AI-Powered Digital Twins for In Silico Clinical Trials: A Computational Methods Approach to Accelarate Drug Discovery, Optimize Dosing and Predict Patient-Specific Outcomes

Mikiyas Hailu, Dinaol Belina, and Michael Abdi

Contents

Introduction

The process of drug [1] discovery and clinical development is notoriously time consuming, costly, and sometimes burdened with higher failure rates, particularly when we plan to make a transition from preclinical studies to human trials. traditional clinical trials, despite their rigor, are often limited by population heterogeneity, ethical concerns, rigid design. this ineffeciency has sparked global interest in insilico clinical trials (ISTs), where computer-simulated models-especially AI-Powered digital twins [3] mimic patient-specific biological and pharmacological responses to therapies, enabling faster, safer, and more cost-effective decision-making in drug development().

Digital twins [2], originally used in engineering, are now redefining healthcare by enabling real-time, individualized simulations. when powered by machine learning and system biology, digital twins can integrate multi-omics, clinical, and lifestyle data to simulate disease disease progressionand thrapeutic responses in virtual patients. this patient-specific approach, capable of iterating thousands of tria scenarios, offer a paradigm shift on how we test drug efficacy, optimize dosing and stratify resopnders all before a drug is ever administered to a human. as shown bu (), this reduces the trial costs, enhances precision medicine ,and mitigates adverse outcomes in at-risk subpopulations.

Several recent studies have validated this approach for instance, the Avicenna Allience and the InSilicoTrials platform have demonstrated the regulatory potential of digital twins, accelerating drug appproval pipelines (). Additionally, () used a digital twin of type 1 diabetes petients to successfully simulate glucose responses to insulin thrapies. Similarly, recent research by () highlighted the use of AI and digital twins for predicting outcomes in oncology trials. these examples provide a growing evidence base and show that regulatory bodies like the FDA and EMA are increasingly receptive to ISTs as complementary to traditional trials.

Despite these momentum, there remain scientific and technical challenges. Building clinically relevant digital twins requires robust integration of diverse datasets, validation against real-world evidence, and alignment with ethical and regulatory frameworks. The proposed project aims to fill these gaps by developing an end-to-end digital twin frameworks that incorporates machine learning, mechanistic modeling and clinical trial simulation tailored to specific therapeutic areas.

This project will be highly relevant to both academia and industry stakeholders. Biopharmaceutical companies seek to reduce the cost and duration of clinical trials, and regulators aim to promote patient safety and innovtion. through this interdisciplinary approach, combining AI, Bioinformatics, Pharmacokinetics, and Clinical Data Science, the research is poised to contribute to a new era of AI-enabled precision

medicine and more ethical, efficient drug development pipelines.

Objectives

- **Framework Development:** Design an AI powered multi-organ digital twin framework integrating PK/PD,multi-omics, and real world data for insilico clinical trials.
- **Drug Modeling:** Develop and validate GNN models for drug-target-organ interaction prediction using CHEMBL, TG-GATEs, and DrugBank datasets.
- **Toxicity Prediction:** Impelement models for hepatotoxicity and nephrotoxicity using TG-GATEs toxicogenomics data and FAERS adverse event reports.
- **Dose Optimization:** Apply reinforcement learning for adaptive dose scheduling and therapy optimization in virtual cohorts.
- **Explainable AI:** Incorpoarate SHAP and Lime frameworks to ensure transparecy and regulatory grade interpretability of model outputs.
- **Validation and Translation:** Conduct retrospective validation against ClinicalTrials.gov oncology datasets and explore regulatory pathways for insilico clinical trial adoption.
- **Industry Alignment:** Assess economic and translational benefits of digital twin integration in pharmaceutical development pipelines.

Materials and Methods

- **Datasets:**
  - **Drug and Molecular Data :** DrugBank, ChEMBL, Pubchem
  - **Toxicity and Safety Data:**TG-GATEs, FAERS, SIDER
  - **Multi-Omics Data:** TCGA, LINCS, GTEx for transcriptomic and methabolomic integration
  - **Clinical Trials Data:** CinicalTrials.gov and MIMIC-IV for retrospective validation.
- **Computational Framework:**
  - **Hardware:** High-performance balanced workstation (RTX 4090 GPU, 128 GB RAM) supplemented with Azure A100 cloud resources for large scale simulation.
  - **AI Models:**
    - **GNN Models:** For Molecular Property and interaction prediction.
    - **Reinforcement Learning:** For optimization of drug-response simulations.
    - **Hybrid Models:** Combining AI and PK/PD models for organ-specific simulations.
  - **Software:** Pythorch, TensorFlow, DeepChem, and GROMACS for molecular simulations, Docker/MLflow for reproducibility, and Ray for distributed computing.
  - **Explainable AI:** SHAP, LIME, and counterfactual explanations for model interpretability.
- **Model Development:**
  - **Data Preprocessing:** Standardize molecular structures(SMILES), Normalize omics data, and harmonize adverse event reports.
  - **Digital Twin Construction:** Build organ-specific twins(liver and kidney) integrating AI predictions with mechanistic PK/PD models.
  - **Simulation Pipeline:** Generate synthetic patient cohorts using statistical variation and real-world clinical data distributions.
  - **Toxicity and Dose Trials:** Run Reinforcement Learning driven virtual trials to identify safe and effective dosing regimens.
  - **Validation and Testing:** Compare virtual outcomes against historical clinical trial data to assess predictive accuracy.
  - **Deployment:** Package Framework for reproducibility, industry collaboration and regulatory review.

Expected Outcomes

This project will deliver both scientific and practical tools for AI-driven drug discovery.

- **Novel AI Framework:** A validated multi-organ digital twin platform capable of simulating drug efficacy, toxicity and optimal dosing insilico.
- **Publications:** At least 8 high-impact publications including methodology, papers, validation studies, and regulatory alignment articles.
- **Datasets and Pipelines:** Curated and Preprocessed datasets (DrugBank, TG-GATEs, FAERS, TCGA) AND Reproducible pipelines made available under FAIR principles.
- **AI Models:**
  - **GNN Model:** For drug-organ interaction.
  - **Reinforcement Learning** For adaptive dosing
  - **Explainable AI:** To enhance trust and regulatory acceptance.
- **Regulatory Insights:** Framework for aligning digital twin outputs with FDA standards, supporting early regulatory adoption of in silico trials.
- **Economic Evaluation:** Quantitative assessment showing potential reductions in trial cost and duration, positioning digital twins as cost-effective complements to Phase I/II trials.
- **Industry Relevance:** Prototype collaboration-ready Software package facilitate translation into pharma ready pipelines.

Ultimately, this project will demonstrate that AI-powered digital twins can reduce trial failure rates, accelarate therapy development, and support precision medicine initiatives, bridging academic innovation with industry and regulatory needs.

Work Plan

- **Year 1: Foundational Work and Framework Setup**
  - Dataset collection from DrugBank, TG-GATEs, FAERS, TCGA, ClinicalTrials.gov.
  - Preprocessing Pipelines for omics and drug data.
  - Build a baseline PK/PD mechanistic models.
  - Publication target for Review and AI on Digital Twins.
- **Year 2: Model Development and Organ-Specific Twins**
  - Implement GNN for drug-target-organ prediction.
  - Develop liver and kidney digital twins using TG-GATEs and FAERS data'
  - Begin toxicity prediction pipelines.
  - Initial validation against retrospective trial data.
  - Publication target for GNN-powered Digital Twins.
- **Year 3: Reinforcement Learning and Personalization**
  - Implement Reinforcement Learning for Dose Optimization.
  - Integrate multi-omics data for (TCGA, LINCS) for patient specific twins.
  - Build explainability modules (SHAP, LIME)
  - Large scale simulateions for oncology and metabolic diseases.
  - Publication target for RL dose optimization paper and multi-omics personalization paper.
- **Year 4: Validation, Translation, and Dissemination**
  - Retrospective validation using CinicalTrials.gov oncology datasets.
  - Industry collaboration for regulatory alignment.
  - Economic alignment of trial cost/time reduction.
  - Publication target for Regulatory pathways paper and Economic impact study.

Budget

Table 1: Project Budget Breakdown

Category	Details	Amount (USD)
Personnel		Empty
	• 2 PhD Students (3 years)	
	• 1 Data Scientist (2 years)	
Software		Empty
	• ML Frameworks (Open-source)	
	• Commercial PK/PD tools (Schrodinger, Simcyp)	
Computational Resources		Empty
	• High-performance workstation (RTX 4090, 128GB RAM, 12TB Storage)	
	• Azure A100 cloud compute (simulation)	
	• Tuxedo Stellaris 17 PC for Portability	
	• Networking (VPN, Cloud Backup and Switches)	
Data Acquisition		Empty
	• DrugBank, ChEMBL, PubChem licenses	
	• Clinical trial data access	
Dissemination		Empty
	• Conference travel and registration	
	• Open-access publication fees	
Total		**None**

Bibliography

[1] Bergthor Björnsson, Carl Borrebaeck, Nils Elander, Thomas Gasslander, Danuta R Gawel, Mika Gustafsson, Rebecka Jörnsten, Eun Jung Lee, Xinxiu Li, Sandra Lilja, David Martínez-Enguita, Andreas Matussek, Per Sandström, Samuel Schäfer, Margaretha Stenmarker, X F Sun, Oleg Sysoev, Huan Zhang, Mikael Benson, and Swedish Digital Twin Consortium. Digital twins to personalize medicine. *Genome Med.*, 12(1):4, December 2019.

[2] Vishwa S Parekh and Michael A Jacobs. Deep learning and radiomics in precision medicine. *Expert Rev. Precis. Med. Drug Dev.*, 4(2):59–72, April 2019.

[3] Marco Viceconti, Adriano Henney, and Edwin Morley-Fletcher. In silico clinical trials: How computer simulation will transform the biomedical industry. *International Journal of Clinical Trials*, 3(2):37, May 2016.