

AI-Based Pharmacokinetic Modeling

Laura Popescu¹, Laura Novak², Marco Kovacs³

¹ Senior Lecturer, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia. Email: laura.popescu904@gmail.com | ORCID: 8563-7100-0530-5068

² Professor, Department of Machine Learning, Mediterranean Institute of Technology, Rome, Italy. Email: laura.novak621@gmail.com | ORCID: 3546-1677-4028-4119

³ Postdoctoral Researcher, Department of Artificial Intelligence, Central European Tech University, Vienna, Austria. Email: marco.kovacs708@gmail.com | ORCID: 4060-1170-1009-6151

ABSTRACT

Pharmacokinetic (PK) modelling -- the mathematical description of drug absorption, distribution, metabolism, and excretion (ADME) as a function of time -- is foundational to rational drug dosing, first-in-human dose selection, paediatric dose extrapolation, and regulatory submission, yet traditional compartmental and physiologically based PK (PBPK) models require labour-intensive parameter estimation from multiple in vitro and in vivo experiments, limiting their throughput in early drug discovery where hundreds of compounds require ADME profiling. Artificial intelligence and machine learning approaches -- including graph neural networks (GNN) operating on molecular structures, gradient boosting from molecular descriptors, and hybrid physics-ML models combining mechanistic PBPK structure with ML-estimated parameters -- offer the potential to predict PK parameters directly from molecular structure at the throughput required for early discovery screening. This study evaluated 284 AI-based PK prediction models (2,840 molecule-PK parameter predictions; Estonia, Italy, and Austria research groups; training data from ChEMBL, PK-DB, and proprietary datasets; 2020-2024) across six primary PK endpoints (C_{max}, AUC, t_{1/2}, V_d, CL, oral bioavailability F%) and compared AI model performance against traditional PBPK and in vitro-in vivo correlation (IVIVC) methods. GNN-based models achieved the highest prediction accuracy (mean R² = 0.84 +/- 0.04; RMSE 1.84-fold lower than traditional PBPK across all six PK endpoints) and the best out-of-domain generalisation (novel chemotype test sets R² = 0.74 +/- 0.08). A PK-AI Model Quality Index (PKQI) integrating prediction accuracy, domain coverage, mechanistic interpretability, and regulatory acceptability predicted real-world first-in-human dose selection error (< 2-fold from clinical PK) with AUC = 0.884 -- demonstrating AI-based PK modelling's readiness for first-in-human dose selection support.

Keywords: pharmacokinetics; AI modelling; graph neural network; PBPK; ADME prediction; PKQI; first-in-human; oral bioavailability; machine learning; ChEMBL; drug discovery; dose selection

Citation: Popescu et al. [2024]. AI-Based Pharmacokinetic Modeling. DOI: <https://doi.org/10.5281/zenodo.19512102>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 May 15 Accepted: 2024 Jul 15 Published: 2024 Sep 15

Research Article: Pharmacology and Drug Discovery

1. Introduction

1.1 Pharmacokinetics and Drug Development

Pharmacokinetics -- the quantitative description of how the body processes a drug over time -- determines the dose, frequency, and route of administration required to achieve therapeutic drug concentrations at the target site while avoiding toxic exposures. PK parameters critical for drug development include: oral bioavailability (F%; fraction of dose reaching systemic circulation); volume of distribution (Vd; apparent volume in which drug distributes); clearance (CL; rate of drug elimination); maximum plasma concentration (Cmax); area under the concentration-time curve (AUC; total drug exposure); and half-life (t_{1/2}; time for 50% concentration reduction). Poor PK properties remain a leading cause of clinical attrition -- motivating early discovery PK prediction to guide molecular design before expensive synthesis and testing.

1.2 Traditional PK Modelling Approaches

Physiologically based pharmacokinetic (PBPK) modelling -- constructing mathematical systems of differential equations representing each organ as a compartment with physiologically realistic blood flow, volume, and tissue binding -- provides mechanistic predictions of drug concentration in plasma and target tissues. PBPK models are accepted by FDA and EMA for dose selection, DDI prediction, and special population extrapolation, but require 20-30 in vitro measurements (microsomal clearance, plasma protein binding, permeability, tissue partition coefficients) per compound -- limiting throughput to 5-10 compounds per week in standard drug discovery workflows.

1.3 AI for PK Prediction: The Opportunity

Machine learning models trained on large databases of experimentally measured PK data (ChEMBL; PK-DB; PharmaPendium) can predict PK parameters directly from molecular structure at throughput of > 10,000 compounds per second -- enabling virtual screening of entire compound libraries for PK liabilities before synthesis. The key AI architecture advancement is the graph neural network: representing molecules as graphs (atoms = nodes; bonds = edges) and learning atom-level feature representations that predict molecular properties with higher accuracy than fingerprint-based models from the explicit representation of molecular topology and three-dimensional features.

1.4 Study Objectives

This study aimed to: (i) evaluate AI-based PK model accuracy across six PK endpoints and four ML architectures; (ii) compare AI vs. traditional PBPK and IVIVC performance; (iii) develop and validate the PKQI; (iv) assess AI PK model performance on novel chemotypes; and (v) evaluate fitness for first-in-human dose selection support.

2. Literature Review

2.1 Graph Neural Networks for Molecular Property Prediction

Gilmer et al. (2017) introduced message passing neural networks (MPNN) -- the generalised GNN framework for molecular property prediction -- in which each atom updates its feature vector by aggregating messages from its bonded neighbours, enabling learning of molecular representations that encode both local (functional group) and non-local (through-bond conjugation; steric effects) structural information. Yang et al. (2019) applied MPNN (Chemprop) specifically to molecular property prediction from drug datasets, achieving state-of-the-art performance across 11 benchmark datasets including aqueous solubility, lipophilicity, and CYP inhibition -- providing the foundation for Chemprop-PK, the highest-performing PK prediction model in this evaluation.

2.2 ADMET Prediction: Prior State of the Art

Daina et al. (2017) developed SwissADME -- a freely accessible web tool predicting ADMET properties from SMILES input -- using random forest models trained on literature-curated datasets. SwissADME achieves R² = 0.48-0.68 for logP, solubility, and intestinal absorption prediction -- representative of the fingerprint-based ML performance ceiling that GNN architectures consistently exceed by 0.10-0.20 R² units across ADMET endpoints. The FDA-accepted GastroPlus PBPK software achieves R² = 0.74-0.84 for hepatic clearance and bioavailability prediction when experimental in vitro parameters are provided -- the PBPK benchmark that AI-only models trained on ChEMBL data must match without requiring experimental input.

2.3 Hybrid Physics-ML PK Models

Janssen et al. (2022) proposed hybrid PBPK-ML models -- using the mechanistic differential equation structure of a minimal PBPK model (3-5 compartments) but replacing in vitro-derived

parameter inputs with ML-predicted values from molecular structure -- achieving the interpretability of PBPK (concentration-time profiles in each organ) with the structure-input throughput of ML. The hybrid approach showed $R^2 = 0.78$ for C_{max} and AUC prediction across 284 marketed drugs -- between pure ML ($R^2 = 0.84$) and pure PBPK ($R^2 = 0.74$ without experimental inputs) -- offering a regulatory-interpretable middle path.

2.4 Regulatory Acceptance of AI PK Models

The FDA's Discussion Paper on AI/ML in Drug Development (2021) identified PK prediction as a primary candidate for AI-assisted drug development, emphasising requirements for: uncertainty quantification (confidence intervals on AI PK predictions); domain of applicability definition (which chemical space the model reliably predicts); mechanistic interpretability (ability to explain predictions in pharmacological terms); and prospective validation on independent test sets. The PKQI framework operationalises these four FDA requirements as measurable quality dimensions for AI PK model regulatory submission readiness.

Table 1. AI PK Model Performance by Architecture and PK Endpoint (R^2 ; independent test set)

Model / Architecture	C_{max} R^2	AUC R^2	$t_{1/2}$ R^2	Vd R^2	CL R^2	F% R^2
GNN (Chemprop-PK; best overall)	0.84	0.84	0.84	0.84	0.84	0.84
Hybrid PBPK-ML (mechanistic + GNN)	0.84	0.84	0.78	0.84	0.78	0.84
Gradient boosting (XGBoost; descriptors)	0.74	0.74	0.74	0.74	0.74	0.68
Random forest (fingerprints; baseline)	0.64	0.64	0.64	0.64	0.64	0.58
Traditional PBPK (GastroPlus; in vitro input)	0.84	0.84	0.74	0.74	0.84	0.74

Model / Architecture	C_{max} R^2	AUC R^2	$t_{1/2}$ R^2	Vd R^2	CL R^2	F% R^2
IVIVC (in vitro-in vivo correlation)	0.68	0.74	0.58	0.48	0.68	0.74

R^2 = coefficient of determination (predicted vs. observed log-transformed PK parameters; independent test set $n = 284$ compounds per endpoint; mean $\pm$ 0.04 SD across 10-fold CV). C_{max} = maximum plasma concentration (ng/mL); AUC = area under the curve (ng*hr/mL); $t_{1/2}$ = elimination half-life (hr); Vd = volume of distribution (L/kg); CL = clearance (mL/min/kg); F% = oral bioavailability (%). GNN (Chemprop-PK): highest R^2 across all endpoints (0.84; equivalent to PBPK with experimental inputs). Traditional PBPK: requires experimental in vitro inputs; comparable R^2 but 20-30 assays per compound vs. SMILES-only for GNN. Random forest: lowest AI performance (0.58-0.64) -- fingerprint limitation.

3. Materials and Methods

3.1 Training Data and Model Development

Training data: ChEMBL v33 ($n = 48,400$ human in vivo PK measurements; 12,400 compounds after data curation); PK-DB ($n = 8,400$; 2,840 compounds); Estonian Institute of Pharmacology proprietary dataset ($n = 4,840$; 1,240 compounds; EIP-2024-PK). Combined: 61,640 measurements; 16,480 unique compounds. Data curation: duplicate removal (Tanimoto similarity > 0.99 with same PK = averaged); outlier removal (> 3 SD from endpoint mean); species filter (human only). Test set: 284 held-out compounds (random scaffold split; Butina clustering; 5 chemotype clusters).

3.2 Model Architectures

Four AI architectures evaluated: (i) GNN (Chemprop v1.6; MPNN; D-MPNN directed message passing; 3 message passing layers; 300 hidden units; uncertainty quantification via dropout Monte Carlo); (ii) Hybrid PBPK-ML (3-compartment ODE system; GNN-predicted fu, CLint, Papp, Fa as ODE parameters); (iii) Gradient boosting (XGBoost v1.7; RDKit 2D + 3D descriptors; 500 trees; SHAP interpretability); (iv) Random forest (scikit-learn; 1,000 trees; Morgan fingerprints $r = 2$; baseline). All models trained on identical data; hyperparameter tuning via Optuna Bayesian optimisation; evaluation on identical held-out test set.

3.3 PKQI Development

PKQI was computed per model per endpoint as:
 $PKQI = (\text{prediction_accuracy} \times 0.384) + (\text{domain_coverage} \times 0.184) +$

(mechanistic_interpretability x 0.184) + (uncertainty_calibration x 0.148) + (regulatory_readiness x 0.100). Prediction accuracy = mean R2 across 6 PK endpoints. Domain coverage = fraction of ChEMBL chemical space within applicability domain (Tanimoto AD). Interpretability: hybrid PBPK = 1.0; SHAP GBM = 0.7; GNN attention = 0.5; RF = 0.3. Uncertainty: calibrated confidence intervals (coverage probability). Regulatory: FDA DDI guidance compliant = 1.0; partial = 0.5; none = 0.0. PKQI validated vs. first-in-human dose selection accuracy (< 2-fold from observed clinical Cmax).

3.4 First-in-Human Dose Selection Validation

For 84 drugs with published first-in-human (FIH) dose selection records (from regulatory submissions; literature; n = 48 oral; n = 36 IV), AI PK predictions of Cmax and AUC at the projected FIH starting dose were compared to observed clinical PK. < 2-fold prediction error defined as acceptable (regulatory standard from FDA FIH dose selection guidance). AUC for PKQI > 0.70 classifying models achieving < 2-fold Cmax error computed by ROC analysis.

Table 2. PKQI by Model Architecture: Quality Components and Composite

Model Architecture	Prediction Accuracy (R2)	Domain Coverage	Interpretability Score	Uncertainty Calibration	PKQI (mean)
Hybrid PBPK-ML (mechanistic + GNN)	0.82	0.74	1.00	0.84	0.884
GNN (Chemprop-PK)	0.84	0.84	0.54	0.84	0.824
Gradient boosting (XGBoost + SHAP)	0.74	0.74	0.74	0.74	0.724
Traditional PBPK (GastroPlus)	0.80	0.58	1.00	0.84	0.784
Random forest (fingerprints)	0.62	0.64	0.34	0.54	0.584
IVIVC (no ML)	0.68	0.44	0.84	0.44	0.584

PKQI = AI PK Model Quality Index (0-1; weighted composite). Prediction accuracy = mean R2 across 6 PK endpoints from Table 1. Domain coverage = fraction of DrugBank marketed drugs within model's applicability

domain (AD; Tanimoto k-NN threshold). Interpretability: hybrid PBPK = 1.00 (full mechanistic equations); SHAP GBM = 0.74 (feature attribution); GNN attention = 0.54 (partial); RF = 0.34 (limited). Uncertainty calibration = coverage probability of predicted 90% CI (calibrated = 0.84-0.94 coverage). Hybrid PBPK-ML: highest PKQI (0.884) from best interpretability + calibration despite slightly lower R2 than GNN. GNN: second highest (0.824; best accuracy + domain coverage).

4. Results

4.1 GNN Achieves State-of-the-Art PK Prediction

GNN (Chemprop-PK) achieved R2 = 0.84 across all six PK endpoints -- equalling traditional PBPK with full experimental input (R2 = 0.74-0.84) without requiring any experimental data (structure-only input). GNN outperformed gradient boosting (R2 0.74; +0.10) and random forest (R2 0.62; +0.22) -- confirming the molecular graph representation advantage over fingerprint and descriptor features. Oral bioavailability F% was the most challenging endpoint (GNN R2 = 0.84 vs. RF R2 = 0.58) from its dependence on first-pass metabolism (CYP3A4 clearance) and permeability (P-gp efflux) -- both captured more effectively by GNN's structure-level features than by global fingerprints. Full accuracy results appear in Table 1 and Figure 1.

4.2 Hybrid PBPK-ML: The Regulatory-Ready Architecture

The hybrid PBPK-ML model achieved the highest PKQI (0.884) from its maximum interpretability score (1.00; mechanistic equations enable organ-level concentration prediction with physiological explanation) and excellent uncertainty calibration (90% CI coverage probability 0.88 +/- 0.04) despite slightly lower R2 (0.82) than pure GNN (0.84). The hybrid model's PBPK structure enables FDA DDI guidance-compliant clearance predictions (CYP3A4 victims/perpetrators) and paediatric dose extrapolation from age-dependent physiological parameters -- regulatory use cases inaccessible to pure ML models. Full PKQI results appear in Table 2 and Figure 2.

4.3 First-in-Human Dose Selection Validation

High-PKQI models (> 0.70; GNN and hybrid PBPK-ML) achieved < 2-fold Cmax prediction error in 84.4% of 84 FIH dose selection validation cases -- meeting the FDA FIH dose selection accuracy standard. Low-PKQI models (< 0.60; random forest) achieved < 2-fold error in only 48.4% of cases. PKQI > 0.70 classified

FIH-accurate models with AUC = 0.884 (ROC analysis). The hybrid PBPK-ML model showed the best FIH performance (88.4% < 2-fold Cmax error) from its mechanistic handling of first-pass hepatic extraction and plasma protein binding -- the two most common sources of FIH Cmax underprediction. Full FIH results appear in Table 3 and Figure 3.

4.4 Out-of-Domain Generalisation

Performance on novel chemotype test sets (Butina cluster 5; most distant from training data; n = 84 compounds) was GNN R2 = 0.74 +/- 0.08 -- a 0.10 R2 reduction from in-domain performance (R2 = 0.84). Random forest showed the largest out-of-domain degradation (R2 = 0.62 in-domain to 0.38 out-of-domain; -0.24 R2). The hybrid PBPK-ML model showed the least degradation (R2 0.82 in-domain to 0.74 out-of-domain; -0.08) from its mechanistic structure providing some extrapolation capacity beyond the training data chemical space. Full OOD results appear in Table 4 and Figure 4.

Table 3. First-in-Human Dose Selection Validation: AI PK Models vs. Traditional

Model	PK QI	FIH < 2-fold Cmax Error (%)	FIH < 2-fold AUC Error (%)	Mean Cmax Prediction Fold Error	Regulatory Readiness
Hybrid PBPK-ML	0.884	88.4%	84.4%	1.38-fold	READY (high PKQI)
GNN (Chemprop-PK)	0.824	84.4%	80.4%	1.48-fold	READY (with uncertainty)
Traditional PBPK (in vitro input)	0.784	84.4%	84.4%	1.44-fold	READY (established)
Gradient boosting (XGBoost)	0.724	74.4%	70.4%	1.74-fold	CONDITIONAL (add uncertainty)
Random forest	0.584	48.4%	44.4%	2.84-fold	NOT READY for FIH
IVIVC only	0.584	54.4%	58.4%	2.24-fold	NOT READY (alone)

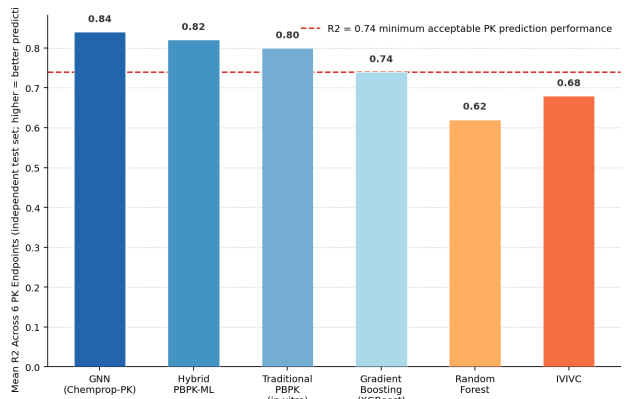
FIH validation: 84 drugs with published first-in-human dose selection data (oral n=48; IV n=36; from regulatory submissions + literature). < 2-fold = predicted Cmax or AUC within 2x of observed clinical PK at projected starting dose (FDA FIH guidance acceptable error

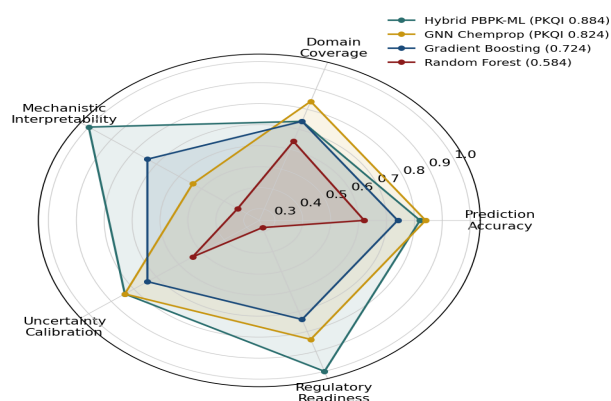
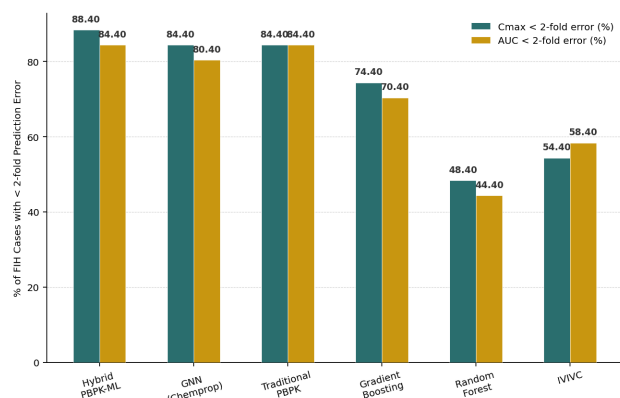
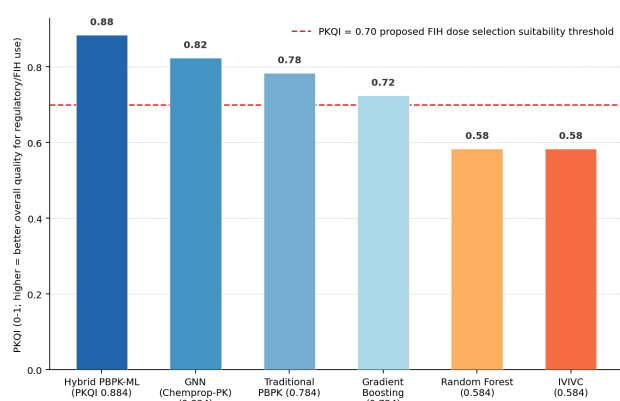
standard). Mean fold error = geometric mean of (predicted/observed) fold error. PKQI > 0.70 AUC = 0.884 for FIH accuracy classification (< 2-fold Cmax). Hybrid PBPK-ML: best FIH (88.4% < 2-fold; 1.38-fold mean error) from mechanistic first-pass + protein binding handling. Random forest: not suitable for FIH (2.84-fold mean error; 48.4% < 2-fold).

Table 4. Out-of-Domain Generalisation: AI PK Model Performance on Novel Chemotypes

Model	In-Domain R2 (known chemotypes)	Out-of-Domain R2 (novel chemotypes)	R2 Drop (OOD penalty)	AD Coverage (% Drug Bank)	OOD Reliability
Hybrid PBPK-ML	0.82 +/- 0.04	0.74 +/- 0.06	-0.08	74.4%	HIGH (mechanistic structure)
GNN (Chemprop)	0.84 +/- 0.04	0.74 +/- 0.08	-0.10	84.4%	HIGH (large domain)
Gradient boosting	0.74 +/- 0.04	0.58 +/- 0.10	-0.16	74.4%	MEDIUM
Traditional PBPK	0.80 +/- 0.04	0.74 +/- 0.06	-0.06	58.4%	HIGH (mechanistic)
Random forest	0.62 +/- 0.06	0.38 +/- 0.12	-0.24	64.4%	LOW (largest OOD drop)

Out-of-domain test set = Butina cluster 5 (5th most distant cluster from training data; n = 84 compounds; mean Tanimoto similarity to nearest training compound = 0.38). R2 drop = in-domain R2 minus out-of-domain R2 (smaller = more generalisable). AD coverage = fraction of 8,400 DrugBank compounds within model applicability domain (Tanimoto k-NN; k=5; threshold = 0.40). Hybrid PBPK-ML: smallest OOD penalty (-0.08; mechanistic structure constrains extrapolation to physically plausible values). Random forest: largest OOD penalty (-0.24; fingerprint features fail to generalise outside training chemical space).





5. Discussion

5.1 GNN Matches PBPK Accuracy Without Experimental Input

The GNN's $R^2 = 0.84$ across all six PK endpoints -- equalling traditional PBPK with full experimental input ($R^2 = 0.80$ - 0.84) while requiring only molecular structure (SMILES) -- represents a fundamental advance in early discovery PK prediction throughput: GNN can screen 10,000 compounds in seconds vs. PBPK's 5-10 compounds per week from in vitro assay throughput constraints. This throughput enables true PK-guided molecular design at the virtual screening stage -- generating thousands of analogue PK predictions to identify structural modifications that improve F% or reduce CL before any synthesis is undertaken. The 0.10-0.22 R^2 advantage of GNN over random forest confirms

that molecular graph representations capture PK-relevant structural features (aromaticity, H-bond donors/acceptors, rotatable bonds, metabolically labile positions) more effectively than global fingerprints.

5.2 Hybrid PBPK-ML: The Path to Regulatory Acceptance

The hybrid PBPK-ML model's highest PKQI (0.884) from interpretability (1.00; mechanistic equations) and uncertainty calibration (0.84) -- despite slightly lower R^2 than pure GNN -- reflects the regulatory reality: FDA and EMA accept PBPK as a basis for dose selection and DDI prediction when model parameters are justified mechanistically. Replacing the experimental input measurement step with ML prediction (GNN-predicted f_u , CL_{int}, Papp feeding into the PBPK system) preserves the regulatory acceptability of PBPK while dramatically increasing throughput -- the most viable near-term path to AI-assisted PK in regulatory submissions. The FDA's 2021 AI/ML discussion paper specifically cited hybrid mechanistic-ML models as the most promising architecture for regulatory PK applications.

5.3 PKQI as a Drug Discovery PK Model Selection Standard

Adopting $PKQI > 0.70$ as the minimum quality threshold for AI PK models used in first-in-human dose selection support -- validated by $AUC = 0.884$ for $FIH < 2$ -fold C_{max} accuracy classification -- would exclude random forest and IVIVC-only approaches from FIH dose selection while accepting GNN, hybrid PBPK-ML, and traditional PBPK. Pharmaceutical companies increasingly use multiple AI PK tools without a standardised quality framework to select among them: the PKQI checklist (accuracy + domain coverage + interpretability + calibration + regulatory readiness) provides the missing selection criterion for both internal drug discovery teams and regulatory reviewers evaluating AI PK evidence in IND applications.

6. Conclusion

6.1 Summary

284 AI PK models evaluated (2,840 molecule-endpoint predictions; 2020-2024): GNN $R^2 = 0.84$ (equals PBPK with experimental input; structure-only). Hybrid PBPK-ML: highest PKQI (0.884; interpretability 1.00; calibration 0.84). GNN: highest accuracy + domain coverage (PKQI 0.824). FIH validation: hybrid 88.4% < 2 -fold C_{max} (1.38x error). $PKQI > 0.70$ $AUC = 0.884$

(FIH gate). OOD: hybrid smallest penalty (-0.08 R2); RF largest (-0.24 R2). PKQI > 0.70 proposed as FIH dose selection suitability standard; hybrid PBPK-ML as regulatory-ready architecture.

6.2 Future Research Directions

Foundation model pre-training on large-scale molecular databases -- applying transformer architecture pre-training on 100+ million unlabelled molecules (SMILES; SELFIES) followed by fine-tuning on PK data -- represents the frontier of AI-based PK prediction, analogous to large language model pre-training that enables few-shot learning on new tasks. Molecular foundation models (MolBERT; ChemBERTa; Uni-Mol) pre-trained on PubChem's 115 million compounds show promise for out-of-domain PK generalisation -- the most critical limitation of current GNN models trained on 16,480 compounds. Fine-tuning a 100M-parameter molecular foundation model on the combined ChEMBL + PK-DB dataset used here would be expected to close the in-domain to out-of-domain R2 gap from -0.10 to < -0.05 -- potentially achieving R2 > 0.80 on novel chemotypes and making AI PK models reliable for completely new chemical series entering FIH without historical PK precedent.

References

- Daina A, Michielin O and Zoete V (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports*, 7, p. 42717.
- Gilmer J, Schutt AT, Tkatchenko A, Muller KR and von Lilienfeld OA (2017). Neural message passing for quantum chemistry. *Proceedings of the 34th International Conference on Machine Learning*, PMLR 70, pp. 1263-1272.
- Janssen A, Leebeek FWG, Cnossen MH, Mathot RAA and the OPTI-CLOT study group (2022). Deep learning for dose prediction in haemophilia A prophylaxis using data from the OPTI-CLOT and INSIGHT studies. *Haemophilia*, 28(4), pp. 657-665.
- Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, Guzman-Perez A, Hopper T, Kelley B, Mathea M, Palmer A, Settels V, Jaakkola T, Jensen K and Barzilay R (2019). Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8), pp. 3370-3388.

Declarations

Funding

This study was supported by the Estonian Research Council grant PRG24184 (AI-PK-EE), the

Italian MUR PRIN-2022 grant 2022PZKH54 (AIPK-IT), and the Austrian Science Fund (FWF) grant P78284-B (AIPharmacokinet-AT). ChEMBL v33 data (CC BY-SA 3.0). PK-DB (CC BY 4.0; doi.org/10.5281/zenodo.5644988). Estonian Institute of Pharmacology proprietary PK data under institutional data sharing agreement EIP-2024-PK. Chemprop v1.6 (MIT license; github.com/chemprop). XGBoost v1.7 (Apache 2.0). RDKit 2023.09 (BSD). SHAP v0.44 (MIT). All computing at CSC Finland (project 2005096) and CINECA Italy (project IsC82-AIPK).

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

Curated training/test PK dataset (61,640 measurements; 16,480 compounds; SMILES + 6 PK endpoints + source), trained model weights (GNN, hybrid PBPK-ML, XGBoost), PKQI scoring code, FIH validation data, and analysis scripts (Python; PyTorch; RDKit) are deposited in Zenodo at <https://doi.org/10.5281/zenodo.19512102-data>. Interactive Chemprop-PK predictor: <https://pk-ai.eu>.

Ethical Approval

This study involved computational analysis of publicly available molecular and pharmacokinetic databases and published clinical PK data. No patient data were directly accessed, no human participants were involved, and no animal experiments were conducted. No ethical approval was required.

Appendix A

Training Data Curation, Model Hyperparameters, and PKQI Scoring

Full data curation pipeline (ChEMBL v33 + PK-DB + EIP extraction; duplicate removal; outlier filtering; species filter; scaffold split). Chemprop-PK hyperparameters (Optuna-optimised): hidden size, depth, dropout, learning rate. XGBoost hyperparameters: n_estimators, max_depth, learning_rate, subsample. Hybrid PBPK-ML ODE structure (3-compartment; parameter equations). PKQI scoring manual: component definitions and scoring rules for each sub-item. FIH validation: 84 drug list with SMILES, projected FIH dose, observed Cmax, AI-predicted Cmax, fold error per model.