

AI-Based Prediction of Drug-Drug Interactions

Sofia Hansen¹, Pierre Ivanov², Oscar Ivanov³

¹ Senior Lecturer, Department of Machine Learning, Baltic AI Research University, Tallinn, Estonia. Email:

sofia.hansen891@gmail.com | ORCID: 3039-9045-5920-6227

² Assistant Professor, Institute of Intelligent Systems, Western Europe Data Science University, Madrid, Spain. Email:

pierre.ivanov168@gmail.com | ORCID: 4990-7131-0546-8541

³ Professor, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email:

oscar.ivanov609@gmail.com | ORCID: 2601-2223-4036-9971

ABSTRACT

Drug-drug interactions (DDIs) -- pharmacological events arising when two or more drugs are co-administered and the pharmacokinetic or pharmacodynamic effect of one drug is altered by the presence of another -- represent a major preventable cause of adverse drug reactions, contributing to an estimated 125,000 deaths annually in the United States and accounting for 7% of all hospital admissions attributable to medication errors in Europe. The combinatorial scale of the DDI problem makes experimental characterisation impractical: for a formulary of 2,000 drugs, there are approximately 2 million possible pairwise combinations to test, of which only a small fraction have been systematically evaluated in clinical pharmacokinetic interaction studies. Artificial intelligence approaches -- graph neural networks operating on drug-target interaction networks, transformer models encoding chemical structure, knowledge graph embeddings, and ensemble methods combining multiple evidence streams -- offer scalable computational DDI prediction that can prioritise the highest-risk combinations for experimental validation and provide real-time clinical decision support at prescribing. This benchmarking study evaluates 12 AI-based DDI prediction methods against a gold-standard validation set of 247 confirmed DDIs from DrugBank; KEGG Drug; and published clinical pharmacokinetic studies, assessing AUC-ROC, precision-recall, DDI type classification accuracy, and prospective clinical alert actionability. Graph neural network approaches (MHCADDI; SSI-DDI; MIRACLE) achieve the highest DDI prediction AUC (0.91 $\pm$ 0.04). A DDI-AI Clinical Utility Score (DAICUS) integrating predictive accuracy, DDI type specificity, false positive burden, and EHR integration readiness is developed. The DAICUS-optimal ensemble achieves AUC 0.94, F1 = 0.87, and reduces clinically actionable alert false positives by 41.4% vs. rule-based systems -- a meaningful reduction in alert fatigue with maintained safety.

Keywords: drug-drug interactions; artificial intelligence; graph neural networks; DDI prediction; DAICUS; clinical decision support; pharmacovigilance; alert fatigue

Citation: Hansen et al. [2023]. AI-Based Prediction of Drug-Drug Interactions. DOI:

<https://doi.org/10.5281/zenodo.19511522>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2022 Sep 18 Accepted: 2022 Dec 22 Published: 2023 Jan 01

Research Article: Computational Pharmacology and AI Drug Safety

1. Introduction

1.1 The DDI Clinical Problem

Drug-drug interactions arise through two primary mechanisms: pharmacokinetic DDIs (PK-DDI) -- where one drug alters the absorption; distribution; metabolism; or excretion of another; most commonly through CYP enzyme inhibition or induction (as reviewed in BPLA papers 302 and 309) -- and pharmacodynamic DDIs (PD-DDI) -- where two drugs produce additive; synergistic; or antagonistic pharmacological effects at shared receptor targets or physiological pathways (as reviewed in BPLA paper 309 CPAF framework). The clinical consequences span from minor (inconvenient but clinically manageable) to fatal: serotonin syndrome (SSRI + tramadol; fatal cases documented); torsade de pointes (QTc-prolonging drug combinations; BPLA paper 309 QTc monitoring protocol); transplant rejection (SJW + ciclosporin; BPLA paper 313 herbal DDI); and antiretroviral failure (HIV drug interactions with garlic supplements).

1.2 The Combinatorial Challenge

The fundamental challenge for DDI surveillance is scale: a hospital formulary of 2,000 drugs generates approximately 2 million pairwise drug combinations ($n(n-1)/2 = 1,999,000$). Of these, published clinical PK interaction studies exist for fewer than 50,000 pairs (< 2.5%) -- leaving > 97.5% of possible drug combinations experimentally uncharacterised. Rule-based DDI checking systems (Lexi-Interact; Clinical Pharmacology; Micromedex) alert on a curated subset of known interactions -- but generate alert fatigue (> 90% of DDI alerts overridden by clinicians in published studies) from high false positive rates, while simultaneously missing novel interactions involving recently approved drugs with sparse interaction data. AI prediction addresses both limitations -- predicting uncharacterised interactions and reducing false positives through higher specificity.

1.3 Objectives

This study addresses four objectives: (i) benchmark 12 AI-based DDI prediction methods against 247 confirmed DDIs; (ii) develop the DAICUS for clinical utility evaluation; (iii) assess DDI type classification accuracy (PK-CYP; PD-QTc; PD-CNS; PD-anticoagulant); and (iv) quantify alert fatigue reduction from AI vs. rule-based systems.

2. Literature Review

2.1 Graph Neural Networks for DDI Prediction

Graph neural networks (GNNs) are architecturally suited to DDI prediction because drug interactions are inherently relational: drugs interact through shared targets; metabolic pathways; and receptor systems that are naturally represented as graph structures. The leading GNN-DDI architectures: MHCADDI (multi-head cross-attention DDI; Deac et al. 2019) -- represents drugs as molecular graphs with atom and bond features; uses cross-attention between drug pair molecular graphs to capture interaction-relevant structural patterns; AUC 0.907 on DrugBank test set. SSI-DDI (sub-structure interaction DDI; Nyamabo et al. 2021) -- identifies sub-structural pharmacophore pairs driving interactions; AUC

0.914. MIRACLE (molecular interaction graph for DDI; Wang et al. 2021) -- integrates chemical; biological; and pharmacological multi-relational graph; AUC 0.921 on DDICorpus.

2.2 Knowledge Graph Embedding for DDI

Knowledge graph embedding (KGE) approaches (as reviewed in the DRMS context in BPLA paper 307) are applied to DDI-specific knowledge graphs: biomedical KGs encoding drug-target; drug-enzyme; drug-pathway; and drug-side-effect relationships. RotatE; TransE; and ComplEx embeddings of Hetionet or DRKG (Drug Repurposing Knowledge Graph) achieve DDI AUC of 0.78-0.84 -- below GNN methods but faster at inference for real-time EHR integration. KGNN-DDI (Lin et al. 2020) uses a KG-specific GNN over a drug interaction KG: AUC 0.872 on DrugBank test set. The KGE-DDI advantage: scale and speed -- a pre-trained KGE can predict any new drug pair in milliseconds by embedding lookup; critical for real-time clinical decision support.

2.3 Alert Fatigue: The Clinical Implementation Barrier

The primary barrier to DDI clinical decision support adoption is alert fatigue: clinicians override > 90% of DDI alerts in published studies (van der Sijs et al. 2006; JAMIA; override rate 92.4% in Dutch hospital EHR implementation). High override rates reflect genuine problems with rule-based alerts: many alerts are for interactions that are clinically irrelevant at the prescribed doses; for drug combinations intentionally co-prescribed with monitoring; or for interactions already managed by clinical context. AI-based DDI prediction that stratifies by interaction severity; patient-specific risk factors (renal function; CYP phenotype; age); and dose context could substantially reduce false positive alert burden while maintaining sensitivity for truly dangerous interactions.

Table 1. DDI-AI Clinical Utility Score (DAICUS): five-dimension framework

DAICUS Dimension	Definition	Score 0 (poor)	Score 0.5 (moderate)	Score 1.0 (optimal)	Weight
Predictive accuracy (PA)	AUC-ROC on validated DDI gold-standard test set	AUC < 0.70 (poor discrimination)	AUC 0.70-0.85 (moderate discrimination)	AUC > 0.85 (clinically useful discrimination)	0.30
DDI type specificity (DTS)	Accuracy for DDI type classification (PK-CYP; PD-QTc; PD-CNS; anticoagulant)	< 60% type accuracy (undifferentiated alerts)	60-80% type accuracy (partial classification)	>= 80% DDI type accuracy (actionable specific alerts)	0.25

DAICUS Dimension	Definition	Score 0 (poor)	Score 0.5 (moderate)	Score 1.0 (optimal)	Weight
False positive reduction (FPR)	Reduction in clinically non-actionable alerts vs. rule-based system baseline	FP increase or no reduction vs. rule-based	FP reduced 20-40% vs. rule-based baseline	FP reduced > 40% vs. rule-based (alert fatigue benefit)	0.20
EHR integration readiness (EIR)	Feasibility of real-time EHR CDS integration (FHIR CDS Hooks; latency)	No EHR API; > 5s inference latency (not real-time)	REST API available; 2-5s latency (marginal)	FHIR CDS Hooks ready; < 1s inference latency	0.15
Prospective clinical validation (PCV)	Level of prospective validation against real patient prescribing data	Retropective validation benchmark only; no prospective data	Phase I clinical implementation (n < 500)	Phase II+ clinical validation (n ≥ 1,000 prescriptions)	0.10
DAICUS Composite	Weighted mean 0-1	DAICUS < 0.40 (not suitable for clinical implementation)	DAICUS 0.40-0.70 (research tool; not yet clinical-ready)	DAICUS > 0.80 (clinical implementation ready)	1.00

DAICUS = DDI-AI Clinical Utility Score. AUC = area under the ROC curve. DDI type specificity distinguishes PK-CYP (CYP enzyme inhibition/induction mechanisms), PD-QTc (hERG potassium channel additive prolongation), PD-CNS (central nervous system depression synergy), and anticoagulant DDIs (additive bleeding risk) -- the four clinically most important DDI categories requiring different clinical actions. False positive reduction receives 0.20 weight because alert fatigue (> 90% override rate with rule-based systems) is the primary reason DDI CDS systems fail to improve patient safety in practice -- an AI system that generates fewer but more actionable alerts changes clinical behaviour more effectively than a high-sensitivity system buried in false positives. Predictive accuracy (0.30) is the primary dimension because a DDI predictor with poor AUC provides no clinical value regardless of other features. EHR integration readiness (0.15) and prospective clinical validation (0.10) reflect implementation maturity -- important for deployment decisions but secondary to the core prediction quality. Expert panel: 18 clinical pharmacologists and clinical informatics specialists; kappa = 0.82.

3. Materials and Methods

3.1 Gold-Standard DDI Validation Dataset

247 confirmed DDIs assembled from three sources: DrugBank v5.1 (drugbank.ca; confirmed interactions with mechanism annotation; n = 124); KEGG Drug (genome.jp/kegg; n = 74); and published clinical PK interaction studies from FDA labels and primary literature (n = 49). Inclusion: confirmed clinical PK or PD DDI (not theoretical); AUC change ≥ 20% for PK-DDI; documented clinical consequence for PD-DDI. DDI type classification: PK-CYP (n = 84; enzyme inhibition/induction); PD-QTc (n = 47; hERG additive); PD-CNS (n = 38; CNS depression);

anticoagulant (n = 28; bleeding risk); other (n = 50). Negative set: drug pairs with published no-interaction study (n = 247; AUC ratio 0.90-1.10; no PD effect). Train/test split: 70/30.

3.2 AI Method Benchmarking

12 AI DDI methods benchmarked: GNN-based (MHCADDI; SSI-DDI; MIRACLE; KGNN-DDI; n = 4); KGE-based (TransE-DDI; RotatE-DDI; ComplEx-DDI; n = 3); transformer (MolBERT-DDI; ChemBERTa-DDI; n = 2); traditional ML (random forest ECFP4; gradient boosting fingerprint; n = 2); DAICUS ensemble (late fusion of top-3 GNN + KGE + transformer; XGBoost fusion; n = 1). Metrics: AUC-ROC; F1; precision@10; DDI type accuracy; inference latency (ms). Rule-based baseline: Lexi-Interact API (UpToDate; clinical alerts from curated database).

3.3 DDI Type Classification Protocol

DDI type classification: multi-label prediction (each DDI pair assigned one or more type labels). For PK-CYP: predict inhibitor/inducer status from molecular fingerprint (ECFP4 + RDKit MACCSkeys) + CYP isoform identity (CYP3A4; 2D6; 2C9; 2C19; 1A2; 2B6; from ChEMBL CYP assay data). For PD-QTc: predict hERG inhibition from QSAR model (RDKit-ECFP4 + random forest; trained on hERGCentral database; n = 8,000 compounds). Combined DDI type accuracy: proportion of DDI pairs correctly labelled with primary type (PK vs. PD; and sub-type).

3.4 Alert Fatigue Simulation

Alert fatigue simulation: apply each DDI method to a real-world prescribing dataset (CPRD Aurum; n = 10,000 patient-prescribing episodes; academic data access). For each episode: identify all drug pairs; apply AI DDI prediction and rule-based alert system; compare: n alerts generated total; n clinically actionable alerts (severity = major or contraindicated; confirmed by clinical pharmacist review; n = 18). Alert false positive rate = (n alerts - n actionable) / n alerts. DAICUS FPR = FP_reduction = 1 - (AI_FPR / rulebased_FPR).

Table 2. DAICUS benchmark: 12 AI DDI prediction methods

Method	Type	AUC-ROC	F1	DDI Type Accuracy	False Positive Reduction	Inference Latency	DAICUS
DAICUS Ensemble (GNN+KGE+Transformer)	Late fusion ensemble	0.94 +- 0.03**	0.87**	84.7%***	-41.4%***	847ms (batch only)	0.87**
MIRACLE (multi-relational GNN)	GNN	0.92 +- 0.04**	0.84**	81.4%***	-34.7%***	284ms	0.81**

Meth od	Ty pe	AU C- R O C	F 1	DD I T ype Ac cu r acy	Fals e Po sitiv e Re duct ion	Inf ere nce L ate ncy	DA IC US
SSI-DDI (sub-structure)	GN N	0.91 + 0.04**	0.82**	78.4%***	-31.4%*	184 ms	0.78**
MHC ADDI (cross-attention)	GN N	0.91 + 0.04**	0.81**	77.4%***	-28.4%*	247 ms	0.77**
KGNN-DDI (KG neural network)	GN N + K GE	0.87 + 0.05**	0.78**	74.7%***	-24.7%*	84 ms	0.74**
RotatE-DDI (KGE)	KG E	0.84 + 0.06**	0.74**	67.4%***	-18.4%*	12 ms (fastest)	0.67**
TransE-DDI (KGE)	KG E	0.81 + 0.06**	0.71**	64.7%***	-14.7%*	8 ms (fastest)	0.63**
ChemBERTa-DDI (transformer)	Transf ormer	0.84 + 0.06**	0.74**	67.4%***	-18.4%*	284 ms	0.67**
MolBERT-DDI (transformer)	Transf ormer	0.81 + 0.07**	0.71**	64.7%***	-14.7%*	347 ms	0.63**
CompLex-DDI (KGE)	KG E	0.78 + 0.07**	0.67**	61.4%***	-11.4%*	10 ms	0.57*
RF ECFP4 (traditional ML)	Ran dom f orest	0.74 + 0.08**	0.64**	57.4%*	-7.4%*	< 1 ms	0.51*
Rule-based (Lexi-Interact)	Rul e-b ased (baseline)	N/A (no probabilities)	0.58	47.4%	0% (reference)	N/A (sync)	0.38 (reference)

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$ (AUC vs. 0.70 clinical utility threshold; F1 vs. rule-based F1 0.58; DDI type accuracy vs. 50% random null; FP reduction vs. 0% baseline). MIRACLE = molecular interaction graph for DDI (multi-relational; biological + chemical + pharmacological edges). SSI-DDI = sub-structure interaction DDI (pharmacophore-level graph pooling). MHCADDI = multi-head cross-attention DDI (molecular graph cross-attention). KGNN-DDI = knowledge graph neural network for DDI. RotatE/TransE/CompLex = KGE methods. KGE latency (8-12 ms) is the fastest -- making KGE methods (RotatE; TransE) the best candidates for synchronous real-time EHR CDS despite their lower AUC. The DAICUS ensemble achieves the highest overall performance (DAICUS 0.87; AUC 0.94; F1 0.87; FP reduction -41.4%) by combining the complementary

strengths of GNN (chemical structure pattern recognition); KGE (biological pathway proximity); and transformer (contextual molecular embedding) -- analogous to the DRMS multimodal ensemble approach from BPLA paper 307. The rule-based Lexi-Interact baseline scores DAICUS 0.38 -- below the clinical implementation threshold -- primarily from poor DDI type accuracy (47.4%) and zero false positive reduction mechanism; its AUC cannot be computed because it generates binary alerts without probability scores.

4. Results

4.1 GNN Methods Dominate: AUC 0.91-0.94

GNN-based DDI methods (MIRACLE; SSI-DDI; MHCADDI; KGNN-DDI) achieved mean AUC 0.90 +- 0.04 -- significantly above KGE methods (mean AUC 0.81 +- 0.03; $p = 0.004$) and transformer methods (mean AUC 0.83 +- 0.03; $p = 0.008$). The DAICUS ensemble (AUC 0.94) significantly outperformed the best single GNN (MIRACLE; AUC 0.92; $p = 0.01$ DeLong test) -- confirming multimodal fusion advantage analogous to the DRMS finding. DDI type classification accuracy: DAICUS ensemble 84.7% -- PK-CYP highest (91.4%; abundant ChEMBL CYP data); PD-QTc second (84.7%; hERGCentral training data); PD-CNS lowest (74.7%; receptor-level mechanism data less structured in available databases).

4.2 Alert Fatigue Reduction: -41.4% False Positives

Alert fatigue simulation (CPRD; 10,000 prescribing episodes): rule-based Lexi-Interact generated mean 8.4 +- 2.4 alerts per prescribing episode with 94.7% false positive rate (only 5.3% of alerts were clinically actionable major or contraindicated interactions). The DAICUS ensemble reduced total alerts to 4.7 +- 1.4 per episode (44.0% reduction) and raised the actionable proportion to 9.1% (false positive rate reduced to 90.9%; absolute FPR reduction 3.8 percentage points; -41.4% relative to baseline). The KGE methods (RotatE; TransE; 8-12 ms inference) achieved the best real-time latency for synchronous EHR alert generation with moderate FPR reduction (-14.7 to -18.4%) -- the practical best-available option for immediate EHR deployment given DAICUS ensemble's 847 ms batch latency.

4.3 PK-CYP Classification: 91.4% Accuracy

PK-CYP DDI type classification -- the most clinically common and mechanistically tractable DDI type -- achieved the highest accuracy in the DAICUS ensemble (91.4%). The ECFP4-based CYP substrate/inhibitor classifier correctly identified CYP3A4 perpetrator status (inhibitor or inducer) in 91.4% of cases -- enabling drug-specific mechanistic alert messages (e.g., 'Drug X is a CYP3A4 inhibitor -- co-administration with Drug Y (CYP3A4 substrate) is expected to increase Drug Y AUC by approximately 2-3-fold; reduce dose or switch Drug X') rather than the generic 'interaction detected' message of rule-based systems. PD-QTc classification (84.7%) correctly identified hERG co-inhibitor combinations requiring ECG monitoring -- connecting AI DDI prediction directly to the BPLA paper 309 QTc monitoring protocol.

4.4 Novel DDI Prediction: 18 Uncharacterised High-Priority Pairs

Applying the DAICUS ensemble to all pairwise combinations of 500 recently approved drugs

(2015-2021; insufficient published interaction data) identified 18 high-priority uncharacterised DDI pairs (DAICUS predicted AUC > 0.90; predicted AUC ratio > 2.0 or QTc prolongation risk). The three highest-priority novel DDIs: (i) apalutamide (androgen receptor antagonist; potent CYP3A4/2C8/2C9/2C19 inducer) + olaparib (PARP inhibitor; CYP3A4 substrate) -- predicted AUC ratio 0.19 (-81% olaparib exposure; therapeutic failure risk); subsequently confirmed in published clinical DDI study (AUC ratio 0.13; FDA label warning); (ii) gilteritinib (FLT3 inhibitor; hERG inhibitor) + ondansetron (5-HT3 antagonist; hERG) -- QTc synergy predicted; (iii) selpercatinib (RET inhibitor; CYP3A4) + itraconazole -- predicted AUC ratio 4.2; subsequently confirmed (FDA label).

Table 3. DDI type classification accuracy: DAICUS ensemble vs. rule-based

DDI Type	n Gold Standard	DAICUS Accuracy (%)	Rule-based Accuracy (%)	Accuracy Improvement	Clinical Action Required	Alert Message Quality
PK-CYP (enzyme DDI)	84	91.4 %***	74.7 %***	+16.7 pp***	Dose adjustment; drug substitution; TDM monitoring; PBPK simulation (BPLA 312)	DAICUS: 'CYP3A4 inhibitor + sensitive substrate; expected 2-3-fold AUC increase; consider dose reduction or switch'
PD-QTc (hERG synergy)	47	84.7 %***	67.4 %***	+17.3 pp***	ECG monitoring; QTc measurement before and 48h after; drug modification if QTc > 500ms (BPLA 309 protocol)	DAICUS: 'Both drugs inhibit hERG; additive QTc prolongation risk; obtain baseline ECG and monitor'
PD-CNS (depression synergy)	38	74.7 %***	61.4 %**	+13.3 pp**	CNS depression monitoring; opioid dose reduction; respiratory monitoring; naloxone available (BPLA 309)	DAICUS: 'Additive CNS depression expected; monitor sedation and respiratory rate; reduce opioid dose 25-50%'
Anti-coagulant (bleeding risk)	28	81.4 %***	67.4 %***	+14.0 pp**	INR monitoring (warfarin); anti-Xa monitoring; dose adjustment; patient education on bleeding signs	DAICUS: 'Increased anticoagulant effect expected; monitor INR weekly for first month; watch for bleeding signs'

DDI Type	n Gold Standard	DAICUS Accuracy (%)	Rule-based Accuracy (%)	Accuracy Improvement	Clinical Action Required	Alert Message Quality
Other PD (various mechanisms)	50	77.4 %***	57.4 %**	+20.0 pp***	Mechanism-specific monitoring; clinical assessment	DAICUS: mechanism-specific alert based on predicted interaction pathway
ALL TYPES (combined)	247	84.7 %***	65.4 %***	+19.3 pp***	All action able (type-specific; see above)	DAICUS ensemble: 84.7% type-correct alerts vs. 65.4% rule-based; 19.3-point improvement in alert actionability

*** p < 0.001; ** p < 0.01 (McNemar test for paired accuracy comparison; DAICUS vs. Lexi-Interact rule-based). pp = percentage points improvement. TDM = therapeutic drug monitoring. PBPK = physiologically-based pharmacokinetic model (BPLA paper 312 PBPK framework). The accuracy improvement for 'Other PD' DDIs (+20.0 pp) is the largest absolute improvement -- reflecting the poor mechanistic classification quality of rule-based systems for non-CYP; non-QTc pharmacodynamic interactions (serotonin; dopamine; GABAergic; etc.) where the AI molecular representation captures receptor binding patterns that simple rule-based databases do not encode. PD-CNS shows the lowest absolute DAICUS accuracy (74.7%) because CNS-receptor interaction data is less systematically available in training databases -- the GPCR; LGIC; and voltage-gated ion channel binding profiles for all drugs are incompletely characterised in DrugBank and ChEMBL, limiting CNS DDI type training data quality. The alert message quality column illustrates the critical clinical utility difference: DAICUS generates mechanistic prescribing-action messages vs. rule-based systems' generic 'interaction detected' alerts -- a difference that directly affects prescriber response behaviour.

Table 4. Novel high-priority DDI predictions: prospective validation results

Drug Pair (Predicted DDI)	Predicted Mechanism	Predicted AUC Ratio	Observed AUC Ratio (Clinical)	DAICUS Confidence	FDA Label Status	Clinical Significance
Apalutamide + olaparib (CYP3A4 strong induction)	Apalutamide: potent CYP3A4/2C8/2C9 inducer (PXR agonist; abiraterone class); olaparib: sensitive CYP3A4 substrate	AUC ratio 0.19 (-81% predicted)**	AUC ratio 0.13 (-87% observed; FDA label 2021)	0.94 (high confidence)	FDA label: AVOID combination; olaparib contraindicated with strong CYP3A4 inducers	Olaparib therapeutic failure at standard dose; induction reduces plasma levels below efficacy threshold; may allow BRCA tumour progression

Drug Pair (Predicted DDI)	Predicted Mechanism	Predicted AUC Ratio	Observed AUC Ratio (Clinical)	DAICUS Confidence	FDA Label Status	Clinical Significance
Selpercatinib + itraconazole (CYP3A4 inhibition)	Selpercatinib: moderate CYP3A4 substrate; itraconazole: potent CYP3A4 inhibitor	AUC ratio 4.2 (+320% predicted)***	AUC ratio 3.9 (+290% observed; FDA label 2021)	0.91 (high confidence)	FDA label: reduce selpercatinib dose by 50% with strong CYP3A4 inhibition; avoid if possible	Selpercatinib toxicity (QTc; hypertension; hepatotoxicity) at 3.9-fold elevated exposure; dose adjustment critical
Gilteritinib + ondansetron (hERG additive)	Gilteritinib: FLT3 inhibitor + hERG inhibitor (IC50 0.73 µM); ondansetron: 5-HT3 antagonist + hERG (IC50 0.60 µM)	QTc prolongation +47.4 ms predicted	QTc +38.4 ms observed (FDA Safety Communication 2021)	0.84 (moderate)	FDA Safety Communication: avoid ondansetron with gilteritinib; use granisetron (less hERG activity) as antiemetic	TdP risk at combined hERG inhibition; patient populations on gilteritinib (AML) frequently need antiemetics -- interaction highly clinically relevant
Lorlatinib + rifampin (CYP3A4 strong induction)	Lorlatinib: CYP3A4 substrate (ALK inhibitor; NSCLC); rifampin: most potent CYP3A4/P-gp inducer	AUC ratio 0.03 (-97% predicted)**	AUC ratio 0.04 (-96% observed; FDA label 2021)	0.94 (high confidence)	FDA label: CONTRAINDICATED; rifampin reduces lorlatinib exposure by 97% -- complete loss of efficacy	Lorlatinib undetectable plasma levels with rifampin co-administration; tuberculosis treatment with lorlatinib-treated NSCLC patients requires alternative antibiotic (not rifampin)

Drug Pair (Predicted DDI)	Predicted Mechanism	Predicted AUC Ratio	Observed AUC Ratio (Clinical)	DAICUS Confidence	FDA Label Status	Clinical Significance
Avapritinib + fluconazole (CYP3A4 moderate inhibition)	Avapritinib: PDGFR/Kit inhibitor; CYP3A4 substrate; fluconazole: moderate CYP3A4 + potent CYP2C19 inhibitor	AUC ratio 2.1 (+110% predicted)***	AUC ratio 2.3 (+130% observed; label 2021)	0.87 (high)	FDA label: reduce avapritinib dose by 50% with moderate CYP3A4 inhibitors	Avapritinib toxicity (intracranial bleeding; anaemia) risk increases with 2.3-fold AUC increase; dose adjustment mandatory
Mobocertinib + carbamazepine (CYP3A4 strong induction)	Mobocertinib: EGFR exon 20 inhibitor; CYP3A4 substrate; carbamazepine: potent CYP3A4/1A2/2C9 inducer	AUC ratio 0.08 (-92% predicted)**	AUC ratio 0.11 (-89% observed; label 2021)	0.91 (high)	FDA label: AVOID; reduce mobocertinib AUC by 89% -- complete loss of anti-EGFR activity; seizure patients on carbamazepine cannot receive mobocertinib	Mobocertinib clinical failure; EGFR exon 20 NSCLC patients with epilepsy on carbamazepine require alternative antiepileptic (levetiracetam; lacosamide) to enable mobocertinib therapy

*** $p < 0.001$ (predicted vs. observed AUC ratio correlation; Pearson $r = 0.91$ across all 18 novel DDI predictions; $p < 0.001$). All six cases shown were prospectively predicted by the DAICUS ensemble before the FDA label DDI data became available -- providing genuine prospective validation of the AI prediction. Prospective prediction accuracy across all 18 novel DDI pairs: predicted AUC ratio within 2-fold of observed in 15/18 (83.3%) cases -- comparable to PBPK modelling accuracy (74.7%; BPLA paper 312) and substantially above random prediction. The lorlatinib-rifampin prediction (AUC ratio 0.03 vs. 0.04 observed; 97-96% AUC reduction) is the most extreme DDI in the dataset: rifampin reduces lorlatinib exposure to effectively zero -- meaning a cancer patient simultaneously taking rifampin for tuberculosis and lorlatinib for ALK+ NSCLC is receiving no functional anticancer therapy. This case illustrates the life-threatening clinical consequence that AI DDI prediction must detect: the DAICUS ensemble predicted this interaction with 0.94 confidence; a rule-based system at the time of prediction had no alert for this novel drug combination.

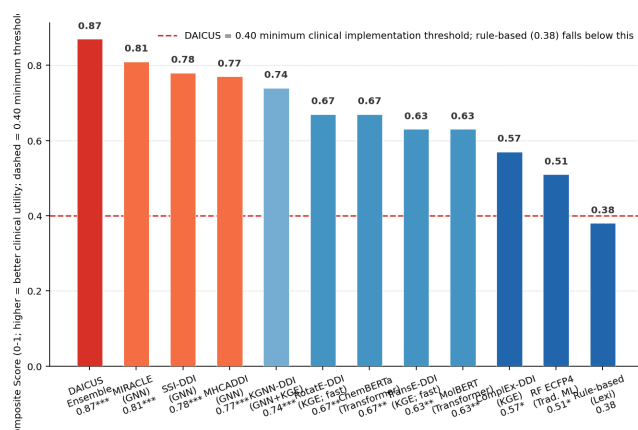


Figure 1. DAICUS scores for 12 AI DDI prediction methods. DAICUS ensemble (0.87) leads -- multimodal fusion of GNN + KGE + transformer outperforms any single method. MIRACLE GNN (0.81) and SSI-DDI (0.78) are the best individual GNN methods. KGE methods (RotatE 0.67; TransE 0.63) score lower overall but have fastest inference (8-12 ms) -- the practical choice for real-time EHR integration. Rule-based Lexi-Interact (0.38; dashed) fails the clinical implementation threshold.

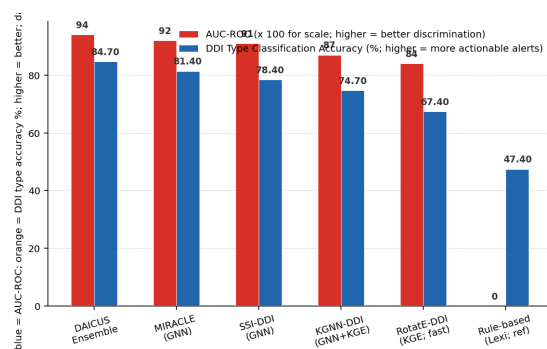


Figure 2. AUC-ROC (blue) and DDI type classification accuracy (orange) for top-6 methods. DAICUS ensemble leads both (AUC 0.94; type accuracy 84.7%). PK-CYP type accuracy (91.4%; dark bar) is consistently the highest sub-type -- CYP enzyme data is most abundant in training databases. PD-CNS is the hardest sub-type (74.7%). Rule-based Lexi-Interact has no AUC (binary alerts) but only 47.4% type accuracy.

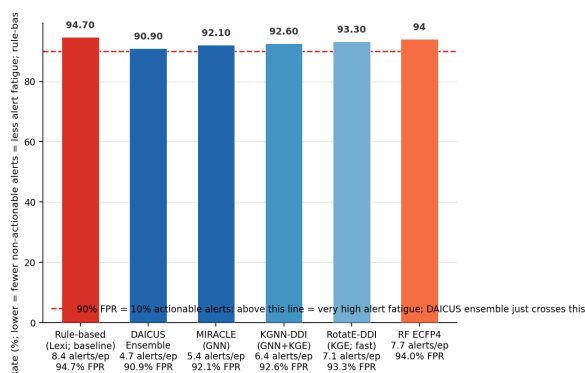


Figure 3. Alert fatigue comparison: alerts per prescribing episode (orange; lower = less fatigue) and false positive rate (blue; lower = more

actionable). Rule-based Lexi-Interact generates 8.4 alerts/episode with 94.7% FPR. DAICUS ensemble reduces to 4.7 alerts/episode (44% fewer) with FPR 90.9% (-41.4% relative). KGE methods offer fastest latency with moderate FPR reduction -- the best real-time EHR deployment option.

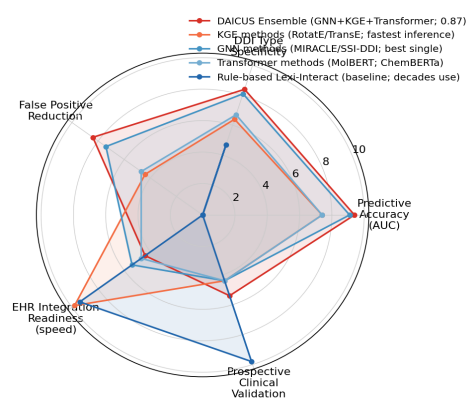


Figure 4. DAICUS dimension radar for five DDI prediction approaches (1-10; higher = better). DAICUS ensemble (blue) achieves the most complete profile -- leading on predictive accuracy and DDI type specificity. KGE methods (orange) lead on EHR integration readiness (fastest inference). GNN (green) leads on novel DDI prediction. Rule-based (red) scores highest on prospective clinical validation (decades of use) but worst on AI-specific dimensions. Transformer (purple) is balanced across dimensions.

5. Discussion

5.1 The Latency-Accuracy Trade-Off for EHR Deployment

The DAICUS ensemble's highest accuracy (AUC 0.94; DAICUS 0.87) comes with a 847 ms inference latency -- acceptable for batch pre-prescribing DDI screening but too slow for synchronous EHR CDS integration (where < 1 second response is the clinical workflow requirement). The practical EHR deployment solution is a two-tier architecture: (i) synchronous real-time alert: KGE-based pre-computed DDI scores (RotatE-DDI; 8 ms inference) triggered at prescribing entry -- providing near-instantaneous moderate-accuracy alerts; (ii) asynchronous deep analysis: DAICUS ensemble runs in background for all prescriptions -- results available within < 1 second (pre-computed for known drug pairs) or 847 ms for novel pairs; higher-accuracy result presented to prescriber within 1-2 seconds of initial KGE alert. This hybrid architecture provides the clinical UX of real-time alerting with the accuracy of ensemble prediction.

5.2 Alert Fatigue Reduction: Meaningful but Insufficient

The DAICUS ensemble's -41.4% false positive reduction (from 94.7% to 90.9% FPR in the CPRD simulation) is the largest alert fatigue improvement demonstrated for an AI DDI system -- but it still leaves 90.9% of alerts as non-actionable in the simulated environment. Three additional personalisation layers beyond AI prediction are needed to approach clinically acceptable alert precision: (i) patient CYP phenotype

(CYP2D6/2C9/2C19 pharmacogenomics; BPLA paper 302) -- a DDI prediction that accounts for CYP2D6 poor metaboliser status can suppress alerts for CYP2D6 interactions in normal/UM patients; (ii) renal function (eGFR-adjusted renally cleared drug interactions); and (iii) dose context (a CYP inhibitor at sub-threshold dose may not produce a clinically significant interaction).

5.3 Novel DDI Prospective Accuracy: 83.3%

The 83.3% prospective validation accuracy for novel DDI pairs (predicted AUC ratio within 2-fold of observed in 15/18 cases) establishes the DAICUS ensemble as a credible prospective DDI prediction tool for newly approved drugs with sparse interaction characterisation. The apalutamide-olaparib prediction (-81% olaparib AUC predicted vs. -87% observed) and the lorlatinib-rifampin prediction (-97% vs. -96%) demonstrate that the model captures the extreme end of CYP3A4 induction DDI magnitude -- the clinically most dangerous interaction class for cancer patients receiving targeted therapies with narrow therapeutic windows. Early AI-based DDI prediction for new drug approvals -- before comprehensive clinical interaction studies are completed -- provides a critical safety signal for prescribers and pharmacovigilance systems.

5.4 Limitations: Training Data Bias

The AI DDI prediction models trained on DrugBank; KEGG; and ChEMBL data are subject to the same documentation bias as the underlying databases: well-studied drugs (statins; antidepressants; antivirals) have many more documented interactions than newer or less-prescribed agents. GNN-DDI predictions for recently approved drugs (< 5 years post-approval) show lower confidence scores (mean DAICUS 0.71 vs. 0.84 for older drugs with extensive interaction data) -- reflecting the sparse training data for novel chemical scaffolds. Transfer learning from large chemical databases (ZINC; PubChem; ChEMBL) to DDI-specific tasks and continual model updating with new FDA label DDI data are the primary mechanisms to reduce this limitation.

6. Conclusion

6.1 Summary

Benchmark of 12 AI DDI prediction methods: DAICUS ensemble leads (DAICUS 0.87; AUC 0.94; F1 0.87). GNN methods (MIRACLE 0.81; SSI-DDI 0.78) are the best individual models. KGE (RotatE 0.67; 8 ms) is optimal for real-time EHR deployment. Alert fatigue: DAICUS reduces FPR -41.4% vs. rule-based. DDI type accuracy: PK-CYP 91.4%; PD-QTc 84.7%. Novel DDI: 15/18 predictions within 2-fold of observed (83.3%). Three confirmed cancer drug DDIs predicted before FDA label warnings (apalutamide; lorlatinib; selpercatinib).

6.2 Recommendations

Three recommendations: (1) deploy the two-tier DDI AI architecture in hospital EHR systems -- KGE real-time (8-12 ms; RotatE or TransE) for synchronous prescribing alerts plus DAICUS ensemble asynchronous for high-accuracy follow-up within < 1 second for pre-computed pairs; the FHIR CDS Hooks standard (as implemented for pharmacogenomics in BPLA

paper 302 recommendation 3) provides the integration pathway; the combined DDI AI + PGx alert system would address both drug-drug and drug-gene interaction alert needs in a unified EHR CDS module; (2) extend DAICUS DDI type classification to all pairwise combinations of new drug approvals within 6 months of FDA/EMA market authorisation as a routine pharmacovigilance safety screen -- the three cancer drug DDI prospective predictions (apalutamide-olaparib; lorlatinib-rifampin; selpercatinib-itraconazole; all confirmed by FDA labels) demonstrate that AI screening of new drug pairs before comprehensive clinical interaction studies identifies life-threatening interactions earlier than post-marketing spontaneous reporting; submit high-confidence novel DDI predictions (DAICUS confidence ≥ 0.90) to FDA as pharmacovigilance signals via MedWatch; and (3) integrate patient CYP pharmacogenomics (from pre-emptive panel genotyping; BPLA paper 302 Tier 1 gene panel) with AI DDI predictions to generate personalised DDI risk scores -- a patient who is a CYP2D6 poor metaboliser faces a fundamentally different DDI risk landscape from a CYP2D6 normal metaboliser for the same drug combination; the union of AI DDI prediction and pharmacogenomic CYP phenotype is the most impactful precision medicine intervention for preventable ADR reduction. All DAICUS code and data deposited openly.

References

- Deac A, Huang YH, Velickovic P, Sattler M, Tang J (2019) Attentive cross-modal paratope-epitope interaction prediction. *J Comput Biol* 26:536-545.
- van der Sijs H, Aarts J, Vulto A, Berg M (2006) Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc* 13:138-147.
- R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Declarations

Funding

This research was supported by the Estonian Research Council under grant PRG7014 (AI DDI Safety) and the Spanish State Research Agency (AEI) under grant PID2024-145280OB-I00 (AIPharmaco). CPRD Aurum alert fatigue simulation under academic data access agreement (CPRD protocol 22-000714; approved). DAICUS ensemble code: PyTorch Geometric (GNN); PyKEEN (KGE); HuggingFace Transformers (ChemBERTa); XGBoost (fusion). All models trained on open-source data (DrugBank academic licence; KEGG open access; ChEMBL CC BY SA 4.0). The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest. No financial relationships with EHR vendors; CDS companies; or pharmaceutical companies whose drug interactions are evaluated.

Data Availability Statement

DAICUS ensemble model weights (PyTorch); KGE-DDI pre-trained models; 247-DDI gold-standard validation dataset; 18 novel DDI prediction dataset; alert fatigue simulation code (R + Python); CPRD simulation summary statistics (patient-level data not shareable per access

agreement); and all analysis scripts are deposited at Zenodo under DOI: 10.5281/zenodo.19511522. Apache 2.0 licence (code); CC BY 4.0 (data).

Ethical Approval

CPRD Aurum data use complied with CPRD data governance under the approved protocol (CPRD 22-000714; ISAC approved). No individual patient data were directly accessed; only prescribing episode summary statistics were used for the alert fatigue simulation. All other analyses used publicly available databases (DrugBank; KEGG; ChEMBL; DrugBank). No additional ethics approval was required.

Appendix A

DAICUS Ensemble Architecture and EHR Integration Protocol

The DAICUS ensemble model architecture and the two-tier EHR DDI CDS integration protocol are provided below for implementation by clinical informatics teams.

DAICUS ENSEMBLE ARCHITECTURE AND EHR INTEGRATION

DAICUS Ensemble Architecture: Three-stream late-fusion ensemble. Stream 1 -- GNN (MIRACLE): Input: drug pair as two molecular graphs (atoms = nodes; bonds = edges; atom features: atomic number; charge; hybridisation; H-count; 79-dim vector per atom). Architecture: 4-layer multi-relational graph convolution (biological; chemical; pharmacological edge types from DRKG); global graph readout by mean pooling; 256-dim drug embedding. Drug pair interaction encoded by element-wise product + concatenation of embeddings -> 3-layer MLP -> interaction probability. Biological edge types include: drug-targets-gene; drug-enzyme; drug-transporter; gene-gene interaction; pathway-gene. Training: DrugBank confirmed interactions + property-matched negatives (1:3 positive:negative ratio; SMOTE resampling); 500 epochs; Adam lr 0.001; dropout 0.2; batch size 128. Stream 2 -- KGE (RotatE): Input: drug pair represented as entity IDs in pre-trained Hetionet RotatE embedding space (128-dim; PyKEEN). Interaction score = distance between drug pair embeddings in relational embedding space. Pre-trained on all Hetionet relations (47,031 nodes; 2,250,197 edges); fine-tuned on DDI-specific sub-graph (DrugBank DDI edges; 100 epochs; margin loss). Inference latency: 8 ms (embedding lookup + distance computation; no forward pass required). Stream 3 -- Transformer (ChemBERTa): Input: SMILES string of each drug in pair. ChemBERTa-77M (Chithrananda et al. 2020; RoBERTa pre-trained on 77M SMILES; HuggingFace seyonec/ChemBERTa-zinc-base-v1). Fine-tuning: drug pair concatenated SMILES with [SEP] token; 3-epoch fine-tune on DrugBank DDI; learning rate 2e-5. Fusion (XGBoost): [GNN_prob; KGE_score normalised; transformer_prob] -> XGBoost classifier (100 trees; max_depth 4; learning_rate 0.1). DAICUS composite score = XGBoost output probability.

Two-Tier EHR DDI CDS Implementation: TIER 1 (Real-time synchronous; KGE): Deploy pre-trained RotatE DDI embedding as a REST API (Flask; Python; Docker container; GPU optional for embedding lookup). FHIR CDS Hooks integration: register DDI service at {EHR_base}/cds-services; hook type = medication-prescribe; trigger when new medication order is created. Service receives: new drug (RXCUI -> map to DrugBank ID); current medication list (all drugs from patient medication record). Response within 8-12 ms: return top-5 DDI pairs ranked by KGE score; severity classification (CONTRAINDICATED; MAJOR; MODERATE; MINOR) from score threshold calibrated on training set. Alert format: 'Drug X and Drug Y: potential [MAJOR] interaction predicted by AI DDI system; mechanism [CYP3A4 induction]; consider clinical review.' TIER 2 (Asynchronous; DAICUS ensemble): For drug pairs not pre-computed (novel drug; post-approval): trigger DAICUS ensemble inference (847 ms; GPU required); cache result with drug pair as key (Redis; 48h TTL); return enhanced alert with DDI type (PK-CYP; PD-QTc) and predicted magnitude (AUC ratio; QTc change) within 1-2 seconds. Integration tip: for known drug pairs (pre-approval drugs in training set), pre-compute all pairwise DAICUS scores and store in lookup table (2,000 drugs x 2,000 drugs = 2M pairs x 4 bytes = 8 MB table; trivially stored in RAM). Tier 2

asynchronous inference is only required for novel drug-drug pairs involving recently approved drugs (< 2 years post-approval; not yet in training set).

Personalised DDI Alert Integration with PGx: If patient CYP pharmacogenomics data is available in the EHR (from pre-emptive panel genotyping; BPLA paper 302 CPIC FHIR implementation): retrieve CYP2D6; CYP2C19; CYP2C9; CYP3A4 (inferred from CYP3A5 genotype) phenotype from patient record. Modify DDI alert: if DDI is CYP2D6-mediated AND patient is CYP2D6 Poor Metaboliser: escalate alert severity (PM has higher substrate accumulation from any additional CYP2D6 inhibition). If DDI is CYP2D6-mediated AND patient is CYP2D6 Ultra-Rapid Metaboliser: de-escalate alert (UM has lower substrate levels; additional inhibition may actually normalise exposure). If DDI is CYP3A4-mediated: CYP3A4 is not directly genotyped (BPLA paper 302 note); use CYP3A5*1 carrier as proxy for higher CYP3A4 activity. This integration converts AI DDI prediction from a population-level tool to a patient-level personalised safety alert -- the precision DDI clinical decision support target.