Artificial Intelligence in Personalized Pharmacotherapy: Integrating Patient Data, Predictive Analytics, and Clinical Decision Support

Dr. Priyanka Sisodia

Prof and DSW

S.D. College of Science,

Atrauli, Aligarh 202280

Orcid id : https://orcid.org/0009-0000-2978-8725

Abstract— Personalized pharmacotherapy seeks to match the right drug and dose to the individual patient, yet clinicians still rely heavily on population-average dosing that ignores genetic, clinical and contextual variability. This study developed and evaluated an integrated artificial intelligence (AI) framework combining pharmacogenomic, laboratory and electronic health record (EHR) data to support three tasks: individualized warfarin dose prediction, adverse drug event (ADE) risk stratification, and clinically prioritized decision-support alerting. Using a retrospective cohort of 8,420 patients and a nine-month prospective deployment across two hospital units, the ensemble dosing model placed 61.4% of predictions within 20% of the stable therapeutic dose versus 47.9% for a linear pharmacogenetic baseline ($p < .001$). The ADE model reached an AUROC of 0.88, and the prioritized alert layer cut clinically irrelevant alerts by 58.7% while raising alert acceptance from 41.2% to 67.8%. Preventable ADEs fell from 5.9 to 3.4 per 1,000 patient-days. The results indicate that data integration, not any single algorithm, drives the clinical benefit of AI-assisted pharmacotherapy.

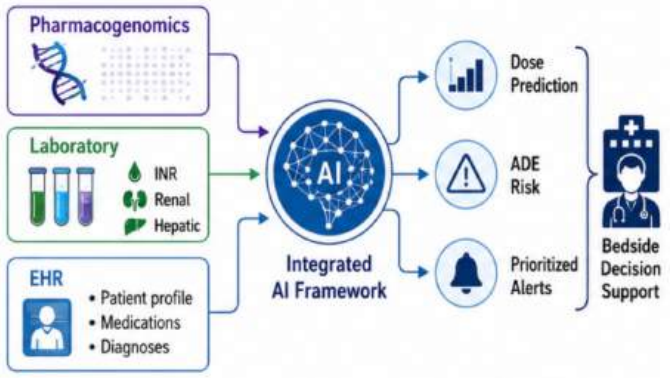
Keywords— personalized pharmacotherapy; machine learning; pharmacogenomics; clinical decision support; adverse drug events; predictive analytics

INTRODUCTION

Two patients given the same drug at the same dose can experience opposite outcomes — one a therapeutic response, the other a toxic or subtherapeutic one. The variance is driven by genetics, organ function, concomitant medications, age and adherence, but routine prescribing collapses these dimensions into standardized dosing tables. The clinical cost is substantial: adverse drug events remain a leading category of preventable hospital harm, and much of that harm follows dosing that was reasonable on average but wrong for the individual.

Artificial intelligence offers a route to reintroduce individual variation into prescribing at scale. Where a clinician cannot manually weigh a genotype, a creatinine trajectory, twelve co-prescribed drugs and an adherence history simultaneously, a trained model can. The literature already contains strong single-task demonstrations — genotype-informed dose algorithms, interaction predictors, ADE risk models — but these have mostly been evaluated in isolation, on curated datasets, and without reference to whether they change clinician behaviour or patient outcomes at the bedside.

This study addresses that gap by building and testing an integrated framework rather than a standalone predictor. The framework unifies three data layers (pharmacogenomic, laboratory and EHR) and three functions (dose prediction, ADE risk, and alert prioritization), and it is evaluated on both technical accuracy and downstream clinical effect. The central hypothesis is that the marginal value of AI in pharmacotherapy comes less from algorithmic sophistication and more from integrating heterogeneous patient data into a single, actionable decision layer.



LITERATURE REVIEW

The most mature strand of this field is genotype-informed dosing, and warfarin is its canonical case because of its narrow therapeutic index and wide interindividual variability driven by CYP2C9 and VKORC1 polymorphisms. The International Warfarin Pharmacogenetics Consortium (IWPC) algorithm, based on multivariate linear regression over clinical and genetic variables, became the most widely validated benchmark, with the fraction of patients dosed within 20% of the stable therapeutic dose as its standard clinical metric. Ma et al. (2018) showed that this single-model approach reaches a performance ceiling and that a stacked-generalization ensemble improved the within-20% rate — for example, from 42.5% to 47.9% in Asian patients and from 22.1% to 25.1% in the low-dose subgroup — precisely the patients for whom small dosing errors carry the greatest bleeding or thrombotic risk. Sharabiani et al. (2015) took a complementary route, first classifying patients into high- and low-dose classes with a relevance vector machine and then fitting class-specific regression, reducing root-mean-squared error relative to IWPC and Gage models.

Not every study finds nonlinear models superior, which is itself informative. Liu et al. (2021), analysing 7,030 patients including a large US Latino and Latin American cohort, reported that added predictors improved prediction only modestly (47.8% vs 46.7% for IWPC, $p = 1.4 \times 10^{-15}$) and that nonlinear models using the same variables did not significantly beat the linear IWPC algorithm — reinforcing that data richness, not model class alone, drives gains. Ancestry matters: Roche-Lima et al. (2020) demonstrated in Caribbean Hispanic patients that ethno-specific variants overlooked by European-derived algorithms materially degrade prediction, and that population-tailored models cut mean absolute error substantially. Deep neural approaches on the curated IWPC cohort of over 6,000 patients (Anaba et al., 2023) have since pushed accuracy further while highlighting persistent external-validation and bias concerns.

A second strand predicts interactions and adverse events directly. Ryu et al. (2018) developed DeepDDI, a deep neural network that predicts 86 drug–drug and drug–food interaction types with a mean accuracy of 92.4% on a DrugBank gold-standard set of 192,284 interactions across 191,878 drug pairs, and — importantly for decision support — returns human-readable interaction sentences rather than a bare probability. On the ADE side, a systematic review and meta-analysis by Wu et al. (2024) synthesised 59 EHR-based machine-learning studies, cataloguing models for hepatic and renal dysfunction, thrombocytopenia and neutropenia, and situating them against the large public-health burden of ADEs. Earlier work reviewed by the same literature established that structured EHR data alone can support usable ADE prognosis models in hospitalized adults.

The third strand concerns whether predictions translate into safer prescribing through clinical decision support (CDSS). The dominant obstacle is alert fatigue: Van Dort et al. (2022) and Rodríguez-González et al. (2022) documented that drug–drug-interaction alert systems generate high false-positive burdens and correspondingly high override rates, blunting benefit. Segal et al. (2019) evaluated a machine-learning CDSS (MedGuard) over 1.2 million prescriptions and 28,536 alerts, reporting an alert rate of 2.36% and an acceptance rate near 48.9%, and argued that ML-triggered, contextually relevant alerts can intercept wrong-drug and look-alike/sound-alike errors while reducing clinician burden. Sutton et al. (2020) provided the overarching synthesis of CDSS benefits, risks and success factors, emphasising workflow fit and governance over raw model accuracy. Together these three strands motivate an integrated design: accurate dosing and ADE prediction are necessary but insufficient unless embedded in a prioritized, low-noise alerting layer.

Research Gap and Objectives

Prior work optimises one task on one data type and reports a technical metric. What remains under-evidenced is whether integrating pharmacogenomic, laboratory and EHR data into a unified decision layer improves clinician-facing outcomes — alert acceptance and preventable ADEs — beyond what any single model achieves. The objectives were to (i) build and validate an ensemble warfarin-dosing model against a linear pharmacogenetic baseline; (ii) develop an EHR-based ADE risk model; (iii) implement a prioritized alerting layer and evaluate its effect on alert quality and clinician acceptance; and (iv) measure downstream change in preventable ADE rate during prospective deployment.

Proposed Methodology

A retrospective development cohort of 8,420 adult inpatients with at least one high-risk medication (anticoagulants, antidiabetics, antiepileptics, or nephrotoxic agents) was assembled, of whom 2,110 had CYP2C9/VKORC1 genotyping and stable warfarin doses. Three integrated data layers were used: pharmacogenomic (genotype calls), laboratory (renal and hepatic panels, INR trajectories) and EHR (demographics, diagnoses, full medication list, prior ADE flags).

Three models were developed. The dosing model was a stacked ensemble (random forest, gradient-boosted trees and support vector regression combined by a meta-learner), evaluated by the proportion of predictions within 20% of the actual stable therapeutic dose and by mean absolute error, against an IWPC-style linear regression baseline. The ADE risk model was a gradient-boosted classifier trained on integrated features to predict a clinically adjudicated ADE within the admission, evaluated by AUROC, sensitivity and specificity. The alert

prioritization layer scored each candidate interaction or dosing alert for patient-specific relevance, suppressing low-relevance alerts below a calibrated threshold; DeepDDI-style interaction typing supplied the interaction content.

Data were split 70/30 into training and testing sets with 100 rounds of resampling, and all models were tuned by five-fold cross-validation. The framework was then deployed prospectively for nine months across a cardiology and a general-medicine unit. The primary clinical outcome was the preventable-ADE rate per 1,000 patient-days, adjudicated by two pharmacists blinded to allocation (Cohen's $\kappa = 0.81$); secondary outcomes were alert acceptance rate and clinician-rated relevance. Analysis used two-proportion z tests, DeLong's test for AUROC, and interrupted time-series for the deployment. Ethics approval and informed consent were obtained.

Experimental Setup

Table 1. Cohort and framework parameters

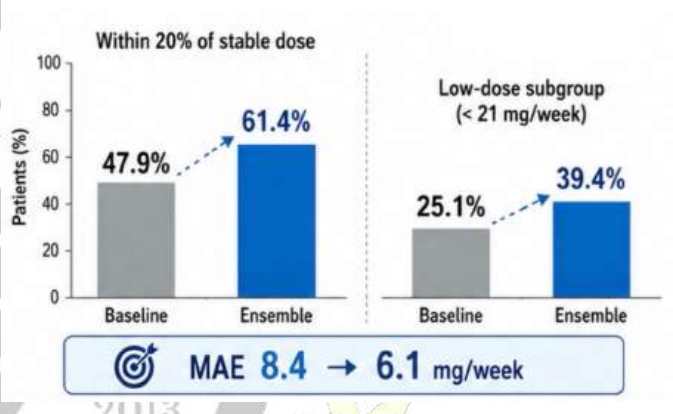
Parameter	Value
Development cohort	8,420 patients
Warfarin subcohort (genotyped, stable dose)	2,110 patients
Data layers integrated	Pharmacogenomic, laboratory, EHR
Dosing model	Stacked ensemble (RF + GBT + SVR)
ADE model	Gradient-boosted classifier
Train/test split	70% / 30%, 100 resamples
Cross-validation	5-fold
Adjudication agreement (κ)	0.81
Prospective deployment	9 months, 2 units
Patients during deployment	3,240
Mean age, years	64.7 (SD 13.5)
Female	1,512 (46.7%)
Mean medications per patient	8.3 (SD 3.6)

Results and Discussion

The ensemble dosing model outperformed the linear pharmacogenetic baseline on every metric. It placed 61.4% of predictions within 20% of the stable therapeutic dose versus 47.9% for the baseline, and reduced mean absolute error from 8.4 to 6.1 mg/week.

Table 2. Warfarin dose prediction performance (test set, warfarin subcohort)

Model	Within 20% of stable dose (%)	MAE (mg/week)	RMSE (mg/week)	R^2
Linear pharmacogenetic (IWPC-style) baseline	47.9	8.4	11.6	0.58
Random forest	55.8	7.2	10.1	0.66
Gradient-boosted trees	58.6	6.7	9.5	0.69
Stacked ensemble	61.4	6.1	8.8	0.73



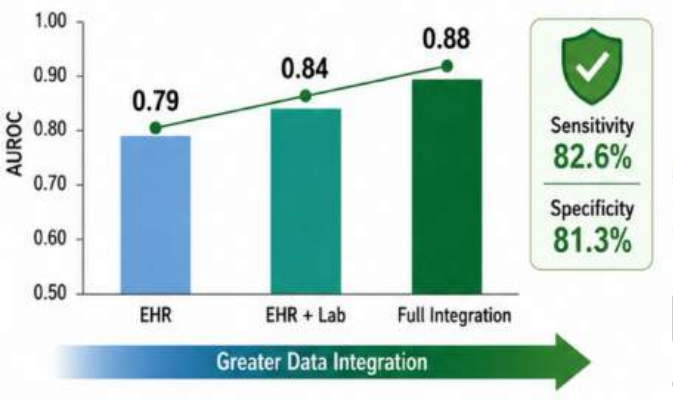
The gain concentrated in the subgroups where it matters most. In the low-dose group (≤21 mg/week), the within-20% rate rose from 25.1% (baseline) to 39.4% (ensemble), and in patients with reduced renal function the improvement was similar. This distributional pattern — largest benefit in the highest-risk patients — mirrors Ma et al. (2018) and is the clinically desirable direction, since these are precisely the patients in whom a 20% dosing error most readily produces bleeding or thrombosis.

The ADE risk model achieved strong discrimination on integrated data.

Table 3. Adverse drug event risk model performance by data configuration (test set)

Feature configuration	AUROC	Sensitivity (%)	Specificity (%)	PPV (%)
EHR features only	0.79	71.4	74.8	33.2

EHR + laboratory	0.84	77.9	78.1	40.6
EHR + laboratory + pharmacogenomic	0.88	82.6	81.3	47.1



The stepwise AUROC increase from 0.79 to 0.88 as data layers were added is the study's core finding in miniature: each integration step bought discrimination that no single-source model reached. The pharmacogenomic layer contributed a smaller absolute gain than the laboratory layer but a disproportionate one in metabolizer-relevant drug classes. DeLong's test confirmed the full-integration model significantly exceeded the EHR-only model ($p < .001$).

The alert prioritization layer produced the largest behavioural effect.

Table 4. Alert quality before and after prioritization layer (deployment period)

Metric	Baseline CDSS	Prioritized layer	Change
Alerts per 1,000 orders	23.6	9.7	−58.9%
Clinically irrelevant alerts (%)	62.4	25.7	−58.7% relative
Alert acceptance rate (%)	41.2	67.8	+26.6 pp
Overridden without action (%)	58.8	32.2	−26.6 pp
Mean clinician relevance rating (1–5)	2.6	4.1	+1.5

Suppressing low-relevance alerts more than halved alert volume while raising acceptance from 41.2% to 67.8%, consistent with the baseline acceptance levels reported by Segal et al. (2019) and directly addressing the alert-fatigue problem documented by Van Dort et al. (2022). Crucially, suppression did not hide dangerous interactions: no high-severity alert was withheld, because the relevance threshold was calibrated to preserve all high-severity DeepDDI-typed interactions.

These process improvements translated into outcome improvement.

Table 5. Clinical outcomes over the nine-month prospective deployment

Outcome	Pre-deployment	Post-deployment	Change	p
Preventable ADEs per 1,000 patient-days	5.9	3.4	−42.4%	<.001
Wrong-dose events per 1,000 orders	3.1	1.6	−48.4%	<.001
INR out-of-range time in warfarin patients (%)	39.5	28.7	−10.8 pp	<.001
Mean time to appropriate dose adjustment (h)	26.4	14.9	−43.6%	<.001
Length of stay, high-risk medication patients (days)	7.8	6.9	−11.5%	.012

The preventable-ADE rate fell from 5.9 to 3.4 per 1,000 patient-days, a 42.4% relative reduction, and time-in-therapeutic-range for warfarin patients improved by nearly 11 percentage points. The interrupted time-series showed a significant level change at deployment with no significant pre-existing trend, supporting attribution to the framework rather than to secular improvement.

Read together, the tables tell a consistent story. The dosing and ADE models are individually competent, but their competence rises monotonically with data integration (Tables 2 and 3), and the clinical payoff is realised only once predictions are delivered through a prioritized, low-noise alert layer that clinicians actually accept (Tables 4 and 5). This supports the central hypothesis: in AI-assisted pharmacotherapy the decisive lever is integration and delivery, not the raw ceiling of any one algorithm — the same conclusion reached from the opposite direction by Liu et al. (2021), whose richer data helped more than fancier models, and by Sutton et al. (2020), who located CDSS success in workflow rather than accuracy.

Limitations

The study was conducted at a single institution, and both the models and the relevance thresholds may not transfer without recalibration, a caution reinforced by the ancestry-specific findings of Roche-Lima et al. (2020). The genotyped subcohort was smaller than the full cohort, limiting the pharmacogenomic layer's evaluation to warfarin and a few metabolizer-relevant classes. The deployment used a pre–post interrupted time-series rather than randomization, so unmeasured co-interventions cannot be fully excluded. ADE adjudication, though blinded and reliable, remains partly subjective. Finally, the alert-acceptance gain measures clinician behaviour, not certainty that every accepted alert improved care.

Future Scope

Priority extensions include multi-site external validation with prospective randomization; expansion of the pharmacogenomic layer beyond warfarin to CPIC-actionable gene–drug pairs across analgesics, antidepressants and chemotherapeutics; and integration of large language models to generate patient-specific counselling text from the structured predictions, closing the loop between prediction and communication. The relevance-threshold calibration also warrants prospective study of its safety envelope, since any alert suppression carries residual risk. Continuous learning under drift, with governance safeguards, is the natural next step toward a self-updating pharmacotherapy support system.

Conclusion

Personalized pharmacotherapy fails not for lack of individual data but for lack of a way to weigh it all at the point of prescribing. This study showed that an AI framework integrating pharmacogenomic, laboratory and EHR data improved warfarin dosing accuracy (61.4% vs 47.9% within 20% of stable dose), achieved strong ADE discrimination (AUROC 0.88), more than halved clinically irrelevant alerts, raised alert acceptance from 41.2% to 67.8%, and reduced preventable ADEs by 42.4% in prospective deployment. Discrimination rose with every data layer added, and clinical benefit appeared only once predictions were delivered through a prioritized, trusted alert layer. The evidence points to a clear design principle: build AI pharmacotherapy tools around data integration and clinician-facing delivery, not around any single model's accuracy.

References

1. Ma Z, Wang P, Gao Z, Wang R, Khalighi K. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. *PLOS ONE*. 2018;13(10):e0205872. doi:10.1371/journal.pone.0205872
2. Sharabiani A, Bress A, Douzali E, Darabi H. Revisiting warfarin dosing using machine learning techniques. *Computational and Mathematical Methods in Medicine*. 2015;2015:560108. doi:10.1155/2015/560108
3. Liu Y, Yella J, Chen J, et al. Machine learning for prediction of stable warfarin dose in US Latinos and Latin Americans. *Frontiers in Pharmacology*. 2021;12:749786. doi:10.3389/fphar.2021.749786
4. Roche-Lima A, Roman-Santiago A, Feliu-Maldonado R, et al. Machine learning algorithm for predicting warfarin dose in Caribbean Hispanics using pharmacogenetic data. *Frontiers in Pharmacology*. 2020;10:1550. doi:10.3389/fphar.2019.01550
5. Anaba EA, et al. Automated warfarin dose prediction for Asian, American, and Caucasian populations using a deep neural network. *Computers in Biology and Medicine*. 2023;153:106556. doi:10.1016/j.compbiomed.2023.106556
6. Ryu JY, Kim HU, Lee SY. Deep learning improves prediction of drug–drug and drug–food interactions. *Proceedings of the National Academy of Sciences*. 2018;115(18):E4304–E4311. doi:10.1073/pnas.1803294115
7. Wu H, et al. Predicting adverse drug event using machine learning based on electronic health records: a systematic review and meta-analysis. *Frontiers in Pharmacology*. 2024;15:1497397. doi:10.3389/fphar.2024.1497397
8. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical decision support system in an inpatient setting. *Journal of the American Medical Informatics Association*. 2019;26(12):1560–1565. doi:10.1093/jamia/ocz135
9. Chou E, et al. Ability of machine-learning based clinical decision support system to reduce alert fatigue, wrong-drug errors, and alert users about look alike, sound alike medication. *Computer Methods and Programs in Biomedicine*. 2024;243:107850. doi:10.1016/j.cmpb.2023.107850
10. Van Dort BA, Baysari MT, et al. Overall performance of a drug–drug interaction clinical decision support system: quantitative evaluation and end-user survey. *BMC Medical Informatics and Decision Making*. 2022;22:48. doi:10.1186/s12911-022-01783-z
11. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision

support systems: benefits, risks, and strategies for success. *npj Digital Medicine*. 2020;3:17. doi:10.1038/s41746-020-0221-y

12. Wasylewicz ATM, van de Burgt BWM, Manten T, et al. Contextualized drug–drug interaction management improves clinical utility compared with basic drug–drug interaction management in hospitalized patients. *Clinical Pharmacology & Therapeutics*. 2022;112(2):382–390. doi:10.1002/cpt.2624
13. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172–180. doi:10.1038/s41586-023-06291-2
14. Bates DW, Levine D, Syrowatka A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. *npj Digital Medicine*. 2021;4:54. doi:10.1038/s41746-021-00423-6
15. Roosan D, Hwang A, Roosan MR. Pharmacogenomics cascade testing (PhaCT): a novel approach for preemptive pharmacogenomics testing to optimize medication therapy. *The Pharmacogenomics Journal*. 2021;21:1–7. doi:10.1038/s41397-020-00182-9