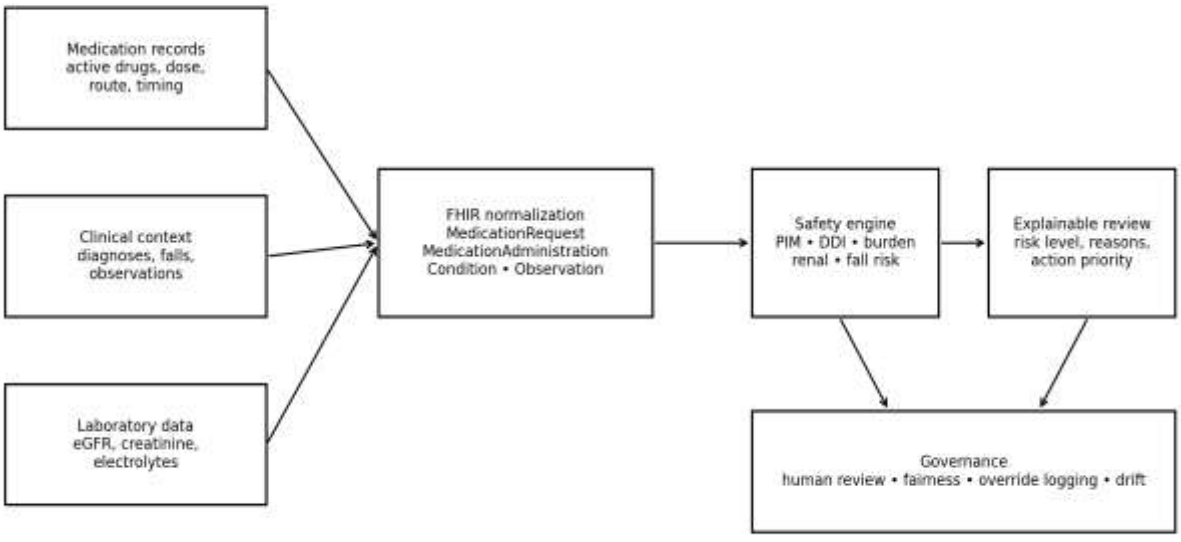

Figure 1. FHIR-based medication-safety clinical decision-support architecture.

4.3 Medication-safety feature engineering

Six explicit safety domains were constructed. Potentially inappropriate medication count represented selected high-risk classes. Drug-interaction count represented clinically important co-exposures. Excessive medication burden was defined as ten or more concurrent medicines. Fall-risk-increasing drug count included psychotropic, opioid, diuretic and gabapentinoid exposure. Renal mismatch indicators represented selected medicines used below prespecified eGFR thresholds. Anticholinergic and sedative burden scores captured cumulative pharmacodynamic effects. Clinical-note concepts represented dizziness, confusion, somnolence, poor intake and bleeding.

Table 2. Medication-safety constructs used by the framework

Safety construct	Operational definition
Potentially inappropriate medication	Benzodiazepine, antipsychotic, anticholinergic, selected NSAID or sulfonylurea exposure
High-risk combination	Opioid-benzodiazepine; anticoagulant-NSAID; multiple CNS-active drugs; renal/electrolyte combinations
Excessive medication burden	Ten or more active medicines
Medication-related fall risk	FRID count plus prior falls, orthostasis, dizziness and dementia
Renal-sensitive risk	Low eGFR with NSAID, gabapentinoid, sulfonylurea, ACE/ARB or diuretic exposure
Pharmacist review priority	Combined predicted risk, rule severity, trend and clinical observations

4.4 Outcomes and modelling

The three outcomes were: medication-related harm within 30 days, medication-related fall within 30 days and pharmacist or clinical review required. The synthetic outcome functions incorporated medication burden, interactions, renal mismatch, comorbidity, prior falls and clinical observations. Facilities were divided into a 75% training set and 25% held-out test set so that performance reflected transportability to unseen organisational settings.

Logistic regression, random forest and histogram gradient boosting were compared. Missing continuous values were median-imputed, categorical variables were mode-imputed and one-hot encoded, and continuous variables were standardised for logistic regression. AUROC was the primary discrimination metric. AUPRC, Brier score, sensitivity, specificity, positive predictive value and F1 score were also reported.

4.5 Explainability and fairness

Permutation importance measured the reduction in held-out AUROC after shuffling each variable. A deployment system would produce resident-level explanations identifying the medicines, laboratory values and observations that increased risk. Fairness was assessed across sex, race and CKD status at a top-quintile review threshold. The analysis reported event rate, alert rate, sensitivity, false-positive rate, positive predictive value and Brier score. These metrics were interpreted jointly because equal alert rates are not necessarily clinically fair when baseline risk differs.

V. RESULTS

Table 3. Characteristics of the synthetic older-adult cohort

Characteristic	Value
Residents	10,000
Facilities	30
Mean age	81.7 (SD 8.0)
Female	62.2%
Mean medication count	9.9
Polypharmacy (≥5)	96.1%
Excessive polypharmacy (≥10)	52.9%
At least one PIM indicator	56.7%
At least one DDI pattern	18.3%
Renal mismatch indicator	7.9%
30-day medication harm	33.5%
30-day medication-related fall	7.6%
Pharmacist review required	37.8%

The mean medication count was 9.9. Polypharmacy was intentionally common in this high-risk cohort,

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

with 52.9% meeting the excessive-polypharmacy threshold. More than half had at least one potentially inappropriate medication indicator, 18.3% had at least one modelled drug-interaction pattern and 7.9% had a renal mismatch indicator. These are synthetic calibration values, not national estimates.

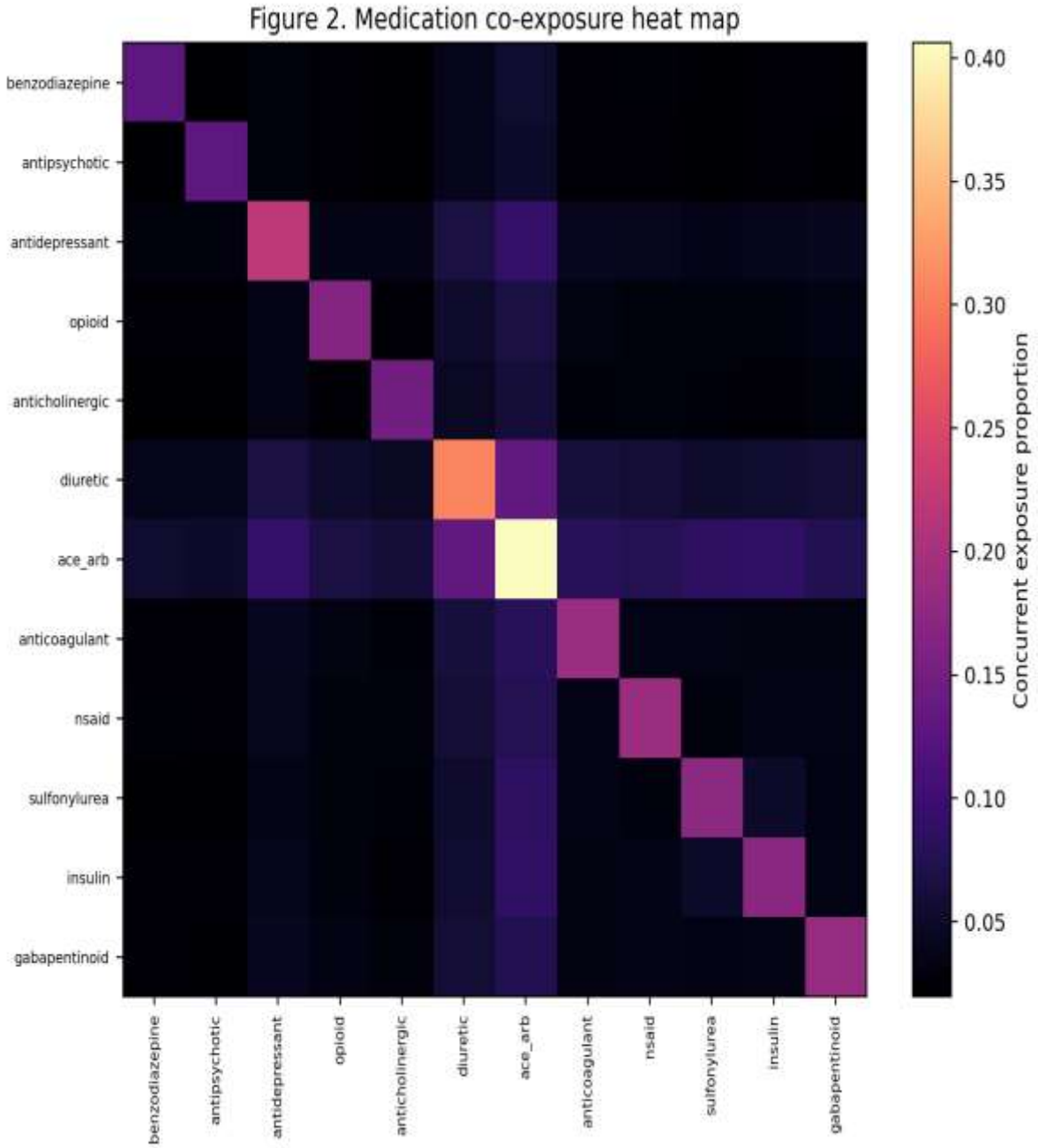

Figure 2. Medication co-exposure heat map. Darker cells indicate more frequent concurrent exposure.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

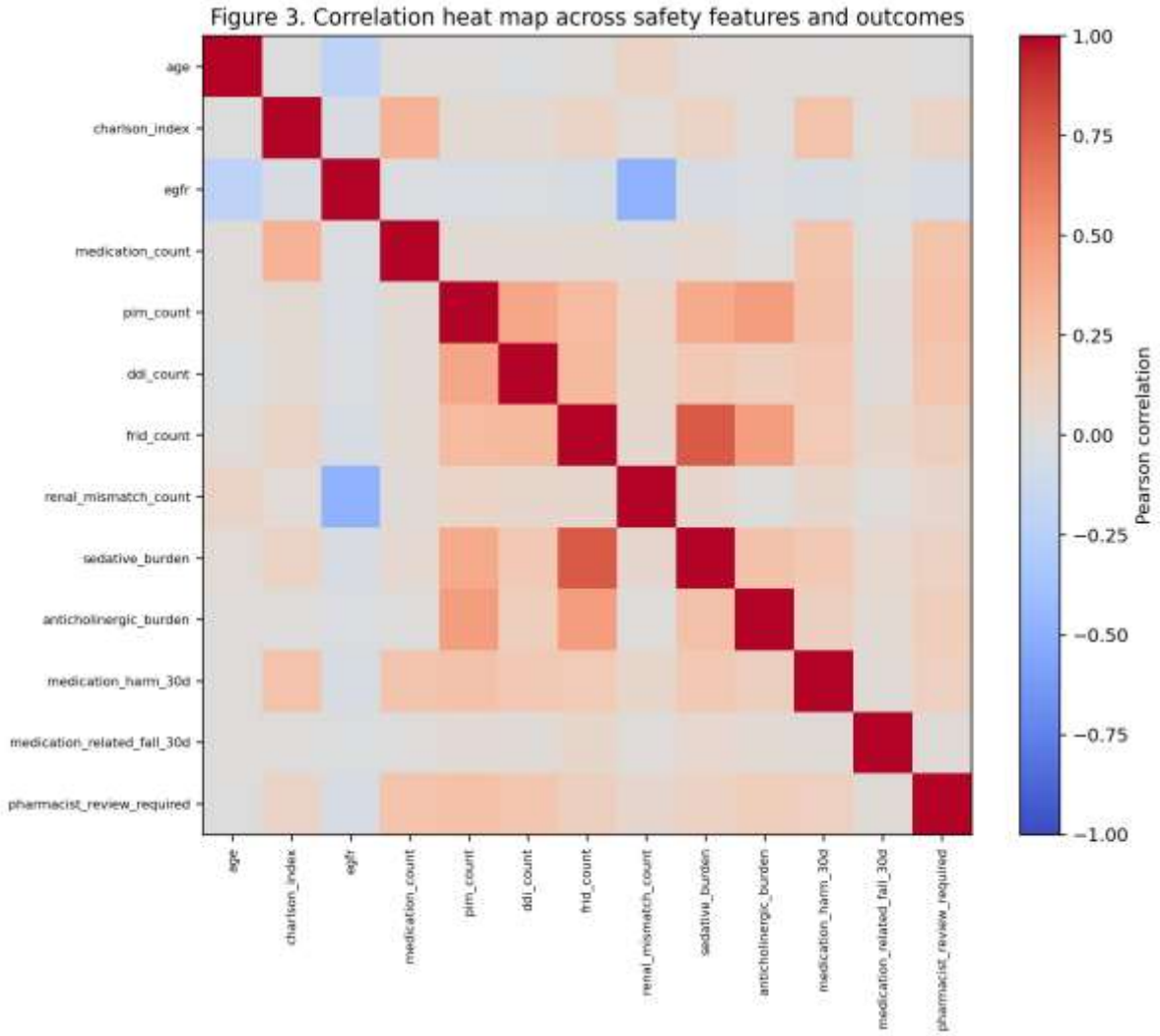

Figure 3. Correlation heat map across medication-safety features and outcomes.

Table 4. Out-of-facility model performance

Outcome	Model	AUROC	AUPRC	Brier	Sensitivity	Specificity	PPV	F1
Medication Harm 30D	Logistic regression	0.728	0.606	0.209	0.643	0.693	0.526	0.579
Medication Harm 30D	Random forest	0.714	0.587	0.208	0.581	0.734	0.536	0.557
Medication Harm 30D	Gradient boosting	0.714	0.584	0.196	0.372	0.887	0.636	0.469
Medication Related Fall 30D	Logistic regression	0.614	0.129	0.236	0.520	0.632	0.106	0.175
Medication Related Fall 30D	Random forest	0.583	0.115	0.138	0.020	0.989	0.129	0.034
Medication Related Fall 30D	Gradient boosting	0.571	0.110	0.072	0.005	0.999	0.333	0.010
Pharmacist Review Required	Logistic regression	0.710	0.599	0.217	0.624	0.683	0.539	0.578

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

Pharmacist Review Required	Random forest	0.701	0.588	0.216	0.589	0.703	0.541	0.564
Pharmacist Review Required	Gradient boosting	0.693	0.572	0.210	0.418	0.828	0.591	0.490

Logistic regression achieved the highest AUROC for all three outcomes. The result indicates that clinically structured safety features carried stable signal across facilities and that increased algorithmic complexity did not improve transportability. Random forest produced similar performance for medication harm and review need but lower sensitivity at the default threshold. Medication-related fall prediction was more difficult because falls are influenced by environment, mobility and chance in addition to medication exposure.

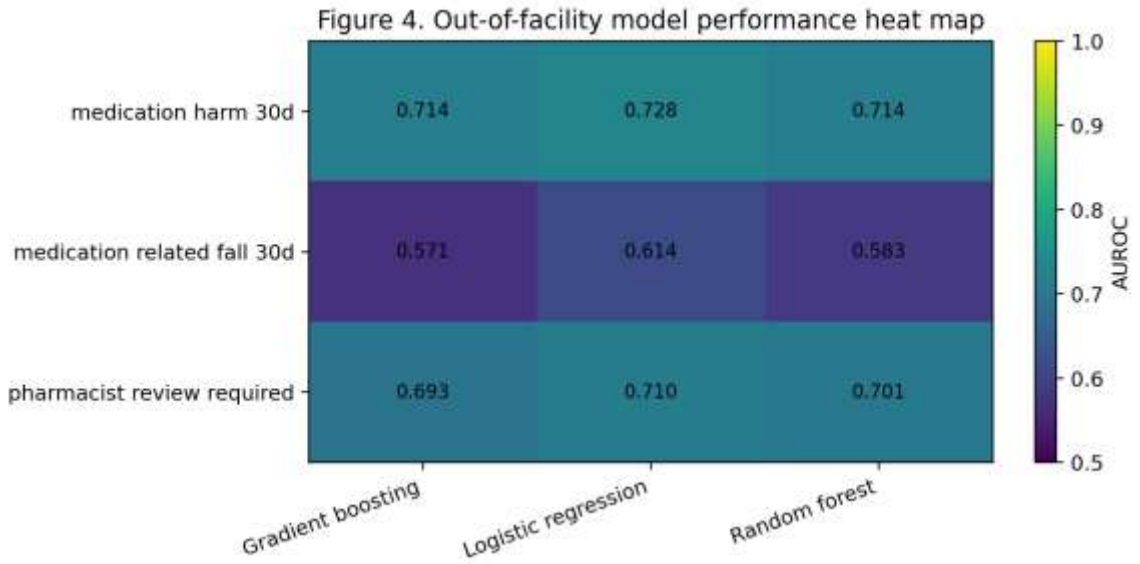

Figure 4. Out-of-facility model performance heat map.

Table 5. Best-performing model for each outcome

Outcome	Model	AUROC	AUPRC	Brier	Sensitivity	Specificity	PPV	F1
Medication Harm 30D	Logistic regression	0.728	0.606	0.209	0.643	0.693	0.526	0.579
Medication Related Fall 30D	Logistic regression	0.614	0.129	0.236	0.520	0.632	0.106	0.175
Pharmacist Review Required	Logistic regression	0.710	0.599	0.217	0.624	0.683	0.539	0.578

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

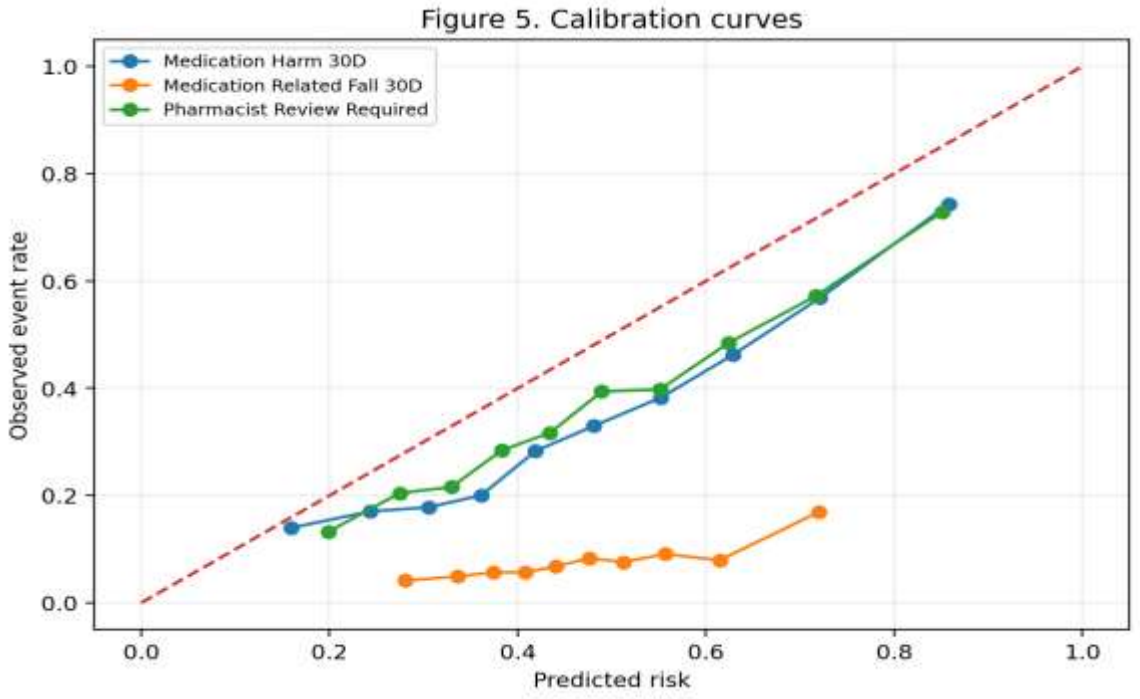

Figure 5. Calibration curves for the best-performing models.

Calibration was broadly monotonic, but higher-risk strata showed outcome-specific deviations. This supports recalibration before local deployment.

Review thresholds should be based on clinical capacity and expected benefit rather than a universal probability cut-off.

Table 6. Race-stratified fairness metrics at the top-quintile threshold

Outcome	Group	N	Event rate	Alert rate	Sensitivity	False positive rate	PPV	Brier
Medication Harm 30D	Asian	136	0.360	0.250	0.429	0.149	0.618	0.224
Medication Harm 30D	Black	429	0.338	0.217	0.393	0.127	0.613	0.212
Medication Harm 30D	Hispanic	307	0.303	0.215	0.430	0.121	0.606	0.208
Medication Harm 30D	Other	63	0.413	0.175	0.346	0.054	0.818	0.194
Medication Harm 30D	White	1708	0.352	0.190	0.366	0.095	0.677	0.208
Medication Related Fall 30D	Asian	136	0.088	0.096	0.333	0.073	0.308	0.166
Medication Related Fall 30D	Black	429	0.089	0.168	0.237	0.161	0.125	0.219
Medication Related Fall 30D	Hispanic	307	0.065	0.228	0.250	0.226	0.071	0.250
Medication Related Fall 30D	Other	63	0.063	0.095	0.250	0.085	0.167	0.200
Medication Related Fall 30D	White	1708	0.076	0.215	0.362	0.203	0.128	0.244
Pharmacist Review Required	Asian	136	0.353	0.184	0.417	0.057	0.800	0.206
Pharmacist Review	Black	429	0.385	0.219	0.370	0.125	0.649	0.220

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

Required								
Pharmacist Review Required	Hispanic	307	0.407	0.228	0.376	0.126	0.671	0.219
Pharmacist Review Required	Other	63	0.365	0.302	0.565	0.150	0.684	0.225
Pharmacist Review Required	White	1708	0.366	0.188	0.325	0.109	0.632	0.216

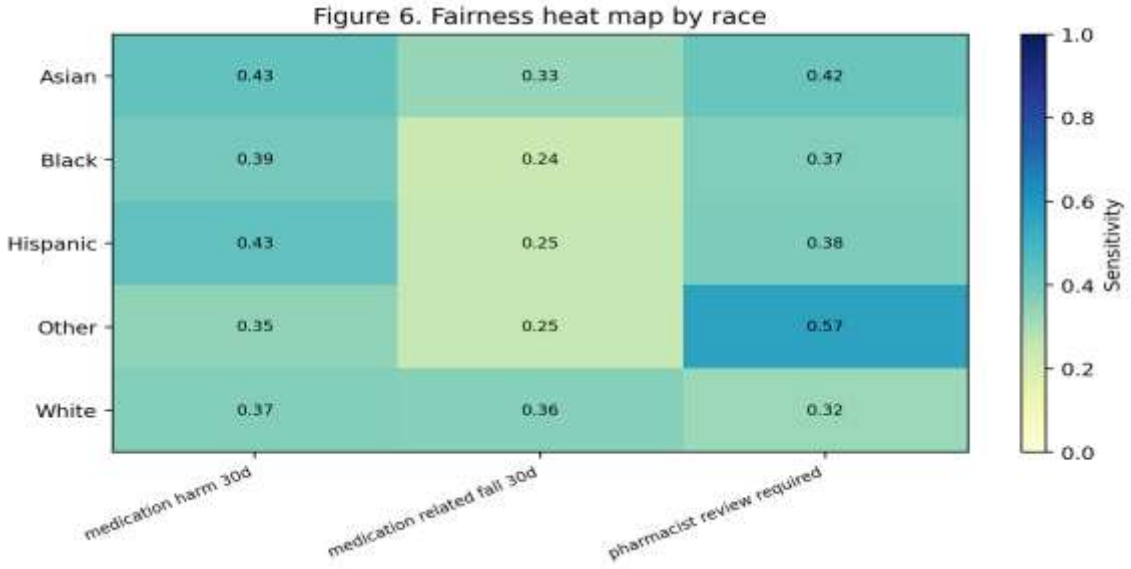

Figure 6. Fairness heat map showing sensitivity by race and outcome.

Sensitivity and false-positive rates varied across subgroups. Differences should trigger data-quality and workflow investigation rather than automatic mathematical equalisation. A lower sensitivity may reflect under-documentation, different disease burden, small sample size or model misspecification. Responsible deployment requires repeated subgroup monitoring and assessment of whether flagged residents actually receive beneficial review.

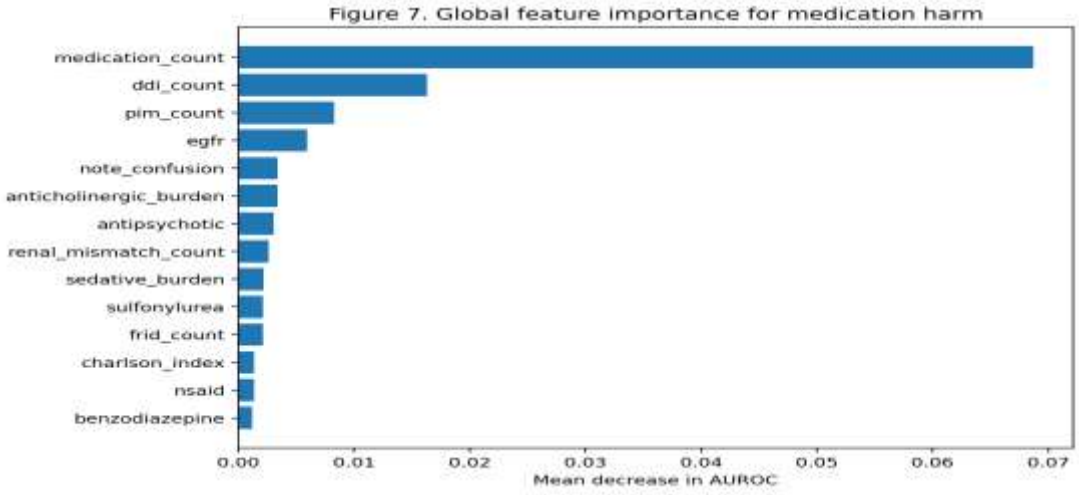

Figure 7. Global feature importance for 30-day medication-related harm.

Medication count, drug-interaction count, potentially inappropriate medication burden, renal mismatch, comorbidity and clinical observations were among the strongest contributors. These features are clinically intelligible and can support a transparent explanation. The model should not simply state that a resident is high risk; it should specify which medicines and physiological conditions create the concern.

Table 7. Illustrative explainable medication-safety alert

Component	Illustrative content
Priority	High: pharmacist review within one business day
Predicted risk	Medication-harm probability 0.61 over 30 days
Principal drivers	Anticoagulant plus NSAID; eGFR 34; 12 active medicines; recent bleeding note
Fall-related drivers	Benzodiazepine, antidepressant, prior fall and orthostatic symptoms
Data-quality warning	Weight recorded 54 days ago; reconciliation incomplete after hospital return
Suggested review	Verify indication and duration; assess bleeding; review NSAID alternative; renal-dose check; reconcile list
Human options	Accept, modify, defer with reason, or close as clinically justified
FHIR output	RiskAssessment, DetectedIssue and Task linked to medication and laboratory sources

VI. CLINICAL DECISION-SUPPORT FRAMEWORK

6.1 Hybrid rules and predictive analytics

A safe system should combine deterministic rules with predictive modelling. Rules are appropriate for explicit contraindications, duplicate therapy, severe interactions and renal thresholds. Predictive models are useful for prioritising among multiple moderate risks and incorporating resident vulnerability. The system should retain the rule trace so that a pharmacist can distinguish an evidence-based contraindication from a statistical risk contribution.

6.2 Review prioritisation

Residents should be ranked by expected clinical benefit of review, not medication count alone. A practical priority score can combine predicted harm, rule severity, recent symptom change, renal decline, previous falls and absence of recent pharmacist review. High-priority cases should enter a shared work queue with target response times. Lower-priority issues can be reviewed during routine medication-regimen review.

6.3 Renal and age-sensitive logic

Renal alerts should display the laboratory value, date, trend and the renal-estimation method. The system should avoid presenting eGFR as an infallible dose recommendation, especially in frail residents with low muscle mass. Age-sensitive alerts should consider indication, dose, duration and alternatives. For example, a benzodiazepine alert should distinguish long-term routine use from short-term treatment under documented palliative goals.

6.4 Pharmacist and prescriber workflow

The preferred workflow is co-pilot review. The model identifies and explains risk; a pharmacist validates the medication list, clinical context and alternatives; and the prescriber makes the final treatment decision. Evidence from medication-review decision support suggests that intelligent visual tools can improve problem identification without increasing review time when they are well integrated (Mouazer et al., 2024).

6.5 FHIR implementation

MedicationRequest and MedicationAdministration should represent ordered and administered therapy. Observation should carry eGFR, creatinine, electrolytes, blood pressure and relevant symptom observations. Condition and Encounter provide diagnoses and care transitions. RiskAssessment records the prediction, horizon and model version; DetectedIssue records the medication-safety concern; Task routes the review; and Provenance records source and transformation. Every alert should link directly to the evidence used to create it.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1720286

VII. RESPONSIBLE-AI GOVERNANCE

Table 8. Minimum responsible-AI governance controls

Governance domain	Minimum requirement
Intended use	Medication-review prioritisation; not autonomous prescribing or deprescribing
Clinical validation	Prospective silent validation followed by controlled pilot deployment
Data governance	Medication reconciliation, terminology mapping, laboratory freshness and provenance
Model governance	Version control, calibration, drift, facility and subgroup performance
Human oversight	Pharmacist or qualified clinician review before medication change
Alert governance	Severity tiers, suppression logic, override reasons and workload monitoring
Fairness	Sensitivity, false positives, calibration and intervention completion by subgroup
Safety monitoring	Missed severe interactions, delayed escalation and unintended discontinuation
Resident-centredness	Goals of care, symptom burden, preferences and treatment burden

Operationalising AI at scale requires repeatable integration, ownership and performance measurement. Mupa et al. (2026a) frame this as a socio-technical implementation challenge, while Mupa et al. (2026b) emphasise alignment between machine-learning outputs and workflow redesign. In medication safety, this means that model performance, pharmacist capacity, prescriber response and resident outcomes must be monitored as one system.

VIII. DISCUSSION

The framework demonstrates that medication-specific feature engineering can support explainable risk prioritisation across facilities. The best-performing model was logistic regression, reinforcing that a transparent and calibrated baseline may be preferable to a more complex model when the principal requirement is transportability and clinical review. The analysis also confirms that medication-related falls are harder to predict than general medication harm because environmental and functional factors contribute substantially.

The central contribution is not a new adverse-event classifier in isolation. It is a structured medication-safety architecture that integrates explicit geriatric criteria, interaction patterns, cumulative burden, renal risk, fall risk and observed symptoms. This allows the system to answer a clinically relevant question: which resident's medication regimen should be reviewed first, and why?

The fairness findings underscore that responsible AI cannot be reduced to excluding protected attributes. Documentation, access, disease prevalence and treatment patterns can create unequal error rates. Facilities should audit both prediction and intervention: a fair alert system is of limited value if some residents are less likely to receive pharmacist review or if recommended alternatives are unavailable.

The FHIR design supports portability and auditability. It also creates a path toward shared decision support across nursing homes, ambulatory practices, pharmacies and hospitals. Medication reconciliation after transitions is especially important because discrepancies, temporary medicines and discontinued drugs can generate false alerts or conceal actual risk.

IX. LIMITATIONS

The cohort is synthetic and the reported metrics cannot establish clinical effectiveness or prevalence.
The Kaggle adverse-event dataset was used as a structural calibration reference, not as a complete long-term-care cohort.

The drug classes and interaction rules are simplified and do not reproduce the full Beers Criteria, STOPP/START criteria or product-labelling requirements.

Medication-related harm was generated from predefined relationships and was not independently adjudicated.

Race categories are simplified and fairness estimates may be unstable for smaller groups.

Renal dosing depends on the medicine, indication, body size and estimation method; the framework must therefore support clinician verification.

The study does not model adherence, over-the-counter medicines, supplements, pharmacogenomics or goals-of-care complexity.

X. PROSPECTIVE VALIDATION AGENDA

Table 9. Proposed prospective evaluation programme

Stage	Purpose
Silent validation	Assess discrimination, calibration, rule accuracy and alert volume without displaying alerts
Pharmacist adjudication	Classify true medication-related problems, severity, preventability and recommended action
Pilot deployment	Measure acceptance, time to review, intervention type and unintended consequences
Pragmatic trial	Compare medication harm, falls, ED transfers and hospitalisation under usual care versus supported review
Fairness analysis	Evaluate model and intervention performance across race, sex, CKD, dementia and socioeconomic groups
Economic analysis	Estimate pharmacist time, avoided harm, hospital use and implementation cost

The primary endpoint should be clinically adjudicated preventable medication harm. Secondary endpoints should include injurious falls, severe drug interactions, renal-dose corrections, deprescribing where appropriate, hospital utilisation, alert burden, time to review and resident-reported treatment burden.

XI. CONCLUSION

Older adults require a medication-safety system that recognises the difference between necessary polypharmacy and hazardous polypharmacy. The proposed framework combines medication records, diagnoses, laboratory values, renal function, fall history and clinical observations to identify potentially inappropriate medications, high-risk combinations, excessive burden, fall risk, renal mismatch and residents who require pharmacist or clinical review. Its distinguishing features are explainability, FHIR interoperability, fairness monitoring and explicit human oversight. The proof-of-concept results support prospective testing, but deployment should proceed only through rigorous validation, transparent governance and pharmacist-prescriber collaboration.

Data availability and ethics

The empirical dataset was synthetically generated and contains no protected health information. A sample dataset, model-performance results and fairness metrics were generated as replication assets. Prospective studies using resident-level records will require appropriate privacy, security and institutional review.

REFERENCES

[1] AlQashqri, H. et al. (2023) ‘Renally inappropriate medications in the old population’, Journal of Clinical Medicine, 12.

[2] 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.

[3] Correia, R.B., de Araújo, L.P., Mattos, M.M. and Rocha, L.M. (2019) ‘City-wide analysis of electronic health records reveals gender and age biases in the administration of known drug-drug interactions’, npj Digital Medicine, 2, 74.

[4] Davies, L.E., Spiers, G., Kingston, A., Todd, A. and Adamson, J. (2020) ‘Adverse outcomes of polypharmacy in older people: systematic review of reviews’, Journal of the American

Medical Directors Association, 21(2), pp. 181–187.

[5] Dhalwani, N.N. et al. (2017) 'Association between polypharmacy and falls in older adults: a longitudinal study from England', BMJ Open, 7, e016358.

[6] Fick, D.M. et al. (2019) 'American Geriatrics Society 2019 updated AGS Beers Criteria', Journal of the American Geriatrics Society, 67(4), pp. 674–694.

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

[8] 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.

[9] 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'.

[10] Marasinghe, K.M. et al. (2015) 'Computerised clinical decision support systems to improve medication safety in long-term care homes: a systematic review', BMJ Open, 5, e006539.

[11] Mouazer, A. et al. (2024) 'A randomized simulation trial evaluating ABiMed, a clinical decision support system for medication reviews and polypharmacy management', preprint.

[12] Munar, M.Y. and Singh, H. (2007) 'Drug dosing adjustments in patients with chronic kidney disease', American Family Physician, 75(10), pp. 1487–1496.

[13] 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.

[14] 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.

[15] Ng, Y. et al. (2025) 'Evaluating the role of clinical decision support systems in improving prescribing for older people: a systematic review', BMC Geriatrics.

[16] 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.

[17] Očovská, Z. et al. (2024) 'Renal drug dosage adjustments and adverse drug events in older adults', Expert Opinion on Drug Safety.

[18] Page, A.T. et al. (2023) 'Polypharmacy and medicine regimens in older adults in residential aged care', Archives of Gerontology and Geriatrics.

[19] 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.

[20] Shah, B.M. and Hajjar, E.R. (2012) 'Polypharmacy, adverse drug reactions, and geriatric syndromes', Clinics in Geriatric Medicine, 28(2), pp. 173–186.

[21] Wang, N. and Taylor, C.O. (2023) 'Drug Interaction Vectors Neural Network: DrIVeNN', arXiv:2308.13891.

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

[23] Xue, L. et al. (2021) 'Persistent polypharmacy and fall injury risk: the Health, Aging and Body Composition Study', BMC Geriatrics, 21, 710.

[24] Zazzara, M.B. et al. (2021) 'Adverse drug reactions in older adults: a narrative review of the literature', European Geriatric Medicine, 12, pp. 463–473.