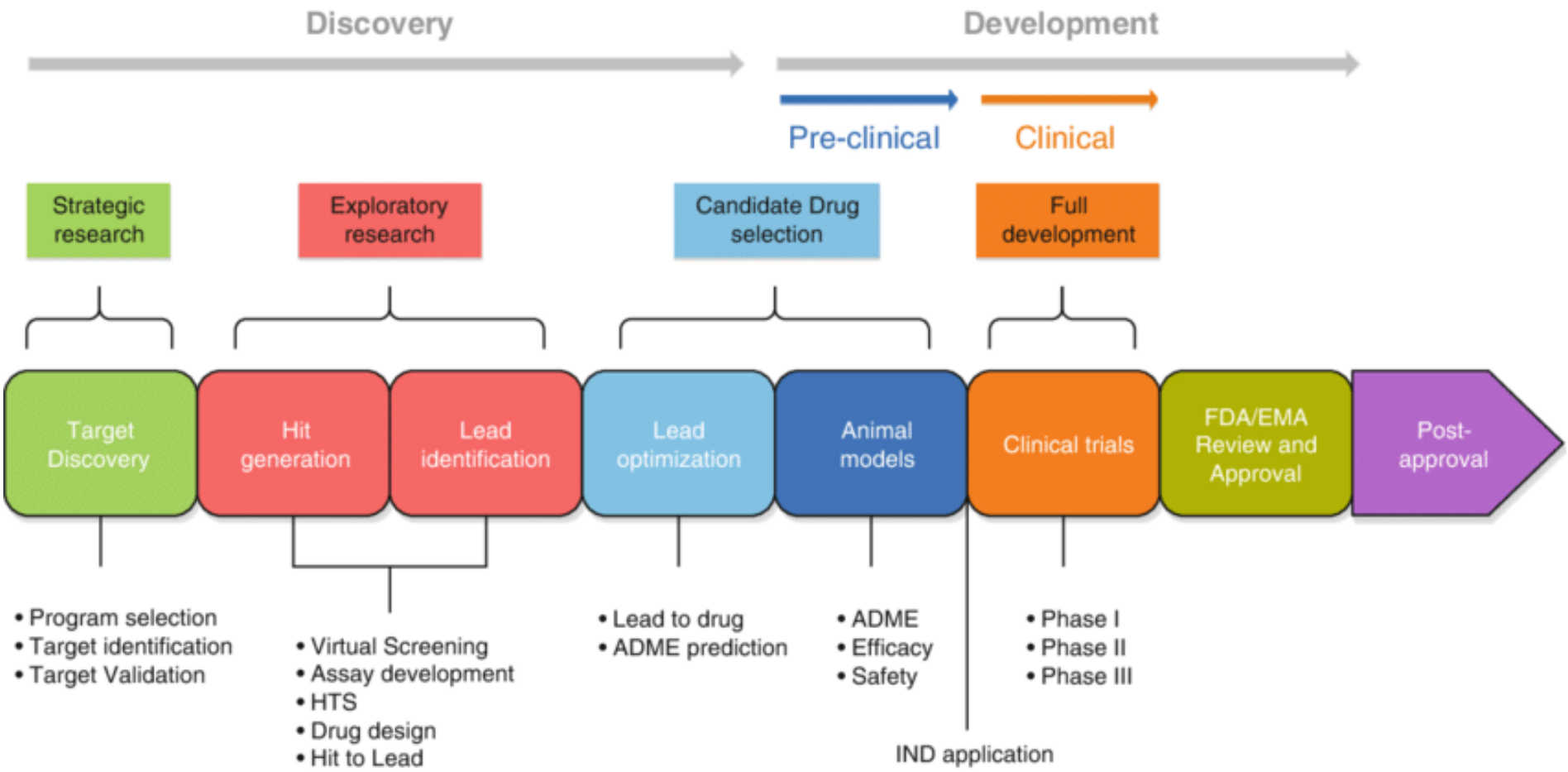
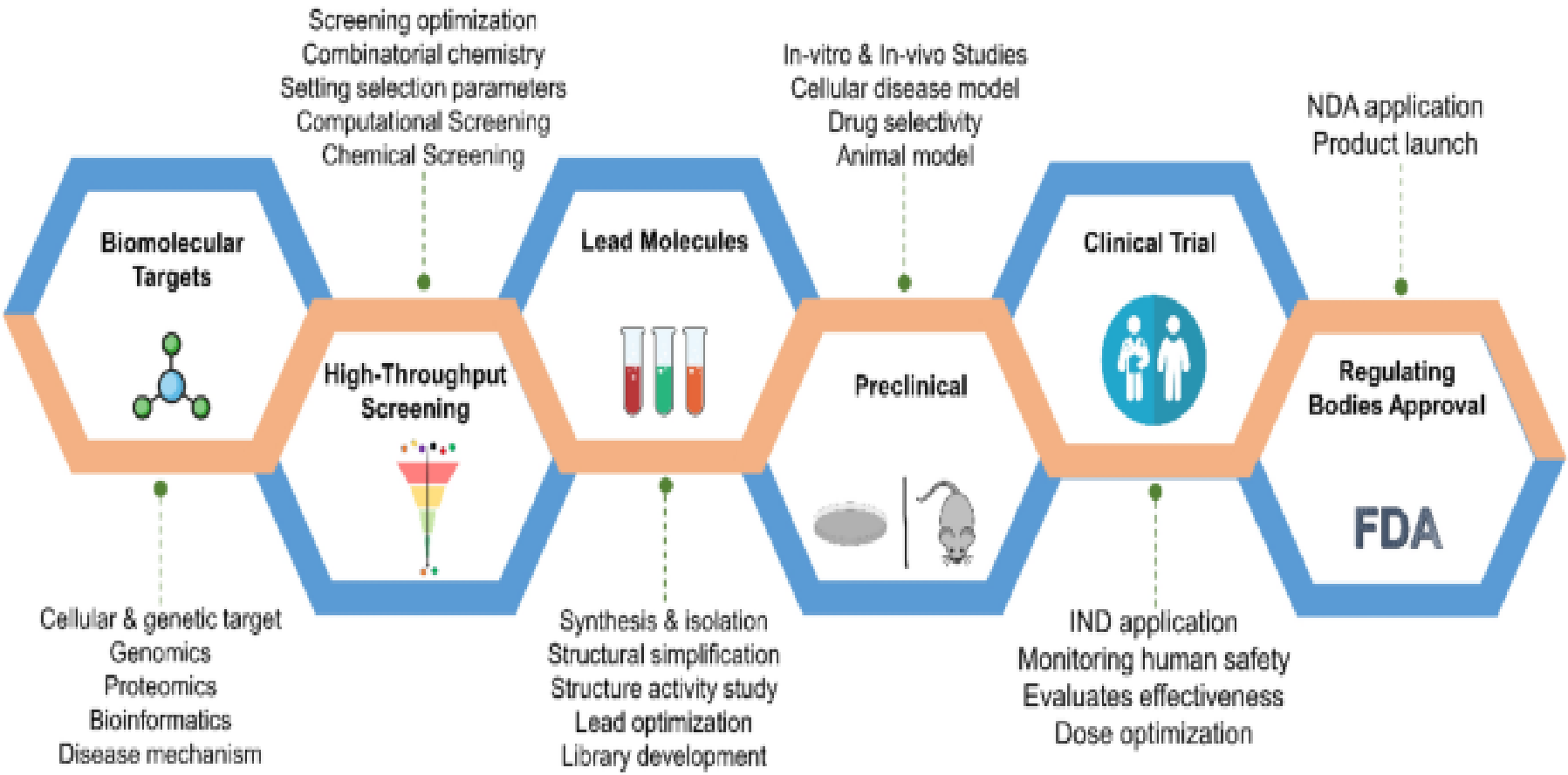
AI and ML in Drug Discovery

Drug Discovery and Development




Screening optimization
Combinatorial chemistry
Setting selection parameters
Computational Screening
Chemical Screening
In-vitro & In-vivo Studies
Cellular disease model
Drug selectivity
Animal model
NDA application
Product launch
Biomolecular Targets
High-Throughput Screening
Lead Molecules
Preclinical
Clinical Trial
Regulating Bodies Approval
FDA
Cellular & genetic target
Genomics
Proteomics
Bioinformatics
Disease mechanism
Synthesis & Isolation
Structural simplification
Structure activity study
Lead optimization
Library development
IND application
Monitoring human safety
Evaluates effectiveness
Dose optimization

Stage 1 — **"Pre-discovery"**

Stages 2 and 3 — **Drug discovery & Preclinical studies**

Stage 4 — **Clinical trials**

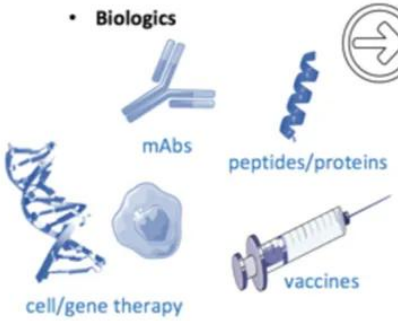
Stage 5 — **Review & Approval** — **Phase IV**

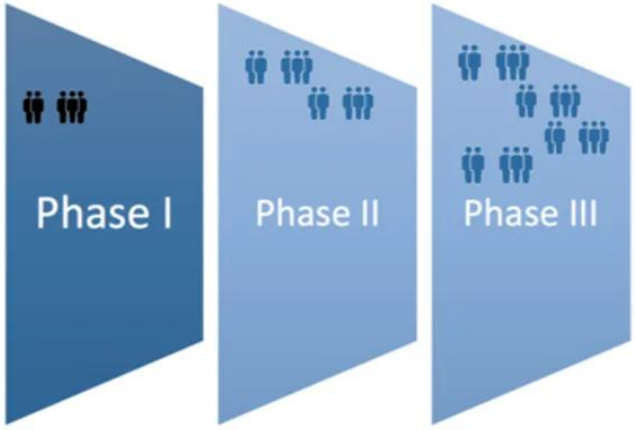
~12-15 years

Basic research
Clinical observations

- **Small molecules**

- **Biologics**





Molecular mechanisms involved in the disease state, identification of putative targets...

Identification of the therapeutic agents modulating the selected target(s), ADMET, efficacy,... first evaluations silico-vitro-vivo

Evaluations in humans PK, dose, efficacy, toxicity...

Reviewing, approval & post-market monitoring

CASE STUDY-1

Cancer Informatics
Volume 24: 1–20

DOI: 10.1177/11769351251371273
journals.sagepub.com/home/cix
Sage

The Impact of Artificial Intelligence on Cancer Diagnosis and Treatment: A Review

Niki Najar Najafi[1], Helia Hajihassani[1] and Maryam Azimzadeh Irani[1]

Abstract

The complexity of cancer has long challenged the medical community, driving the need for improved early detection and treatment. Artificial intelligence (AI) has profoundly impacted oncology research in recent decades, resulting in innovative diagnostic and therapeutic approaches. This review synthesizes the critical applications of AI in oncology, focusing on 4 key areas: medical imaging, digital pathology, robotic surgery, and drug discovery. We highlight the role

The Impact of Artificial Intelligence on Cancer Diagnosis and Treatment: A Review

simplifying the pathologists' tasks. AI-powered robotic surgery provides different levels of automation, leading to precise and minimally invasive procedures that not only improve surgical outcomes but also lower readmission rates, hospital stays, and infection risks. Moreover, AI expedites the process of discovering cancer therapies by identifying potential lead compounds, predicting drug reactions, and repurposing current medications. In the past decade, several AI-developed drugs have successfully entered clinical trials. These significant advancements underscore the expanding role of AI in shaping the future of cancer diagnosis and treatment. Although standardization, transparency, and equitable implementation must be addressed, AI brings hope for more personalized and effective therapies.

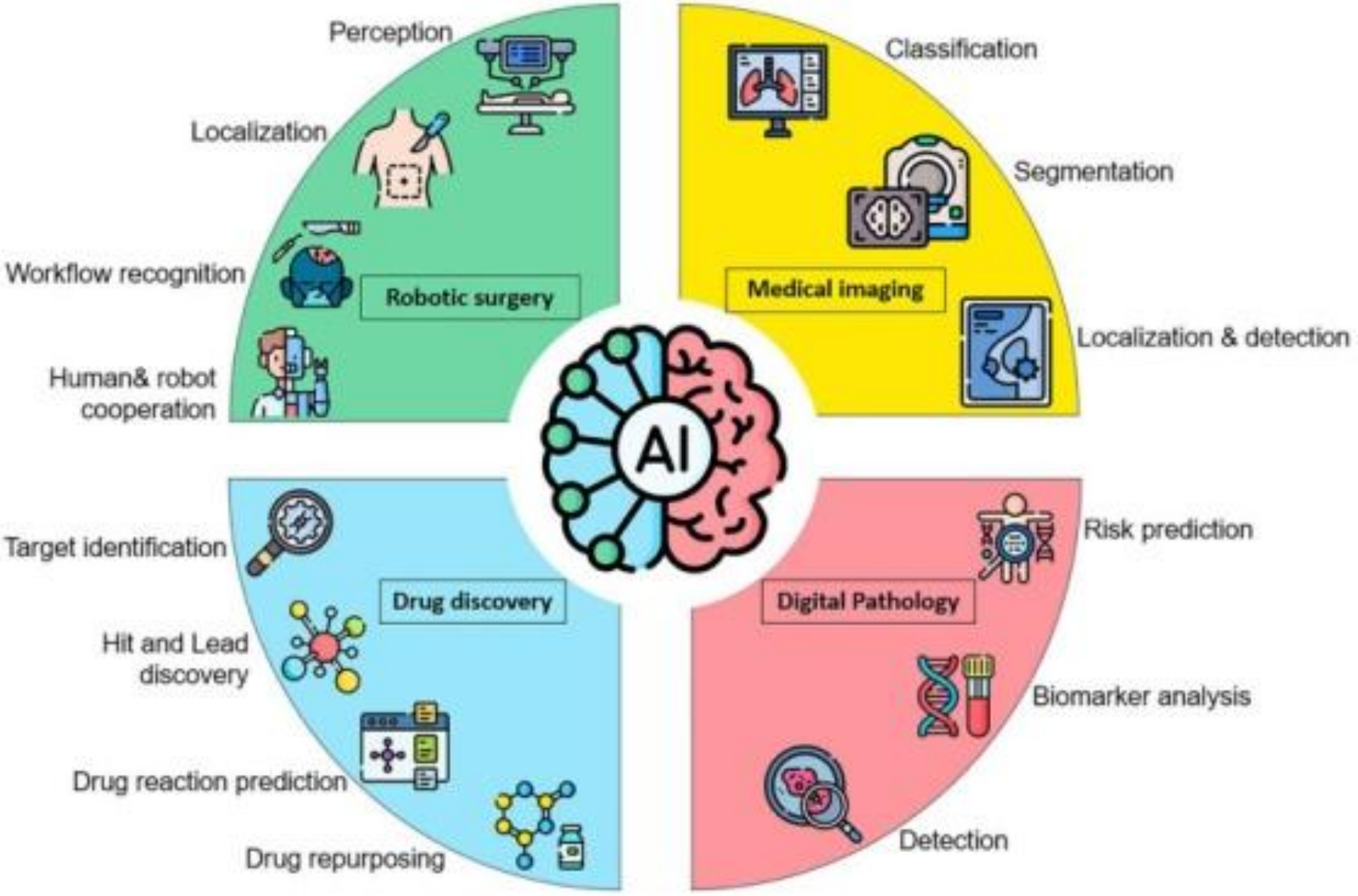
Keywords

artificial intelligence, cancer, diagnosis, treatment, precision medicine

Received: 14 February 2025; accepted: 5 August 2025

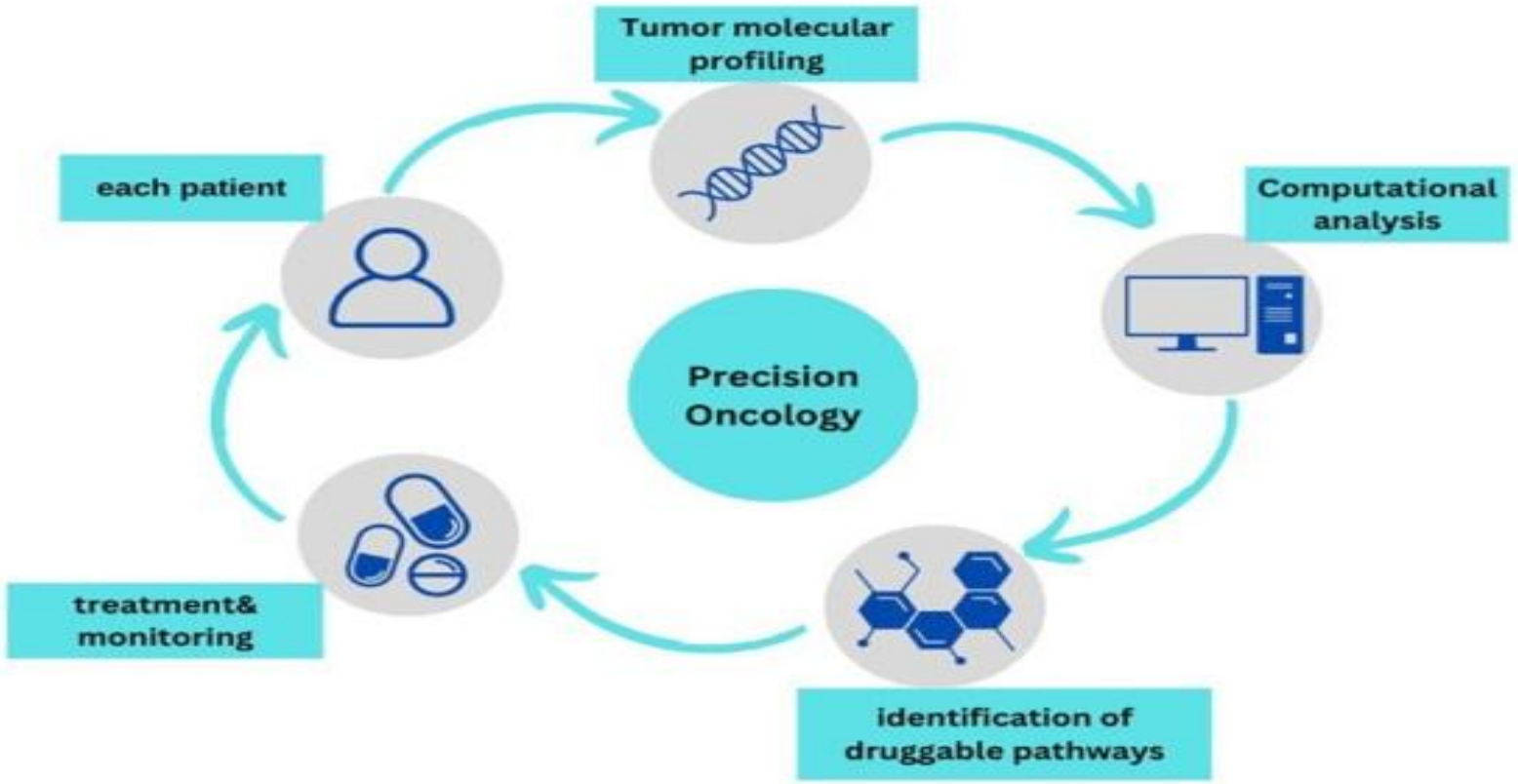
Introduction

- The complexity of cancer has long challenged the medical community, driving the need for improved early detection and treatment.
- Artificial intelligence (AI) has profoundly impacted oncology research in recent decades, resulting in innovative diagnostic and therapeutic approaches.
- The critical applications of AI in oncology, focusing on 4 key areas: medical imaging, digital pathology, robotic surgery, and drug discovery.



The role of AI in cancer care, focusing on its applications in medical imaging, digital pathology, robotic surgery, and drug discovery.

AI in the Field of Medicine



The precision oncology process. Where AI plays a crucial role in each stage. Firstly, AI will analyze patient data, including molecular tests and medical images. Secondly, AI algorithms scrutinize vast amounts of molecular data to identify genetic mutations and tumor biomarkers. Thirdly, AI stores and uncovers new insights about personalized data. Fourthly, AI algorithms create predictive calculators to determine disease risk and recommend the most effective treatment targets and practices. Finally, AI can access medical records to aid in monitoring patients.

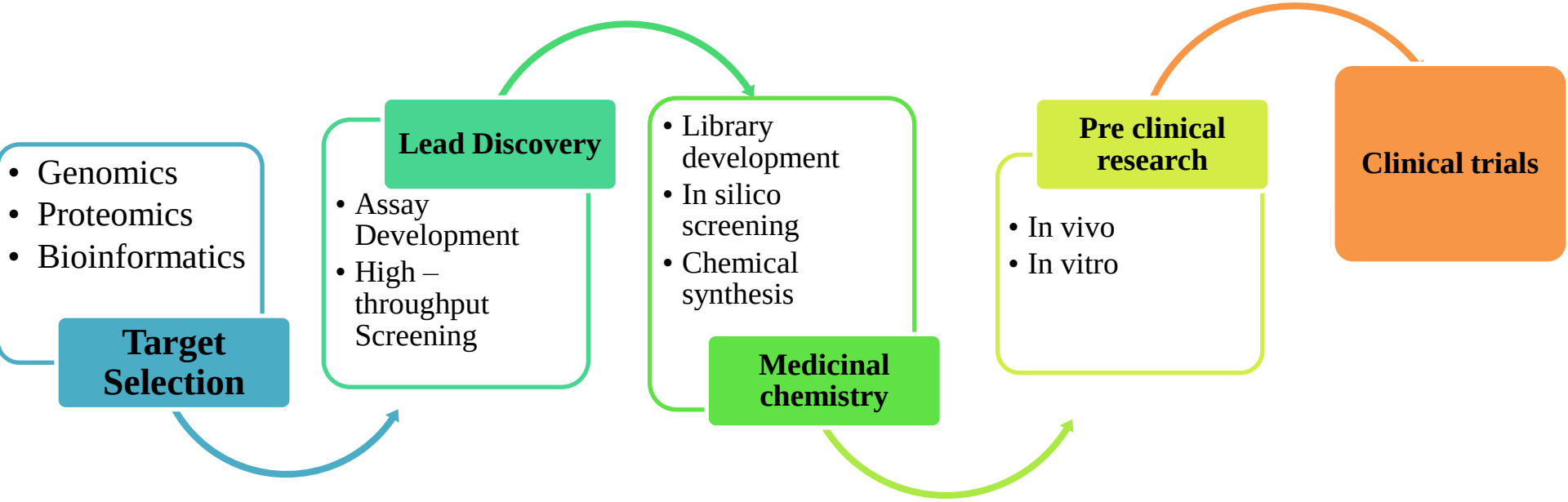
Early Detection and Diagnosis

- Breast Cancer Risk Assessment Tool - **The Gail model**
- The Gail model is widely used for estimating a woman's risk of developing breast cancer.
- The Gail model can provide a more accurate assessment of a woman's risk of developing breast cancer over the next 5 years as well as throughout her lifetime, up to the age of 90.

AI Models in Medical Imaging From 2022 to 2024

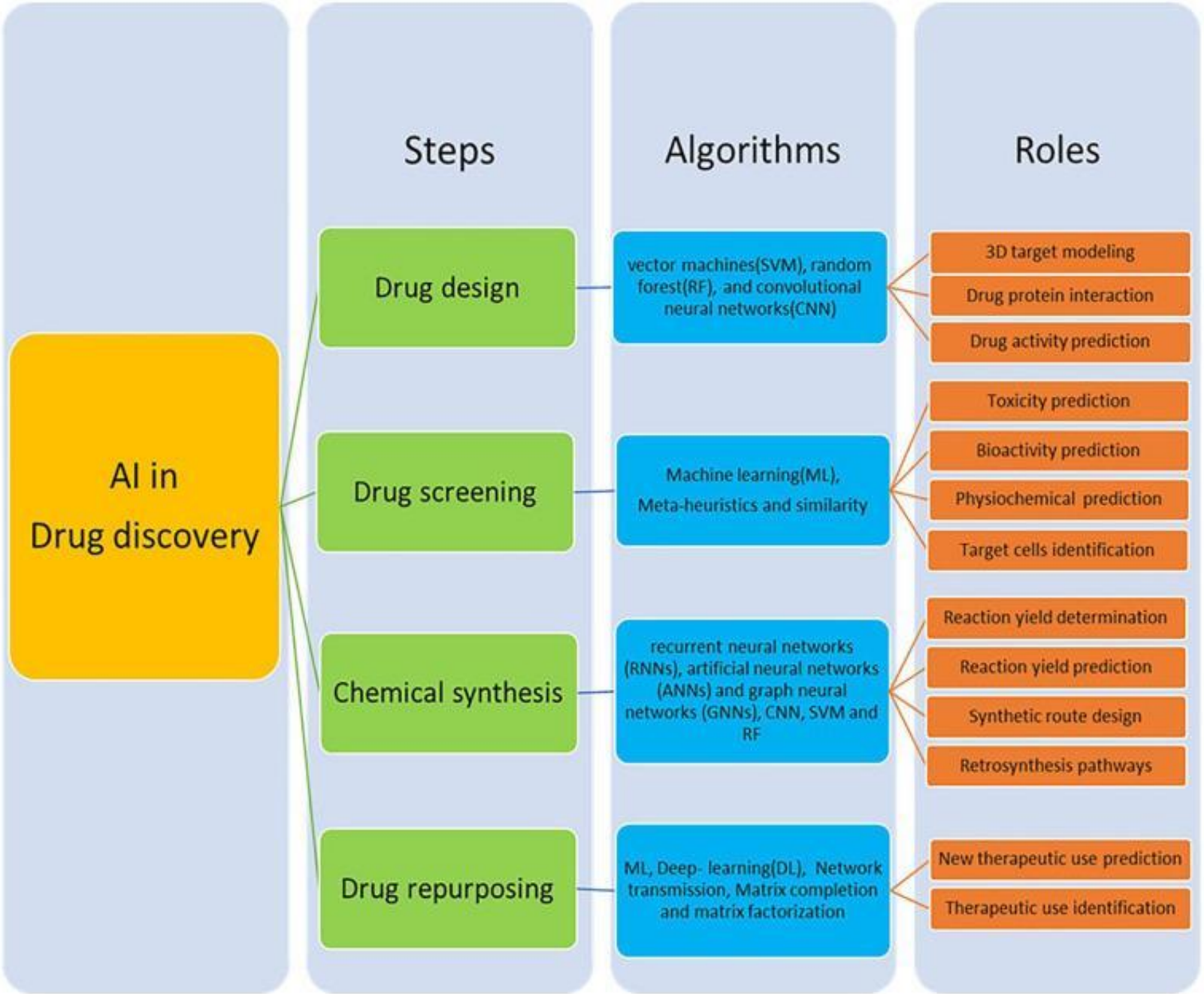
Imaging technique	Method	Description	Clinical status	Year
MRI	Deep Learning (DL)	Segmentation and classification of 18 distinct intracranial tumor types using T1- and T2-weighted images, including T2 contrast MRI sequences.	Research only; multi-center validation.	2022
	Convolutional Neural Network (CNN)	The prediction of tumor grading is combined with segmentation techniques that utilize whole 3D structural preoperative MRI data.	Research only; single-center validation.	2023
	Deep Learning (DL)	Real-time prostate contouring on MRI images from a 1.5 T MR-Linac system to guide radiation beam delivery.	Research only; single-center validation.	2024
CT scan	Deep Learning (DL)	"Sybil" estimates future lung cancer risk from a single low-dose CT (LDCT) scan.	FDA Accelerated Approval Program; investigational use only.	2023
	Convolutional Neural Network (CNN)	Detection of pancreatic ductal adenocarcinoma (PDA) on diagnostic CT scans and identification of visually occult PDA on pre-diagnostic CT scans.	Research only; multi-center validation.	2023

AI in Anti-Cancer Drug Discovery



The typical steps of drug discovery.

The role of AI in different steps of drug discovery



Conclusion

- Over the past decade, AI has profoundly transformed the field of oncology, enhancing early detection, precision diagnostics, robotic-assisted surgery, and expediting drug discovery.
- The AI systems have outperformed traditional methods in critical areas such as tumor classification and biomarker analysis.
- However, the integration of AI into clinical practice is not without challenges. Problems such as dataset biases, insufficient prospective validation, regulatory obstacles, and cost barriers require immediate attention and resolution.

CASE STUDY-2

Research Workx

International Journal of Digital Innovation
https://researchworkx.com/index.php/ijdi
Vol 1 (1) 2020

Accelerating Drug Discovery with Machine Learning: Case Studies and Implications for Future Pharmaceuticals

Kushwanth Gondi[1], Veera Venkata Raghunath Indugu[2], Dedeepya Sai Gondi[3], Vamsi Krishna Reddy Bandaru[4], Hemanth Volikatla[5]

[1]: Software Developer, Computer Science and Technology Company, USA, kushlu. sai @ gmail.com

[2]: Engineer 1, Data Science and Cloud Technologies Company, USA, veerajet1 @ gmail.com

[3]: CTO/Director Department, Artificial Intelligence and Machine Learning Company, USA gmail.com

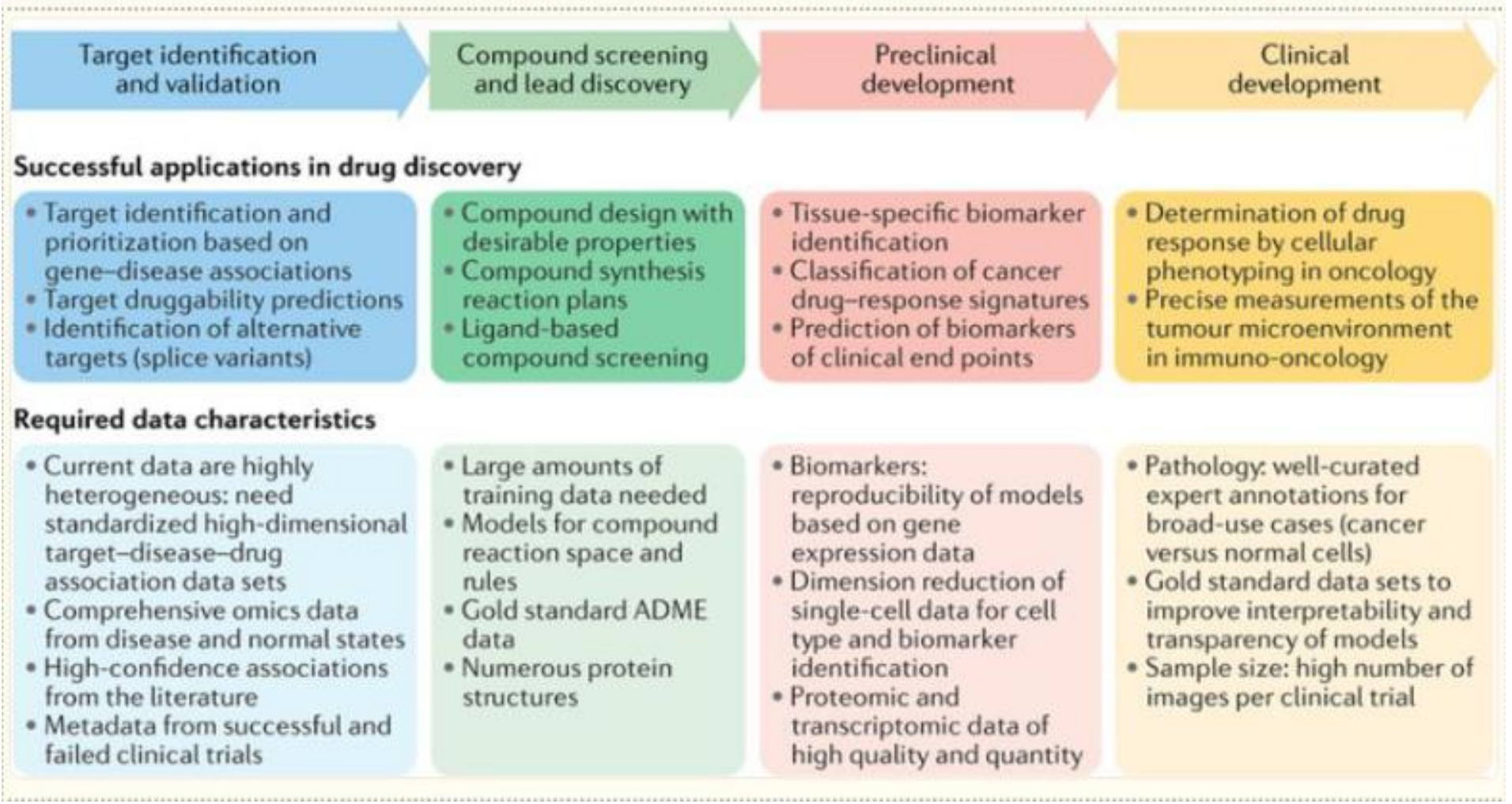
[5]: Independent Researcher, hemanthvolikatla @ gmail.com

Accelerating Drug Discovery with Machine Learning: Case Studies and Implications for Future Pharmaceuticals

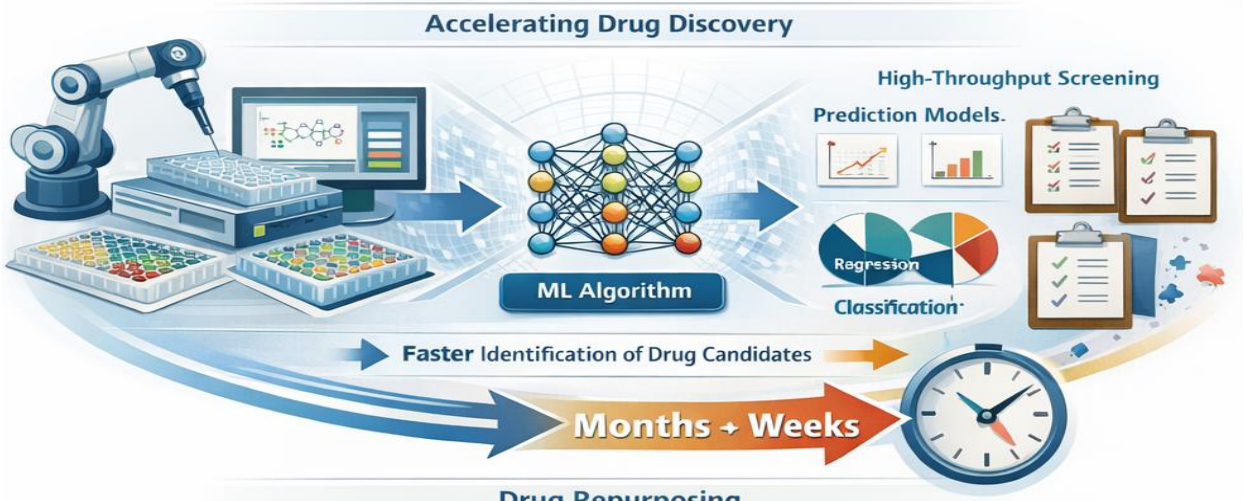
Abstract:

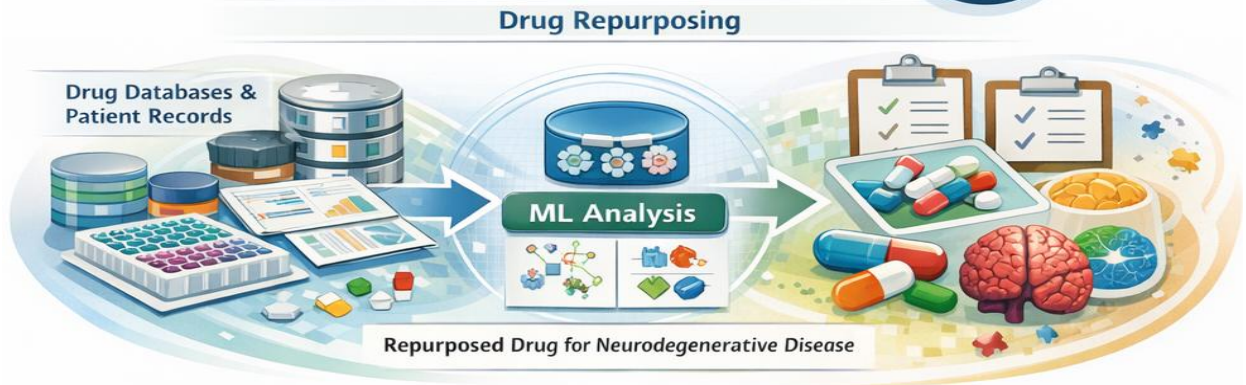
Machine learning (ML) is revolutionizing drug discovery, offering new avenues for identifying potential treatments more efficiently. This paper delves into the application of ML algorithms in drug discovery, with a focus on their role in developing treatments for COVID-19. We present case studies of successful ML-driven discoveries, discuss the benefits and limitations of these technologies, and explore their broader implications for the pharmaceutical industry. The paper

Machine learning applications in the drug discovery pipeline and their required data characteristics



Case Studies: Early-Stage Drug Discovery





Implications for Future Pharmaceuticals

The Role of Machine Learning in Drug Discovery

Accelerating Drug Discovery

- Predictive Modeling
- Rapid Screening of Compounds

Enhancing Safety & Efficacy

- Adverse Reaction Prediction
- Targeted Therapies

Data Challenges

- Data Quality & Availability
- Bias & Incomplete Data

Ethical & Regulatory Issues

- Privacy & Security
- Regulatory Compliance

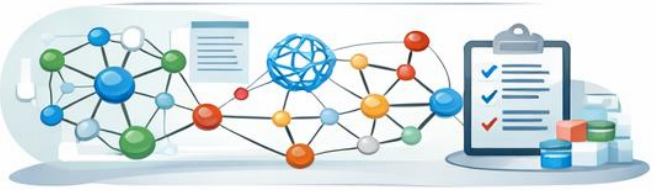
Model Transparency

- Explainable AI (XAI)
- Interpretability of ML Models

Integrating ML with Traditional Methods

- Combining Old & New Approaches
- Workflow Adaptation

Future Solutions



- Advancing Data Curation
- Improving Reliability & Trust

Conclusion

- Machine learning is transforming drug discovery by making the process faster, more efficient, and cost-effective.
- It improves drug safety, efficacy, and supports the development of personalized treatments.
- However, challenges such as data quality, model interpretability, and integration with traditional methods remain.
- As these challenges are addressed, ML will play a crucial role in advancing pharmaceutical innovation and improving patient outcomes.

CASE STUDY-3

Artificial Intelligence Review (2023) 56:5975–6037
https://doi.org/10.1007/s10462-022-10306-1

Deep learning in drug discovery: an integrative review and future challenges

Heba Askr[1] · Enas Elgeldawi[2] · Heba Aboul Ella[4] · Yaseen A. M. M. Elshaier[5] · Mamdouh M. Gomaa[2] · Aboul Ella Hassanien[3]

Accepted: 24 October 2022 / Published online: 17 November 2022

Abstract
Recently, using artificial intelligence (AI) in drug discovery has received much attention since it significantly shortens the time and cost of developing new drugs. Deep learning (DL)-based approaches are increasingly being used in all stages of drug development as DL technology advances, and drug-related data grows. Therefore, this paper presents a sys-

This study to summarize (2000–2022) deep learning applications, datasets, methods, and future directions in AI-based drug discovery.

sets, the databases, and the evaluation measures are also presented. In addition, this paper provides an overview of how explainable AI (XAI) supports drug discovery problems. The drug dosing optimization and success stories are discussed as well. Finally, digital twining (DT) and open issues are suggested as future research challenges for drug discovery problems. Challenges to be addressed, future research directions are identified, and an extensive bibliography is also included.

Keywords Drug discovery · Artificial intelligence · Deep learning · Drug–target interactions · Drug–drug similarity · Drug side-effects · Drug sensitivity and response · Drug dosing optimization · Explainable artificial intelligence · Digital twining

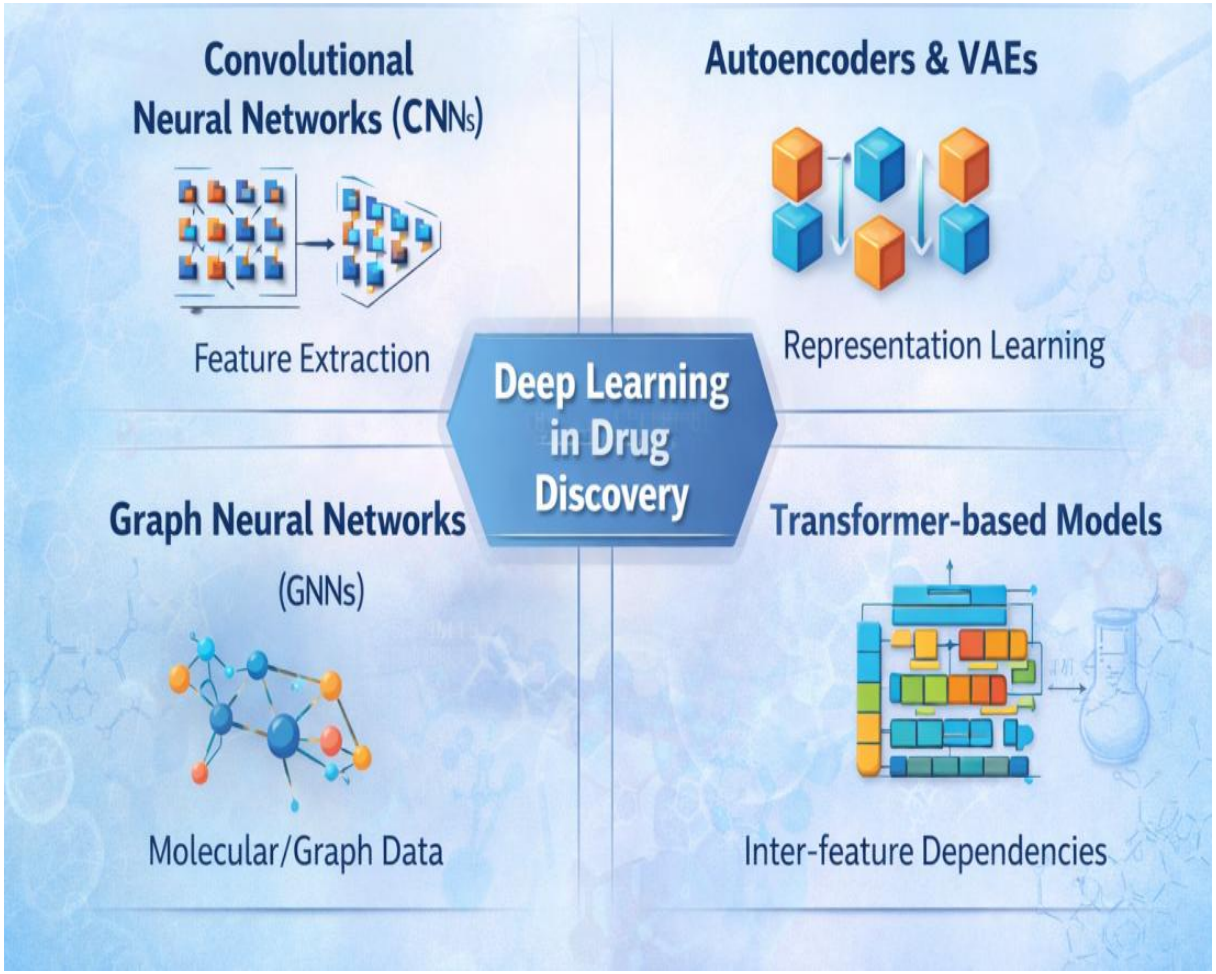
Background

- Drug discovery is **costly and slow** (millions of dollars + years).
- Deep learning (DL) accelerates discovery by learning from large bio-data.
- DL used in all discovery stages: from target interaction to side effects.

Deep Learning Techniques

DL Algorithms used in drug discovery

- ✓ **CNNs** (Convolutional Neural Networks) – feature extraction.
- ✓ **Autoencoders & VAEs** – representation learning.
- ✓ **Graph Neural Networks** (GNNs) – molecular/graph data.
- ✓ **Attention & transformer-based models** – inter-feature dependencies.



AI Algorithms in Drug Discovery

1 Artificial Neural Networks (ANN)

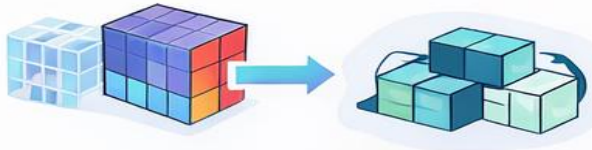
Basic deep learning model.

- Activity prediction
- Toxicity prediction

TensorFlow Keras

2 Convolutional Neural Networks (CNN)

Good for image or spatial data



- Protein–ligand interaction prediction
- Molecular structure analysis

PyTorch DeepChem

3 Recurrent Neural Networks (RNN)

Used for sequence data

- Protein sequences
- Drug molecule sequences (SMILES)

TensorFlow DeepChem

4 Graph Neural Networks (GNN)

Very important in drug discovery

Treats molecules as graphs:

- Atoms = nodes
- Bonds = edges

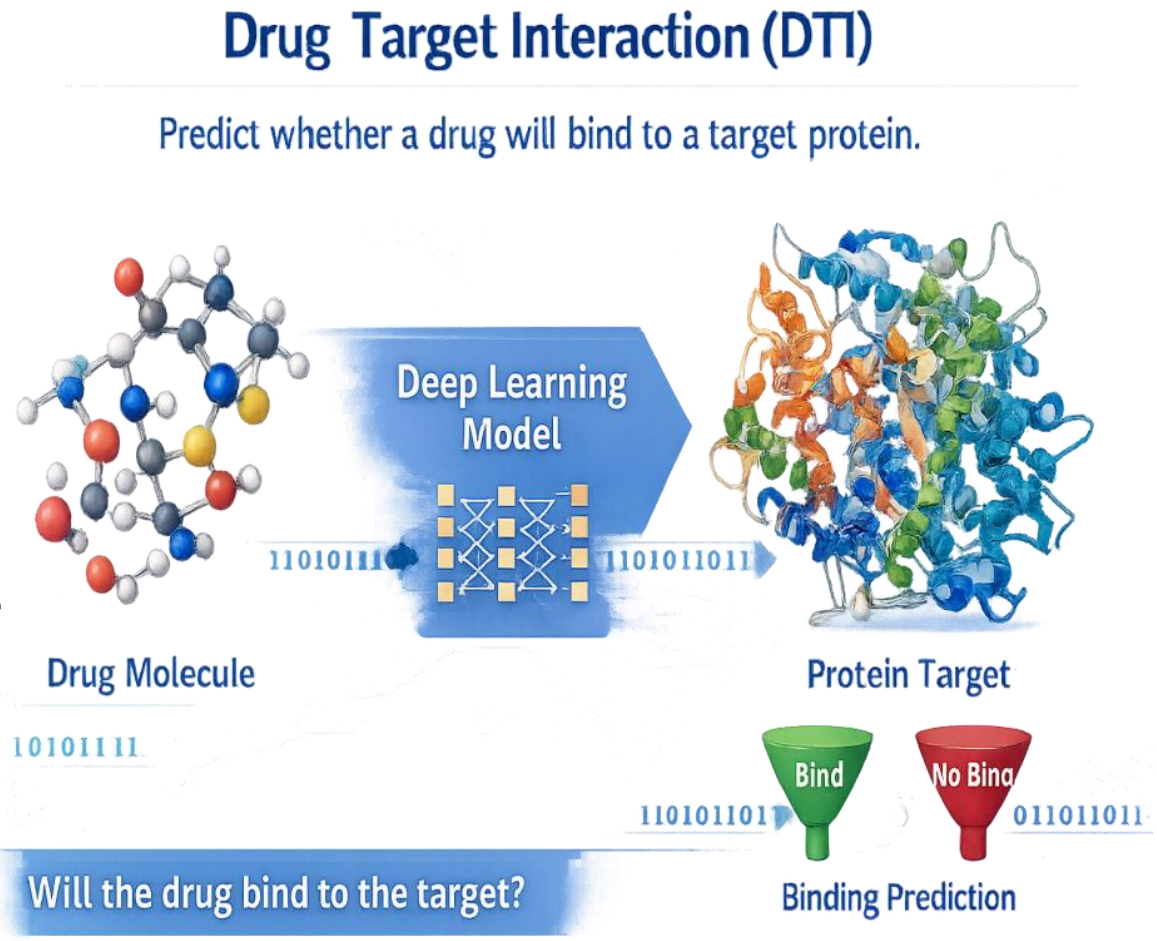
DeepChem DGL

Drug-Target Interaction

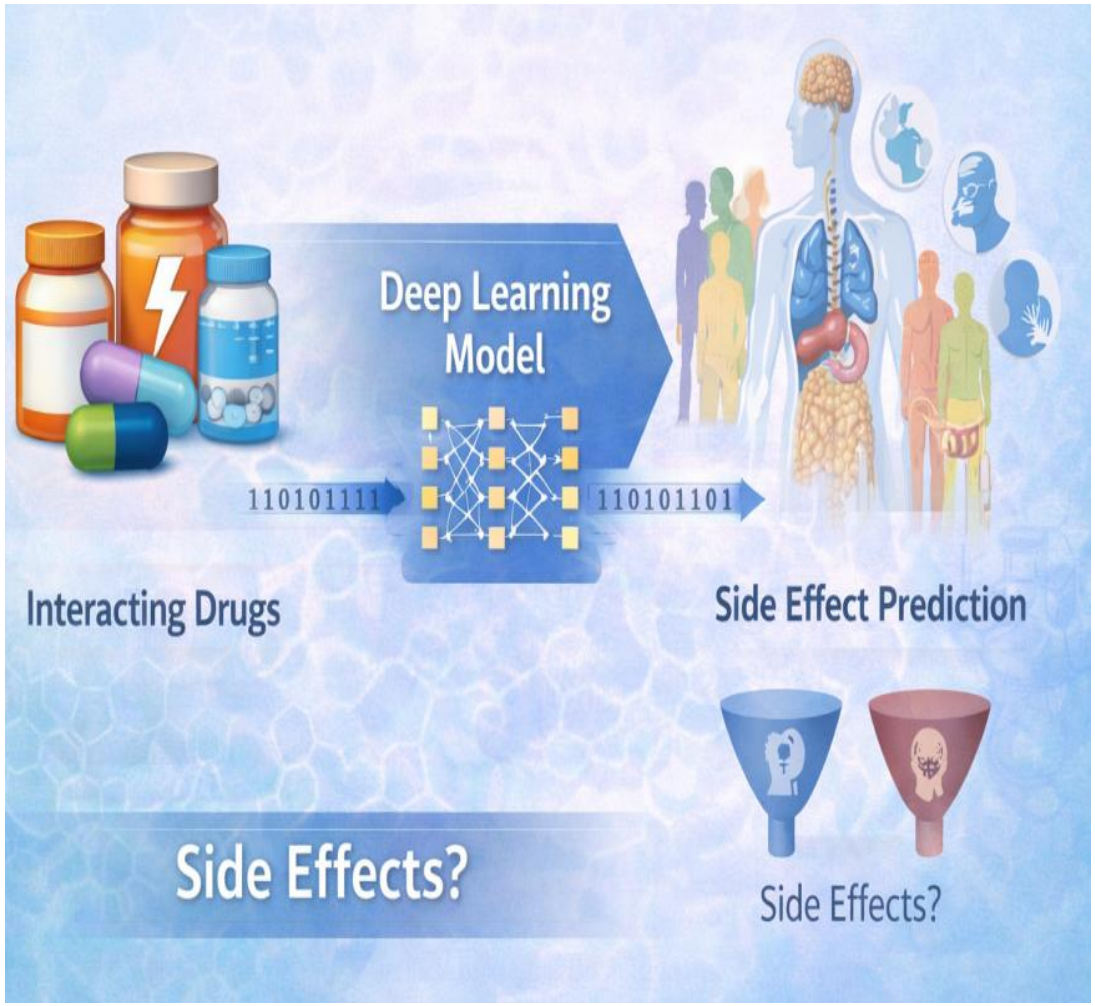
Goal: Predict whether a drug binds to a target protein. DL models don't require 3D structures

Examples:

- CNN/RNN based for sequence embeddings.
- Attention models express interrelations.
- Graph methods for molecular properties.



Drug-Drug Interaction (DDI) Side Effect Prediction



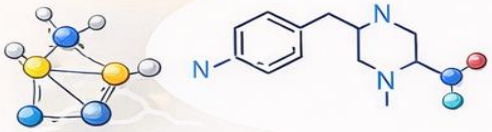
- ✓ **DDI:** how one drug affects another's effect or causes adverse effects.
- ✓ DL frameworks utilize structural, graph, and similarity features.
- ✓ Technologies: GNNs, CNNs, BiLSTM, transformer-based for capturing complex patterns.
- ✓ Performance often measured by AUC, precision & F1.

How **Deep Learning Models Predict DDIs**

DL models use different types of drug information:

1 Structural Features

Chemical structure of the drug



2 Graph Features

Drug molecules are treated as **graphs:**

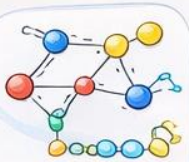
- ☑ Atoms = nodes
- ☑ Bonds = edges

DL Models

3 Similarity Features

Compares drugs based on:

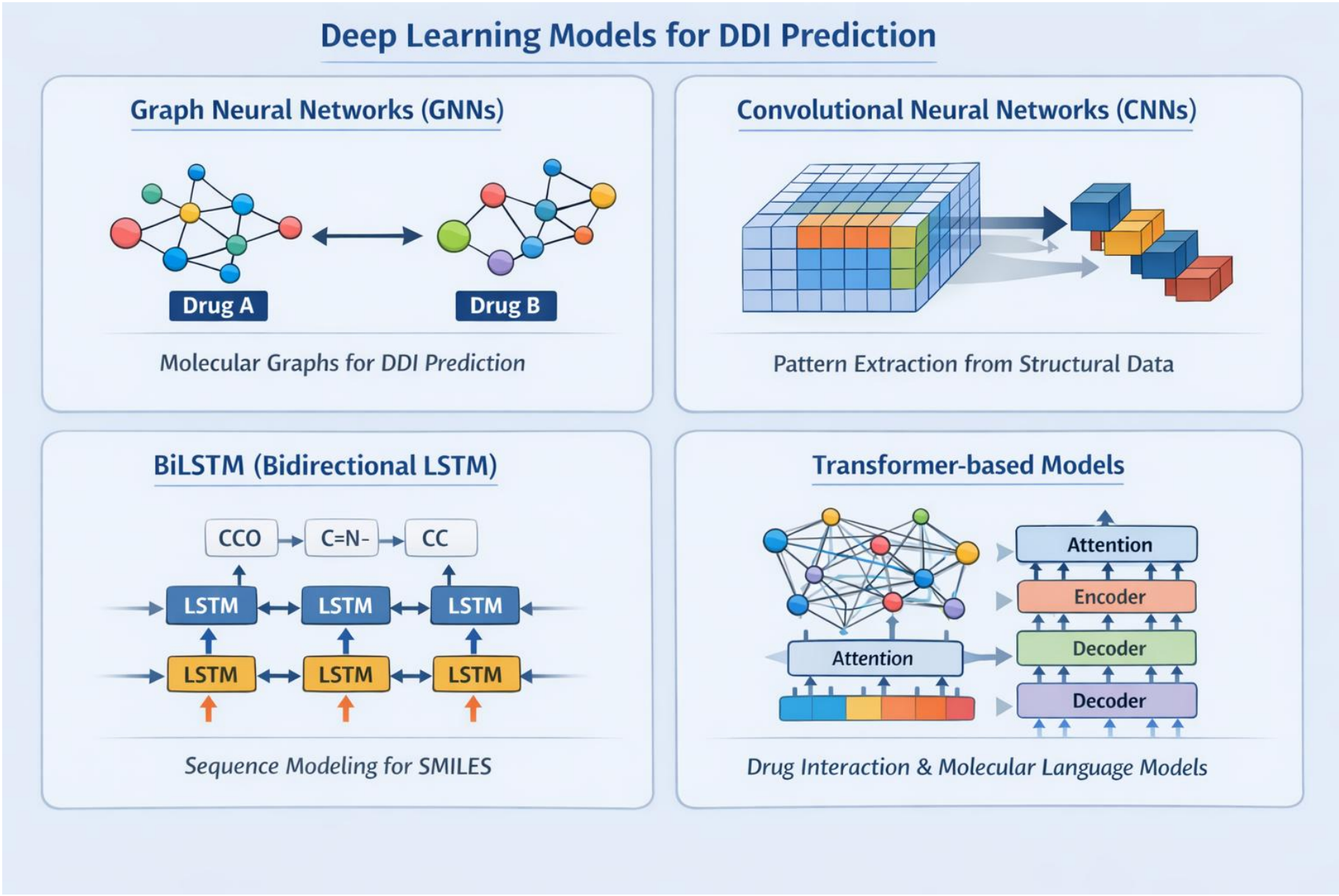
- ☑ Chemical similarity
- ☑ Target proteins
- ☑ Biological pathways



DDI Prediction

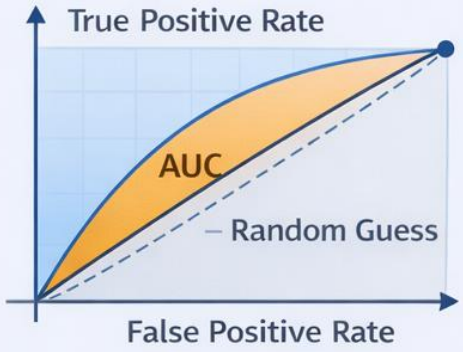
AI technologies used

These are deep learning models used for DDI prediction:



How Model Performance is Measured

AUC (Area Under the Curve)



- Measures overall prediction accuracy
- Higher AUC = better model

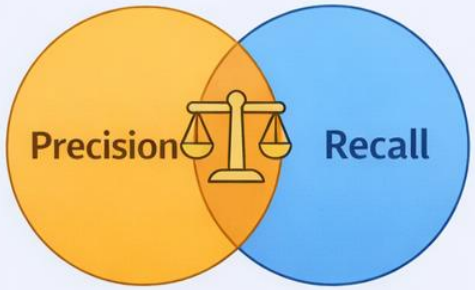
Higher AUC = better model

Precision

- Measures how many predicted interactions are correct

$$\text{Precision} = \frac{TP + FP}{TP + FP}$$

F1-score

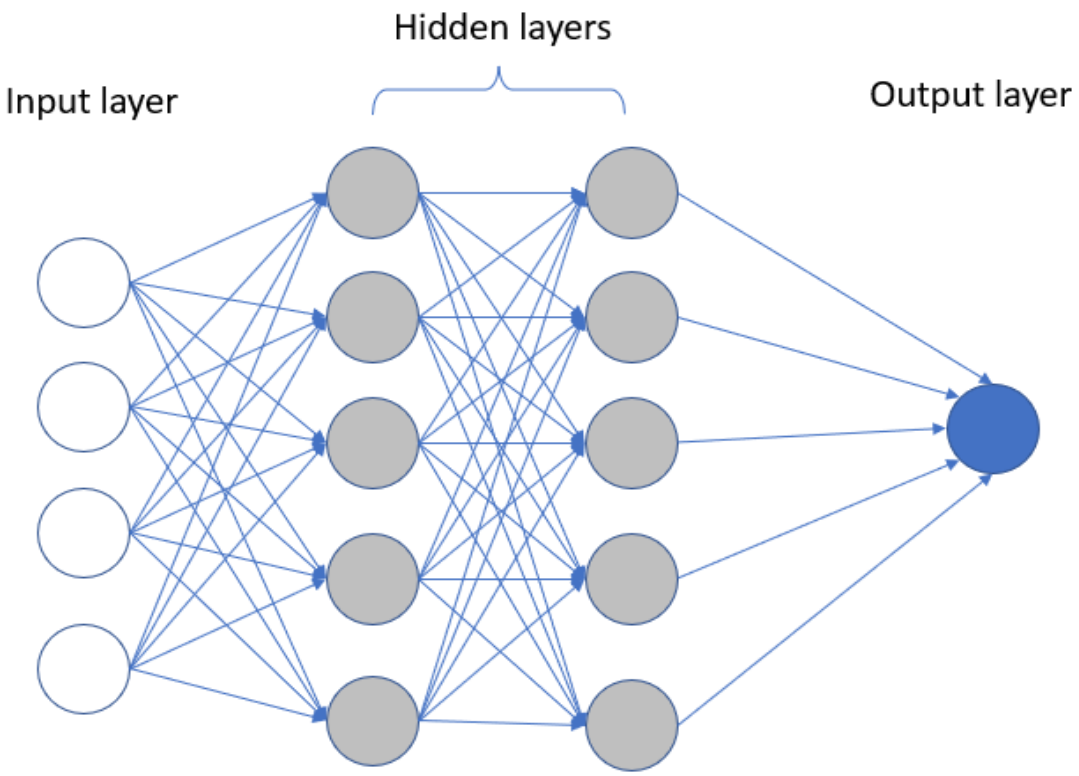


- Combines precision and recall
- Gives overall performance balance

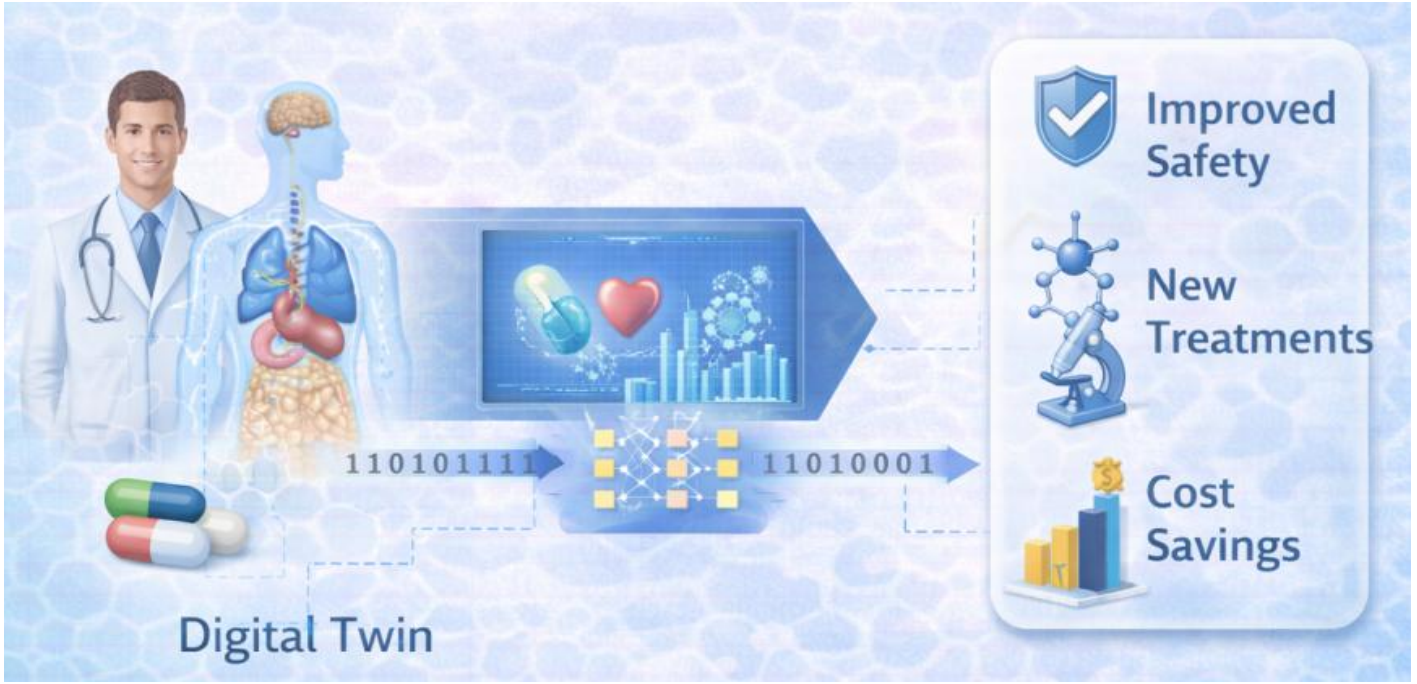
$$\text{F1-score} = 2 \times \frac{(\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Explainable AI (XAI) in Drug Discovery

- ✓ XAI **interprets black-box DL models** for trust & transparency.
- ✓ Important for **scientific insights** and regulatory approval.
- ✓ Current approaches still limited more work needed to produce meaningful explanations.



Digital Twinning & Success Stories



- **Digital Twin (DT):** virtual simulation of drugs, patients, workflows.
- **Benefits:** simulate patient response & reduce real trials.
- **Success Stories:** notable DL-assisted discoveries accelerate drug repurposing and novel targets.

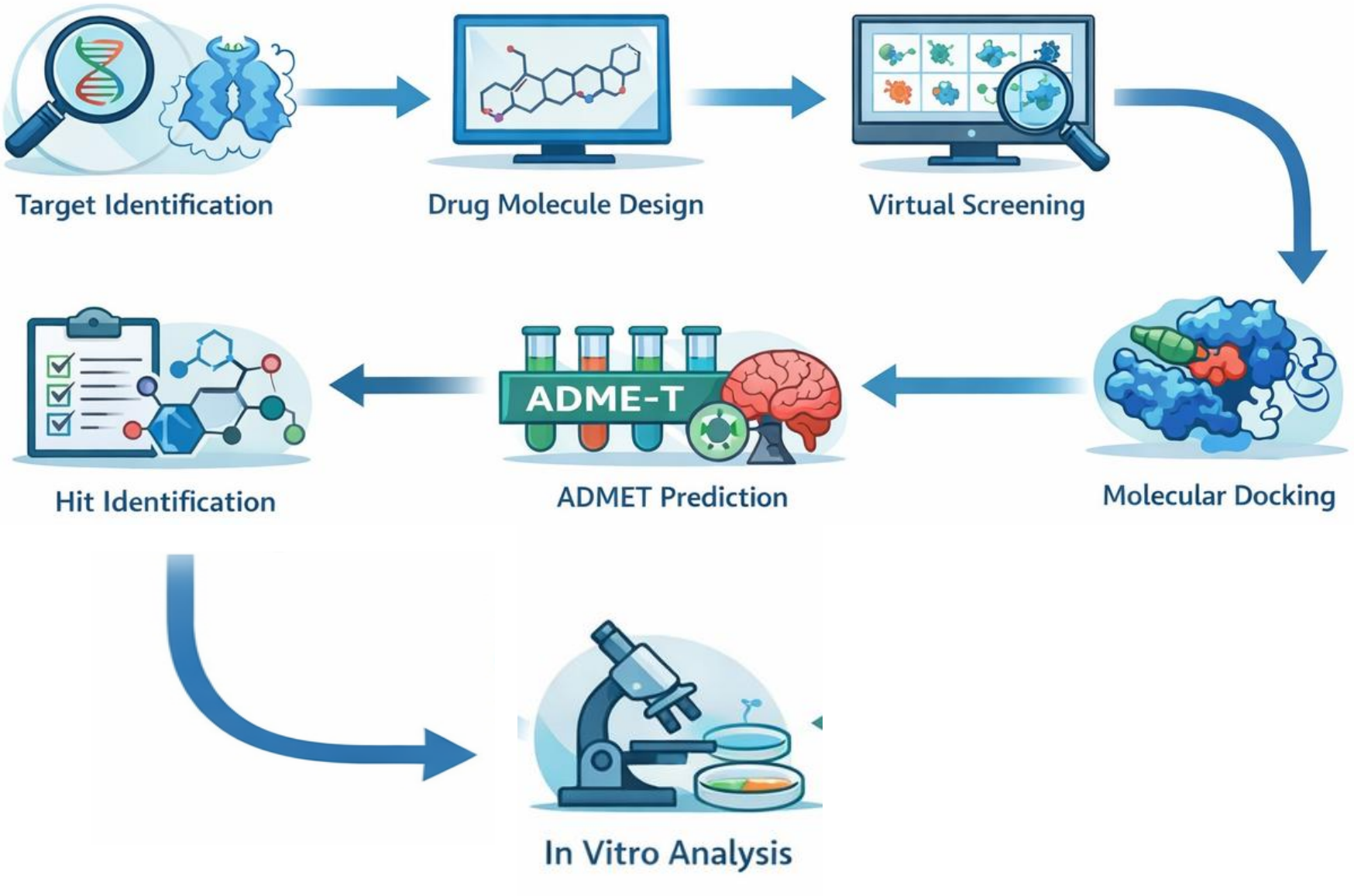
Future Challenges & Open Problems

- Data scarcity & quality issues.
-
- Need better models for complex drug combinations.
- More robust XAI needed for interpretability.
- Standard benchmarks and interoperable tools for research community.

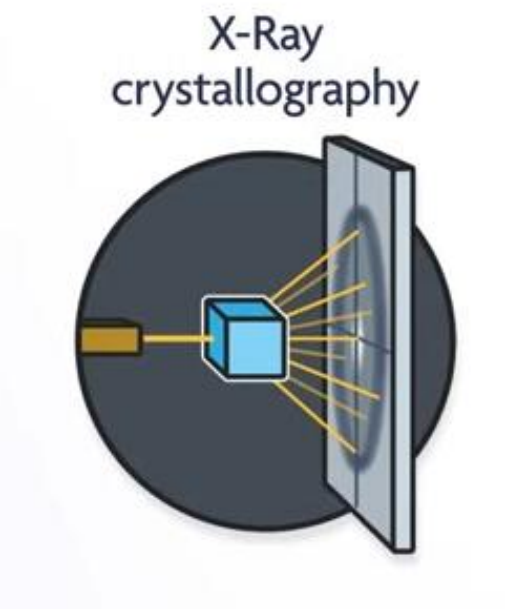
Conclusion

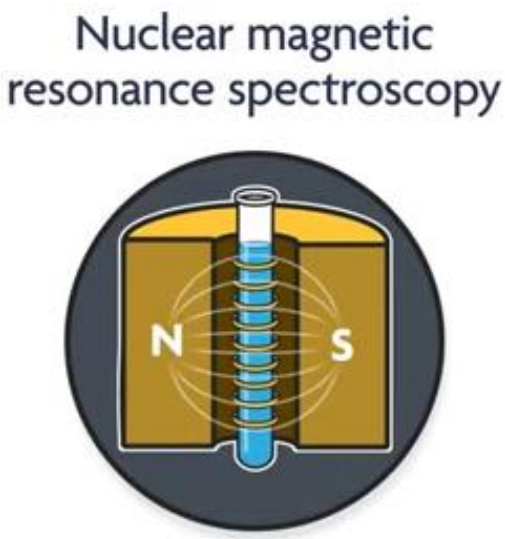
- Deep learning is transforming drug discovery, making it faster, cheaper, and more efficient.
- DL models are effective in predicting drug-target interactions, drug-drug interactions, side effects, and response sensitivity.
- Explainable AI (XAI) is crucial for interpretability, trust, and regulatory compliance.
- Digital twins enable virtual simulations of drugs and patients, improving safety and guiding treatment strategies.
- Challenges remain: data quality, complex combinations, standard benchmarks, and scalable explain ability.

Computational Drug Discovery



Traditional Target Identification Methods





Some Disadvantages

- Time-consuming
- High cost
- Low throughput
- Limited accuracy
- Difficulty in complex diseases
- Experimental errors
- Ethical concerns
- Lack of predictive power

Advantages of
AI & ML Based Target Identification

Faster Identification Process

Cost-Effective

High-Throughput Analysis

Improved Accuracy

Analyzes Large Datasets

Better Prediction of Novel Targets

Target Identification

AI Tools in Target Identification

AlphaFold

- Predicts 3D structure of proteins from amino acid sequence
- Helps identify drug-binding sites
- Developed by DeepMind

DeepSEA

- Uses deep learning to analyze DNA sequences
- Predicts effects of mutations on gene expression
- Helps identify disease-associated genes

Alphafold-Target Identification

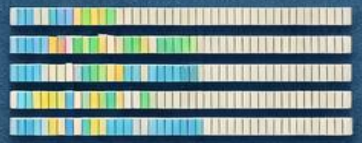
How AlphaFold Works

Input Sequence

MKTWFGVI.

Amino Acid Sequence

Multiple Sequence Alignment



Similar Sequences Found

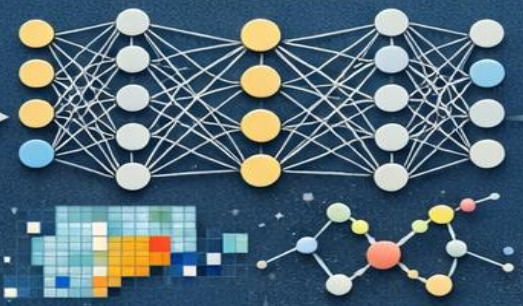
Feature Extraction

Residue Patterns

Co-Evolution Signals

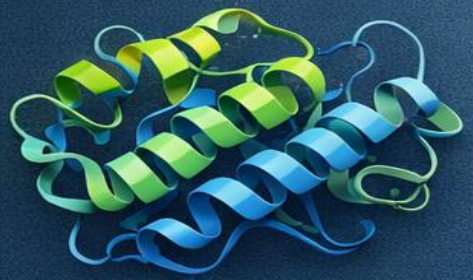
Neural Network Prediction

Distance Map



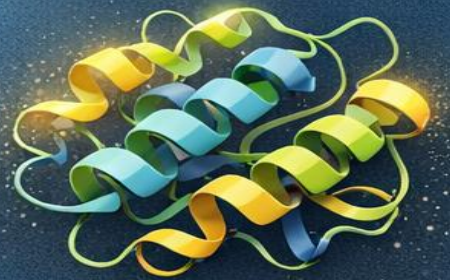
Angle Prediction

Structure Generation



3D Protein Model

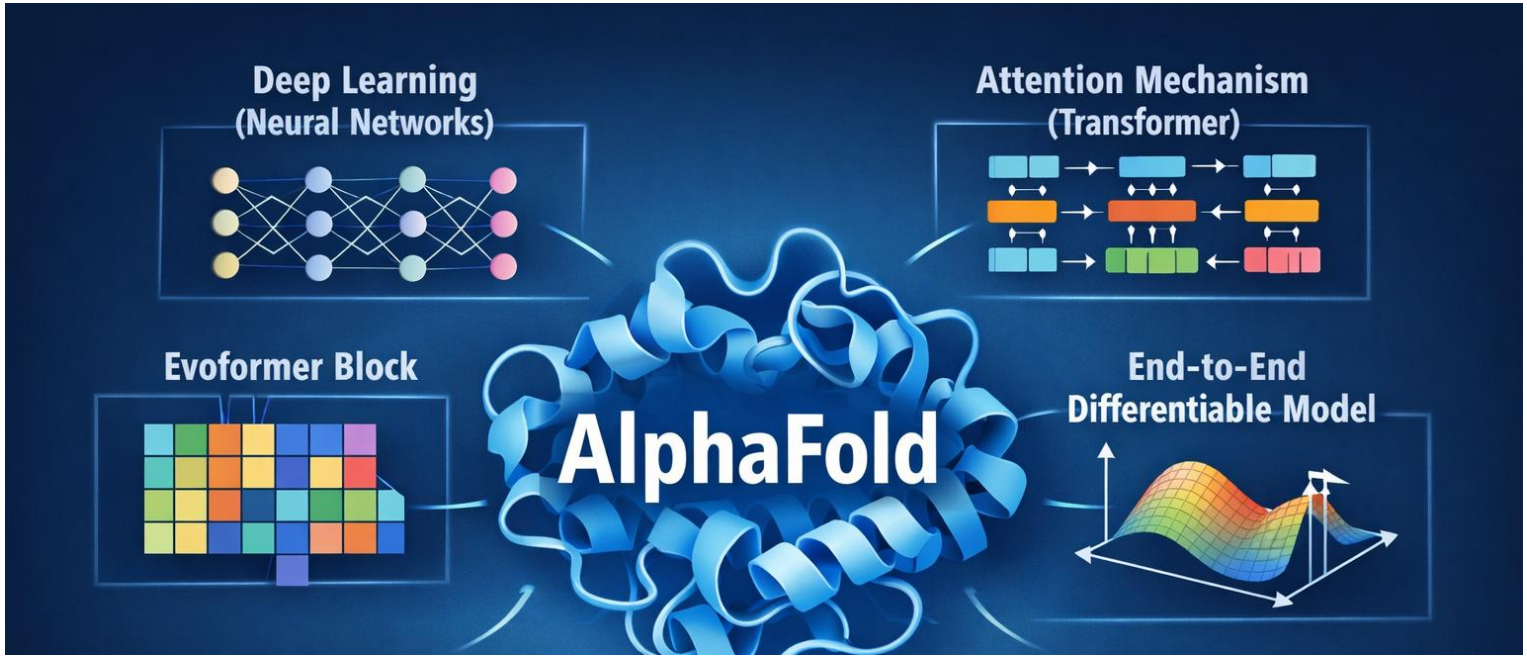
Refinement



Optimized Structure

Confidence Score

Alphafold Algorithms



Algorithm Used

Deep Learning (Neural Networks)

Attention Mechanism (Transformer architecture)

Evoformer Block (core innovation)

End-to-end differentiable model (Graph-based learning , Spatial geometry)

Advantages of AlphaFold

- High accuracy (near experimental level)
- Reduces cost of drug discovery
- Fast compared to lab methods
- Predicts structures for unknown proteins

DeepSea-Target Identification

How DeepSEA Works

1 Input DNA Sequence

Provide raw DNA sequence (A, T, G, C)

Can include mutation or variant

2 Sequence Encoding

A	A	T	G	G	C	C

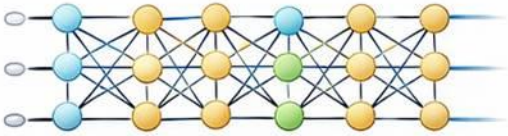
1000 0100 0001
1000 0100 0001
1000 0100 0001

Step 2 Input DNA Sequence

Converts DNA sequence into numerical format (one-hot encoding)

A →	1000	0100	0001	1000
T →	0100	0010	0001	1000
G →	0010	0010	0001	1000

3 Deep Learning Processing



- Passes encoded sequence through neural network
- Learns patterns in DNA

4 Feature Detection

Detects:

- Transcription factor binding sites
- Chromatin accessibility
- Regulatory elements

5 Mutation Effect Prediction

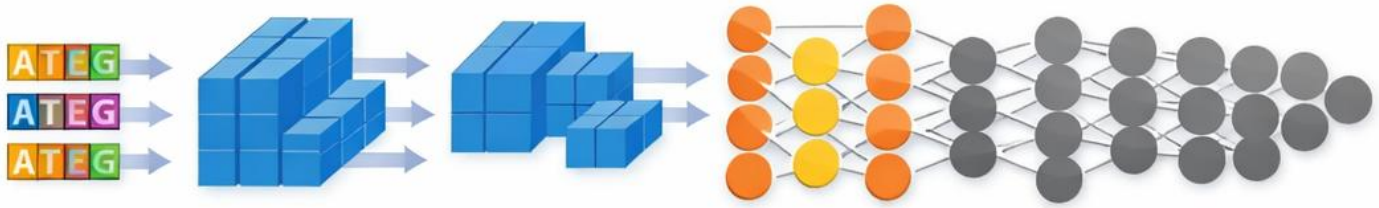
A|CGTTAGC → ACATAGC
A|AGTTAGC → ACATAGC

6 Output Results

- Gives probability scores
- Indicates whether mutation is harmful or neutral

Algorithm Used

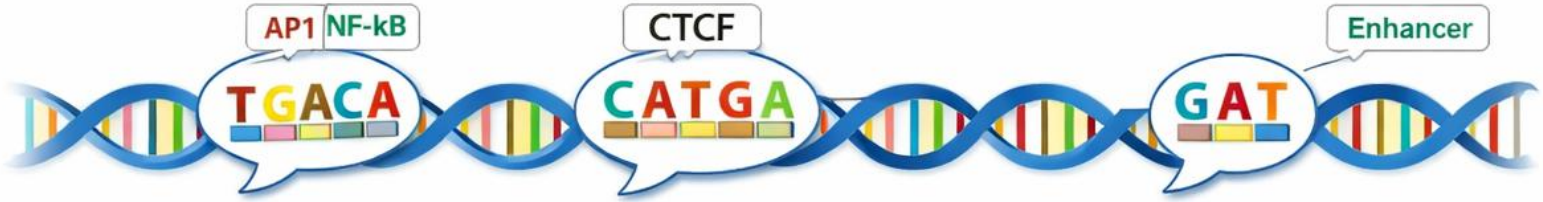
DeepSEA mainly uses:



- Convolutional Neural Networks (CNNs)
- Deep learning architecture with multiple layers
- Pattern recognition in DNA sequences
- Trained on large genomic datasets

Key Concept: **CNN automatically learns sequence motifs**

Patterns like transcription factor binding sites



How to Use DeepSEA

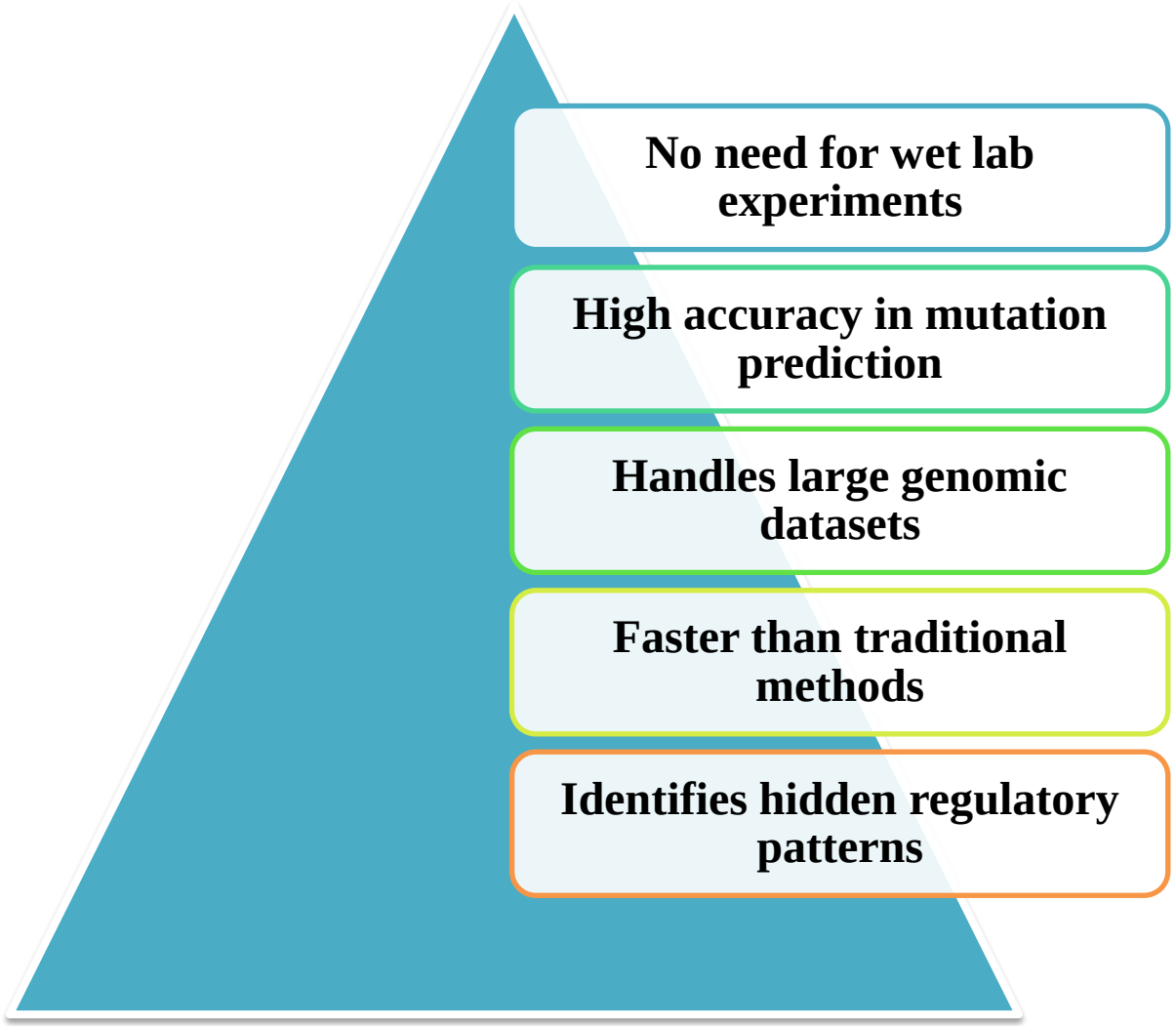
Method 1: Online Server

1. Open DeepSEA web server
2. Paste DNA sequence or upload file
3. Submit job
4. View prediction results

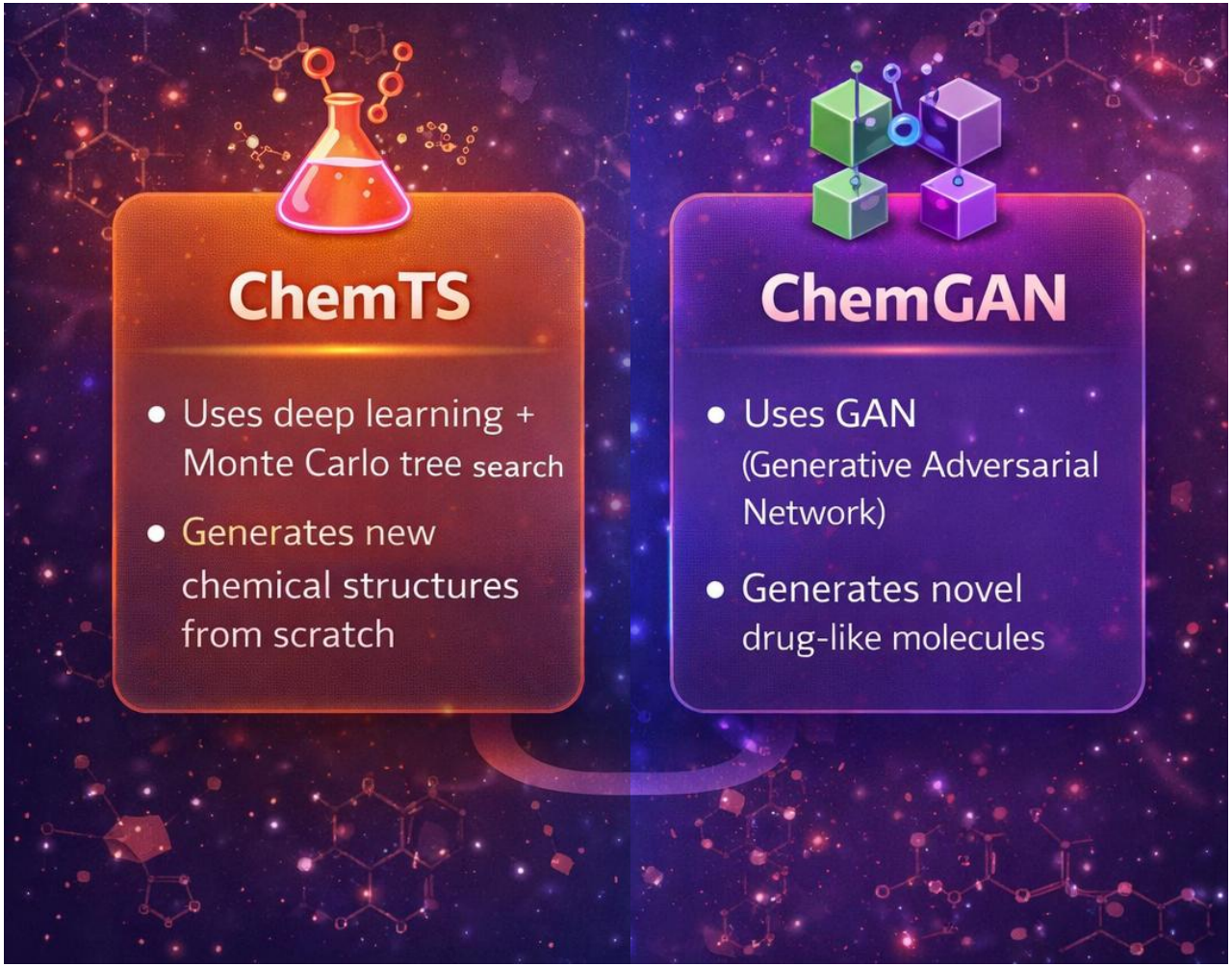
Method 2: Local Usage (Advanced)

1. Install DeepSEA software
2. Provide FASTA file (DNA sequence)
3. Run prediction using command line

Advantages of Deepsea



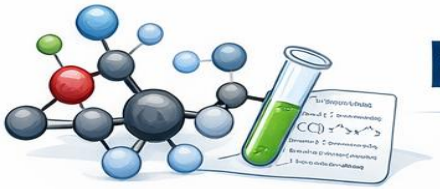
Drug Molecule Design



What is ChemTS?

- ChemTS is an **AI-based drug design tool** that automatically **generates new chemical molecules** from scratch.

- It combines **deep learning with search algorithms** to design novel compounds with desired properties.

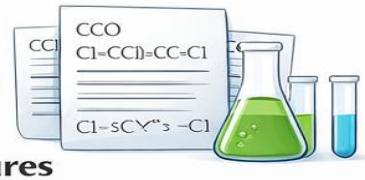


How ChemTS Works

1 Training Data Input

Uses a dataset of known molecules (SMILES format)

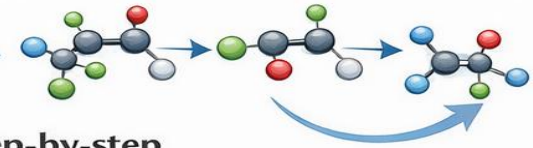
Learns chemical patterns and **structures**



2 Sequence Generation

Treats molecules as sequences (like language)

Generates new molecules **step-by-step**



3 Search Strategy

Uses intelligent search to explore possible molecules

Avoids random generation

4 Property Evaluation

Evaluates generated moledecules for:

- Drug-likeness - Activity
- Toxicity

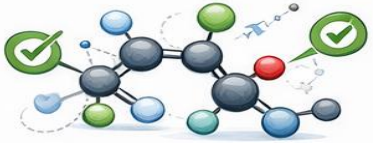
5 Optimization

Iteratively improves molecules

Selects best candidates

6 Output

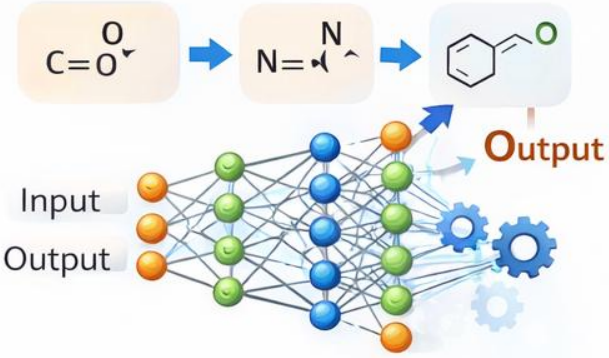
Produces novel, optimized drug-like molecules



Algorithm Used in ChemTS

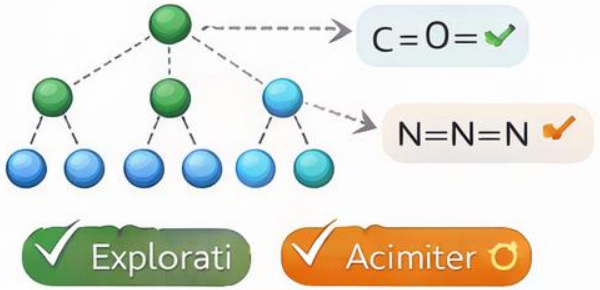
Recurrent Neural Network (RNN)

- Learns SMILES representation of molecules
- Generates valid chemical structures



Monte Carlo Tree Search (MCTS)

- Explores possible molecule structures efficiently
- Balances:
  - Exploration (new molecules)
  - Exploitation (best molecules)



Advantages of ChemTS

Cost-effective

Reduces need for manual design

Completely new molecules

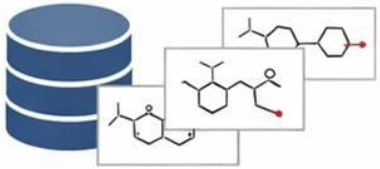
Can optimize multiple properties simultaneously

What is ChemGAN?

- ChemGAN is an **AI-based generative model** that uses **Generative Adversarial Networks (GANs)** to design **new drug-like molecules**.
- It can generate **novel chemical structures** that do not exist in databases, helping accelerate drug discovery.

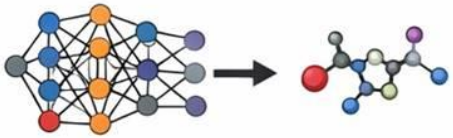
How ChemGAN Works

Step 1: Input Data



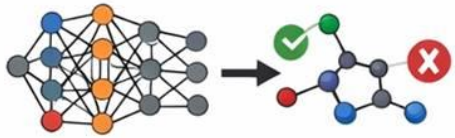
- Chemical datasets (SMILES strings)
- Learns patterns of existing compounds

Step 2: Generator Network



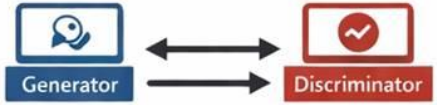
- Generates new molecular structures
- Fake molecules

Step 3: Discriminator Network



- Evaluates molecules
- Real vs Fake

Step 4: Adversarial Training

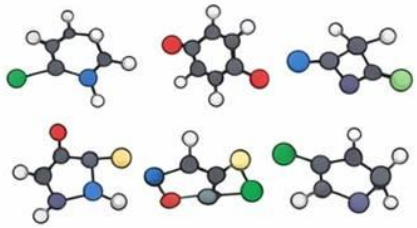


- Generator vs Discriminator
- Improves to fool Discriminator

Step 5: Optimization

- Drug-likeness
- Stability
- Desired properties

Step 6: Output

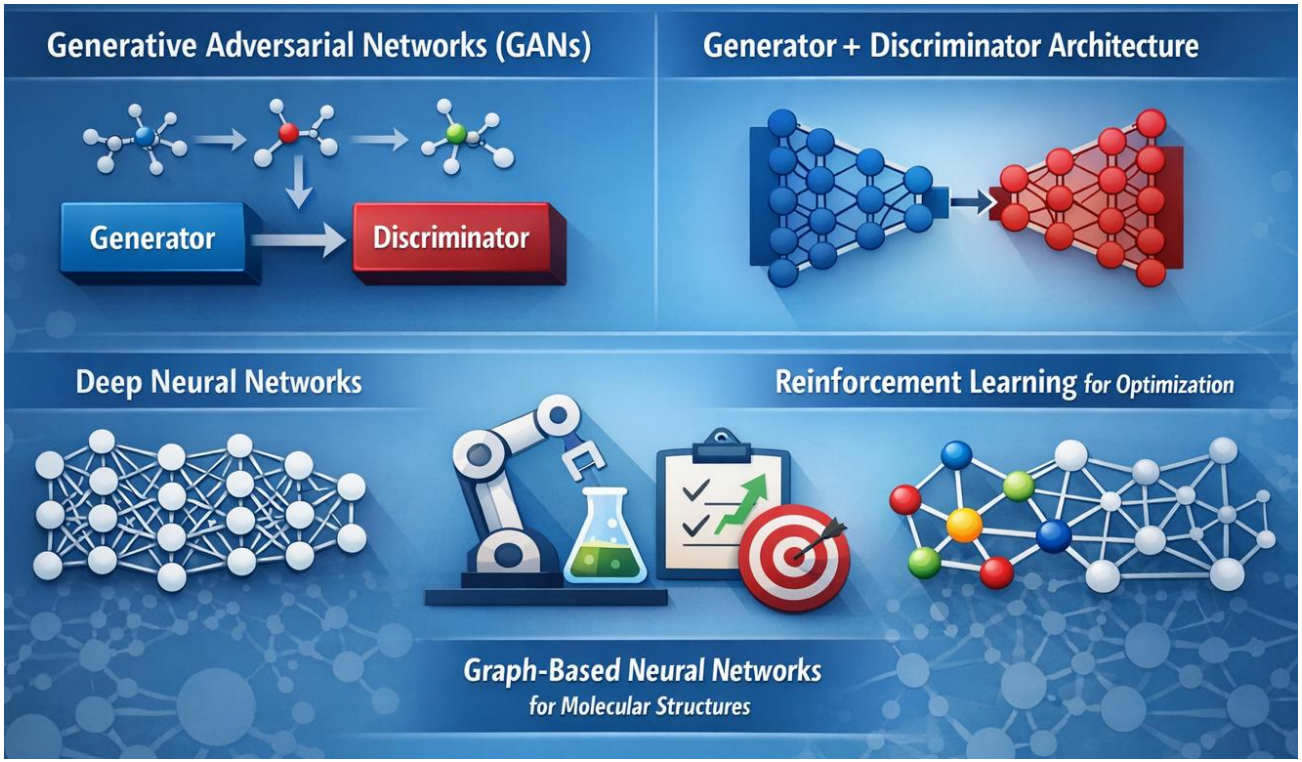


Novel, optimized
drug-like molecules

Algorithm Used in ChemGAN

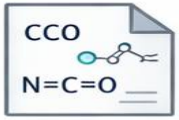
- Generative Adversarial Networks (GANs)
  - Generator + Discriminator architecture
- Deep Neural Networks (DNNs)
- Reinforcement Learning (optional for optimization)
- Graph-based neural networks (for molecule structure representation)

Two networks compete → system learns to generate realistic molecules



How to Use ChemGAN

1. Dataset Preparation



- Collect molecular data (SMILES format)
- Clean and preprocess dataset

2. Model Setup

- Install deep learning libraries (TensorFlow/PyTorch)
- Load ChemGAN model

3. Training

- Train generator and discriminator on dataset
- Adjust parameters (epochs, learning rate)

4. Molecule Generation

- Run generator to create new molecules

5. Evaluation

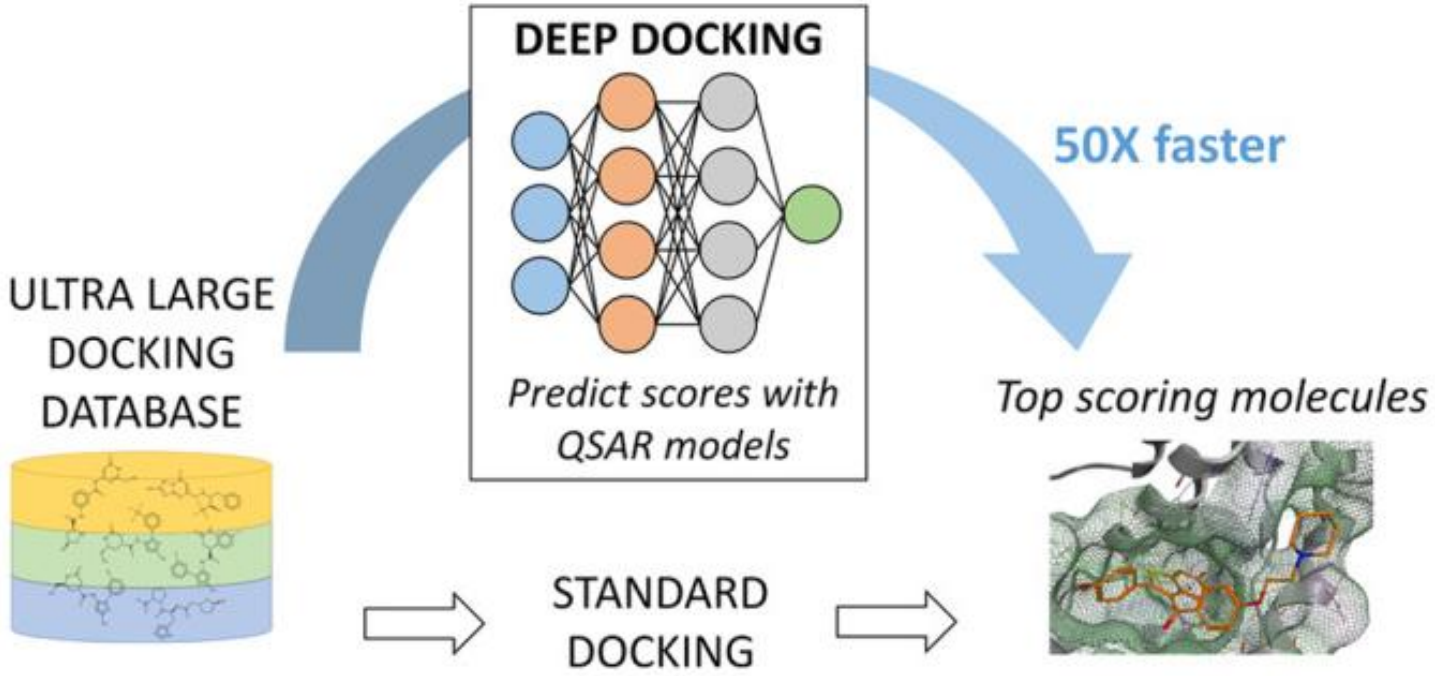
- Check:
- Validity
- Drug-likeness (Lipinski rules)
- Toxicity

6. Selection

- Choose best candidate molecules for further testing

Deep Dock for Virtual Screening

What is DeepDock?



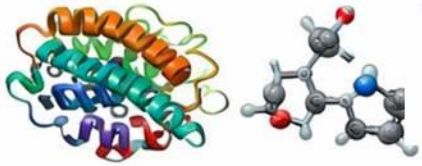
- DeepDock is an **AI-based molecular docking tool** that uses **deep learning** to predict how a drug molecule (ligand) binds to a target protein.

- It improves the **speed and accuracy of virtual screening** compared to traditional docking methods.

How DeepDock Works

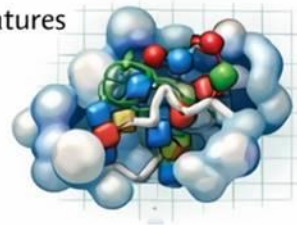
Step 1: Input Data

- Protein structure (PDB format)
- Ligand molecules (SMILES or 3D structures)



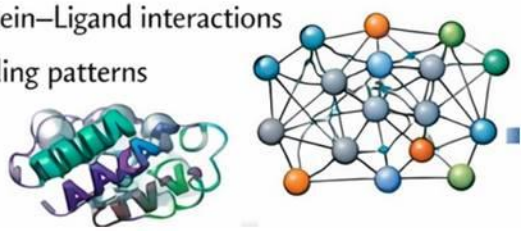
Step 2: Feature Representation

- Graph structures
- 3D spatial features



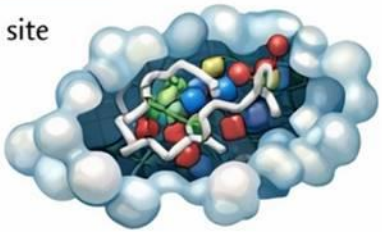
Step 3: Deep Learning Model

- Protein–Ligand interactions
- Binding patterns



Step 4: Binding Pose Prediction

- Ligand docking in binding site
- Best orientation (pose)

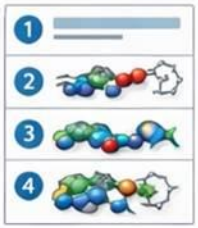


Step 5: Scoring Function

- Binding affinity score
- Molecule ranking

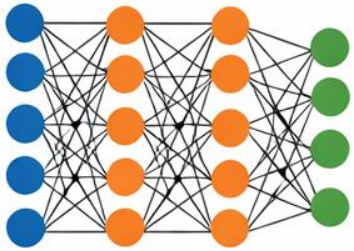
Step 6: Output

- Top-ranked compounds
- Predicted binding poses
- Interaction details



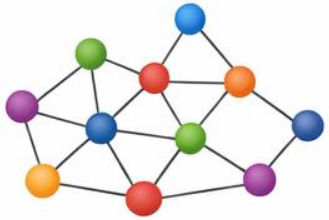
Algorithm Used in DeepDock

Deep Neural Networks (DNNs)



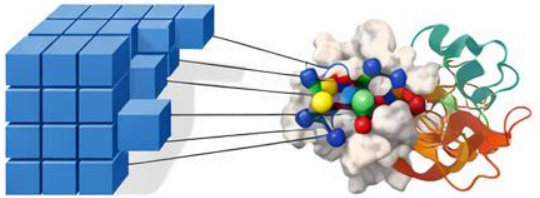
Graph Neural Networks (GNNs)

For Molecular Structure

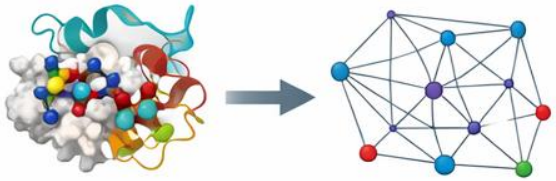


Convolutional Neural Networks (CNNs)

For Spatial Interaction



Geometric Deep Learning

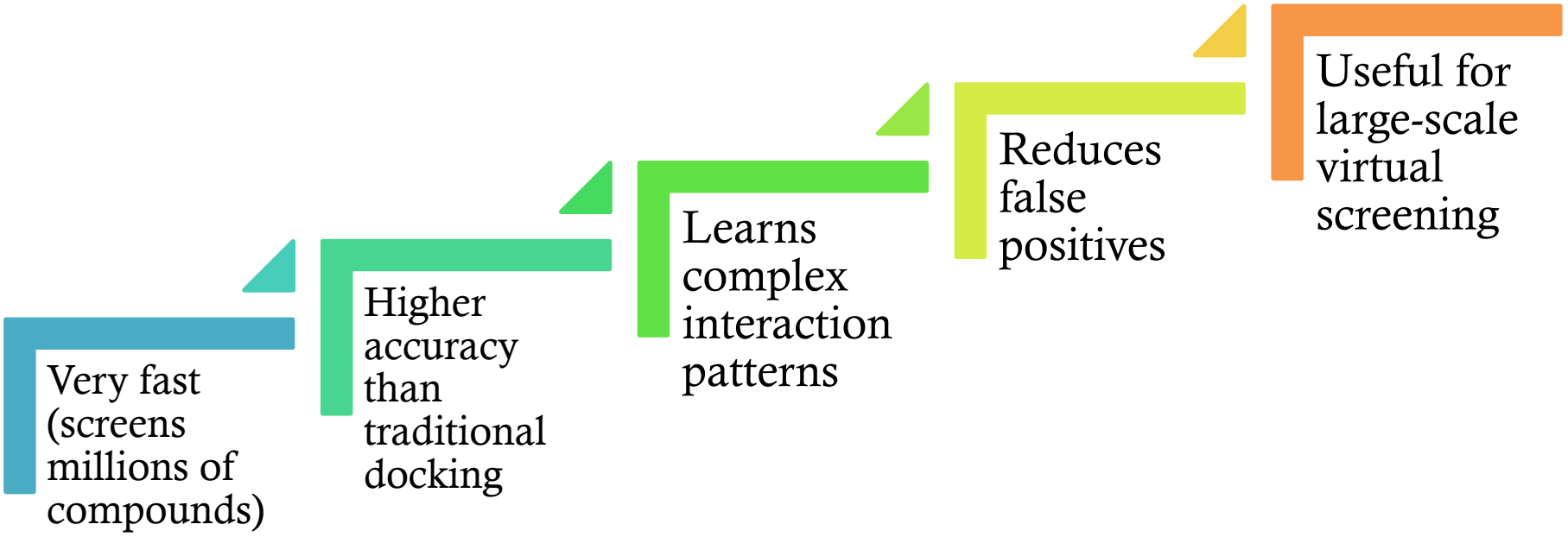


Learned Scoring Functions

Instead of Rule-Based Scoring

AI learns binding patterns directly from data instead of using fixed equations

Advantages of DeepDock



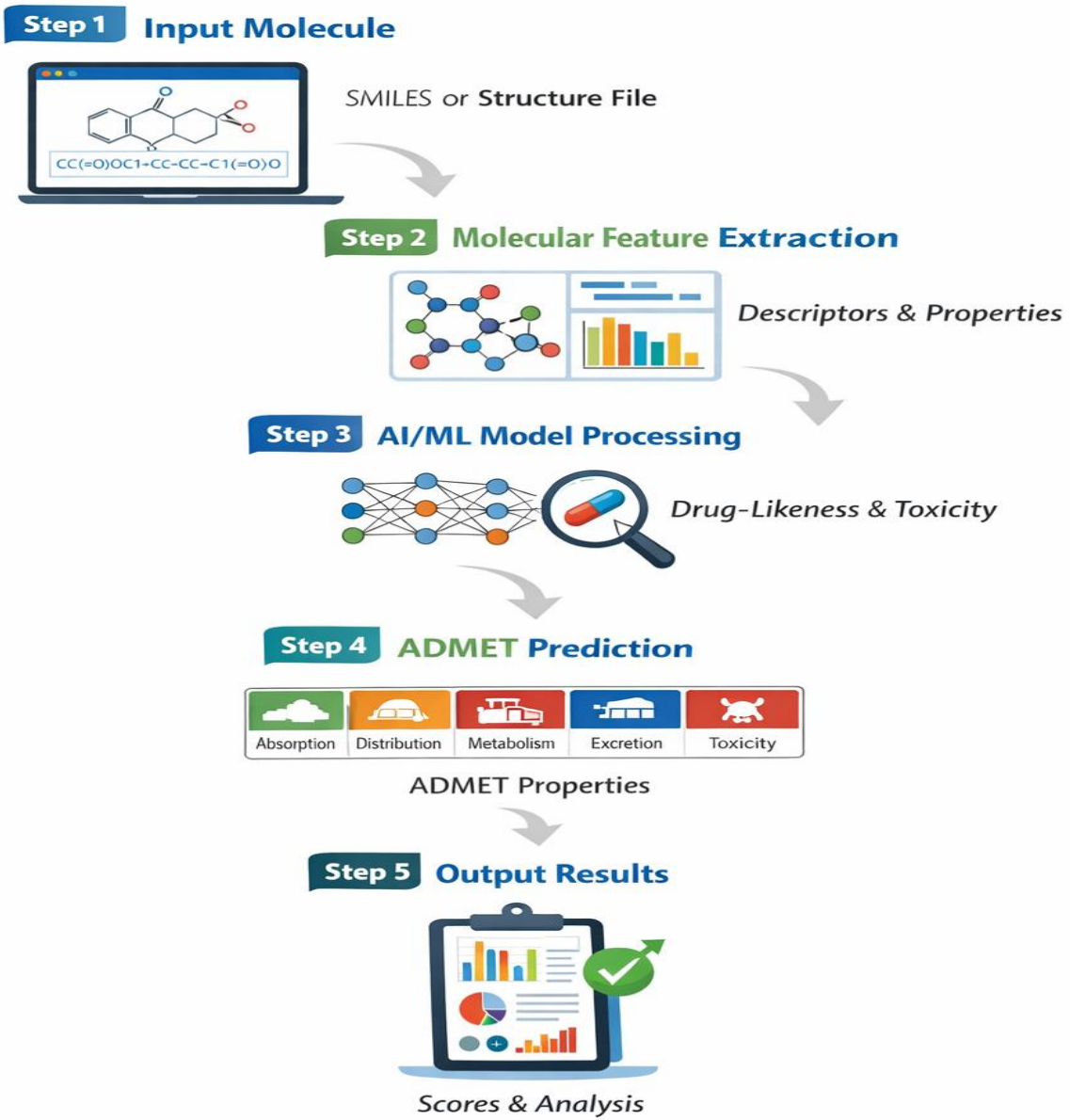
ADMETlab for ADMET Prediction

What is ADMETlab?

ADMETlab is an **AI-based online platform** used to predict **ADMET properties** of drug molecules:

- Absorption
- Distribution
- Metabolism
- Excretion
- Toxicity

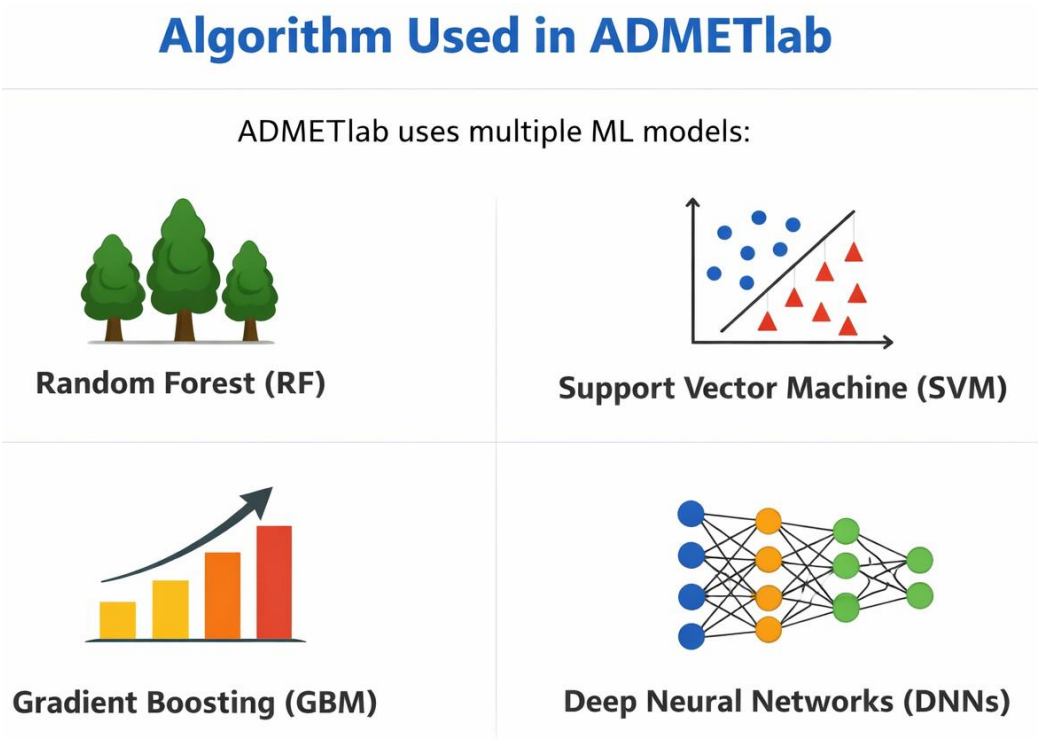
How ADMETlab Works



Molecular Feature Extraction

Converts molecule into descriptors:

- Physicochemical properties
- Structural features



1. Random Forest (RF)

Purpose:

Used for **classification and regression** of ADMET properties
Handles large datasets with high accuracy

In ADMETlab:

Predicts:

- Toxicity (toxic / non-toxic)
- Drug-likeness

2. Support Vector Machine (SVM)

Purpose:

- Separates data into categories using decision boundaries
- Best for **binary classification problems**

In ADMETlab:

Used for:

  - Predicting **BBB penetration (Yes/No)**
  - Identifying **toxic vs non-toxic compounds**

3. Gradient Boosting (GBM)

Purpose:

- Improves prediction accuracy by combining multiple weak models
- Focuses on correcting previous errors

In ADMETlab:

Used for:

  - Predicting **absorption levels**
  - Estimating **metabolic stability**

4. Deep Neural Networks (DNNs)

Purpose:

- Learns complex patterns from large datasets
- Handles **non-linear relationships**

In ADMETlab:

Used for:

  - Predicting **multi-parameter ADMET properties simultaneously**
  - Modeling complex interactions like:
    - Drug metabolism
    - Toxicity pathways

How to Use ADMETlab

1 Open Web Tool

Visit ADMETlab website

2 Input Data

Paste SMILES string or upload molecule

3 Select Prediction

Choose ADMET properties to analyze

4 Run Analysis

Click submit
Wait for processing

5 Run Analysis

Click submit
Wait for processing

6 Analyze Results

Check:

- ADMET values
- Drug-likeness rules

Thank You