AI-enabled Drug Discovery and Development: Opportunities and Challenges

Rajnish Kumar, Ph.D.
Associate Professor
Department of Pharmaceutical Engineering & Technology
Indian Institute of Technology (BHU), Varanasi, India

Nobel prize in Physiology and Medicine

Background

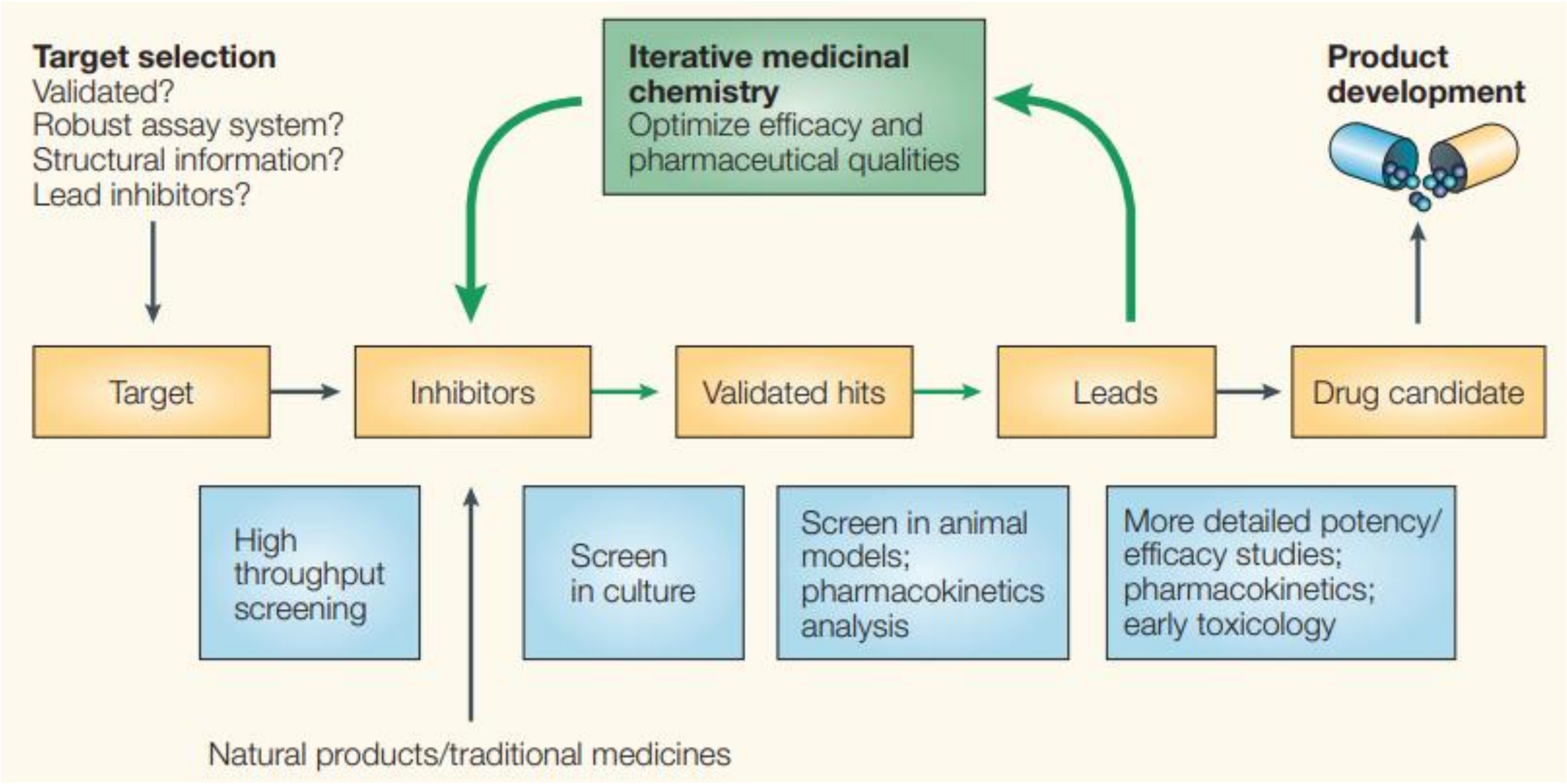
- Associate Professor, IITBHU, Varanasi, India (Since Nov. 2024)
- Assistant Professor, IITBHU, Varanasi, India (2019-2024)
- JSPS Summer Fellow, NIBIOHN, Osaka, Japan (2019)
- Assistant Professor, Karolinska Institutet, (2019)
- Postdoctoral researcher, Karolinska Institutet, Stockholm, Sweden (2016-2019), Winblad group
- Postdoctoral researcher, Karolinska Institutet, Stockholm, Sweden (2014-2016), Darreh-Shori group
- PhD in Pharmaceutical and Medicinal Chemistry from UIPS, Panjab University, Chandigarh, India (2014)

What is a drug?

A substance used to prevent or cure a disease or ailment or to alleviate its symptoms

Discovery

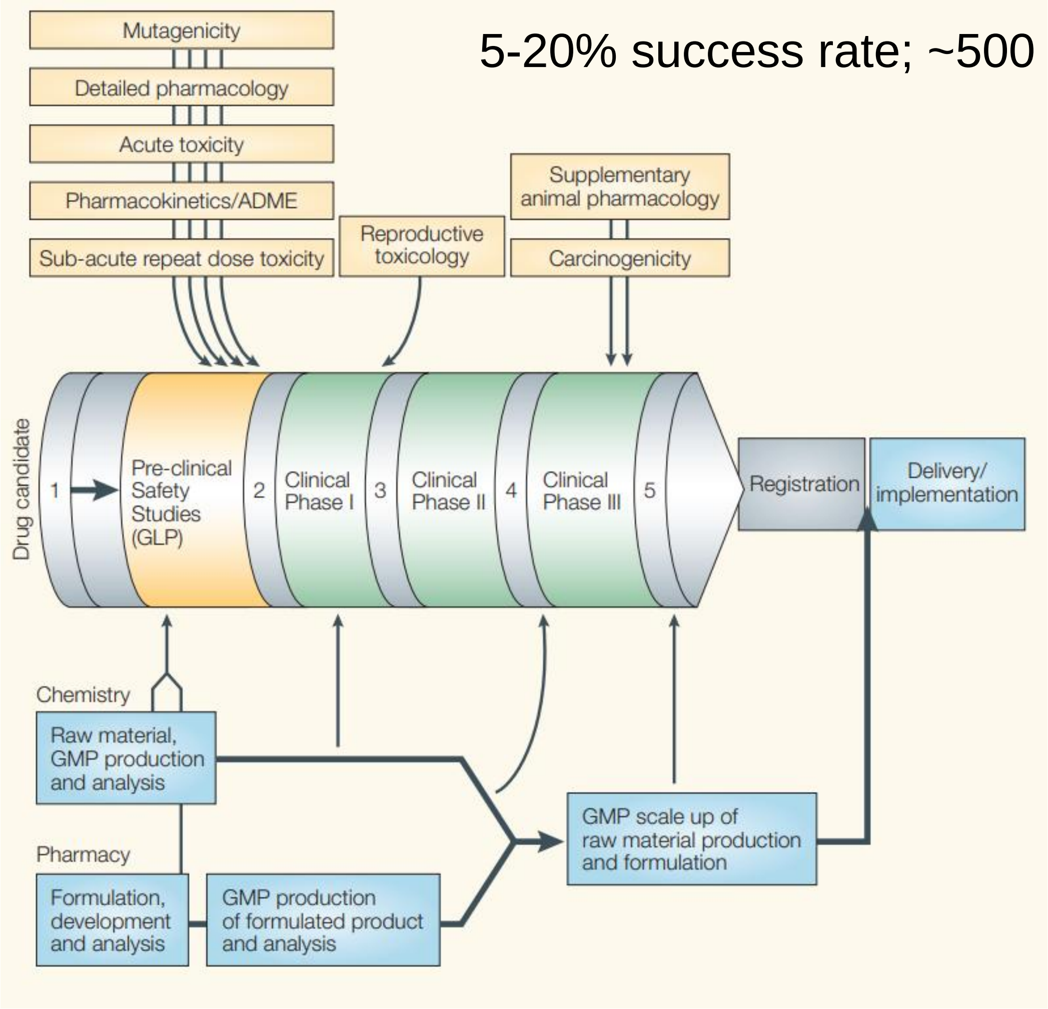
2-5% success rate; ~500 Million USD; 3-5 Years



Nwaka, S., Ridley, R. Virtual drug discovery and development for neglected diseases through public–private partnerships. *Nat Rev Drug Discov* **2,** 919–928 (2003).

Development

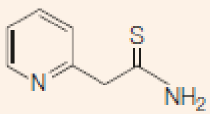
5-20% success rate; ~500 Million USD; 4-6 Years



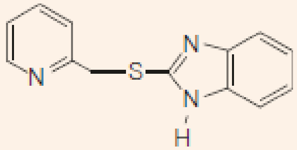
Nwaka, S., Ridley, R. Virtual drug discovery and development for neglected diseases through public–private partnerships. *Nat Rev Drug Discov* **2,** 919–928 (2003).

Omeprazole: a blockbuster drug

Liver toxicity



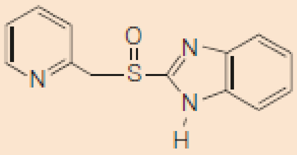
[CMN 131]



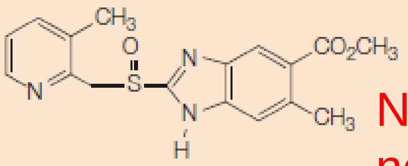
[H 124/26] (1973)

Patent issues

Metabolite of H-124
Inhibit Iodine uptake



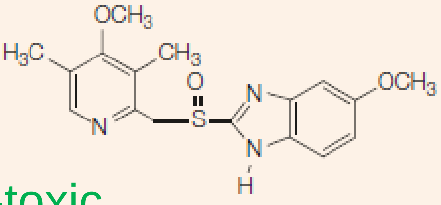
Timoprazole
[H 83/69] (1974)



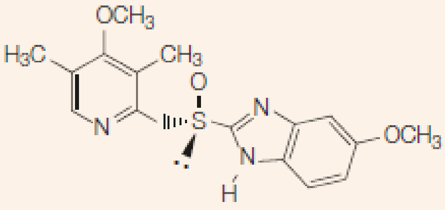
Picoprazole
[H 149/94] (1976)

New lead
necrotizing vasculitis

Potent, non-toxic



Omeprazole
[H 168/68] (1979)



Esomeprazole
(1983)

2

2 4

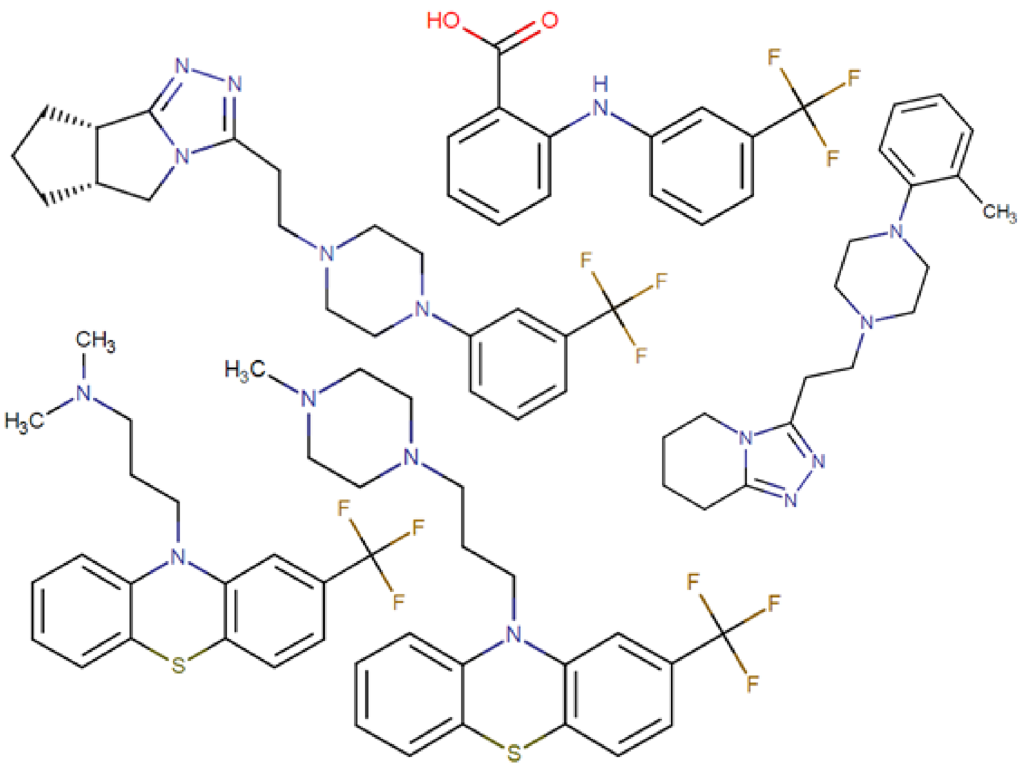
2 4 6

2 4 6 8

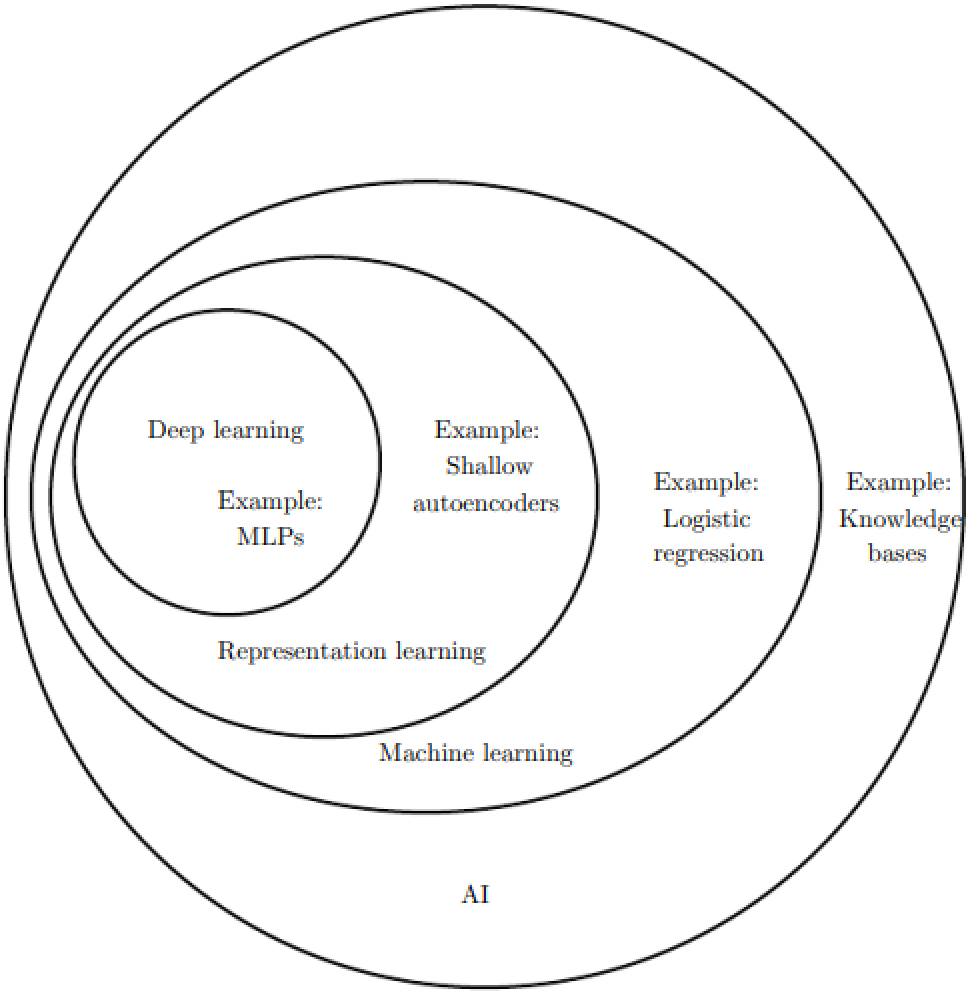
2 4 6 8 ?

2 4 6 8 10

Which of the following crosses BBB?



Artificial Intelligence (AI)



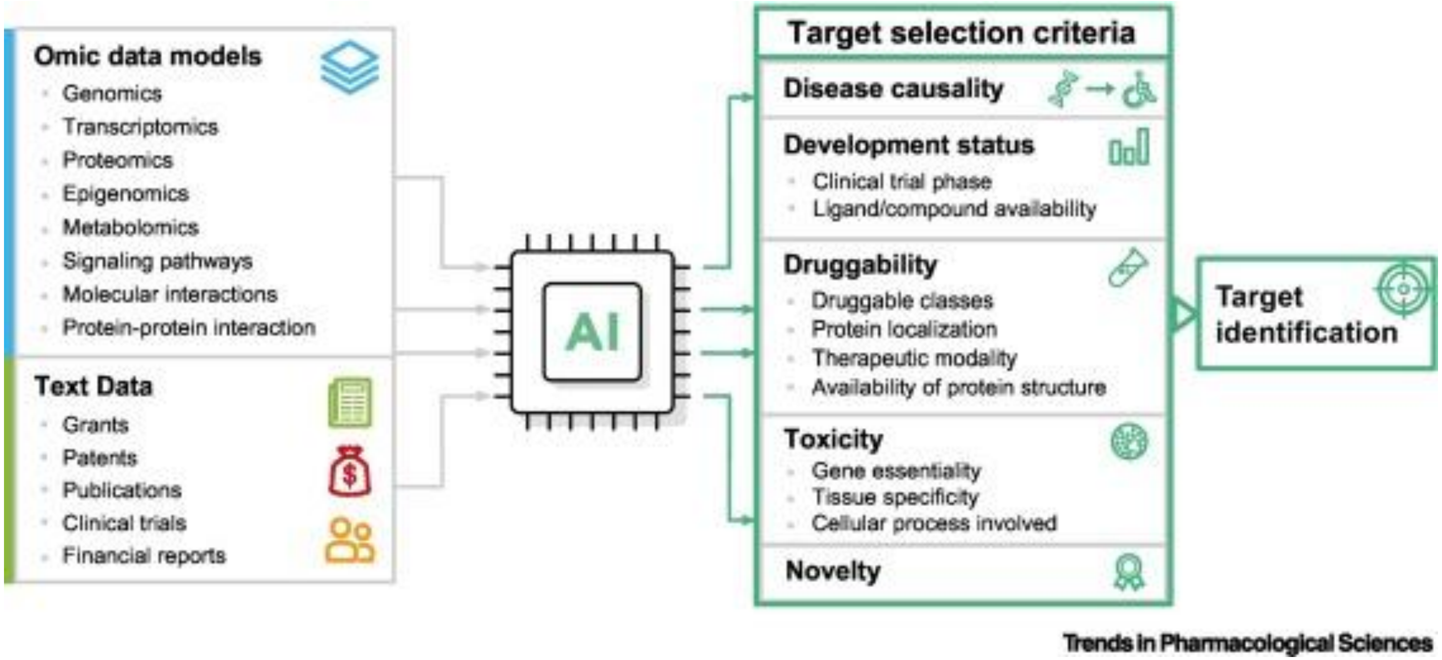
Machine Learning: Using known data develop a model to predict unknown data

- Known Data: Big enough archive, previous observations, past data
- Model: Known data + Algorithms (ML algorithm)
- Unknown data: Missing, unseen, not existing, future data

How AI Can Help to Accelerate Drug Discovery

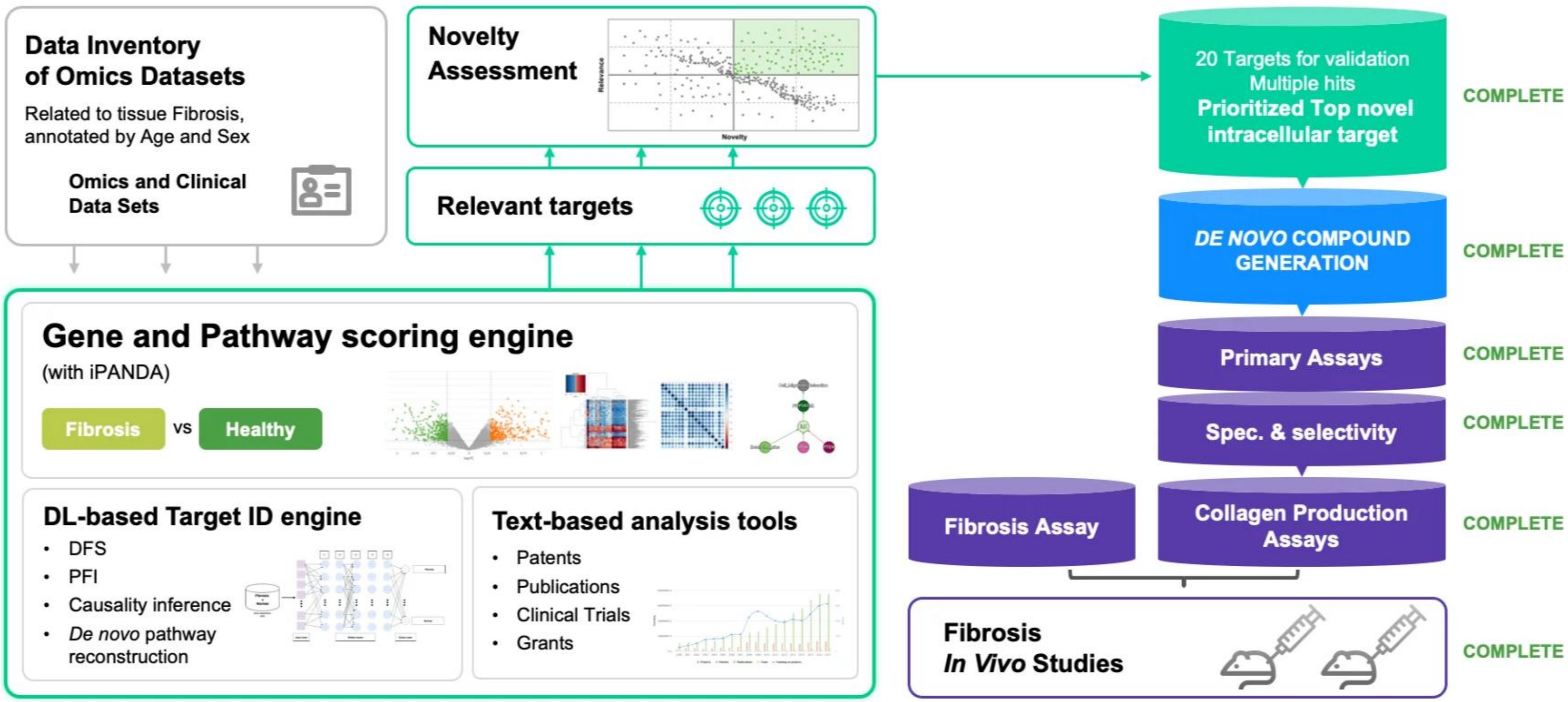
- Drug target identification and validation
  - → Protein-ligand interaction prediction
  - → Target prioritization using AI algorithms
- High-throughput screening and virtual compound libraries
  - → Virtual screening techniques
  - → Generative models for compound design
- Predictive toxicology and safety assessment
  - → In silico ADMET predictions
  - → Toxicity modeling with AI
- Biomarker discovery and patient stratification
  - → AI in genomics and proteomics
  - → Personalized medicine advancements

Target Identification



Pun, Frank W. et al., AI-powered therapeutic target discovery, Trends in Pharmacological Sciences, Volume 44, Issue 9, 561 - 572

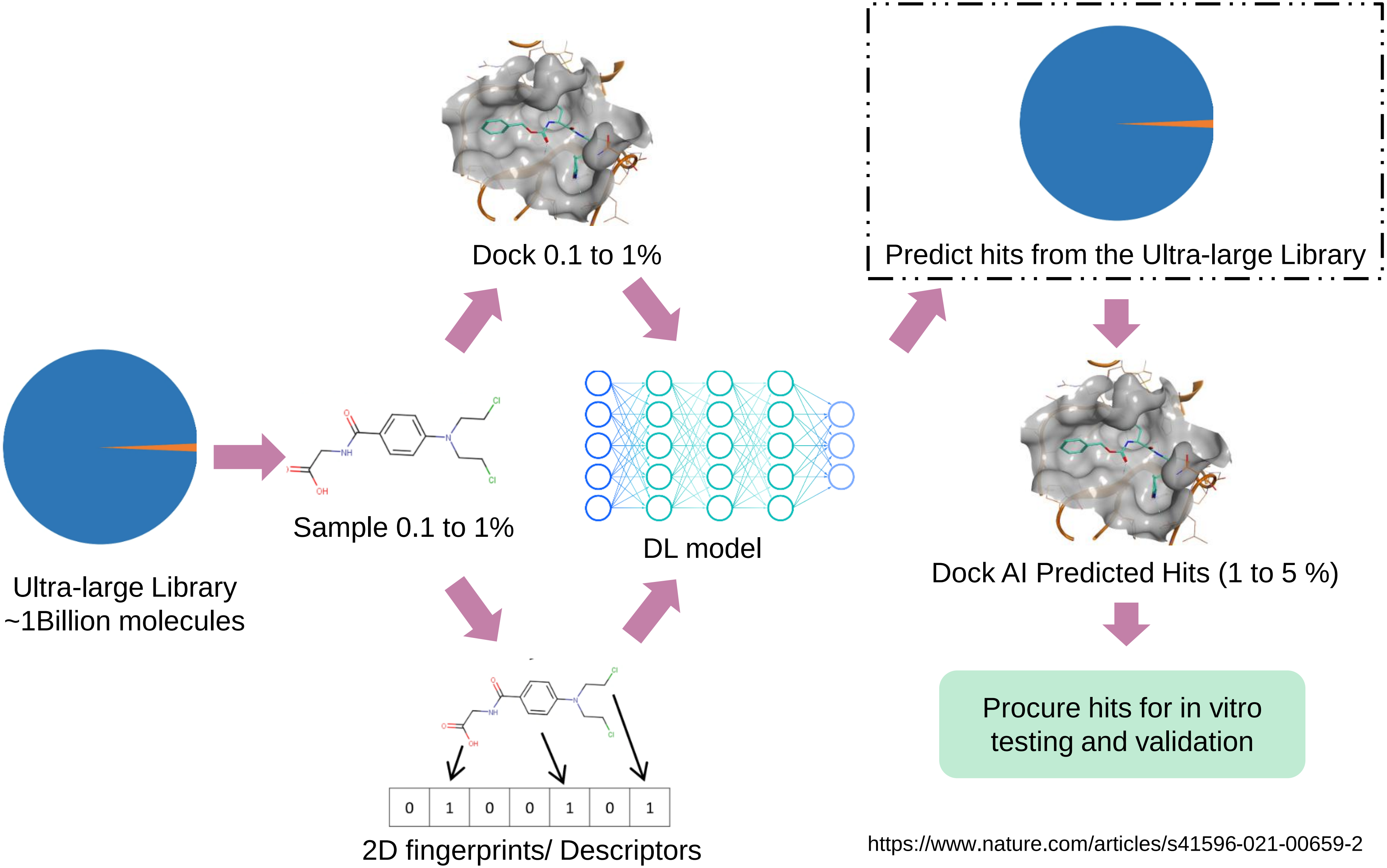
InSilico Medicine's INS018_055 for Idiopathic Pulmonary Fibrosis (IPF)



Pun, Frank W. et al., AI-powered therapeutic target discovery, Trends in Pharmacological Sciences, Volume 44, Issue 9, 561 - 572

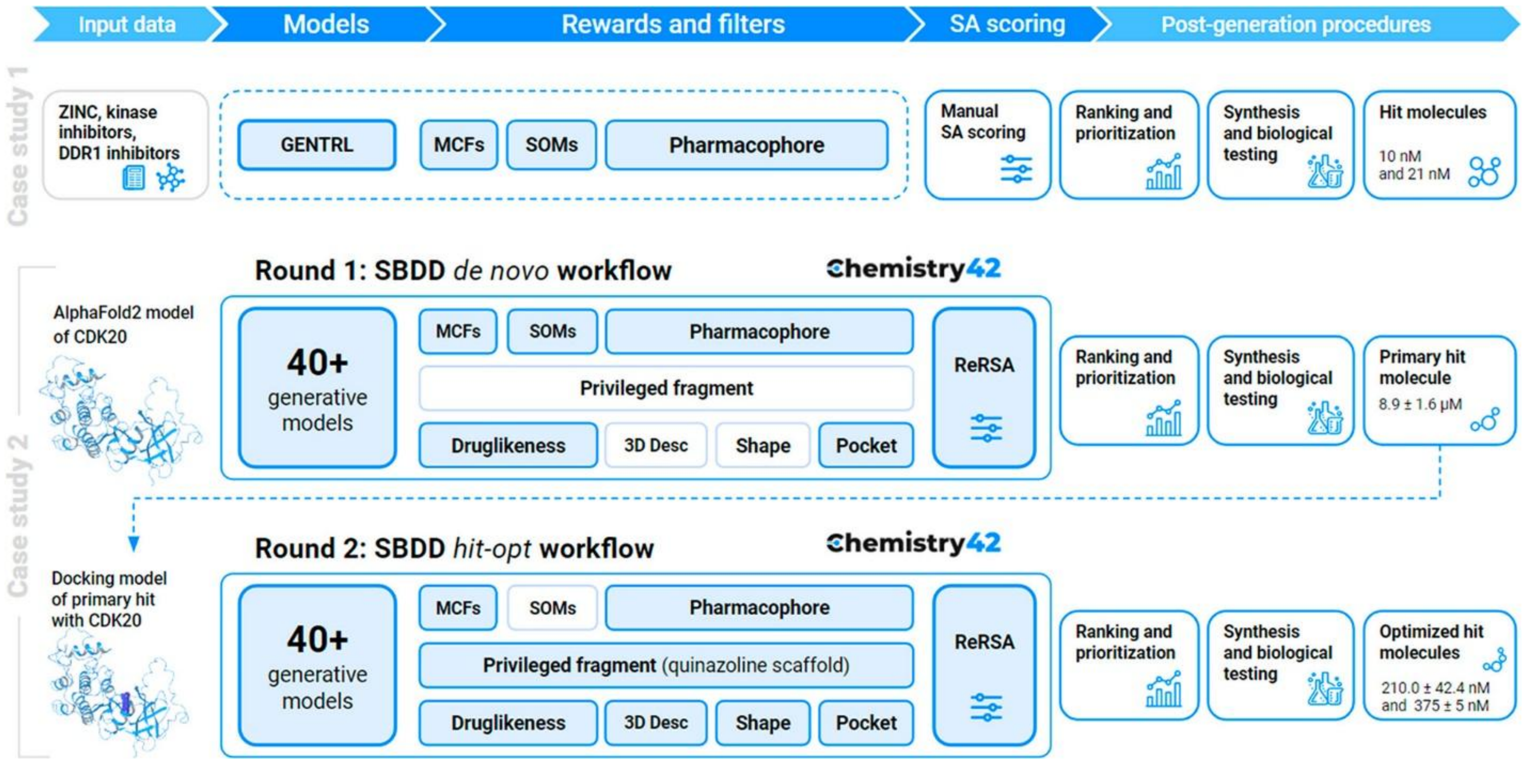
Hit Identification

Deep learning enabled ultra large virtual screening



https://www.nature.com/articles/s41596-021-00659-2

Chemistry42



Yan A. Ivanenkov, et. al., Chemistry42: An AI-Driven Platform for Molecular Design and Optimization, Journal of Chemical Information and Modeling 2023 63 (3), 695-701.

Generative modelling for molecule design

Generative AI

You

Can you suggest some journals in the field of AI and drug discovery?

ChatGPT

Certainly! Here are some reputable journals in the field of AI and drug discovery:

1. *Journal of Chemical Information and Modeling* - Covers research in computational chemistry and molecular modeling, including applications of AI in drug discovery.
2. *Journal of Medicinal Chemistry* - Focuses on the discovery and design of new medicinal compounds, including computational approaches and AI techniques.
3. *Drug Discovery Today* - Publishes research articles and reviews on all aspects of drug discovery, including the application of AI and machine learning.
4. *Journal of Computer-Aided Molecular Design* - Covers research in computational chemistry and molecular modeling, including the use of AI for drug discovery and design.
5. *Expert Opinion on Drug Discovery* - Provides expert reviews and opinions on emerging trends and technologies in drug discovery, including AI.
6. *Bioinformatics* - Publishes research articles on the development and application of computational methods for biological data analysis, including AI approaches in drug discovery.
7. *Nature Reviews Drug Discovery* - Features reviews and perspectives on the latest advances in drug discovery, including the role of AI and machine learning.

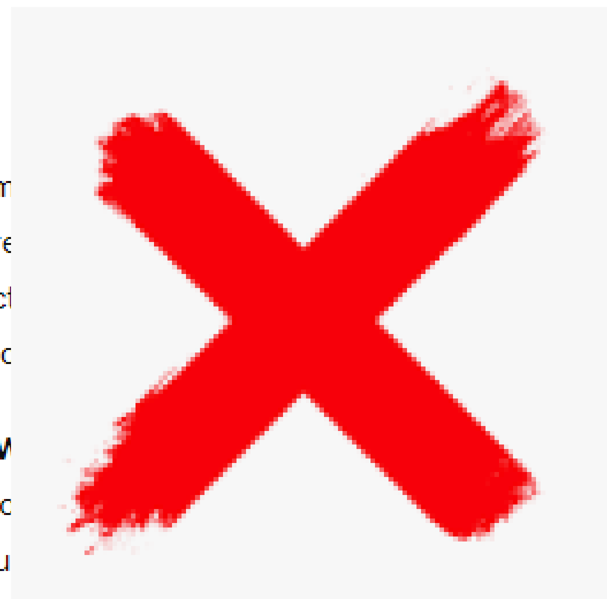
Generative AI for molecular design

You

Can you generate structures of new drug-like molecules that are permeable to blood-brain barrier?

ChatGPT

Designing drug-like m... rain barrier (BBB) is a complex task that requires care... al properties. While I can't provide specific struct... and properties that are associated with molec...

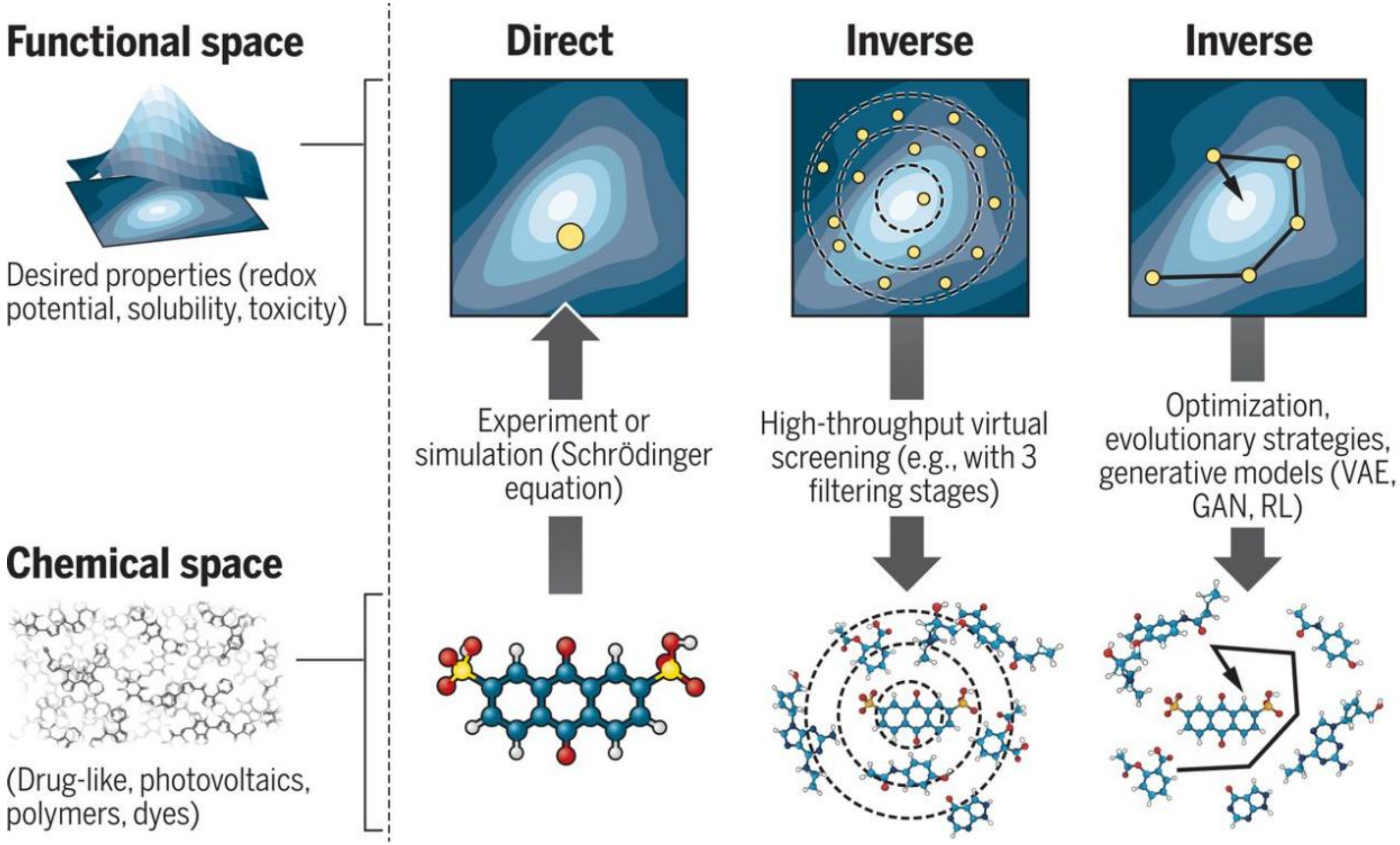


1. **Size and Molecular W...** ave better chances of crossing the BBB compared to ... transporters.
2. **Lipophilicity:** Molecu... n 1 and 3) have higher permeability through the BBB, as lipophilic compounds can dissolve in the lipid bilayer of the BBB.

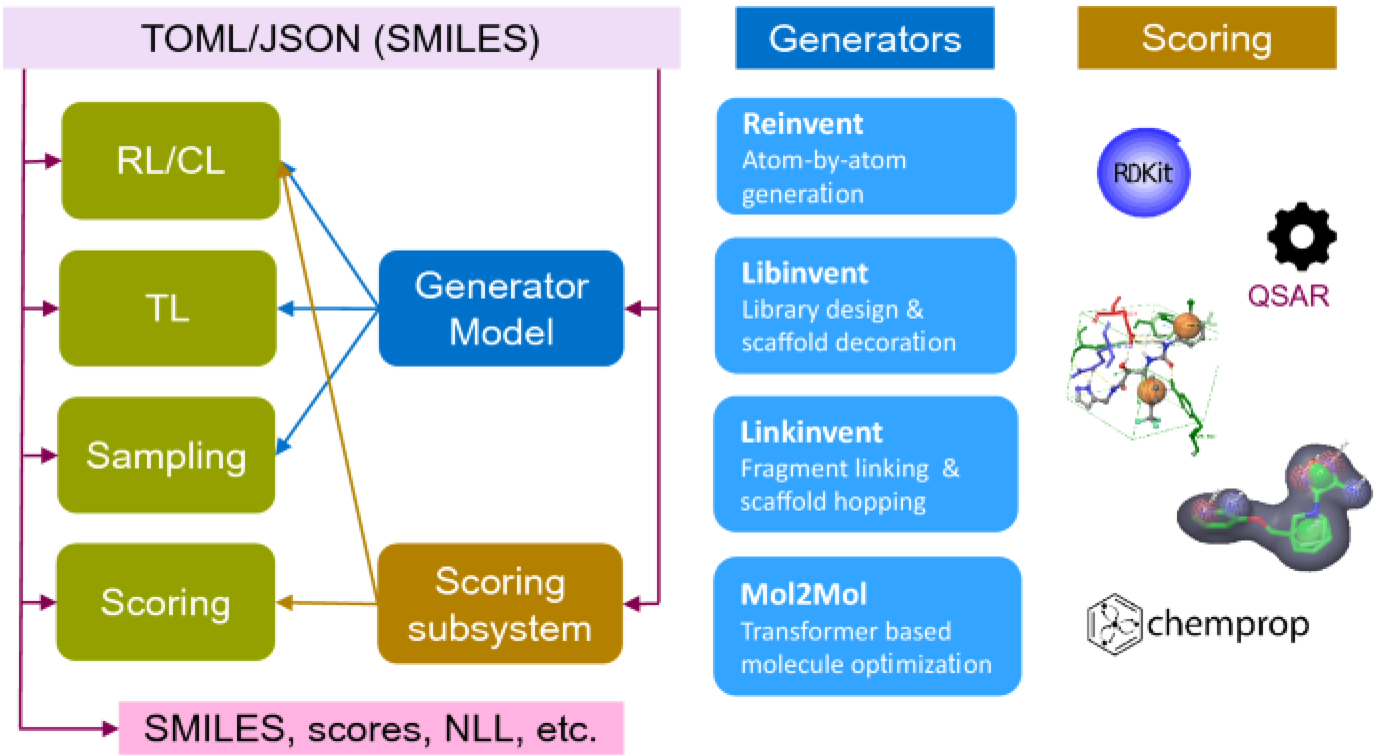
De novo molecular design

- Automatically proposes novel chemical structures meeting desired molecular profiles.
- To elicit a desirable biological response while maintaining acceptable pharmacokinetic properties.
- Also known as generative chemistry which reflects the rise of generative models in AI.
- These models are trained on large datasets and learn to produce output that is similar to the examples they were trained on.

Molecular Design Approaches



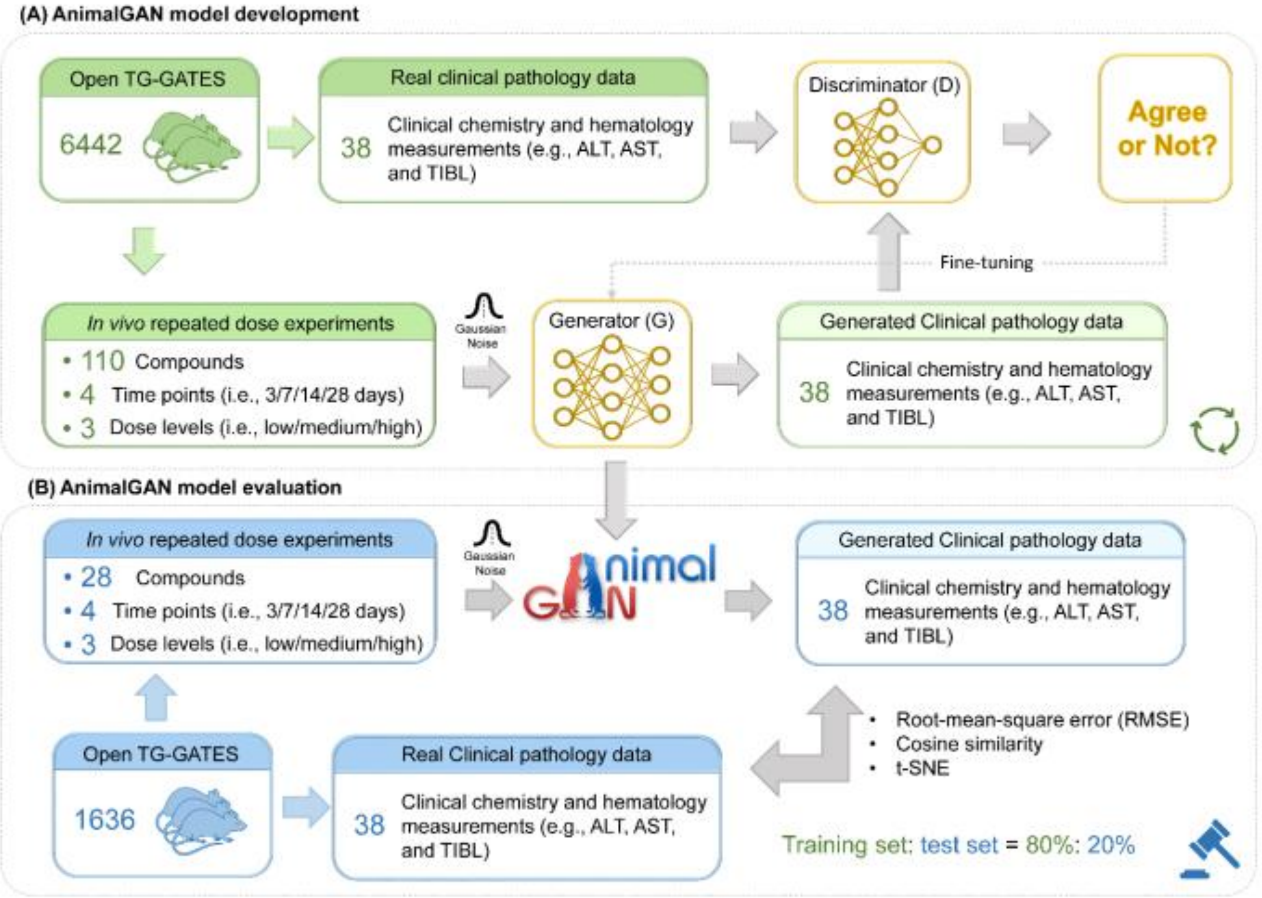
ReInvent4



Loeffler, H.H., He, J., Tibo, A. et al. Reinvent 4: Modern AI–driven generative molecule design. J Cheminform 16, 20 (2024).

Toxicity Prediction

Workflow (Xi chen et.al 2023)



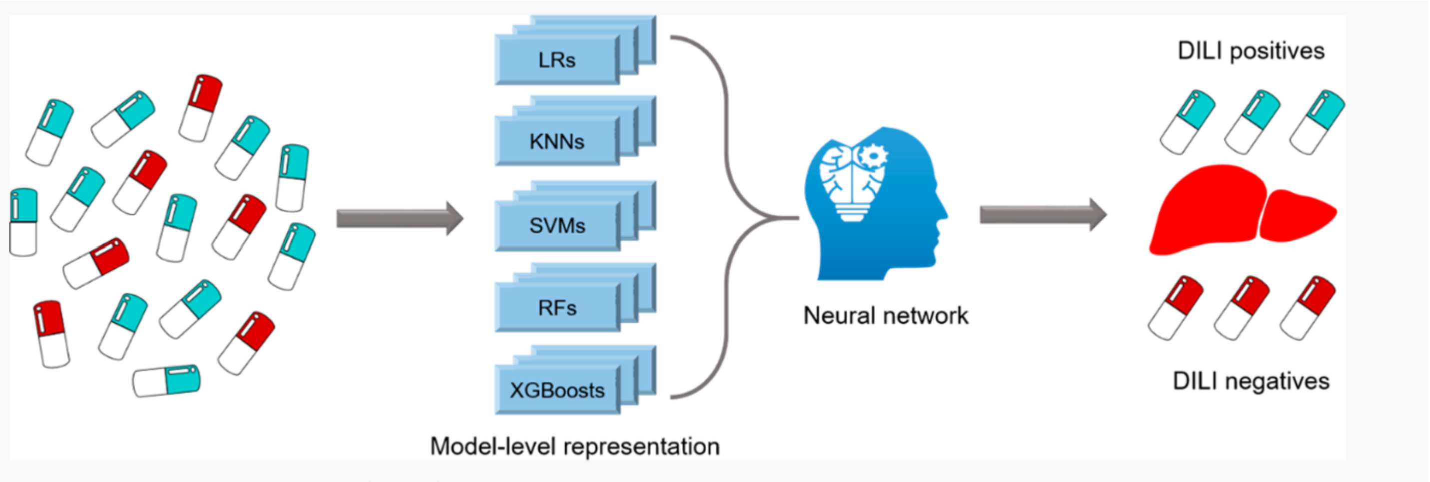
https://doi.org/10.1101/2023.03.25.534230

SafetAI

- SafetAI is a collaborative initiative led by CDER where NCTR is developing a suite of deep learning based QSAR models for various safety endpoints critical to regulatory science and the IND review.
- Focus on five key safety endpoints: Hepatotoxicity, Carcinogenicity, mutagenicity, nephrotoxicity and cardiotoxicity
- SafetAI facilitates drug safety research with a novel deep learning architecture that improves the “precision” in toxicity assessment by tailoring prediction to chemical characteristics.
- Examples of AI tools for assessing drug safety profiles include:-
- DeepDILI ; DeepCarc; DeepAmes

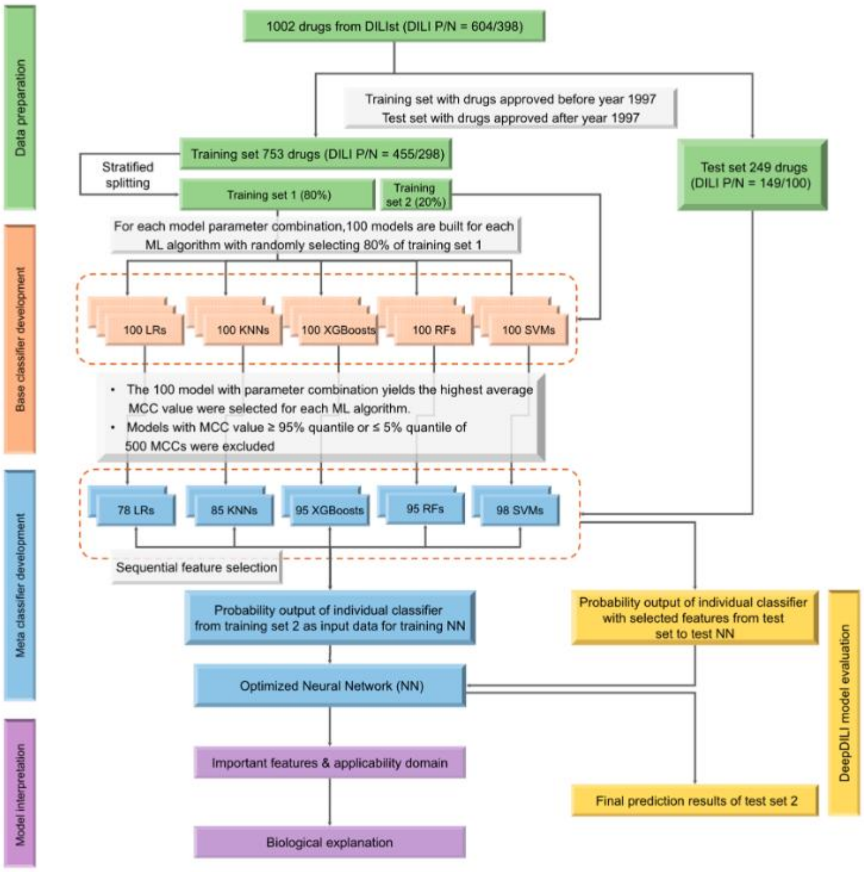
DeepDILI

- Deep Learning framework for prediction of Liver toxicity
- Generated by conventional machine learning algorithm with deep learning framework based on Mold2 descriptors.



https://pubs.acs.org/doi/10.1021/acs.chemrestox.0c00374

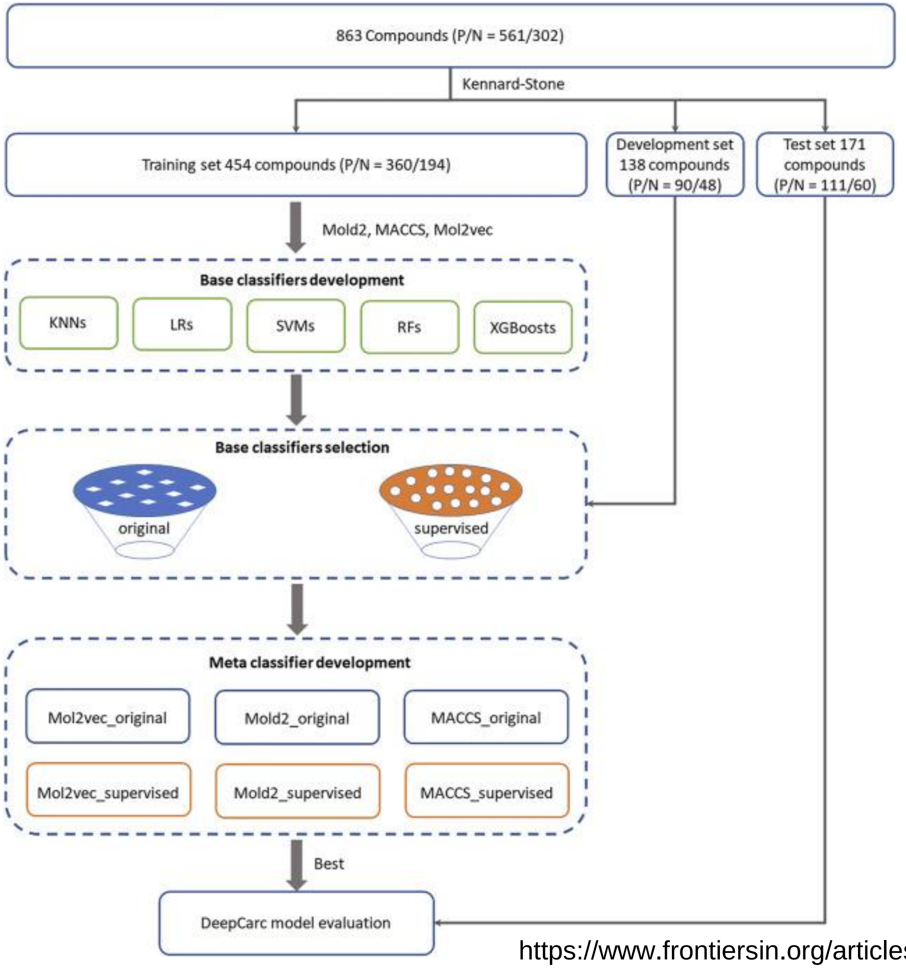
Workflow for DeepDILI model



https://pubs.acs.org/doi/10.1021/acs.chemrestox.0c00374

DeepCarc

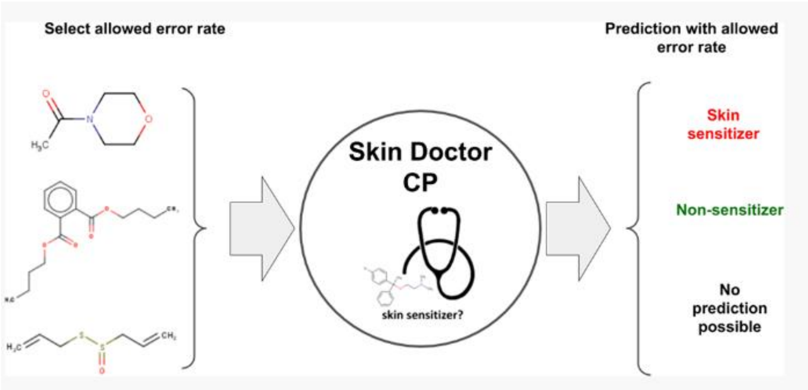
- Deep learning framework for prediction of carcinogenicity.



https://www.frontiersin.org/articles/10.3389/frai.2021.757780/full

Skin Doctor CP

- Machine learning model for classification of small organic compounds into Skin sensitizers and non-sensitizers.
- The core of Skin Doctor CP is a random forest binary classifier that is enveloped in an aggregated Mondrian conformal prediction framework.
- Allows users to define error significance level
- Skin Doctor CP is trained on a curated data set of 1285 compounds measured in the local lymph node assay (LLNA)

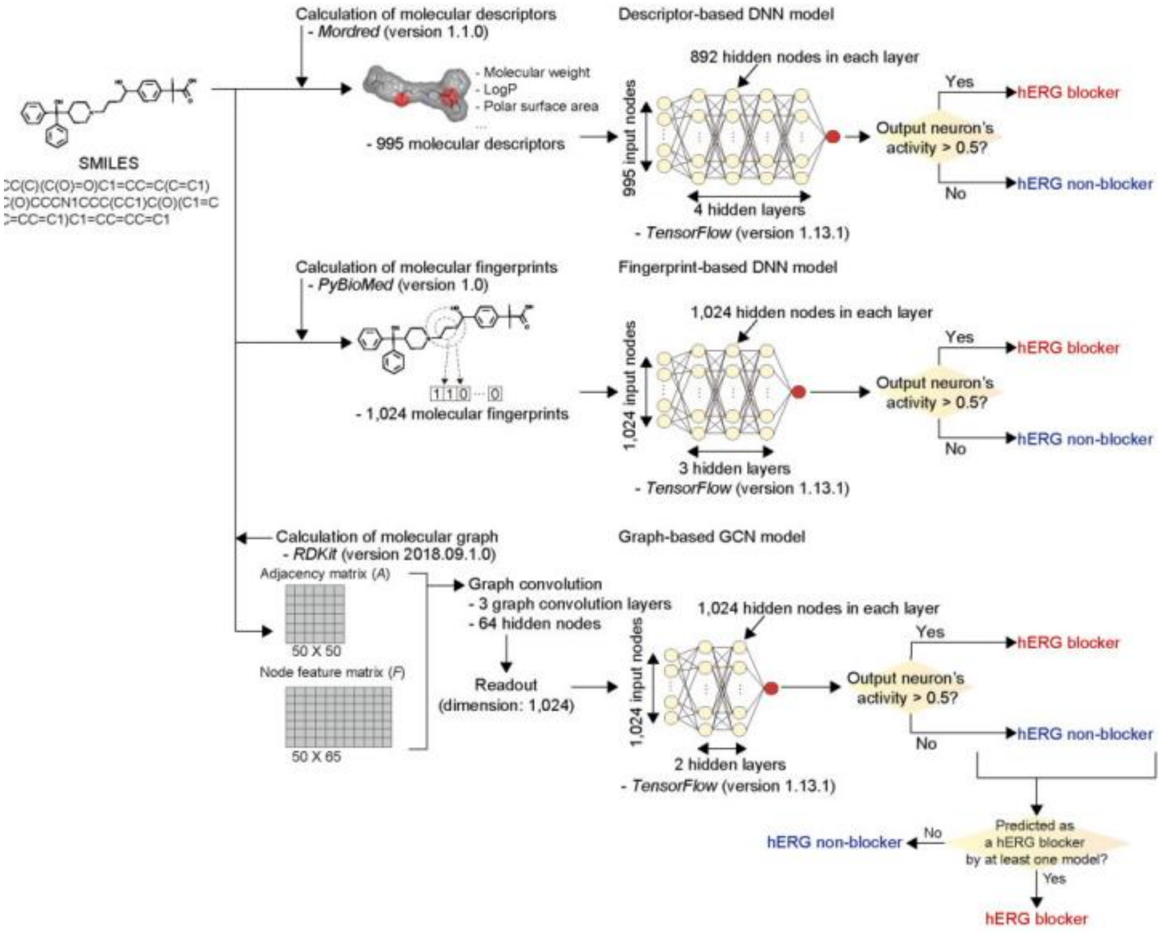


https://pubs.acs.org/doi/10.1021/acs.chemrestox.0c00253

DeepHIT

- Deep learning framework for prediction of hERG-induced cardiotoxicity.
- Consist of 3 optimized deep learning model
  - → Descriptor-based DNN
  - → Fingerprint-based DNN
  - → Graph-based GCN model

Schematic overview of DeepHIT



https://academic.oup.com/bioinformatics/article/36/10/3049/5727757

Predictive ADMET

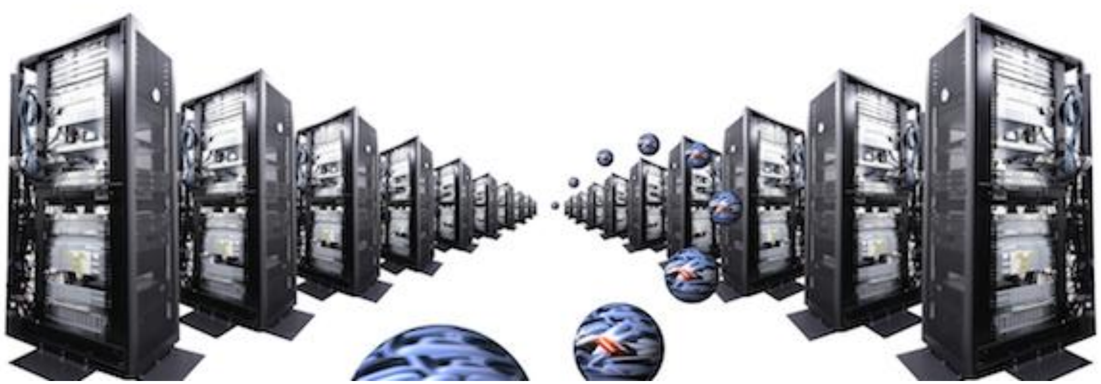
Mizuguchi Laboratory

National Institutes of Biomedical Innovation, Health and Nutrition

Home | News | Research | Members | Tools | Opportunities

Welcome to the Mizuguchi Laboratory

We are based in the National Institutes of Biomedical Innovation, Health and Nutrition, Japan. We carry out bioinformatics and computational biology research into drug discovery and development (with special emphasis on early-stage target discovery and toxicity prediction). Our research involves both the analysis of real-life experimental data and the development of novel computational techniques, aiming to establish systems approaches to drug discovery.



Mizuguchi Lab, NIBIOHN, Osaka, Japan (2019)

Support vector machines

- https://drumap.nibiohn.go.jp/prediction

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: *J. Chem. Inf. Model.* 2019, 59, 3251−3261

pubs.acs.org/jcim

Computational Model To Predict the Fraction of Unbound Drug in the Brain

Tsuyoshi Esaki,[*,†] Rikiya Ohashi,[†,‡] Reiko Watanabe,[†] Yayoi Natsume-Kitatani,[†,§] Hitoshi Kawashima,[†] Chioko Nagao,[†,§] and Kenji Mizuguchi[*,†,§]

[†]Laboratory of Bioinformatics, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Osaka, Ibaraki 567-0085, Japan

[‡]Discovery Technology Laboratories, Mitsubishi Tanabe Pharma Corporation, 2-2-50 Kawagishi, Toda, Saitama 335-8505, Japan

[§]Laboratory of In-silico Drug Design, Center of Drug Design Research, National Institutes of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Osaka, Ibaraki 567-0085, Japan

Supporting Information

Method

- Dataset: ChEMBL (ver. 23), KEGG DRUG
- Descriptor: CDK Descriptor Calculator GUI (ver. 1.4.8), Mordred (ver. 1.0.0), and PaDEL-Descriptor
- ML Algorithms: Random Forest, Gradient Boosting, Support Vector Machine with radial functional kernel, linear kernel and partial least square
- Validation: Squared correlation coefficient (R^2) and the root-mean-square-error (RMSE)

https://pubs.acs.org/doi/pdf/10.1021/acs.jcim.9b00180

DruMAP ver.1.1

Drug Metabolism and pharmacokinetics Analysis Platform

This application may not work on Internet Explorer or Microsoft Edge properly. Please use Google Chrome or Firefox.

DruMAP consists of a database for DMPK parameters and a program that can predict some DMPK parameters based on the chemical structure of a novel compound. The DruMAP database contains data of DMPK parameters from curated public data and newly acquired experimental data obtained by unified conditions. It also contains predicted data using our prediction programs. Users can predict several DMPK parameters at once for novel compounds. Our platform provides a simple method to search and predict DMPK parameters and could contribute to the acceleration of new drug development. DruMAP has following three functions.

Compound search

Activity search

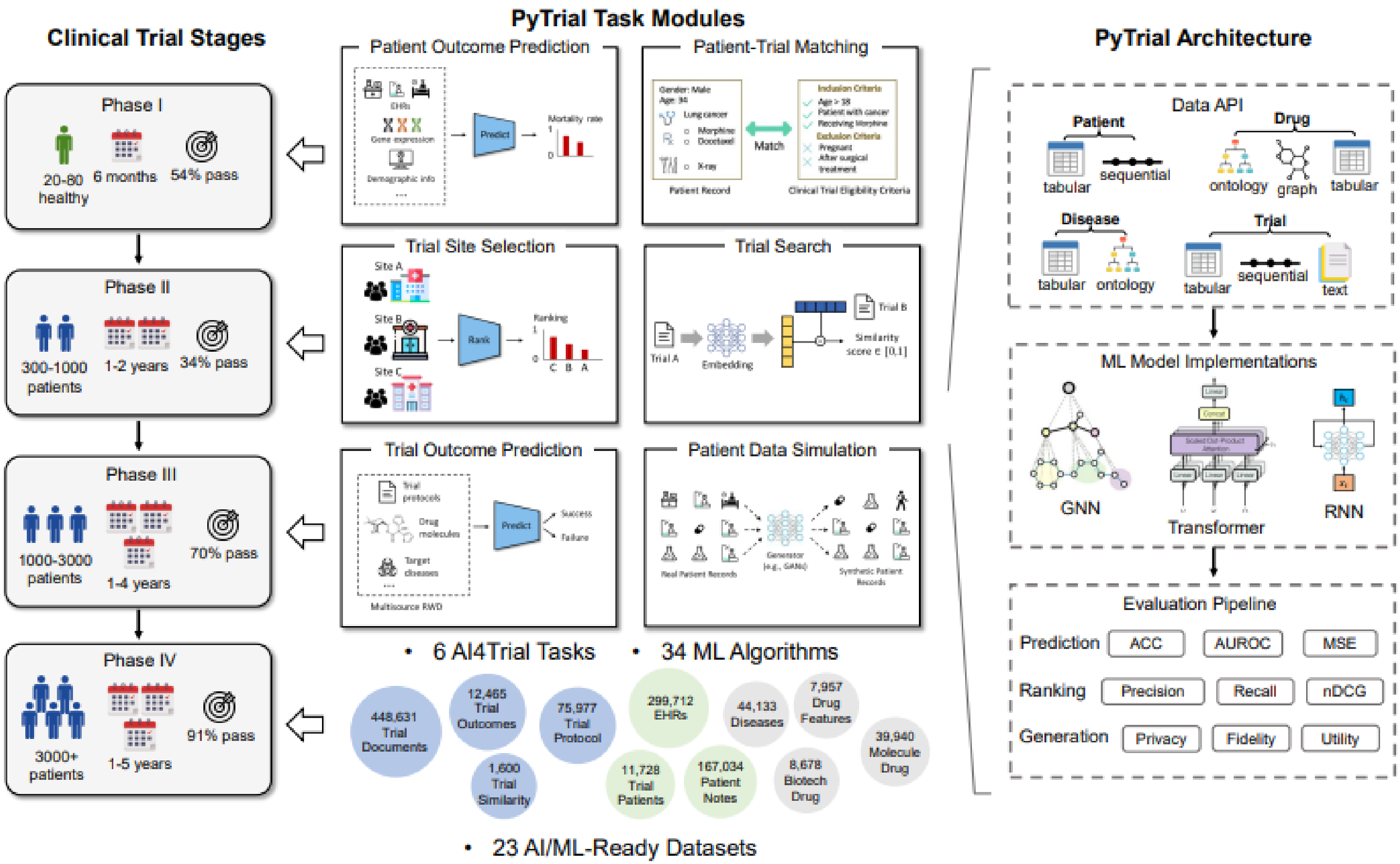
New prediction

AI in Clinical Trials

PyTrial

- Python package that implements various clinical trial tasks supported by AI algorithm.
- Implements 6 essential drug developmental tasks:
  - → Patient outcome prediction
  - → Trial site selection
  - → Trial outcome prediction
  - → Patient trial matching
  - → Trial similarity search
  - → Synthetic data generation

PyTrial Workflow



https://github.com/RyanWangZf/PyTrial

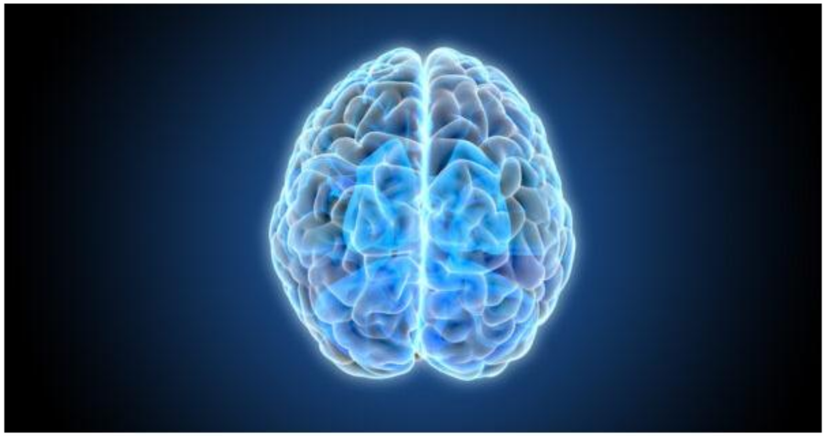
Limitations

ARTIFICIAL INTELLIGENCE, BIOPHARMA

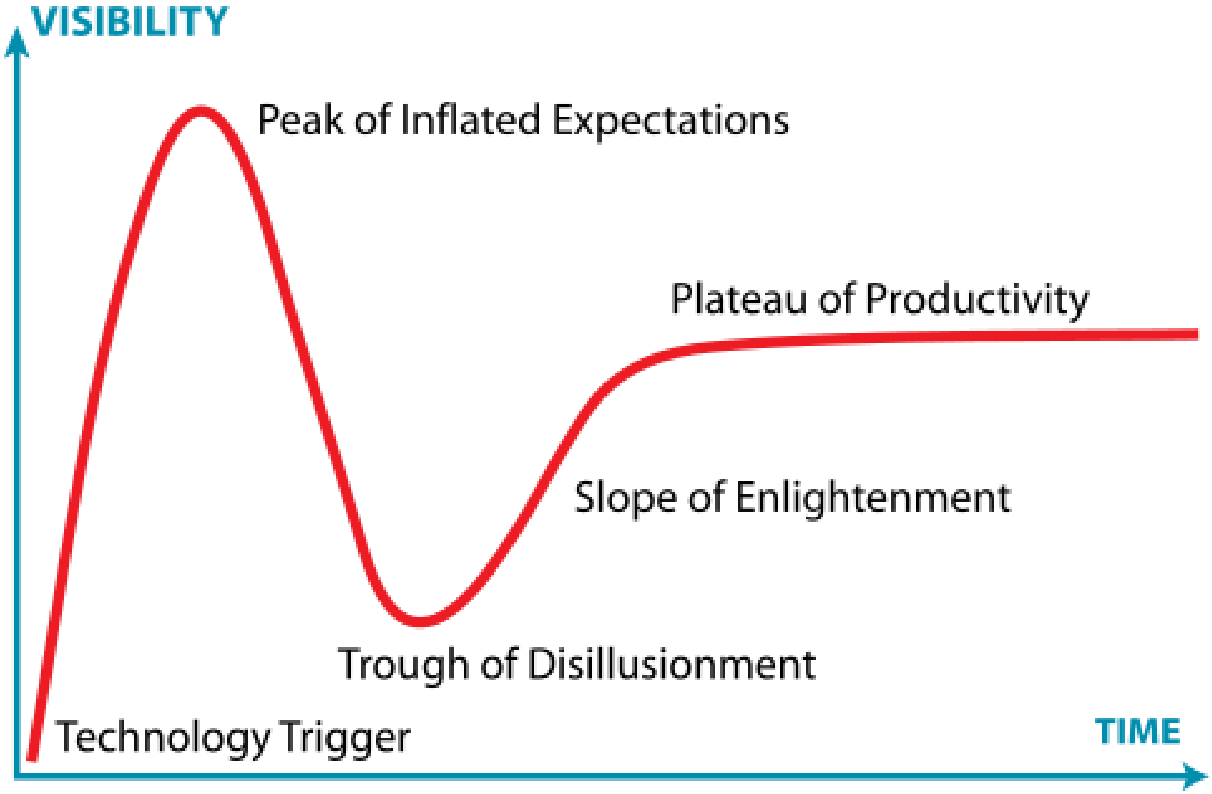
Sumitomo Schizophrenia Drug Discovered With AI Tech Fails in Two Phase 3 Trials

A schizophrenia drug candidate from partners Sumitomo Pharma and Otsuka Pharmaceutical failed to outperform a placebo in two Phase 3 studies. The compound is the most advanced program from a psychiatry and neurology collaboration spanning four drugs discovered by an artificial intelligence technology platform.

By Frank Vinluan on July 31, 2023 Share



AI hype cycle



Limitations

Data Quality and Availability:

- Limited Data: In drug discovery, obtaining comprehensive and high-quality data can be challenging, as experimental data is often limited or expensive to generate.
- Data Heterogeneity: Diverse types of data, including genomic, chemical, and clinical data, need to be integrated. However, these datasets often come in different formats and have varying levels of quality.

Complexity of Biological Systems:

- Biological Complexity: The biological systems involved in drug interactions are highly complex. AI models must account for the intricate relationships within biological pathways, which can be challenging to model accurately.
- Multifactorial Nature: Diseases often have multifactorial causes, making it difficult to identify specific targets for drug intervention.

Limitations

Interpretable Models:

- Black-Box Nature: Many AI models, especially deep learning models, are considered black boxes, meaning their decision-making processes are not easily interpretable. In drug discovery, interpretability is crucial for understanding the rationale behind predictions and gaining insights into potential drug targets.

Transferability of Models:

- Domain Transfer: AI models trained on data from one domain may not generalize well to another. Transferring knowledge between different therapeutic areas or diseases can be challenging.
- Data Drift: Changes in experimental protocols or data sources over time can lead to data drift, affecting the performance of AI models.

Limitations

Computational Resources:

- Computational Intensity: AI models, particularly deep learning models, can be computationally intensive and require significant resources. Access to high-performance computing infrastructure can be a bottleneck, especially for smaller research institutions.

Ethical and Regulatory Concerns:

- Ethical Use: Ensuring ethical use of AI in drug discovery, including issues related to data privacy, consent, and potential biases, is a critical challenge.
- Regulatory Approval: Developing AI-driven drug discovery methods that adhere to regulatory standards and obtaining approval from regulatory agencies pose challenges due to the novel and dynamic nature of these technologies.

Limitations

Collaboration and Integration:

- Interdisciplinary Collaboration: Successful AI applications in drug discovery require collaboration between experts in biology, chemistry, data science, and other fields. Bridging the gap between these disciplines can be a challenge.
- Integration with Experimental Work: Integrating AI predictions into traditional experimental workflows and decision-making processes is essential but can be challenging to implement seamlessly.

Cost and Investment:

- Resource Investment: Developing and maintaining AI-driven drug discovery pipelines requires significant financial investment. Smaller research institutions and startups may face challenges in securing the necessary resources.

Conclusion

- AI is highly promising in accelerating drug discovery by contributing to every step of the drug discovery process.
- AI can improve the clinical success rate of drugs by designing better clinical trials.
- AI can increase productivity and reduce costs by automating routine tasks.
- However, it faces challenges with data quality and model interpretability making it a tough road to travel.

Artificial Intelligence in Drug Discovery and Development

By Prof. Rajnish Kumar | IIT (BHU) Varanasi

Learners enrolled: 10610 | Exam registration: 3169

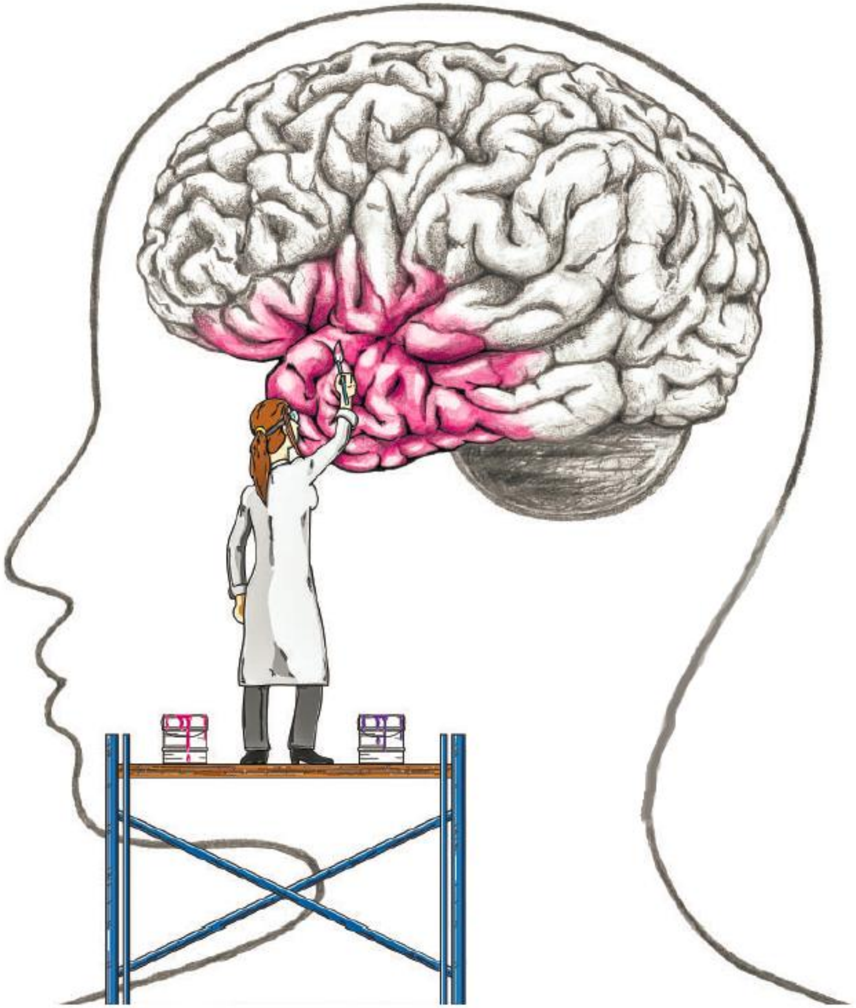
ABOUT THE COURSE:

This 12-week course, *Artificial Intelligence in Drug Discovery and Development*, is designed to equip

Summary

Course Status :	Ongoing
Course Type :	Elective
Language for course content :	English
Duration :	12 weeks
Category :	◦ Chemical Engineering ◦ Computational Biology
Credit Points :	3
Level :	Undergraduate/Postgraduate
Start Date :	21 Jul 2025
End Date :	10 Oct 2025
Enrollment Ends :	04 Aug 2025

http://tinyurl.com/aidd-biopredict



Thank you

Rajnish.phe@iitbhu.ac.in

C & E News, June 1, 2015 Cover, Volume 93, Issue 22