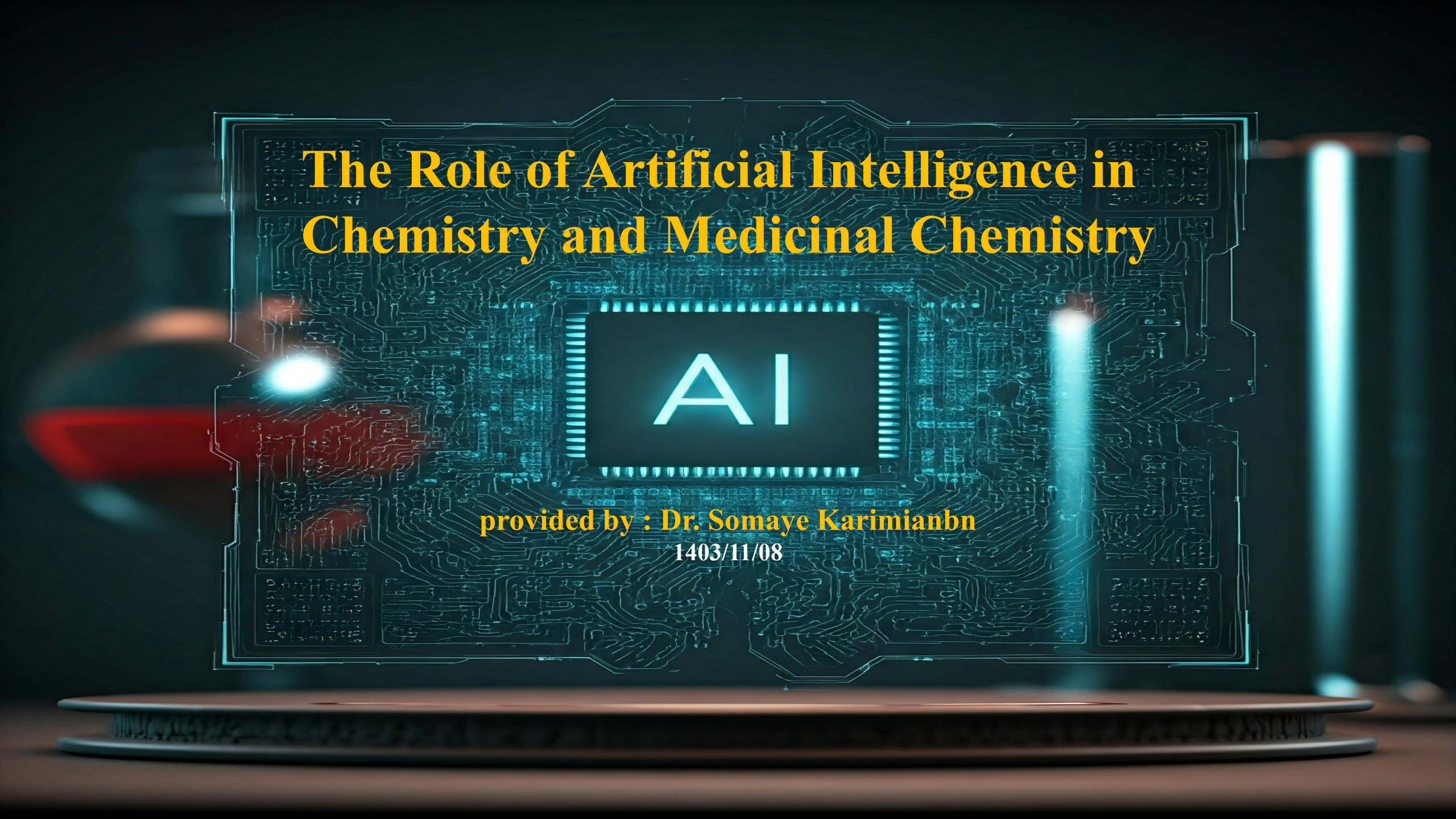
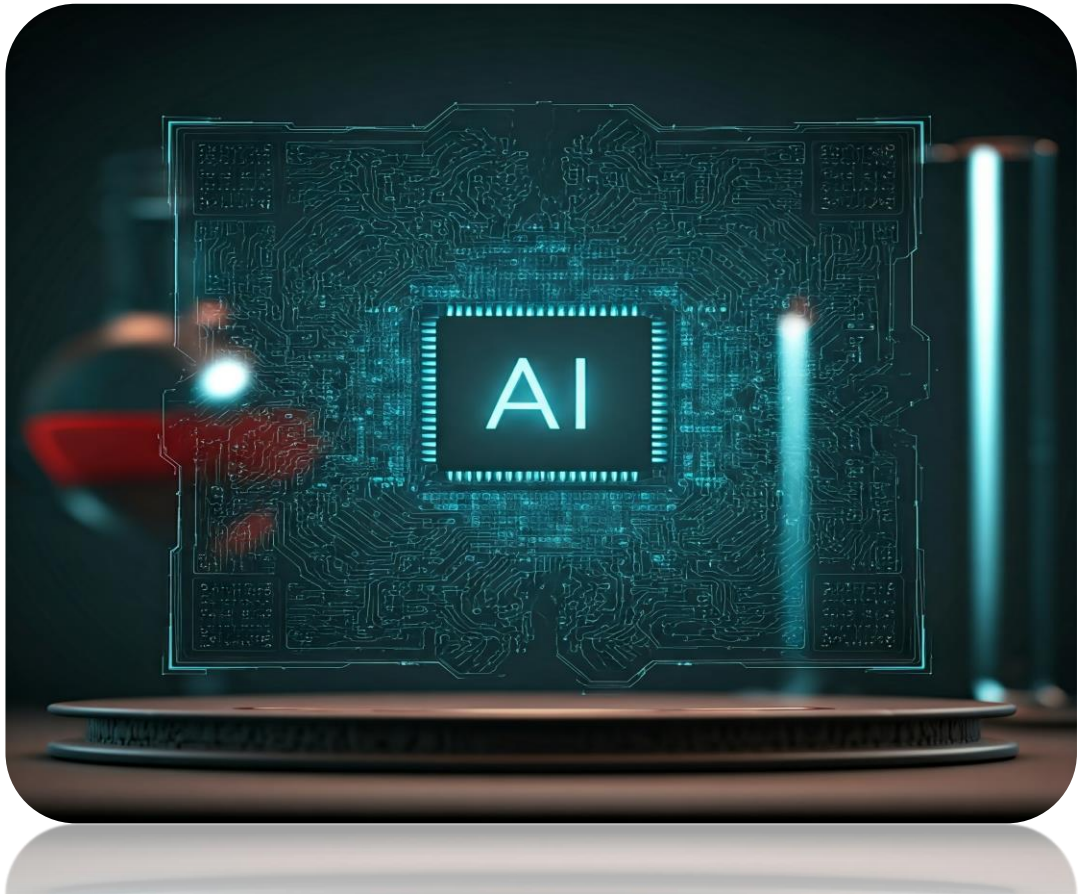
The Role of Artificial Intelligence in Chemistry and Medicinal Chemistry

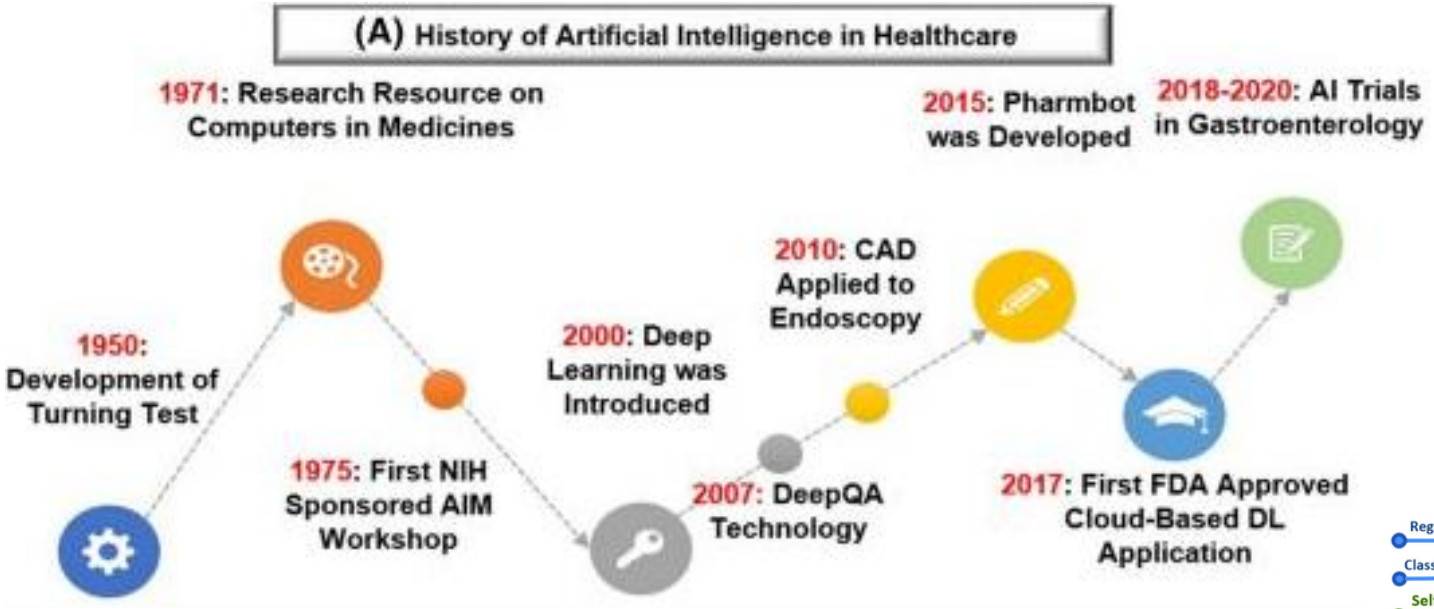


provided by : Dr. Somaye Karimianbn

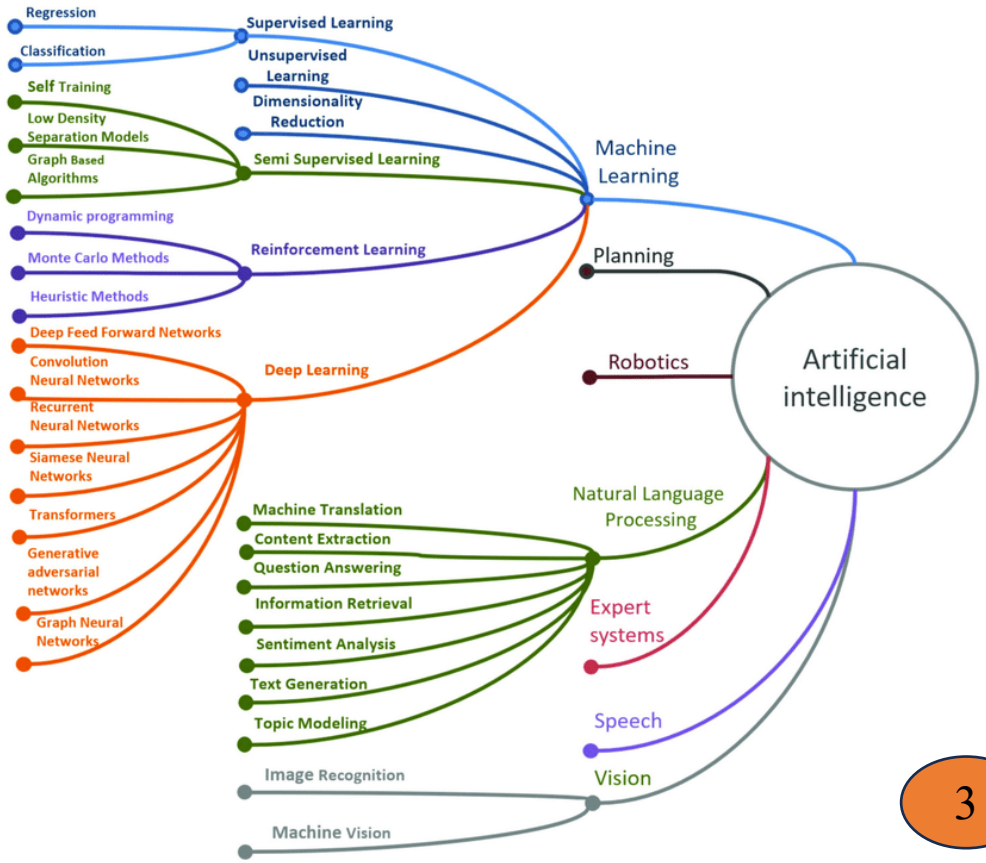
1403/11/08

- Artificial Intelligence (AI) is a field of computer science dedicated to developing systems capable of performing complex tasks such as reasoning, learning, problem-solving, perception, and language comprehension by utilizing algorithms and computational models in a way that is similar to human thinking.

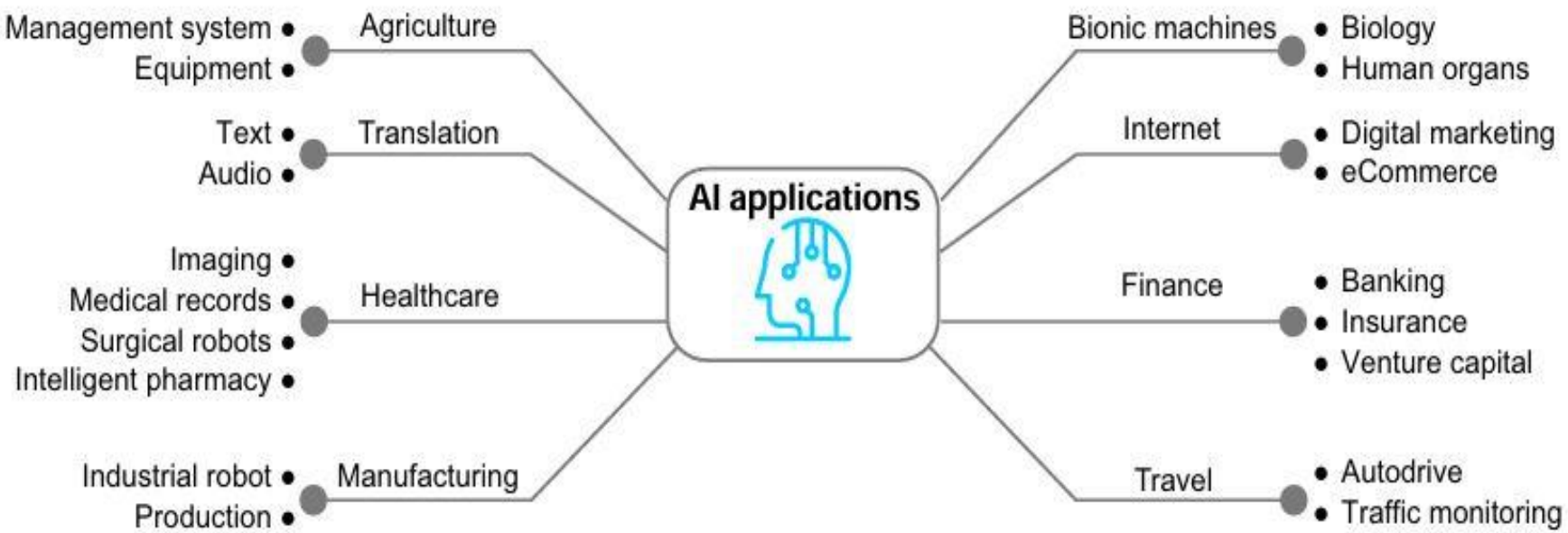


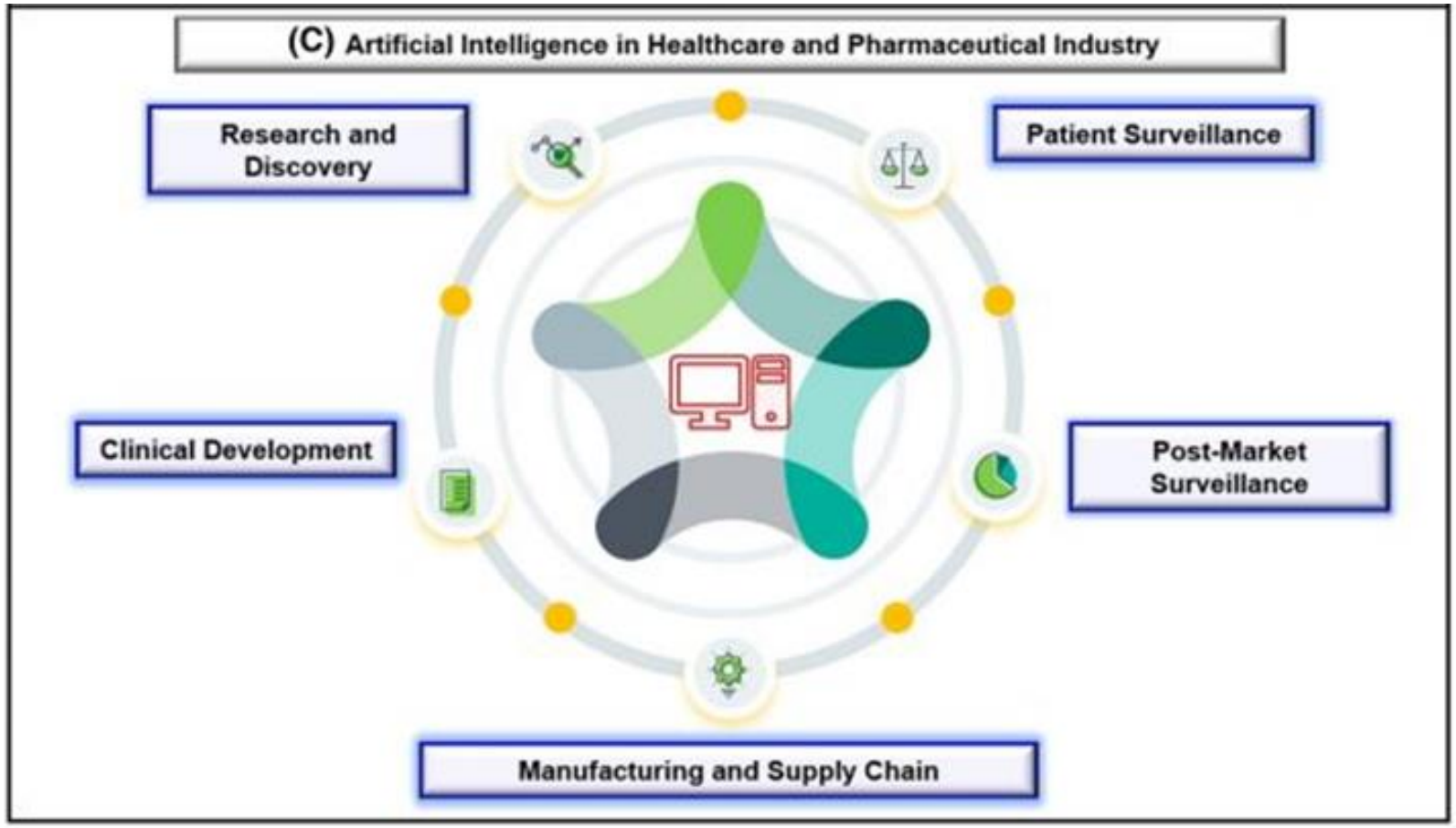


Artificial intelligence uses various algorithms such as machine learning, deep learning techniques, and artificial neural networks to analyze data, identify patterns, and make predictions, and plays an important role in advancing scientific and technological advances.



Schematic showing the diverse applications of AI in different areas.




(C) Artificial Intelligence in Healthcare and Pharmaceutical Industry
Research and Discovery
Patient Surveillance
Clinical Development
Post-Market Surveillance
Manufacturing and Supply Chain

Artificial Intelligence & Chemistry and Medicinal Chemistry

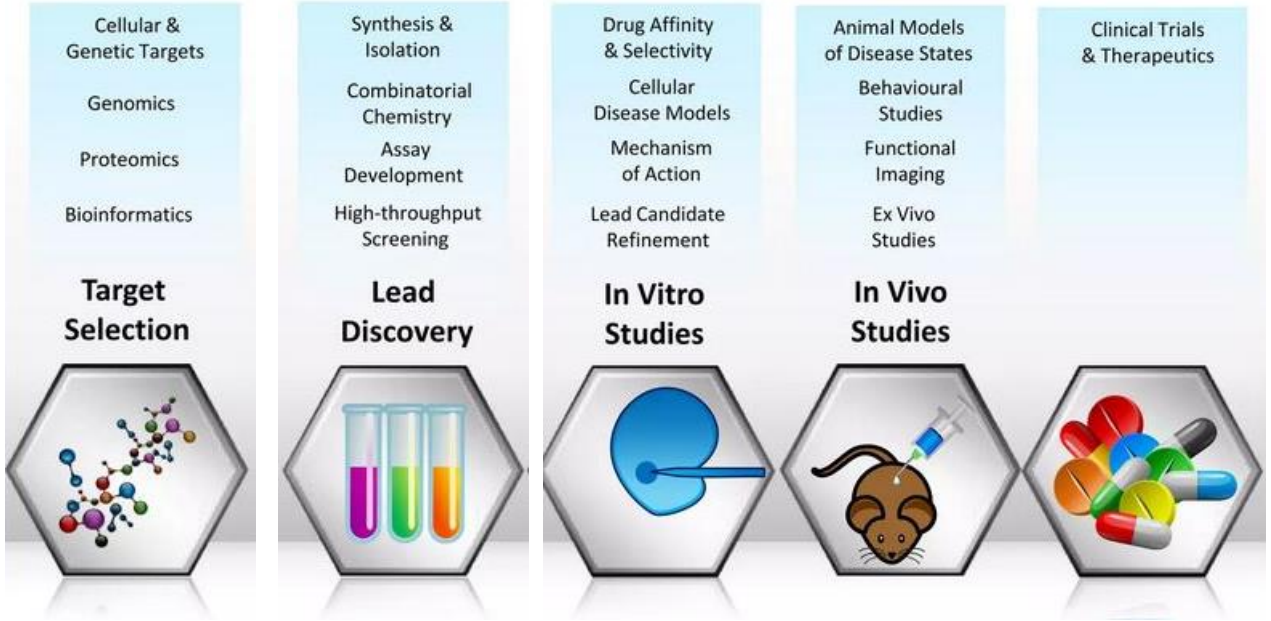
AI, through advanced algorithms and data analysis, has transformed the understanding of molecular interactions, the discovery of novel compounds, and the modeling of chemical processes. This technology has significantly accelerated research and innovation in the chemical and pharmaceutical industries.

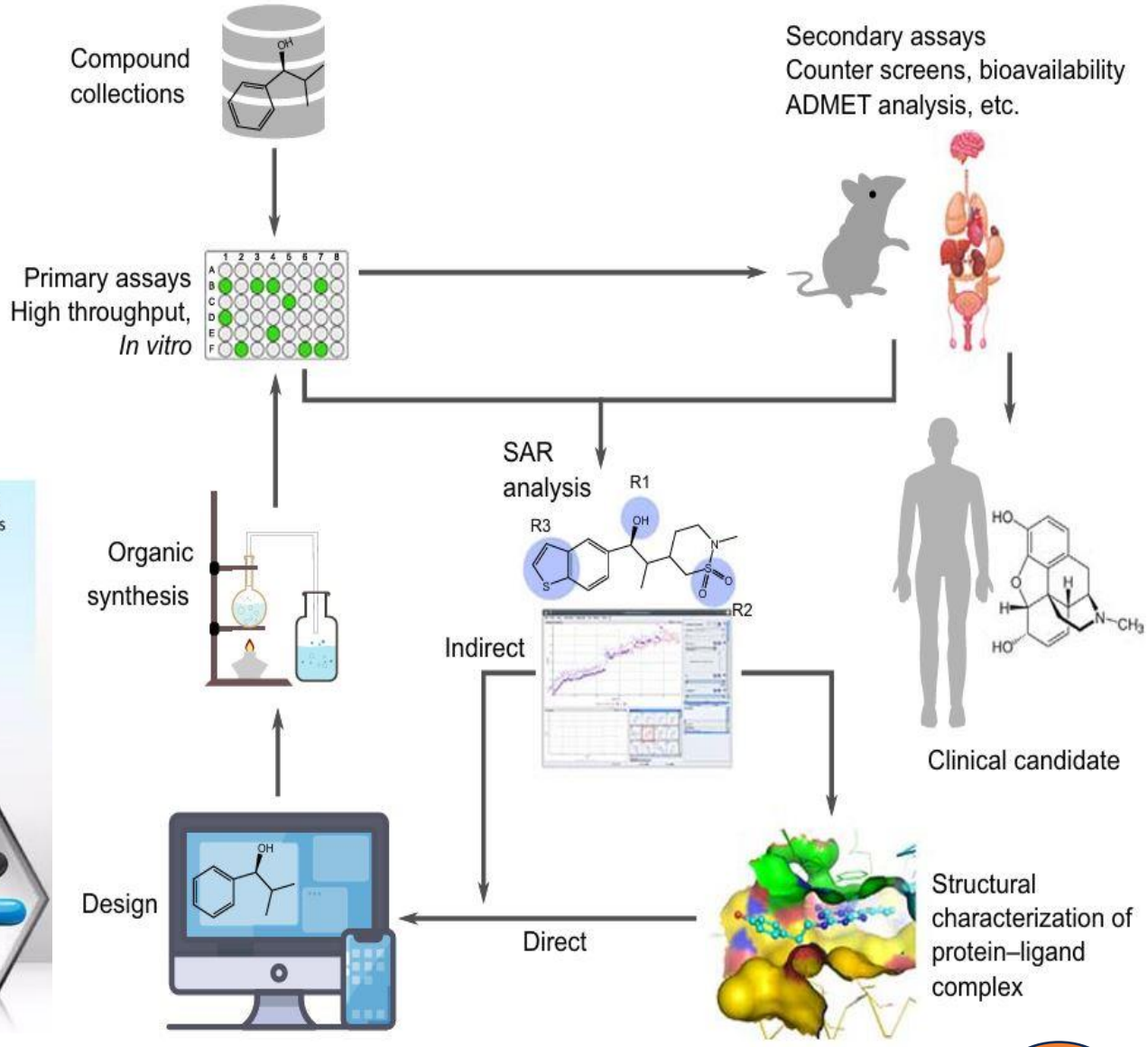
- ✓ Chemical Synthesis and Reaction Prediction
- ✓ Material Discovery
- ✓ Quantum Chemistry and Simulation
- ✓ Spectroscopy and Analytical Chemistry
- ✓ Drug Discovery and Design
- ✓ Structure-Activity Relationship (SAR) Analysis
- ✓ Predictive Toxicology
- ✓ Pharmacokinetics and Pharmacodynamics (ADME)
- ✓ Repurposing Existing Drugs
- ✓ …

Traditional drug discovery

Drug discovery is a long and complex process that can be broadly divided into four major stages:

(i) Target selection and validation
(ii) Compound screening and lead optimization
(iii) Preclinical studies
(iv) Clinical trials





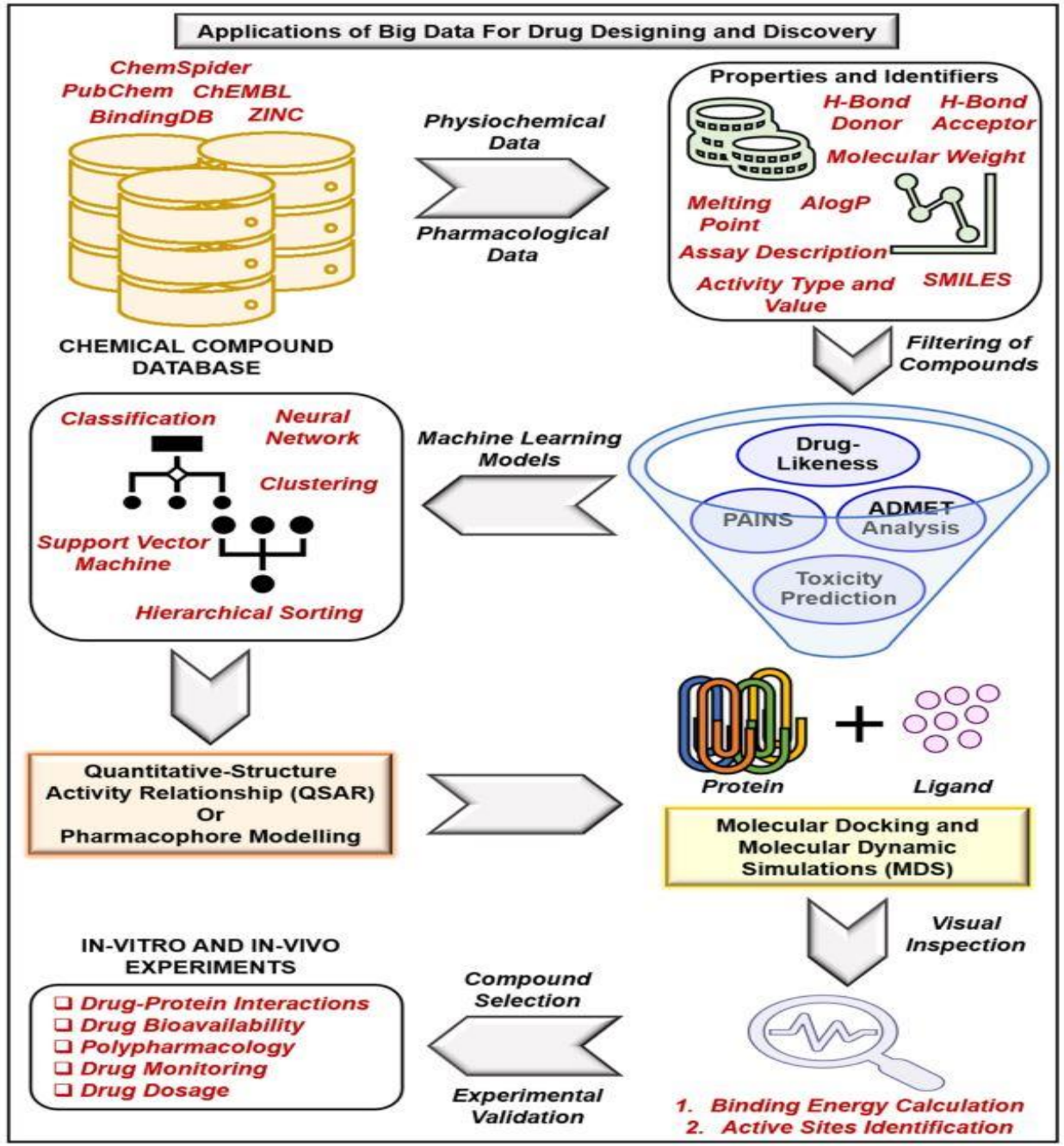
The Traditional drug discovery process

Ai in drug discovery

Big data and navigation in chemical space

- NCI Open Database 250,000 molecules
- Chembl Database 2 million compounds, 13 thousand targets
- and 16 million (compounds –targets)
- PubChem Database 100 million unique compounds
- ZINC Database 230 million commercially
- DNA-based library synthesis (virtual libraries) > reaching billions
- Drug Bank Database 13087 compounds

- **Sequence Read Archive** (https://www.ncbi.nlm.nih.gov/sra)
- The National Cancer Institute Genomic Data Commons (**NCIGDC**) (https://gdc.cancer.gov/), **TCGA** (sequencing data of cancer).
- **DriverML** (https://github.com/ HelloYiHan/DriverML) , ML-based tool, genes related to cancer (Han et al. 2019)



- **AI for Primary Drug Screening**

  Sorting and Classification of Cells by Image Analysis Using AI

  (AI-based convoluted deep neural network (DNN, Box 1) algorithms)

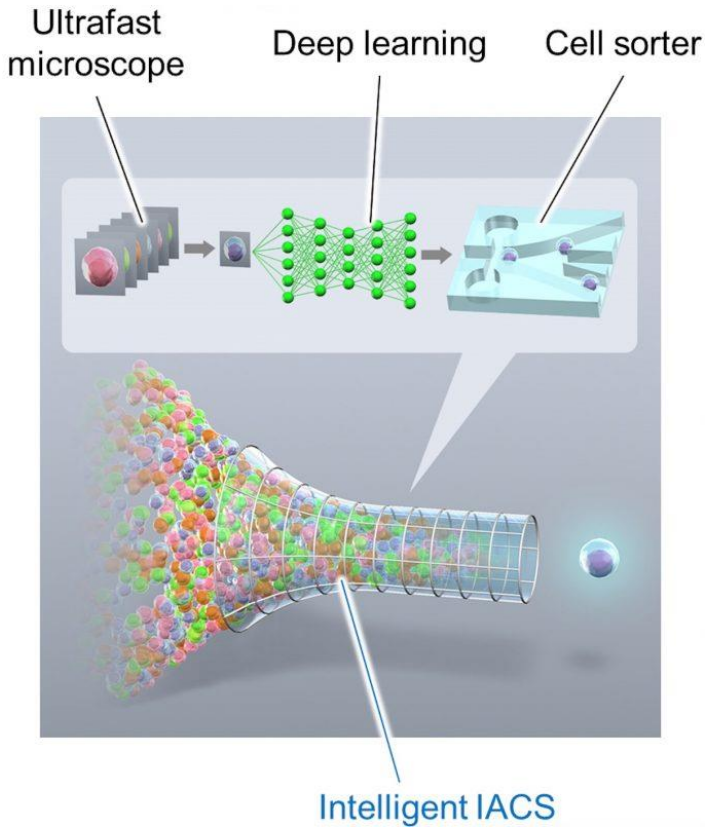


Figure 1. Schematic of the intelligent image-activated cell sorter (intelligent IACS). https://doi.org/10.1016/j.cell.2018.08.028.

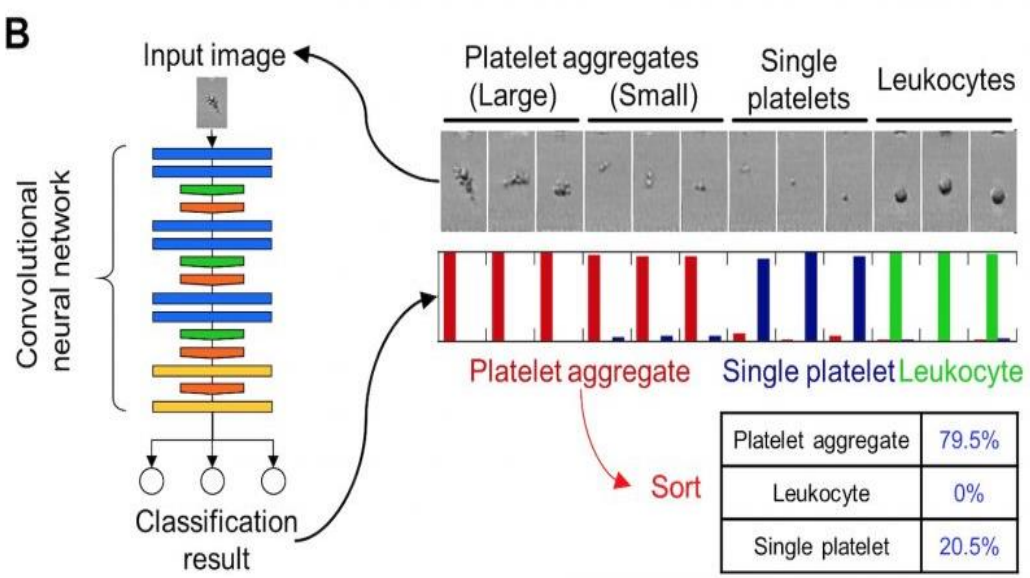


Figure 2. B, Isolation of platelet aggregates. An image classifier based on deep-learning is employed to recognize platelet aggregates. Sorting of the platelet aggregates followed by microscopic enumeration showed a high sorting purity of 79.5%. https://doi.org/10.1016/j.cell.2018.08.028.

- **AI in Secondary Drug Screening**

Predictions of Physical Properties

(AI-based convoluted deep neural network (DNN, Box 1) algorithms)

In drug design, it's crucial to select candidates with desired properties like bioavailability, bioactivity, and toxicity. Physical properties such as melting point and partition coefficient (logP) impact bioavailability and must be considered. AI algorithms use various molecular representations (e.g., SMILES strings, potential energy measurements, molecular graphs) to design drugs. These inputs are processed by deep neural networks (DNNs) in generative and predictive stages, facilitating reinforcement learning to optimize drug properties.

Prediction of Toxicity

(DeepTox algorithm)

The toxicology profile is crucial in drug development, with toxicity optimization being an expensive and time-consuming preclinical task. The DeepTox algorithm, an ML tool, performed exceptionally well in the Tox21 Data Challenge by predicting the toxic effects of 12,000 chemicals. It normalizes chemical representations and computes various static and dynamic descriptors for ML input. Despite a multitude of features, DeepTox accurately predicts compound toxicity in typical test cases.

- **Prediction of physicochemical properties and bioactivity**

Different AI-based tools have been developed such as ALOGPS 2.1 (http://www.vcclab.org/lab/alogps/), ASNN (http://www.vcclab.org/lab/asnn/), E-BABEL (http:// www.vcclab.org/lab/babel/), PCLIENT (http://www.vcclab.org/lab/pclient/), E-DRAGON (http://www. vcclab.org/lab/edragon/), ChemSpider (http://www. chemspider.com/), SPARC (http://sparc.chem.uga. edu/sparc/), and OSIRIS property explorer (https:// www.organic-chemistry.org/prog/peo/)

Fr example:

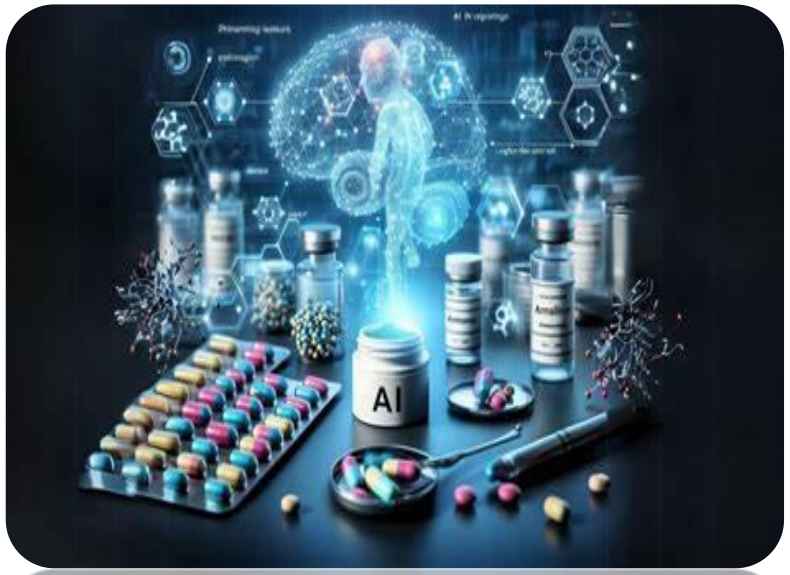
In 2016, Puratchikody et al.used ORISIS property explorer in their study to predict the quantitative structural toxicity of tyrosine derivates intended for safe, potent inflammation treatment. The results concluded that out of 55 potent molecules, only 19 molecules were considered potent cyclooxygenase-2 inhibitors.

- **Prediction of the mode of action and toxicity of compounds**

(LimTox (http://limtox.bioinfo.cnio.es/), pkCSM (http://biosig.unimelb.edu.au/pkcsm/), admetSAR (http://lmmd.ecust.edu.cn/admetsar2/), and Toxtree (http://toxtree.sourceforge.net/), Tox21 (https://ntp.niehs.nih.gov/whatwestudy/tox21/index.html), SEA (http://sea.bkslab.org/), ToxPred (https:// www.brylinski.org/etoxpred-0), TargeTox (https:// github.com/artem-lysenko/TargeTox), DeepTox (http://bioinf.jku.at/research/DeepTox/tox21. html)and PrOCTOR (https://github.com/kgayvert/ PrOCTOR).

For example:

In 2020, Robledo-Cadena et al. predicted the effect of non-steroidal anti-inflammatory drugs on cisplatin, paclitaxel, and doxorubicin efficacy against cervix cancer cells using PrOCTOR.



- **AI in Drug Design**

Predicting the 3D Structure of a Target Protein

The 3D structure of a target protein is crucial for structure-based drug discovery. Traditionally, homology modeling and de novo protein design were used, but AI tools like AlphaFold have improved accuracy. In the Critical Assessment of Protein Structure Prediction contest, AlphaFold accurately predicted 25 out of 43 structures using only protein primary sequences. AlphaFold uses deep neural networks (DNN) to predict amino acid distances and φ-ψ angles, combining these into a score to estimate the accuracy of proposed 3D structures, effectively exploring the protein structure landscape.

Predicting Drug–Protein Interactions

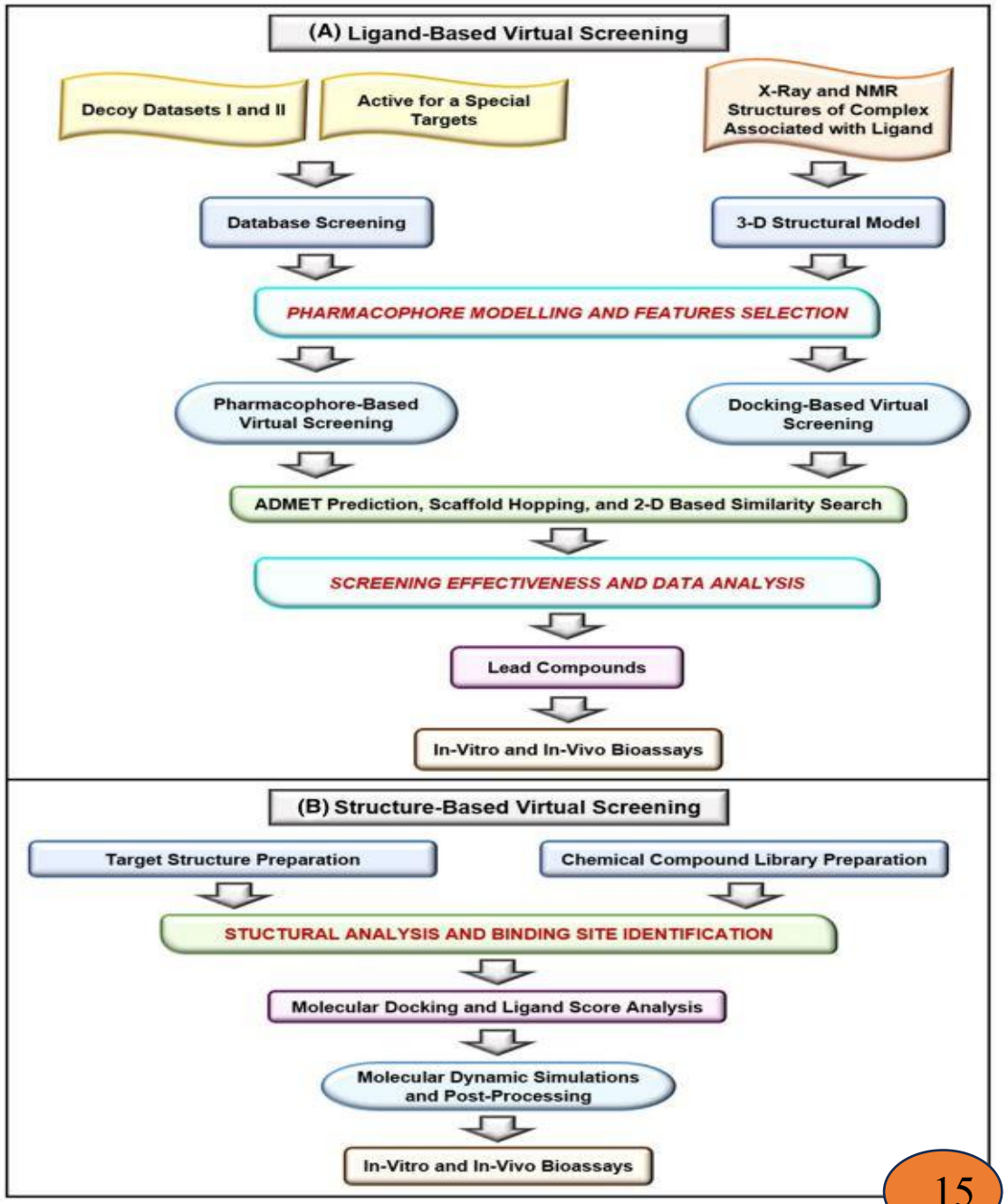
Predicting drug-protein interactions is enhanced by Quantum mechanics QM or QM/MM (molecular mechanics) hybrid methods, which consider atomic-level quantum effects for better accuracy than classical MM methods. However, QM methods are time-consuming. AI models, trained to reproduce QM energies from atomic coordinates, can achieve the speed of MM methods while maintaining accuracy. For large datasets, AI and deep learning (DL) have been used to predict potential energies, significantly improving efficiency in drug discovery.

List of AI-Based computational tools for drug discovery

Tools	Description	Websites
AlphaFold	Protein 3D structure prediction	https://deepmind.com/blog/alphafold
Chemputer	A more standardized format for reporting a chemical synthesis procedure	https://zenodo.org/record/1481731
DeepChem	A python-based AI tool for various drug discovery task predictions	https://github.com/deepchem/deepchem
DeepNeuralNet-QSAR	Molecular activity predictions	https://github.com/Merck/DeepNeuralNet-QSAR
DeepTox	Toxicity predictions	www.bioinf.jku.at/research/DeepTox
DeltaVina	A scoring function for rescoring protein–ligand binding affinity	https://github.com/chengwang88/deltavina
Hit Dexter	ML models for the prediction of molecules which might respond to biochemical assays	http://hitdexter2.zbh.uni-hamburg.de
Neural Graph Fingerprints	Property prediction of novel molecules	https://github.com/HIPS/neural-fingerprint
NNScore	Neural network-based scoring function for protein–ligand interactions	http://rocce-vm0.ucsd.edu/data/sw/hosted/nnscore/
ODDT	A comprehensive toolkit for use in chemoinformatics and molecular modeling	https://github.com/oddt/oddt
ORGANIC	An efficient molecular generation tool to create molecules with desired properties	https://github.com/aspuru-guzik-group/ORGANIC
PotentialNet	Ligand-binding affinity prediction based on a graph convolutional neural network (CNN)	https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507
PPB2	Polypharmacology prediction	http://ppb2.gdb.tools/
QML	A Python toolkit for quantum ML	www.qmlcode.org
REINVENT	Molecular *de novo* design using RNN (recurrent neural network) and RL (reinforcement learning)	https://github.com/MarcusOlivecrona/REINVENT
SCScore	A scoring function to evaluate the synthesis complexity of a molecule	https://github.com/connorcoley/scscore
SIEVE-Score	An improved method of structure-based virtual screening via interaction-energy-based learning	https://github.com/sekijima-lab/SIEVE-Score

- **Structure-based and ligand-based virtual screening**

VS is an efficient method to screen out promising therapeutic compounds from a pool of compounds. Thus, it becomes an important tool in high-throughput screening. In general, two important types of VS are structure-based VS (SBVS) and ligand-based VS (LBVS).



AI algorithms in ligand-based virtual screening (LBVS):

- SwissSimilarity (http://www.swiss similarity.ch/)
- METADOCK, Open-source plat form
- HybridSim-VS (http://www.rcidm.org/Hybri dSim-VS/)
- PyGOLD (http://www. agkoch.de/)
- BRUSELAS (http://bio-hpc.eu/softw are/Bruselas)
- RADER (http://rcidm.org/rader/)
- AutoDock Bias (http://autodockbias.wordp ress.com/)
- GCAC (http://ccbb.jnu.ac.in/gcac)
- …

For examples:
AI algorithms in ligand-based virtual screening (LBVS): aurora kinase A inhibitors, PI3Kα inhibitors, selective histone deacetylase 8 inhibitors, p-Hydroxy phenylpyruvate dioxygenase inhibitors, HIV inhibitors, and potent DNA methyltransferase inhibitors.

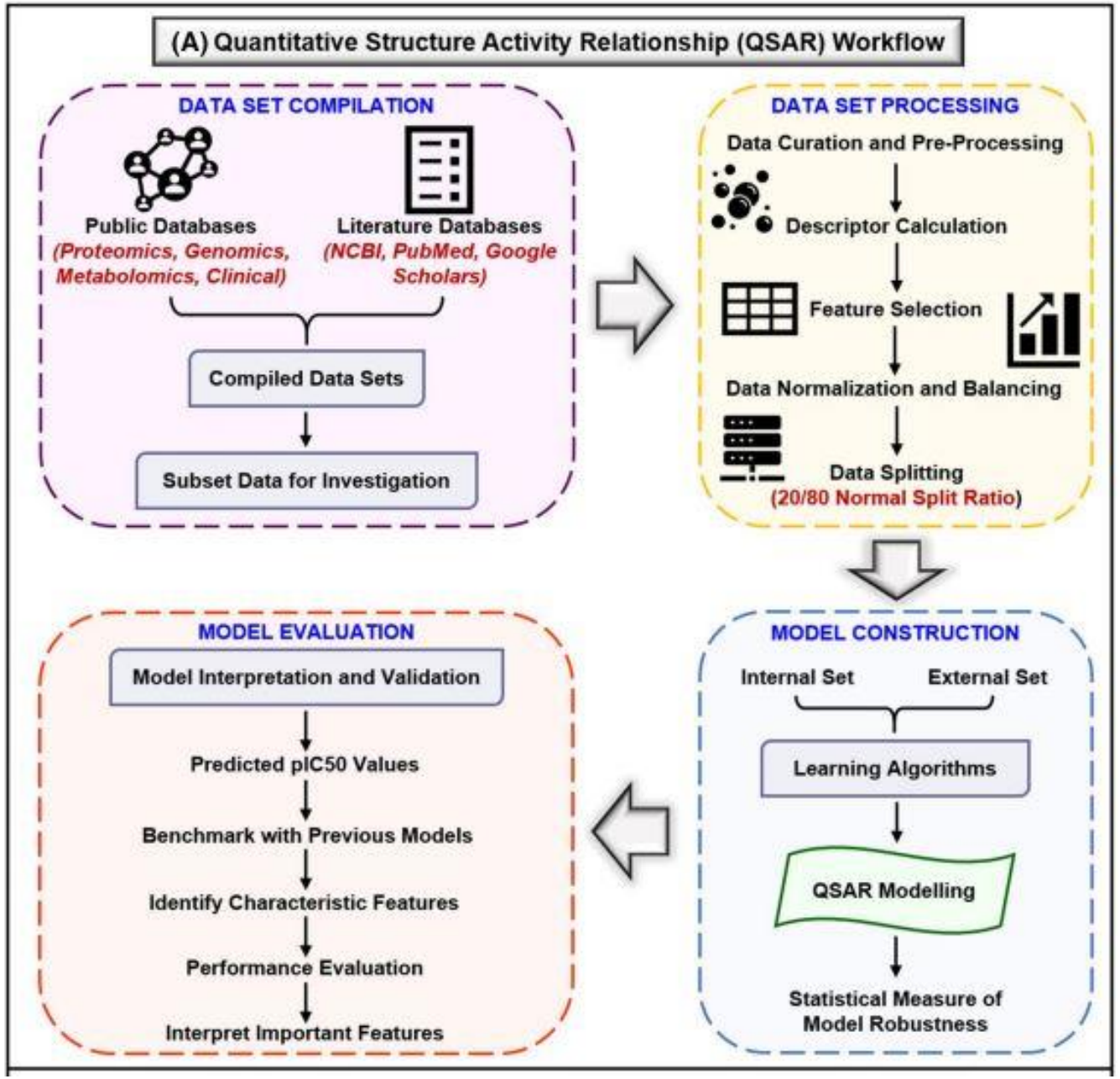
AI algorithms in Structure-based virtual screening (SBVS)

- MTiOpenScreen(http://bioserv.rpbs.univ-paris-diderot.fr/services/MTiOpenScreen/),
- CompScore (http://bioquimio.udla.edu. ec/compscore/)
- PlayMolecule BindScope (PlayMol ecule.org)
- GeauxDock (http://www.brylinski.org/ geauxdock)
- EasyVS (http://biosig.unimelb.edu.au/ easyvs)
- PL-PatchSurfer2 (http:// www.kiharalab.org/plps2/)
- SPOT-ligand 2 (http:// sparks-lab.org/)
- Gypsum-DL (https://durrantlab.pitt. edu/gypsum-dl/)
- …

For examples:
AI algorithms in Structure-based virtual screening (SBVS): glycogen synthase kinase 3 beta (GSK-3β) inhibitors, interleukin-1 receptor-associated kinase-1 inhibitors, identification of vascular endothelial growth factor receptor 2 potent compounds for the treatment of renal cell carcinoma, identification of multi-targeted inhibitors against breast cancer and discovery of Mdm2-p53 inhibitor

- **QSAR modeling**

In drug designing and discovery, it is crucial to develop the relationship between chemical structures and their physiochemical properties with biological activities. Thus, QSAR modeling is a computational approach through which quantitative mathematical models can be created between chemical structure and biological activities. The main advantage of developing a mathematical model is identifying the diverse chemical structure from molecular databases, which can be used as therapeutic compounds against a disease target. Once the most promising compound is selected, it is subjected to laboratory synthesis and in vitro or in vivo testing.



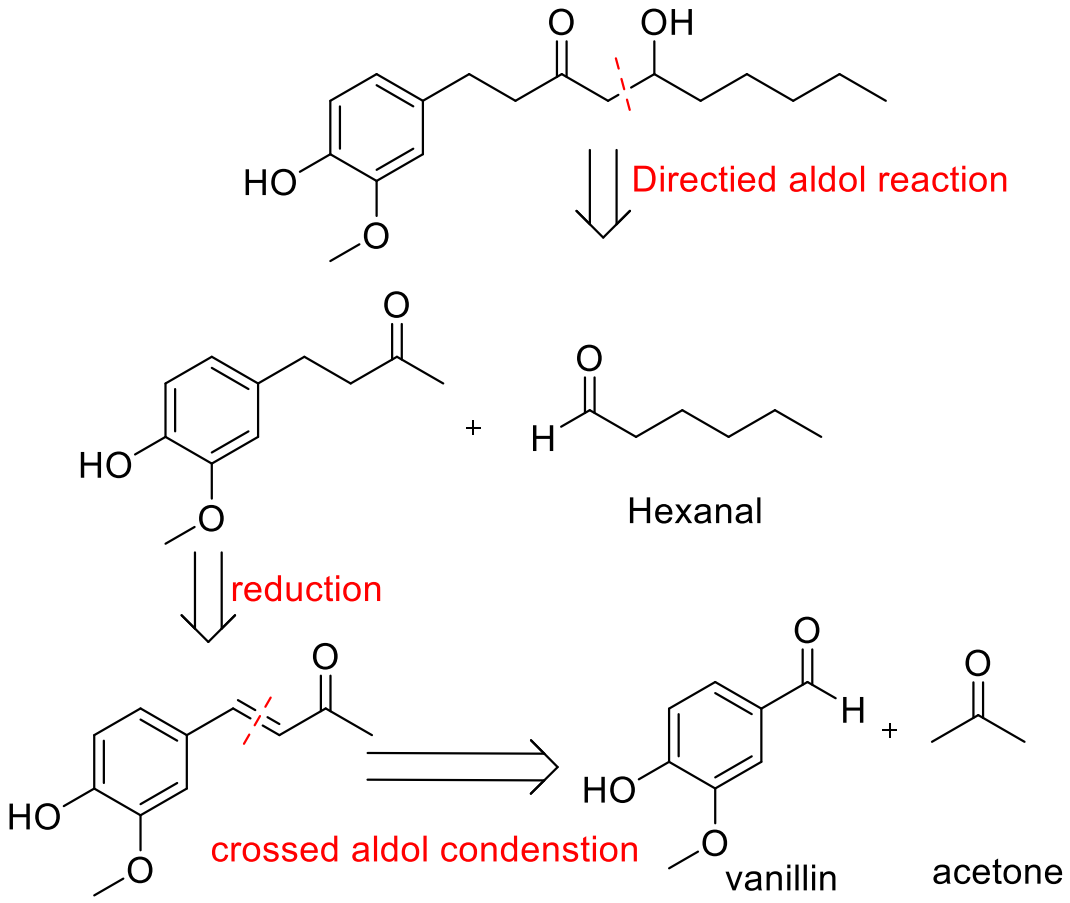
QSAR modeling such as VEGA platform (https:// www.vega-qsar.eu/), QSAR-Co (https://sites.google.com/view/qsar-co), FL-QSAR (https://github.com/ bm2-lab/FL-QSAR), Meta-QSAR (https://github.com/meta-QSAR/simple-tree) (https://github.com/meta QSAR/drug-target-descriptors), DPubChem (www. cbrc.kaust.edu.sa/dpubchem), Transformer-CNN (https://github.com/bigchem/transformer-cnn), Cloud 3D-QSAR (http://chemyang.ccnu.edu.cn/ccb/server/cloud 3dQSAR/), MoDeSuS and Chemception (https:// github.com/Abdulk084/Chemception), PyQSAR (https://github. com/crong-k/pyqsar_tutorial).

For examples:
The QSAR-Co tool: screening of ERK inhibitors as anti-cancer agents, and prediction of antifungal properties of phenolic compounds.

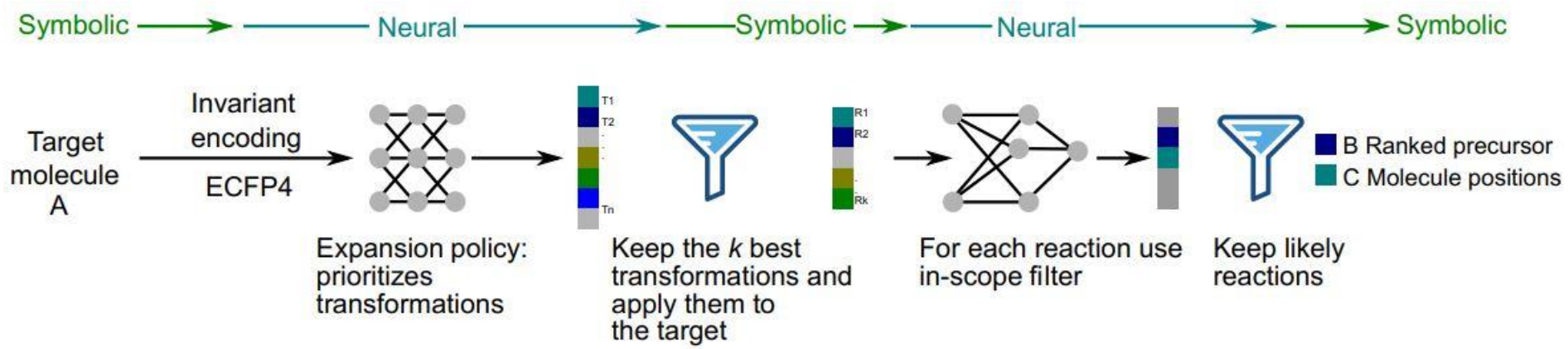
In 2020, A. S. Geofrey et al. the identification of potent drug candidates for novel coronavirus and the development of QSAR of quercetin using PyQSAR.

- Planning Chemical Synthesis with AI

Retrosynthesis is a technique used in organic chemistry for planning the synthesis of complex molecules. It involves breaking down a target molecule into simpler precursor molecules, working backward from the desired product to simpler starting materials. This reverse-engineering approach helps chemists design efficient and practical synthetic routes.



AI & Retrosynthesis Pathway Prediction :



ECFP4, extended-connectivity fingerprint

For example:

- Likewise, Segler et al. used the integration of three distinct neural networks in conjugation with the Monte Carlo tree search to discover novel retrosynthesis routes. ICSYNTH (https://www.deepmatter.io/products/ icsynth/) is another tool that can produce novel chemical synthesis pathways by using a collection of chemical rules which are generated via ML models.

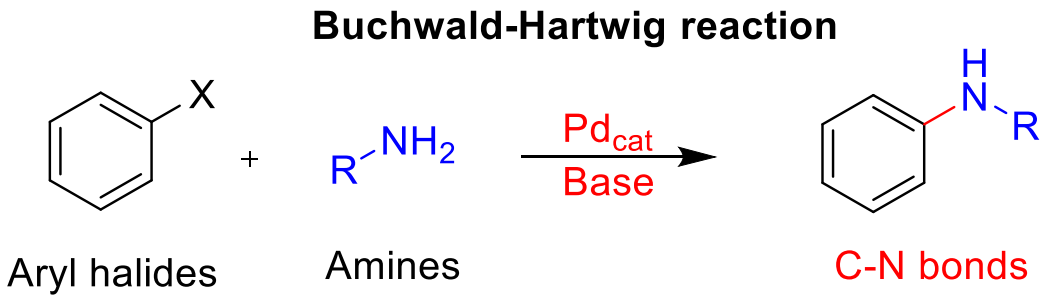
- AiZynthFinder (https://github.com/MolecularAI/aizynthfn der), an open-source tool for retrosynthesis planning built on Monte Carlo tree search-neural network.

- **Planning Chemical Synthesis with AI**

Reaction Yield Prediction and Insights into Reaction Mechanism

AI algorithms can not only design routes of synthesis but also can effectively predict the products and yields of organic reactions on the basis of the molecular properties of the reactants. In the past, predicting the outcome of complex chemical reactions has been a major challenge.

For example, Doyle and Dreher used machine learning to predict the yields of Buchwald-Hartwig coupling reactions. In this study, vibrational frequencies and dipole moments calculated by quantum chemistry were used as descriptors, and the final product yields were obtained from a given set of reactants through high-throughput experimental syntheses. A random forest approach was then used to investigate the relationship between the input descriptors and the product yield. When using a variety of reactants, the algorithm also predicted the yields of other expected products with promising accuracy.



- **Identification of Molecular Pathways and poly pharmacology**

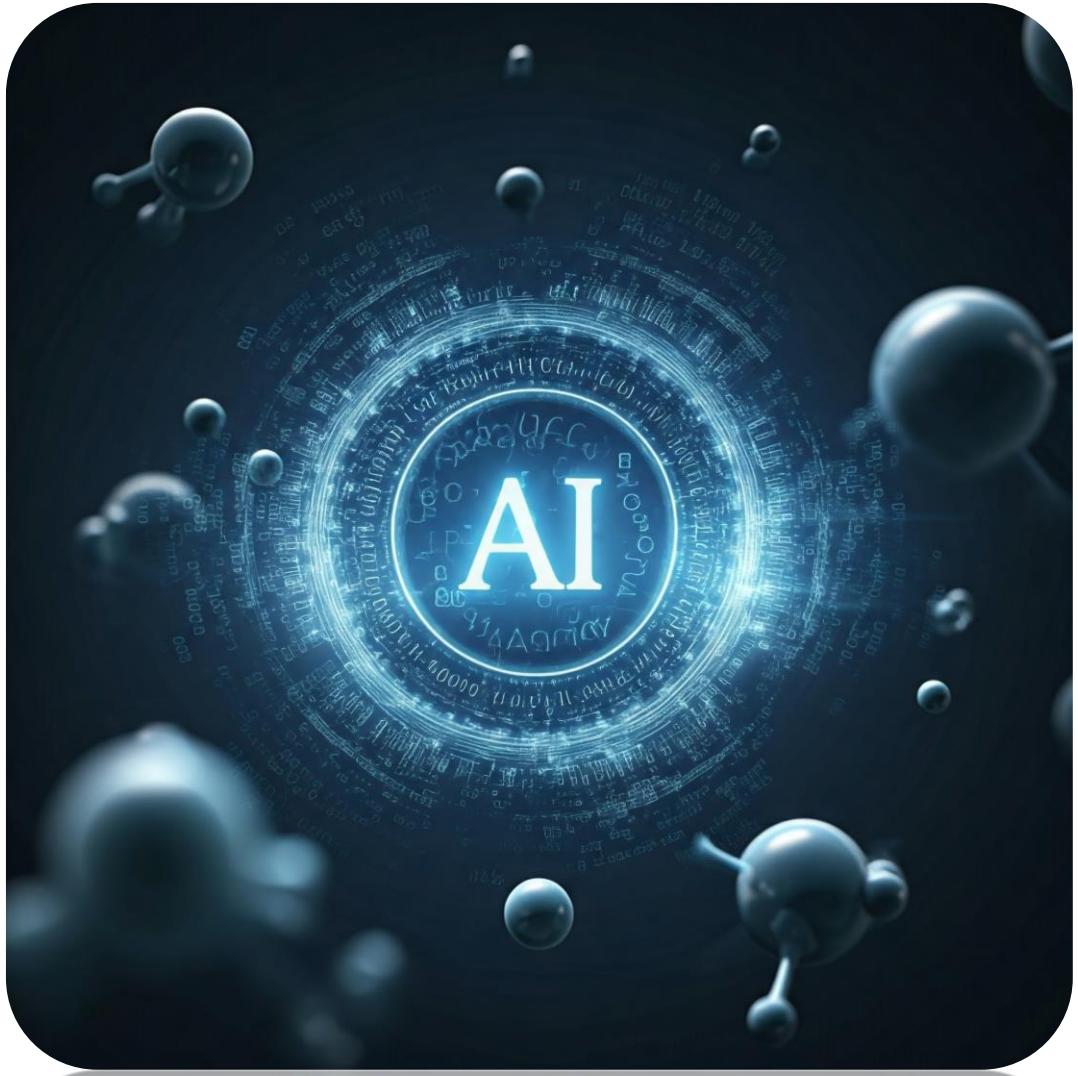
- One of the significant outcomes of AI and ML algorithms in drug discovery and development is the prediction and estimation of the overall topology and dynamics of disease networks drug-drug interaction or drug-target relationships. This methodology offers a vast avenue for the identification of novel molecular therapeutic targets for a particular disease. Text mining-driven databases like DisGeNET, STITCH, and STRING are widely used to ascertain gene-disease associations, drug-target associations, and molecular pathways, respectively.

poly pharmacology browser (PPB) (http://www.gdb.unibe.ch/), TarPred (http://www.dddc.ac.cn/tarpred/), Self-Organizing Map Based Prediction of Drug Equivalence Relationship (SPiDER) (http://modlabcadd.ethz.ch/software/spider), Targethunter (https://www.cbligand.org/TargetHunt er3D/), PharmMapper (http://lilab-ecust.cn/pharm mapper/), ChemMapper (http://lilab.ecust.edu.cn/chemmapper/), and Swiss Target Prediction (Swis sTargetPrediction) (http://www.swisstargetprediction.ch/).

For example:

PharmMapper was used to predict the multiple mechanisms of Hedyotisdifusa Willd on Colorectal Cancer.

- Peptide synthesis and small molecule design
- Identification of drug dosage and drug delivery effectiveness
- Predicting bioactive agents and monitoring of drug release
- Prediction of protein folding and protein-protein interactions
- …



Future challenges and possible solutions

Advantages:
Speed and Accuracy, Cost Reduction, Novel Drug Discovery

Challenges:

- Data Quality: The accuracy and reliability of AI models depend heavily on the quality of the input data, and inaccurate data can lead to incorrect predictions.
- Interpretability: Complex AI models can be difficult to interpret, making it challenging for scientists to understand the underlying processes and results.
- Integration with Traditional Methods: Integrating AI-based methods with traditional pharmaceutical chemistry approaches can be complex and time-consuming.

Concerns:

- Ethics and Privacy: The use of personal data in AI processes raises significant ethical and privacy concerns.
- Incorrect Predictions: AI models can produce false positives or false negatives, which could lead to undesirable outcomes.
- Dependence on Technology: Over-reliance on AI may diminish traditional scientific skills and knowledge, overshadowing human expertise.

References

- Advancing drug discovery via artificial intelligence, HCS Chan, H Shan, T Dahoun, H Vogel, S Yuan, Trends in pharmacological sciences, 2019
- AI for targeted polypharmacology: The next frontier in drug discovery. A Cichońska, B Ravikumar, R Rahman, Current Opinion in Structural Biology, 2024
- Artifcial intelligence to deep learning: machine intelligence approach for drug discovery, R Gupta, D Srivastava, M Sahu, S Tiwari, RK Ambasta, P Kumar, Molecular diversity, 2021
- Impact of Artificial Intelligence on Drug Development and Delivery, C Aundhia, G Parmar, C Talele, N Shah, D Talele, Current Topics in Medicinal Chemistry, 2024
- AI for targeted polypharmacology: The next frontier in drug discovery, A Cichońska, B Ravikumar, R Rahman, Current Opinion in Structural Biology, 2024