

Research Article

Harnessing AI for Natural Drug Discovery: Integrating Big Data and AI-based Algorithms Models in Advancing Drug Discovery and Drug Development

Rajneesh kumar¹, Riya Chaudhary², Kusum Parmar³,
Dr. Sunidhi⁴ and Dr. Shoket Ali^{5*}

Article Info

¹Assistant Professor, Department of Pharmacognosy, Sai School of Pharmaceutical Education and Research, Sri Sai University Palampur, India 176081. rajneeshkumarhp32@gmail.com

²Assistant Professor, Department of Pharmacology, Sai School of Pharmaceutical Education and Research, Sri Sai University Palampur, India 176081. riyachoudhary687@gmail.com

³Assistant Professor, Department of Pharmacognosy, Sai School of Pharmaceutical Education and Research, Sri Sai University Palampur, India 176081. parmarmimmi@gmail.com

⁴Assistant Professor, Department of Biochemistry, Sai School of Pharmaceutical Education and Research, Sri Sai University Palampur, India 176081. sunidhivaid1@gmail.com

⁵Corresponding Author:

Dr. Shoket Ali,

Associate Professor (HOD), Department of Pharmacology, Sai School of Pharmaceutical Education and Research, Sri Sai University Palampur, India 176081.

drshaukatpathan@gmail.com

Received: 09/09/2026

Accepted: 21/10/2026

Published: 26/10/2026

DOI: 10.5281/zenodo.23662063

Publisher's Note: IJABR Press stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: ©2026 by the author(s). Licensee IJABR Press, India. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution- Share Alike (CC BY -SA) license.

Abstract

Artificial intelligence (AI) is a branch of computer science that studies intelligent machine learning, including various advanced tools capable of mimicking human intelligence. AI plays an important role in the research field, especially in new drug discovery. AI can tackle many disciplines, including computer science, psychology, and philosophy. AI can work in various fields including science, technology and the pharmaceutical as well as natural science sector to improve efficacy and productivity. AI is now widely used in many industries and research fields because of the high-performance of computer hardware, ample storage, easily available software, and also be a cost effective. In recent years, the continuous advances in Machine learning (ML) algorithms have led to the development of Deep learning (DL) techniques with stronger capabilities in handling big data, which can further assist in identifying natural products and validation of computer-assisted lead identification. This review mainly summarized the rational application and advantages of AI based models to enhance the drug discovery process, including the Natural drug discovery and optimizing drug structures through ML/DL computational tools.

Keywords: - Artificial intelligence, big data and application of ML/DL based AI models, Natural products, future perspectives, and conclusion.

1. INTRODUCTION

Artificial intelligence (AI) is based on computer science that studies intelligent machine learning and computer programs and provides relative

results for the human process [1]. It is a system-based technology that includes various advanced tools and can mimic human intelligence [2]. AI is based on computers that perform complex tasks,

analyze large datasets, and interpret those using algorithms [3, 4]. It is used in various research fields and industries for easy to processing and reduced to human errors. Additionally, AI can reveal data structures that are difficult to obtain from other sources [4]. AI gained intention and is widely used by chemists to do numerous tasks in a drug development phase, to predict molecular properties, and to estimate the bioactivity of molecules [5]. Now in the rapid development of machine learning and the wide range of data stored in databases and archives in cloud computing, researchers can gain access to large amounts of data and integrate AI/ML approaches into their tasks. A branch of artificial intelligence called machine learning (ML) involves the gathering of complex data and mathematical formulas, which humans implement through algorithms to solve numerous types of problems [6]. ML algorithms can be easily and effectively implemented at very large scales in the cloud at fordable at a low cost. The World Wide Web internet is overflowing called the “data flood”, which is the digitization of various data types like text, images, audio, and biometrics from sensor devices and wearables into online public figures and proprietary [6, 7]. E-VI is an AI platform created by Euler’s that uses machine learning algorithms in a user-friendly interface to generate insights for the pharmaceutical industry. This platform creates analytical roadmaps that focus on rivals, key stakeholders, and market share, helping predict the

main factors influencing sales. Recently, AI has become a core component in the pharmaceutical industry, with applications ranging from drug discovery to patient care. These advancements enable more precise predictions, enhance decision-making, and streamline complex processes, making AI indispensable in both technical and scientific areas [8]. The lack of advanced technology slows down the medication development process, making it costly and time-consuming which can be solved with AI. AI can help to speed up the research, analyze the data faster, and predict outcomes more accurately [9]. AI can identify hit and lead compounds, validate the drug targets more rapidly, and improve drug structure design [1, 10]. For example, AI clustering techniques have helped in de novo molecular design, predicting which proteins will respond to certain drugs, and identifying different cell types in images. A standout example is the U.S. Company Insilco, which found novel medicine compounds using the AI platform GENTRL (Generative Tensorial Reinforcement Learning). In just 46 days, GENTRL discovered six kinase inhibitors that block the discoid receptor 1 and target lung fibrosis [11]. AI work in the field of drug discovery using various computational-based AI tools used in drug design, identification of pharmacological activates AI-based determination of drug target pathway, Identification of drug therapeutic and drug screening mentioned in Fig. 1.

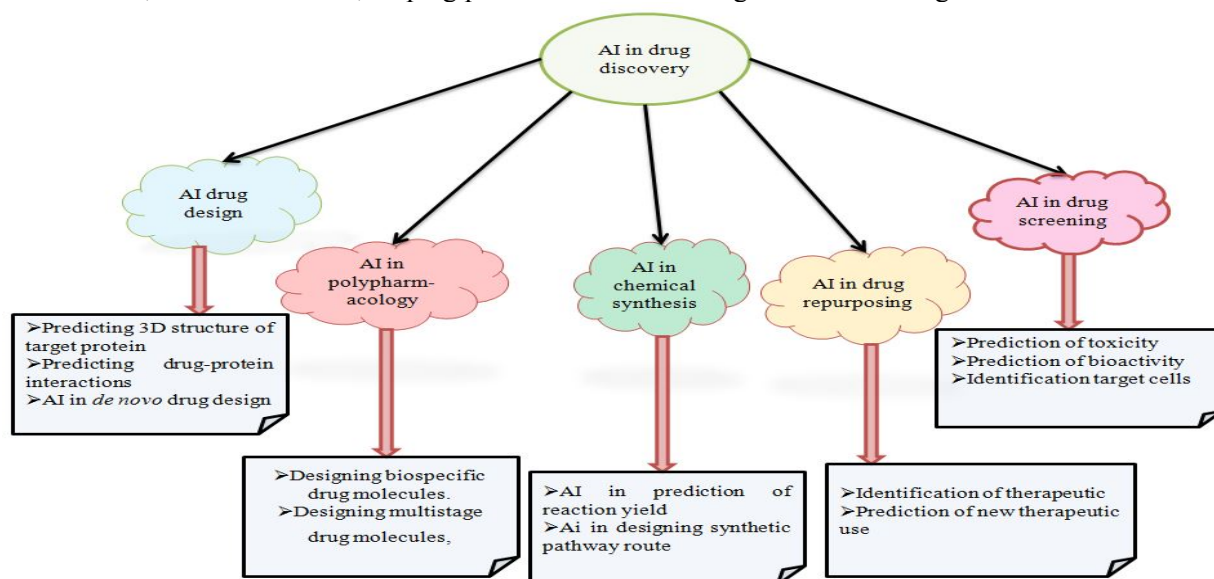


Figure 1: Computational-based AI is used in the drug discovery and identifying medicinal and pathological properties.

AI can effectively identify lead and hit compounds, and provide drug target validation and optimization of drug structure design, chemical synthesis, drug screening, poly pharmacology and drug repurposing. The use of AI in natural product research is a young one that could completely transform the process of finding new medications [12]. Natural products (NPs) that have been used in medicine, cosmetics, dyes, and, perfumes since ancient times. NPs made from natural occurring source such as plants, animals, microorganisms, and marine sources [13]. NPs are primarily understood to be unique structures that interact with therapeutically important for the targeting of protein. Their structural diversity and biological continue to encourage the synthesis of small molecules [14]. AI system can be trained to evaluate large volumes of data, spot trends, identify patterns and make predictions, which helps to speed up the discovery of novel substance with potential therapeutic activities [15]. AI is a better scope for the production of the highest quality of nutraceuticals & Microalgae which is considered as a potential source of protein, carbohydrate, fatty acids, vitamins, amino acids, lute in, antioxidants, other health beneficial product [16]. Over the years, chemo-informatics, bioinformatics, and other informatics have played a major role in drug discovery in natural products, contributing extensively [17, 18].

2. REVIEW METHODOLOGY

This narrative review was developed through targeted searches of PubMed/PubMed Central, ScienceDirect, Nature Portfolio and relevant publisher webpages, supplemented by reference-list checking of key review and primary research articles. Searches were conducted on 26 October 2026 and focused on literature published from January 2000 to October 2026. Search terms were combined using Boolean operators, including: ("artificial intelligence" OR "machine learning" OR "deep learning" OR "generative model") AND ("natural product" OR phytochemical OR "medicinal plant") AND ("drug discovery" OR bioactivity OR "virtual screening" OR metabolomic OR "genome mining"). Additional targeted searches combined terms such as "natural product" AND ("biosynthetic gene cluster" OR dereplication OR "mass spectrometry" OR NMR

OR extraction OR phenotyping). Searches were intended to identify relevant literature for a narrative synthesis and were not designed as an exhaustive systematic search. Publications were considered relevant when they described AI/ML/DL methods applied directly to natural-product discovery, phytochemical analysis, medicinal-plant identification or cultivation linked to phytochemical quality, bioactivity prediction, or natural-product-inspired drug design. General crop-monitoring studies without a clear connection to medicinal plants or natural-product production were not treated as central evidence. Duplicate records and publications lacking a clear connection to the review scope were not discussed. Relevance was judged qualitatively from the study objective, dataset/source description, model or experimental methods, validation strategy, reported outcomes and limitations. Where reported, model performance metrics and experimental confirmation were distinguished from computational predictions. No formal risk-of-bias tool, meta-analysis, or independent dual-reviewer screening protocol was applied; therefore, the manuscript is presented as a narrative review and not as a systematic review.

2.1 BIG DATA AND ARTIFICIAL INTELLIGENCE FOR DRUG DISCOVERY

The term ‘Big Data’ refers to the huge amount of organized and unstructured data that comes from various sources like social media, transactional systems and different devices. Big data as a process used to store, manage and analyze large, complex data in a sets [19]. The healthcare industry is moving towards precision medicine due to the increasing role of big data and biomedical research. Artificial intelligence is also analyze large biological data seats, leading to the development of high-throughput sequencing strategies for precision medicine [20]. The trend of modern medical devices and information technology led to rapidly increase in the amount of data available so the protection of information is required. Know recent advancement of data storage used “Big data technology” [21]. The development of RNA-sequencing, high-throughput sequencing (HTS) and microarray technologies generated enormous amounts of biomedical data. The two tightly integrated databases are makeup Array Express repository,

which area query-optimized Array Express data warehouse and the archive-style. In 2005, Array express contains gene expression and other microarray data almost 35,000 hybridizations, comprising over 1200 studies, covering 70 different species [22]. Furthermore, the development of AI has much easier to analyze large amounts of data because of a various machine learning algorithm. For example, the cancer Genome Atlas (TCGA) is one of the largest cancer research projects, gathering and sharing genetic information on over 11,000 people with more than 30 different cancer types. The TCGA program provides a data on gene sequences, expression levels, chemical changes, and gene copy variations, which help for the scientists worldwide better understand cancer and developed treatments [23]. Gene expression is a key feature that is often used to identify the target genes, understand disease mechanisms and detect the gene which is responsible for the disease. Gene expression data for many diseases have been collected large amounts using microarray and

RNA-seq methods. This data is stored and shared through the NCBI Gene Expression omnibus (GEO) platform [24]. Drug development is a very expensive and long process; in contrast, medications that have been repositioned or repurposed may be introduced into clinical practice more quickly and for a lot less money. The strategies use data from Genome-Wide Association Studies (GWAS), which are created for medication repositioning and also useful to the determination of complex disorders by genomic variants [25]. GWAS technology enable to effective screening of molecular markers that significantly influence the phenotype [26]. Han *et al.*, (2019) created a supervised machine learning tool, which can identify driver genes associated with cancer and conducted, objective benchmark analysis and Driver ML comparisons with 20 other tools in use across 31 different datasets from (TCGA) [27]. Drug discovery using big data and artificial intelligence involves various steps as shown in Fig. 3.

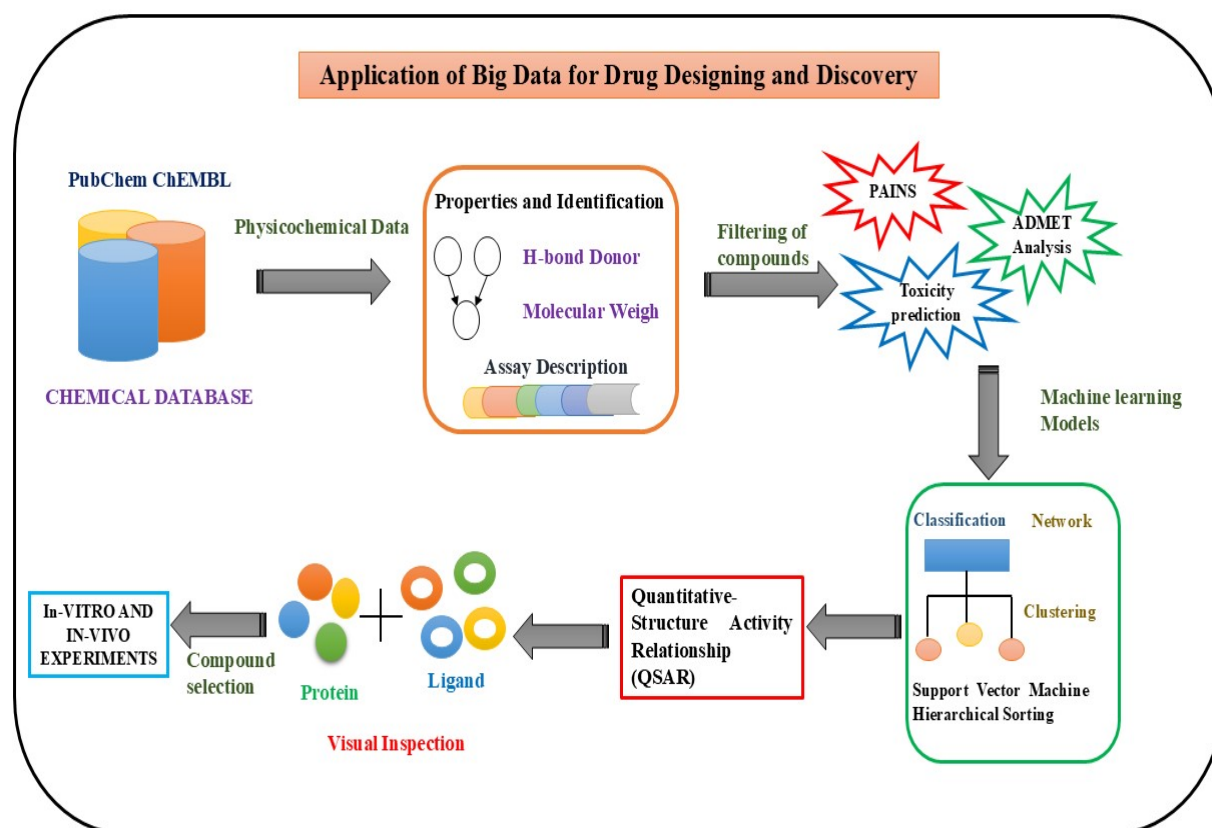


Figure 2: Big data useful for drug discovery and drug design. Many data repositories have been established due to the identification and storage of various publications such as biological and chemical research data, preclinical, clinical experimental data, metabolomics, gene ontology, and pathway data.

For example, Chem Spider, ChEM BL and PubChem databases are used for the synthesis and screening of compounds. In place of quantum mechanical computations, information present in the previously mentioned databases was selected and the information was filtered based on Pharmacokinetics, Pharmacodynamics and physiochemical characteristics. Along with toxicity analysis collected the high-throughput screen data performed the ADMET analysis and PAIN calculations. Additional study, the filtered compounds were analyzed using artificial intelligence models, molecular dynamics simulations, and quantitative models. The final compounds were identified and validated through *in vitro* and *in vivo* experiments.

Protein Data Bank (PDB), those data are often used to study protein-ligand interactions, helping to find out suitable inhibitors for target proteins. For COVID-19 research, PDB data support molecular docking and ML methods to identify inhibitors of the 3CL protease, a crucial protein in the virus. The crystal structure of the protease, which provides a detailed 3D model, was published by Xu et al. (2020) [28]. In drug design and discovery, making connections between chemical structures and their biological activities and understand their physical and chemical characteristics is essential for effective drug formulation and development. Various ML techniques introduced into QSAR modeling enhance the outlier prediction of models, such as the VEGA platform [29-31]. Zebra fish AC50 modeling is the type of QSAR model that determined the embryo toxicity in zebra fish [32]. The QSAR-Co model is an open-source software designed to develop robust classification models for multitasking or multi-target applications [33]. The FL-QSAR model is a federated learning based QSAR prototype, therefore encouraging the growth of cooperative, private-preserving drug discovery that can extend its ability to other privacy-related biomedical domains [34].

Transformer-CNN QSAR model [35], Cloud 3D-QSAR model has been developed to virtually evaluate the activity of multiple compounds, helping to reduce the time and labor costs associated with chemical synthesis and experimental analysis [36]. QSAR models based on AI/ML algorithms are able to overcome the constraints imposed by previous methods. Hologram-based QSAR (HQSAR) and group-based QSAR (G-QSAR) an QSAR model with AI/ML have accelerated the drug discovery process [37].

Nano-QSAR model for predicting the cytotoxicity of metal-based and metal-oxide nanoparticle (MBNPs and MONPs) [38]. Monte Carlo based global QSAR (CORAL) modeling of Bruton tyrosine kinas' inhibitors using hybrid descriptors; the Monte Carlo algorithm of CORAL software using the DCW hybrid descriptors derived from molecular SMILES notation [39]. The modeling of QSAR is a popular computational method with many uses for wide verity of applications.

However, a significant of conventional QSAR modeling is that models is constructed using a small number of theoretical and experimental conditions. To overcome this, new computational techniques known as multitasking or multi target QSAR (mt-QSAR) approaches have surfaced. These tools can combine multiple data points into a single model of biological and chemical data, enhancing and increasing the validity of this type of modeling.

The MT-QSAR differential model was developed using a large and heterogeneous database of compounds. Halder *et al.*,(2021) has been developed a QSAR-Co-X, an open source toolkit that supports mt-QSAR modeling using the Box-Jenkins moving average approach [40]. Various AI-generated tools and models are shown in Table 1.

Table 1: - Use of AI tools and algorithms, such as DL and ML approaches, in drug discovery

Tools and software's	Explanation	Method	Application	References
<u>QSAR model</u>				
Cloud 3D-QSAR	An online tool designed to create quantitative structure-activity relationship models for drug discovery.	Machine-learning	Design molecular structures, arrange molecules, calculate molecular interaction field (MIFs), and analyze results to deliver comprehensive solutions.	[41]
CORAL	The Monte Carlo Method acts as a mechanism for formulating predictive Quantitative Structure-property/Activity relationships (QSPR/QSAR).	Molecular descriptors	Examine the harmful effects of dioxin and related compounds to assist in choosing effective treatments for neurodegenerative diseases.	[42-44]
Chem-Grapher	Optical Character Recognition-Device of chemical compounds.	Machine learning and Deep learning	It is used to automatically analyze images and convert them into chemical graph structures, which can be beneficial for drug discovery applications.	[45]
YZS-model	Predictive model for organic Drug solubility.	Machine Learning, Deep learning	Predict the drug molecule solubility based on graph conventional network.	[46]
Chem-MORT	Automatic ADMET optimizer.	Deep learning	They are used to optimize unfavorable ADMT characteristics.	[47]
Chem SAR	An online pipelining platform for molecular SAR modeling.	RDkit or Chemopy package, scikit-learn package	Producing of SAR classification models for chemo informatics and other biomedical applications.	[48]
<u>Drug repurposing</u>				
DrugNet	Prioritizing drug diseases using a network and combining diverse data.	Machine learning, Deep learning	Important insights into the interactions between drugs, targets, and them, not only within specific drug-target pairs but also through the integration of disease-related data.	[49]
RLFDDA	A meta-path-based graph representation learning model.	Machine learning, meta-path-based learning model	Widely used to predict drug-disease associations.	[50]
wSDTNBI	A novel network-based inference method for virtual screening.	Network based method,	Prediction and validation of Rory ligands, Evaluation of network-based models.	[51]
Binding MOAD	Mother of all databases set a focused on high-quality x-ray crystal structures.	Binding affinity data	It provides a set of high-quality X-ray crystal structures for identifying ligands essential to biological functions.	[52]
SPIDR	Repurposing drugs: the role of small molecule peptides.	Heuristic search method and genetic algorithms	DockoMatic v2.1 now includes a generic and integrated version of SPIDER, during mammalian meiosis, homologous recombination requires SPIDER	[53, 54]
AI-BIND	A pipeline that combines network-based sampling	Deep learning	Improving the generalizability of protein-ligand binding predictions	[55]
EDC-DTI	It is a comprehensive, deep collaborative learning framework.	Deep learning	Propose a novel method for predicting drug-target interactions.	[56]

Tools and software's	Explanation	Method	Application	References
CeDR Atlas	A knowledgebase of cellular drug response	Machine learning, data collection	CeDR provides insights into drug resistance, Pharmacological adverse effects, and the formulation of combinatorial therapies by fine-tuning drug response at the cellular level.	[57]
AMDGT	Attention aware multimodal fusion using a dual graph transformer for drug-disease associations prediction	Molecular docking, drug repositioning	Multi-modal network is constructed for fusing and learning representations, Determine the relationship between drugs and diseases.	[58]
GCNGAT	Drug-disease association prediction based on graph convolution neural network and graph attention network	Neural network,	Deep feature to determination of drug-disease relationship	[59]
HeTDR	Drug repositioning based on heterogonous networks and text mining	Deep learning	Estimation of the ratings of association between drug and diseases	[60]
MiRAGE	Mining relationships for advanced generative evolutionary drug repositioning	Heterogeneous network	Capture the complex relationship between drugs and diseases.	[61]
Qualities of physical-chemistry and bioactivity models				
CSM-Potential2	Deep learning platform for the analysis of protein interacting interfaces.	Deep learning	Prediction of interaction with nucleic acids at the residue level and allows for ligand transplantation based on sequence and structure similarity to experimentally determined structures, also predicts binding sites for protein-protein interactions and classes of biological.	[62]
FP-ADMET	A compendium of fingerprint-based ADMET prediction models.	Fingerprint based, Machine learning	Predict the properties absorption, distribution, metabolism, extraction, and toxicity of drug.	[63]
CSM-AB	Graph-based methods for predicting the binding affinity of antibodies to antigens, along with a scoring function for docking assessments.	Data collection, machine learning	Approaches are essential for creating novel treatments; computational methods can quickly evaluate binding landscapes and inform experimental design.	[64]
GB-score	A minimally designed ML scoring function that relies on distance-weighted interatomic contact features.	Machine learning	Estimating the binding affinity of a generated protein-ligand poses from docking or a crystal structure using ML and DL techniques.	[65]
TB-IECS	Theory-based interaction energy component score.	Machine learning	Used as a coring function for virtual screening.	[66]
Additional AI tools and models utilized in drug discovery				
Space Grow	Optimized size-oriented virtual screening of combinatorial fragment libraries containing billions of compounds.	Molecular docking, virtual screening	A novel approach for 3D method that uses a single CPU too quickly and screen of billions of chemical compounds for virtually using ligands.	[67]
Reinvent4	It is a molecule design command line tool that can read user configuration in TOML or JSON	Machine learning, transfer learning, reinforcement	Used for the making of <i>de novo</i> design, scaffold hopping, R-group replacement, library design, linker design, and	[68]

Tools and software's	Explanation	Method	Application	References
	formats.	learning	molecule optimization easier.	
MDeePred	Novel multi-channel protein factorization.	Deep learning	Technique to be applied and used in computational drug repositioning and discovery, and also predict the <i>in vitro</i> comparative analysis.	[69]
GPCNDDTA	Using cross-attention networks enhanced with graph features and pharmacophore information, this tool forecasts the binding affinity between drugs and targets.	Deep learning	To predict the drug-target binding affinity, constructing cross attention inter and intra molecules	[70]
Affinity2Vesc	Prediction of drug-target binding affinity.	Representation learning, graph mining, and machine learning	A novel approach based on regression that formulates the whole work as a heterogeneous graph that incorporates the information regarding multiple sources, like binding affinities between drugs and targets as well as drug-drug, target-target similarity, and drug-target affinities.	[71]
CNN model	Excellent predictive ability in the computer vision domain.	Deep learning	Identifying and learning complex patterns from molecular data, non- liner transformation can be used to effectively learn the underlying patterns in raw data by improving its representation.	[72]

2.2 UTILIZATION OF AI TECHNOLOGIES IN NATURAL PRODUCTS

Natural products have been used for many centuries, and in recent times they have been used to treat many diseases. For more than half a century, scientists have been able to extract natural compound from plants, microorganisms, or animals that useful for treating the disease. As the contribution of scientists in the work related to natural science is increasing day by day, the amount of data related to natural products is also increasing day by day. In recent years, due to the rapid development of artificial intelligence, it has brought many new opportunities in the discovery of natural products. Artificial intelligence utilized the many advanced technologies, such as deep learning, machine learning and image processing, and has been proven useful in solving challenges in natural product drug discovery and managing the huge data [73, 74]. Development of Chemo informatics, bioinformatics, and other informatics-related fields played major contributions to NPs-based drug development over the years. Their limitation and successful uses have been examined recently [17, 18]. Computational photochemistry is growing field that uses computational tools such as AI, big database and mathematical modeling to solve phytochemical research issues. It has been applied in screening plant materials, metablomics, chemical fingerprinting, biosynthesis, and prediction of pharmacological properties. This technology saves time and money in phytochemical research [75]. Advances in computational omics and artificial intelligence approaches such as machine learning technologies which have been used in the computational *de novo* drug design for molecular targets of interest and predict the biological activity of the natural products, omics technologies developed by revealing the variety of natural products [76]. Other hand, Plant-based natural products and associated microorganisms play a major role in the fight against human cancer disease. For example, Taxol is an anticancer drug that is derived from plants. Major obstacles may prevent identification and application of anticancer lead compounds. Now, new developments in sequence and omics-related technologies have provided a wealth of data for the creation of novel natural product-based medicines. Artificial intelligence and machine learning have provided powerful methods for comprehensive data analysis and predict the

anticancer lead molecule [77]. The risk of Methicillin-resistant *Staphylococcus aureus* (MRSA) infection is continuing to increase globally, requiring new strategies. Dias *et al.*, (2018) introduced a ligand-based strategy that included two quantitative structure-activity relationship (QSAR) studies, utilizing R^2 and RSME models alongside ML techniques. The highest-performing model reached an R^2 of 0.68 and an RSME of 0.59 for the dataset. Additionally, they proposed a novel method for natural product drug discovery that leverages 1D NMR descriptors to evaluate the accuracy of both training and test models [78].

ML/DL algorithm-based NPs drug discovery

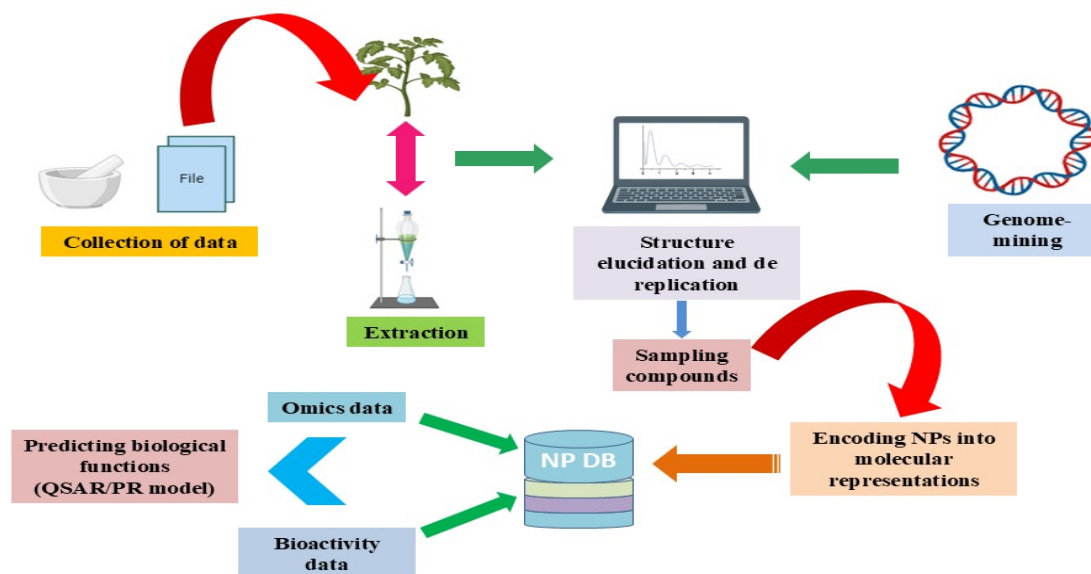


Figure 3: ML/DL based AI algorithm that are used at various phases for the discovery of new natural active compounds. The pipeline is divided into two sections: computer-assisted NPs discovery (genome mining, structure elucidation, data-mining into traditional medicine) and machine learning/deep learning algorithms applied to NPs (encoding omics and bioactive data) for biological function prediction with QSAR/PR model.

2.3 AI tools used in natural products

The research examined the current landscape of intelligent manufacturing in the Traditional Chinese Medicine (TCM) sector to find the ways that generative AI can help improve process and supply chains. It found four main areas where it can be used: a helpful information system for medicines, better training for workers, automated checks for quality in production, and an intelligent supply chain [79]. For TCM, the integration of AI-based research aims to deepen the understanding of TCM's overall material base while adhering to evidence-based principles. The system Function Decoding Model (SFDM) used to identify, quantify, predict, and verify the physical basis of TCM [80]. Botanical and natural drugs show significant potential and are used to treat and prevent complex diseases. Therefore, quality control of the natural product is essential to ensure the safety, purity, potency, and therapeutic stability of drug products. Many methods have been developed to ensure the quality of natural products. But sometimes the methods cannot show sufficient data for the natural product. At the end, artificial intelligence algorithm based models are developing name as conventional neural network (CNN) models and dual-scale anomaly (DSA) model, CNN models are used to determine the phytochemicals and bioactivity-based spectra and DSA is proposed for visual particle inspection based on images [81]. AI-driven drug discovery strategies inspired by natural products can serve as creative tools for the identification and discovery of lead compounds and molecular design. Various models of machine learning produced many kind of yield readily synthesizable

mimetic of the templates of natural goods [82]. During the COVID-19 pandemic, Ayurveda treatments provide personalized herbal remedies for enhancing immunity. Botanists discovered immunity-boosting plants for Ayurvedic medicine by observing their leaf characteristics. Recognizing these plants by their leaves alone can be difficult for an average person. In the medicinal plant analysis, the leaves show significant similarity. A deep learning-based mobile app was developed for generating numerical leaf shape descriptors using the SDAMPI algorithm. The app achieved a 96% accuracy rate with 64×64 pixel images [83]. The computer based neural network model generates synthetic structural copies of natural products. The successful synthesis and *in-vitro* testing with the help of computer-generated molecules confirmed by their functionality towards the retinoid X receptor. They strongly support the use of generative neural networks for drug discovery inspired by natural products, uncovering both opportunities and certain limitations of the current approach, and they also suggest possible future advancements and development [84-90].

2.4 Artificial neural network (ANN) regression models in natural drug discovery

A bioprocess medium was formulated using response surface methodology (RSM) and a combination of artificial neural networks and ant colony optimization (ANN-ACO) to enhance the production of neural protease from *Bacillus cereus* during submerged fermentation. The ANN-ACO model shows the excellent predictive ability and low error rate. Moreover, the biomolecule successfully clarified fruit juice and indicated promising applications in the food industry [91]. Ray *et al.*, (2020) developed an ANN model to predict the most suitable region for cultivation *Hedychium coronarium* plant with high coronarin D content. The model was based on 50 rhizomes from 50 distinct places throughout five Indian states. The results showed that the ANN model correctly predict the coronarin D content, with estimate the 0.891 correlation coefficient and a 0.06 root mean square error. The model is crucial for predicting the optimum coronarin D yield in *H. coronarium* [92].

2.5 AI tools approach for biological drug discovery of natural products

A new deep learning approach known as e-DeepBGC has been created to identify biosynthetic gene clusters (BGCs) in bacterial genomes. This method increases sensitivity and reduces the rate of false positives compared to its predecessor, DeepBGC [93]. NaPDoS (Natural Product Domain Seeker) is an online tool that identifies and classifies keto synthase and condensation domains from various DNA sources, including amplicon sequences, metagenomics data, and genomic data. It detects polyketide synthase (PKS) and non-ribosomal peptide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) genes using a classification method based on their phylogenetic relationships. This tool is particularly valuable for studying incomplete biosynthetic genes or gene clusters. The latest version, NaPDoS2 has added 1,417 new keto synthase sequences and now covers 41 different class and subclass categories [94]. An online platform called Global Natural Product social; molecular networking (GNPS) focused for tandem mass spectrometry data curation and analysis with a small molecule. It offers chemical insight into untargeted MS² datasets and links into biological questions. GnNPS-MassIVE is a public repository for untargeted MS² data, with metadata and annotated spectra. Molecular networking is an important analytical tool that creates a structured data table to reflecting molecular diversity. The platform can be shared or replicated disseminate knowledge, uncover new chemicals, metabolic pathways, and ecosystem [95]. Natural product (NPs) is small, bioactive molecules that have potential uses in industries and pharmacology. Still, there is no online resource that complies all known NPs in one place. This manuscript presents the online version of the collection of Open Natural Products (COCONUT), an aggregated dataset with an accessible web interface for easy browsing and downloading [96]. LOTUS is an online platform initiative that aims to improve knowledge management and analysis in natural products research by consolidating and sharing referenced structure-organism pairings. Over 750,000 referenced structured-organism pairs are to be harmonized, curate, validated, and publicly disseminated as part of this endeavor,

which is maintained on Wiki data. As a result, data access and interoperability will increase, and enable community curation and changing publication formats will be made possible.

The integration of LOTUS data into the Wiki data knowledge network will facilitate for the discovery and create of novel biological and chemical relationships [97]. A deep learning program called NP Classifier is designed to automatically classify natural products (NPs) based on their counted Morgan fingerprints. It improves discovery and improves the analysis of a large number of NPs, particularly when evaluating a big number of NPs, by connecting NP structures to their underlying attributes [98]. Artificial intelligence (AI) was used to predict pharmacological inhibitors of EGLN1, a protein that prevents HIF-1 α proteosomal destruction. A custom-based virtual screening was conducted using Deep Conventional Neural Network (DNN) architecture, docking, and molecular dynamics simulation. The findings support the use of deep learning architecture for identifying EGLN1 inhibitors, demonstrating its potential to accelerate drug discovery to accelerate drug discovery and development [99].

Table2: -Application of other AI tools includes deep learning and machine learning approach for drug discovery of natural products:

Tool and Software	Description	Method	Application	References
DeepBGC	Deep learning-based method focuses on identifying bacteria-derived small molecules.	Deep learning	Detecting the biosynthetic gene clusters and their biosynthetic class in bacterial genomes.	[93]
NaPDos2	The natural product domain seeker version 2.	Deep learning, database expansion web tool	The tool used for analyze amplicon, metagenomic and genomic sequencing data to detect and classify ketosynthase (KS) and condensation domains.	[94]
GnNPS	It is a public repository platform focusing tandem mass spectrometry data analysis.	Molecular networking, data deep analysis, learning	Builds a structured data table-focused tandem mass spectrometry (MS ²) data creation and analysis infrastructure.	[95]
COCONUT	Collection of open natural products database.	Database study, Machine learning	It has been observed multiplication of generalistic and thematic Natural product databases	[96]
LOTUS	Initiative supports open knowledge sharing in natural products research.	Databases analysis, Machine learning	Initiative to the harmonization, curation, validation and open dissemination of natural products.	[97]
NP Classifier	A deep neural network system developed for the structural categorization of natural products.	Deep learning, neural network based	This tool enhances the speed of natural product discovery by associating structural features with their properties. It also automates the classification of natural products using their computed Morgan fingerprints.	[100]

2.6 De Novo design and discovery of natural compounds: Natural products are valuable sites for drug discovery due to their large chemical structure, unique chemical structures and diverse functional groups. These are used in *de novo* and using generic models to inspire pseudo-natural products [101]. The first *de novo* assembly of the pink *Auricularia cornea*, commonly called the genome wood ear mushroom or jelly mushroom, has been achieved. This species, valued as a top edible and medicinal mushroom, will be commercially cultivated starting in 2022. However, research has encountered obstacles due

to the scarcity of genomic data, which has hindered gene discovery and product development. In 2024, Ye Leie and colleagues published a *de novo* assembly of the pink *A. cornea* Fen-A1 genome, featuring detailed functional annotations. The genome measures the size of 73.17 Mb and is contain the predicted gene 19,120, with a BUSCO completeness score of 92.60%. the study identified 789 CA Zyme gene linked to postharvest quality degradation, along with 3 related to antibiotics and 36 that are light-dependent gene [102]. Icarilin is a potent anticancer agent used in nutraceuticals and medicine. However, its structural complexity and

low abundance in plants makes it unstable in chemical synthesis and extraction from plants [103]. A bicycle scaffold inspired by sordarin, a natural antifungal, was synthesized and fused with the fungal protein eukaryotic elongation factor 2 (eEF2). The scaffold constructed using functionalized chiral cyclopentadiene showed high affinity for eEF2, offering a new avenue of medicinal chemistry drive [104]. Limonoids are a class of terpenoids isolated from pyridine-containing alkaloids, some of which have PTP1B-inhibitory activities. Access to several strategies for limonoid alkaloids using *de novo* pyridine synthesis has been developed. Laboratory routes were developed to access these compounds, including Pd-catalyzed transformation and computational investigation of an important pyran formation [105], [106].

Target sampling of natural product space (TSNaP) is designed to discover and evaluate sections of chemical space that are related to polyketide-like macrolides (pMLs). This technique prioritizes specific compounds for targeted synthesis and includes the computational design of pMLs containing tetrahydrofuran. Its effectiveness in drug discovery is underscored by successful validation from whole-cell biological testing, which has shown hit rates that exceed those usually found in small molecule libraries [107]. Using two generative algorithms, Merkus *et al.*, (2012) have successfully designed DOGS (Design of actual structures) software which is a computational tool that can be used to characterize drug-like molecules using a ligand-based approach [108]. DOGS to produce and evaluate NP mimetic from three recognized RXR (retinoid X receptor) agonists: valerenic acid, isopimeric acid, dehydroabietic acids were validated as active *in vitro* and acted as templates for the computational *de novo* design of RXR [109]. A virtual screening protocol was constructed using machine learning models to identify mimics of natural products (-)-glutamine. This automated approach identifies on (-)-glutamine targets, two of which are suitable

for the treatment of Alzheimer's disease. The results suggest multi target drug design using advanced virtual screening protocols [110].

Relevance to natural-product research: AI-based plant phenotyping can contribute to natural-product research by supporting the identification, authentication, cultivation, and quality assessment of medicinal plants. These applications may help monitor plant health and cultivation conditions that influence the availability and consistency of bioactive phytochemicals. However, the relevance of plant-phenotyping methods to natural-product discovery depends on establishing a clear link between the measured plant traits and phytochemical composition, bioactivity, or product quality. General crop-monitoring studies should therefore be interpreted as indirect applications unless their contribution to medicinal-plant research or natural-product production is demonstrated.

2.7 Predicting plant phenotype using Deep Learning and Machine Learning models

Genomics and proteomics are revolutionizing the drug discovery. Genomics is the study of entire genome of an organism, including the DNA sequence and gene expression patterns. Proteomics is the study of proteins in an organism. They help identify new drug targets, which are molecular basis of disease helps scientists develop drugs and improve safety and efficacy [111]. Advances in plant genomics have given rise to high throughput techniques for the identification of molecular phenotypes at low cost. AI based Deep learning has proven effective in predicting and explaining these phenotypes [112]. Applying machine learning techniques to image-based plant phenotyping has proven helpful, as non-destructive approaches to assess plant responses to biotic and abiotic stress have developed quickly [113, 114] and used to the estimation of photosynthetic capacity [115]. Advances in image technology, artificial intelligence algorithms, and data analysis techniques have enabled detailed exploration of genotype, phenotype, and

environment [116]. Plant-phenotyping studies illustrate a potential enabling role for AI, but their relevance to this review depends on a demonstrated link to medicinal plants or natural-product production. Teshome et al. [117] described UAV-based machine learning for crop phenotyping and monitoring, while Elmanawy et al. [118] reported hyperspectral image-processing software for plant-stress assessment. Such methods may help monitor cultivation conditions or plant health; however, general crop-yield or drought-stress prediction is not direct evidence of improved natural-product discovery. These studies should therefore be interpreted as indirect applications and retained only to the extent that they inform medicinal-plant selection, cultivation, authentication or phytochemical consistency. The same distinction applies to spectral disease-detection methods [119]: plant-health classification alone does not establish the identity, abundance or bioactivity of specific phytochemicals.

3. FUTURE PERSPECTIVES

Integration of AI into natural product plays a key role in transforming the discovery and development of bioactive compounds from plants. Advanced machine learning algorithms and predictive modeling will accelerate the identification of promising plant-based compounds, along with AI-powered robotics to automate extraction and synthesis processes. The ability to create herbal formulations through AI will be beneficial to maintain the quality, standard and reduced the risk factors. Additionally, the ability of AI to integrate diverse datasets from genomics and metabolomics will provide a comprehensive understanding of plant compounds, which lead to more effective and targeted treatment. AI can also be useful for documenting and preserving traditional knowledge which can help to future generation.

4. CONCLUSION

In conclusion, this review emphasizes the transformative role of AI and big data together, which have ushered in a new era in drug discovery and development. Using state-of-the-art algorithms based on machine learning and deep learning models, researchers can now mine, explore data sources, identify novel drugs, and accelerate the process of new drug discovery and production. These paradigm shifts are crucial in the discovery of natural products, where the complexity and diversity of natural compounds pose unique challenges. Using AI-powered tools, the identification of bioactive compounds from natural resources can be accelerated and their medicinal properties can be predicted. Additionally, AI models can also help understand the complexity of natural compounds, enabling the development of targeted and effective treatments. It can handle the large datasets from genomics, proteomics and metabolic sources, offering a deeper understanding of plant compounds. Advancements in AI-based machine learning technology can improve the extraction and synthesis processes, ensuring their better quality and efficacy. Overall, the review advocates for the continued exploration and application of AI technologies to advance the field of drug discovery.

Declarations

Competing Interests: The authors declare that they have no competing interests that could have influenced the work reported in this manuscript.

Funding: The authors received no specific funding for this work.

Acknowledgements: The authors thank all researchers whose published work contributed to the preparation of this review.

Use of Artificial Intelligence Tools: During the preparation of this manuscript, the authors used an artificial intelligence (AI) tool solely for language editing. No manuscript content was generated by

AI. The authors reviewed and revised the manuscript and take full responsibility for its accuracy, originality, and integrity.

REFERENCES

1. Mak, K.-K., & Pichika, M. R. (2019). Artificial intelligence in drug development: Present status and future prospects. *Drug Discovery Today*, 24(3), 773–780. <https://doi.org/10.1016/j.drudis.2018.11.014>
2. Yang, Y., & Siau, K. L. (2018). A qualitative research on marketing and sales in the artificial intelligence age. In *Proceedings of the Midwest Association for Information Systems Conference (MWAIS 2018)*. <https://aisel.aisnet.org/mwais2018/41>
3. Skinnider, M. A., Wang, F., Pasin, D., Greiner, R., Foster, L. J., Dalsgaard, P. W., & Wishart, D. S. (2021). A deep generative model enables automated structure elucidation of novel psychoactive substances. *Nature Machine Intelligence*, 3(11), 973–984. <https://doi.org/10.1038/s42256-021-00407-x>
4. Jiménez-Luna, J., Grisoni, F., & Schneider, G. (2020). Drug discovery with explainable artificial intelligence. *Nature Machine Intelligence*, 2(10), 573–584. <https://doi.org/10.1038/s42256-020-00236-4>
5. Baum, Z. J., et al. (2021). Artificial intelligence in chemistry: Current trends and future directions. *Journal of Chemical Information and Modeling*, 61(7), 3197–3212. <https://doi.org/10.1021/acs.jcim.1c00619>
6. Beneke, F., & Mackenrodt, M.-O. (2019). Artificial intelligence and collusion. *IIC—International Review of Intellectual Property and Competition Law*, 50, 109–134. <https://doi.org/10.1007/s40319-018-00773-x>
7. Mekanik, F., et al. (2013). Multiple regression and artificial neural network for long-term rainfall forecasting using large-scale climate modes. *Journal of Hydrology*, 503, 11–21. <https://doi.org/10.1016/j.jhydrol.2013.08.035>
8. Dasta, J. F. (1992). Application of artificial intelligence to pharmacy and medicine. *Hospital Pharmacy*, 27(4), 312–315, 319–322. <https://pubmed.ncbi.nlm.nih.gov/10183640/>
9. Mishra, V. (2018). Artificial intelligence: The beginning of a new era in pharmacy profession. *Asian Journal of Pharmaceutics*, 12(2). <https://doi.org/10.22377/ajp.v12i02.2317>
10. Sellwood, M. A., et al. (2018). Artificial intelligence in drug discovery. *Future Medicinal Chemistry*, 10(17), 2025–2028. <https://doi.org/10.4155/fmc-2018-0212>
11. Zhavoronkov, A., et al. (2019). Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, 37(9), 1038–1040. <https://doi.org/10.1038/s41587-019-0224-x>
12. Thomford, N. E., et al. (2018). Natural products for drug discovery in the 21st century: Innovations for novel drug discovery. *International Journal of Molecular Sciences*, 19(6), Article 1578. <https://doi.org/10.3390/ijms19061578>
13. Nath, R., et al. (2024). The importance of natural products in drug discovery: Opportunities and success. In A. Das Talukdar et al. (Eds.), *Traditional resources and tools for modern drug discovery: Ethnomedicine and pharmacology* (pp. 351–376). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-4600-2_13
14. Rodrigues, T., et al. (2016). Counting on natural products for drug design. *Nature Chemistry*, 8(6), 531–541. <https://doi.org/10.1038/nchem.2479>
15. Liu, Q., Lawrence, A. J., & Liang, J.-H. (2011). Traditional Chinese medicine for treatment of alcoholism: From ancient to modern. *The American Journal of Chinese Medicine*, 39(1), 1–13. <https://doi.org/10.1142/S0192415X11008609>
16. Pattanaik, A., et al. (2019). Artificial intelligence and virtual environment for microalgal source for production of

- nutraceuticals. *Biomedical Journal of Scientific & Technical Research*, 13, <https://doi.org/10.26717/BJSTR.2019.13.002459>
17. Chen, Y., & Kirchmair, J. (2020). Cheminformatics in natural product-based drug discovery. *Molecular Informatics*, 39(12), Article e2000171. <https://doi.org/10.1002/minf.202000171>
18. Romano, J. D., & Tatonetti, N. P. (2019). Informatics and computational methods in natural product drug discovery: A review and perspectives. *Frontiers in Genetics*, 10, Article 368. <https://doi.org/10.3389/fgene.2019.00368>
19. Nuar, A. N. A., & Sen, S. C. (2024). Examining the trend of research on big data architecture: Bibliometric analysis using Scopus database. *Procedia Computer Science*, 234, 172–179. <https://doi.org/10.1016/j.procs.2024.04.010>
20. Joshi, I., et al. (2023). Artificial intelligence, big data and machine learning approaches in genome-wide SNP-based prediction for precision medicine and drug discovery. In S. C. Basak & M. Vračko (Eds.), *Big data analytics in chemoinformatics and bioinformatics* (pp. 333–357). Elsevier. <https://doi.org/10.2174/1389450122999210104205732>
21. Xiao, T., et al. (2024). A review of big data technology and its application in cancer care. *Computers in Biology and Medicine*, 176, Article 108577. <https://doi.org/10.1016/j.compbiomed.2024.108577>
22. Brazma, A., et al. (2006). Data storage and analysis in ArrayExpress. *Methods in Enzymology*, 411, 370–386. [https://doi.org/10.1016/S0076-6879\(06\)11020-4](https://doi.org/10.1016/S0076-6879(06)11020-4)
23. Wang, Z., Jensen, M. A., & Zenklusen, J. C. (2016). A practical guide to The Cancer Genome Atlas (TCGA). In *Statistical genomics: Methods and protocols* (pp. 111–141). Springer. https://doi.org/10.1007/978-1-4939-3578-9_6
24. Chicco, D. (2022). geneExpressionFromGEO: An R package to facilitate data reading from Gene Expression Omnibus (GEO). In *Microarray data analysis* (pp. 187–194). Springer. https://doi.org/10.1007/978-1-0716-1839-4_12
25. Lau, A., & So, H.-C. (2020). Turning genome-wide association study findings into opportunities for drug repositioning. *Computational and Structural Biotechnology Journal*, 18, 1639–1650. <https://doi.org/10.1016/j.csbj.2020.06.015>
26. Shangguan, A., et al. (2024). Genome-wide association study of growth and reproductive traits based on low-coverage whole-genome sequencing in a Chubao black-head goat population. *Gene*, 931, Article 148891. <https://doi.org/10.1016/j.gene.2024.148891>
27. Han, Y., et al. (2019). DriverML: A machine learning algorithm for identifying driver genes in cancer sequencing studies. *Nucleic Acids Research*, 47(8), Article e45. <https://doi.org/10.1093/nar/gkz096>
28. Xu, Z., et al. (2020). Discovery of potential flavonoid inhibitors against COVID-19 3CL proteinase based on virtual screening strategy. *Frontiers in Molecular Biosciences*, 7, Article 556481. <https://doi.org/10.3389/fmolb.2020.556481>
29. Roncaglioni, A., Lombardo, A., & Benfenati, E. (2022). The VEGAHUB platform: The philosophy and the tools. *Alternatives to Laboratory Animals*, 50(2), 121–135. <https://doi.org/10.1177/02611929221090530>
30. Benfenati, E., Manganaro, A., & Gini, G. C. (2013). VEGA-QSAR: AI inside a platform for predictive toxicology. *Proceedings of the Italian Workshop on Artificial Intelligence and Applications*, 1107, 21–28.
31. Banerjee, A., & Roy, K. (2023). Machine-learning-based similarity meets traditional QSAR: “q-RASAR” for the enhancement of

- the external predictivity and detection of prediction confidence outliers in an hERG toxicity dataset. *Chemometrics and Intelligent Laboratory Systems*, 237, Article 104829. <https://doi.org/10.1016/j.chemolab.2023.104829>
32. Lavado, G. J., et al. (2020). Zebrafish AC50 modelling: (Q)SAR models to predict developmental toxicity in zebrafish embryo. *Ecotoxicology and Environmental Safety*, 202, Article 110936. <https://doi.org/10.1016/j.ecoenv.2020.110936>
 33. Ambure, P., et al. (2019). QSAR-Co: An open source software for developing robust multitasking or multitarget classification-based QSAR models. *Journal of Chemical Information and Modeling*, 59(6), 2538–2544. <https://doi.org/10.1021/acs.jcim.9b00295>
 34. Chen, S., et al. (2020). FL-QSAR: A federated learning-based QSAR prototype for collaborative drug discovery. *Bioinformatics*, 36(22–23), 5492–5498. <https://doi.org/10.1093/bioinformatics/btaa1006>
 35. Karpov, P., Godin, G., & Tetko, I. V. (2020). Transformer-CNN: Swiss knife for QSAR modeling and interpretation. *Journal of Cheminformatics*, 12, Article 17. <https://doi.org/10.1186/s13321-020-00423-w>
 36. Wang, Y.-L., et al. (2021). Cloud 3D-QSAR: A web tool for the development of quantitative structure–activity relationship models in drug discovery. *Briefings in Bioinformatics*, 22(4), Article bbaa276. <https://doi.org/10.1093/bib/bbaa276>
 37. Kwon, S., et al. (2019). Comprehensive ensemble in QSAR prediction for drug discovery. *BMC Bioinformatics*, 20, Article 1–12. <https://doi.org/10.1186/s12859-019-3135-4>
 38. Li, J., et al. (2022). Nano-QSAR modeling for predicting the cytotoxicity of metallic and metal oxide nanoparticles: A review. *Ecotoxicology and Environmental Safety*, 243, Article 113955. <https://doi.org/10.1016/j.ecoenv.2022.113955>
 39. Ahmadi, S., et al. (2021). CORAL: Monte Carlo-based global QSAR modelling of Bruton tyrosine kinase inhibitors using hybrid descriptors. *SAR and QSAR in Environmental Research*, 32(12), 1013–1031. <https://doi.org/10.1080/1062936X.2021.2003429>
 40. Halder, A. K., & Dias Soeiro Cordeiro, M. N. (2021). QSAR-Co-X: An open source toolkit for multitarget QSAR modelling. *Journal of Cheminformatics*, 13, Article 1–18. <https://doi.org/10.1186/s13321-021-00508-0>
 41. Wang, Y.-L., et al. (2021). Cloud 3D-QSAR: A web tool for the development of quantitative structure–activity relationship models in drug discovery. *Briefings in Bioinformatics*, 22(4), Article bbaa276. <https://doi.org/10.1093/bib/bbaa276>
 42. Toropov, A. A., & Toropova, A. P. (2020). The Monte Carlo method as a tool to build up predictive QSPR/QSAR. *Current Computer-Aided Drug Design*, 16(3), 197–206. <https://doi.org/10.2174/1573409915666190328123112>
 43. Kumar, P., Kumar, A., & Singh, D. (2022). CORAL: Development of a hybrid descriptor-based QSTR model to predict the toxicity of dioxins and dioxin-like compounds with correlation intensity index and consensus modelling. *Environmental Toxicology and Pharmacology*, 93, Article 103893. <https://doi.org/10.1016/j.etap.2022.103893>
 44. Toropova, A. P., et al. (2020). How the CORAL software can be used to select compounds for efficient treatment of neurodegenerative diseases? *Toxicology and Applied Pharmacology*, 408, Article 115276. <https://doi.org/10.1016/j.taap.2020.115276>
 45. Oldenhof, M., et al. (2020). ChemGrapher: Optical graph recognition of chemical compounds by deep learning. *Journal of Chemical Information and Modeling*, 60(10),

- 4506–4517.
<https://doi.org/10.1021/acs.jcim.0c00459>
46. Wang, C., et al. (2024). YZS-model: A predictive model for organic drug solubility based on graph convolutional networks and transformer-attention. arXiv.
<https://doi.org/10.48550/arXiv.2406.19136>
47. Yi, J.-C., et al. (2024). ChemMORT: An automatic ADMET optimization platform using deep learning and multi-objective particle swarm optimization. Briefings in Bioinformatics, 25(2), Article bbae008.
<https://doi.org/10.1093/bib/bbae008>
48. Dong, J., et al. (2017). ChemSAR: An online pipelining platform for molecular SAR modeling. Journal of Cheminformatics, 9, Article 1–13. <https://doi.org/10.1186/s13321-017-0215-1>
49. Pan, X., et al. (2023). AI-DrugNet: A network-based deep learning model for drug repurposing and combination therapy in neurological disorders. Computational and Structural Biotechnology Journal, 21, 1533–1542.
<https://doi.org/10.1016/j.csbj.2023.02.004>
50. Zhang, M.-L., et al. (2022). RLFDDA: A meta-path-based graph representation learning model for drug–disease association prediction. BMC Bioinformatics, 23(1), Article 516.
<https://doi.org/10.1186/s12859-022-05069-z>
51. Wu, Z., et al. (2022). wSDTNBI: A novel network-based inference method for virtual screening. Chemical Science, 13(4), 1060–1079. <https://doi.org/10.1039/D1SC05613A>
52. Smith, R. D., et al. (2019). Updates to Binding MOAD (Mother of All Databases): Polypharmacology tools and their utility in drug repurposing. Journal of Molecular Biology, 431(13), 2423–2433.
<https://doi.org/10.1016/j.jmb.2019.05.024>
53. King, M. D., et al. (2018). SPIDR: Small-molecule peptide-influenced drug repurposing. BMC Bioinformatics, 19, Article 1–11.
<https://doi.org/10.1186/s12859-018-2153-y>
54. Huang, T., et al. (2023). SPIDR is required for homologous recombination during mammalian meiosis. Nucleic Acids Research, 51(8), 3855–3868. <https://doi.org/10.1093/nar/gkad154>
55. Chatterjee, A., et al. (2023). Improving the generalizability of protein-ligand binding predictions with AI-Bind. Nature Communications, 14(1), Article 1989.
<https://doi.org/10.1038/s41467-023-37572-z>
56. Yuan, Y., Zhang, Y., Meng, X., Liu, Z., Wang, B., Miao, R., Zhang, R., Su, W., & Liu, L. (2023). EDC-DTI: An end-to-end deep collaborative learning model based on multiple information for drug-target interactions prediction. Journal of Molecular Graphics and Modelling, 122, Article 108498.
<https://doi.org/10.1016/j.jmgm.2023.108498>
57. Wang, Y.-Y., et al. (2022). CeDR Atlas: A knowledgebase of cellular drug response. Nucleic Acids Research, 50(D1), D1164–D1171. <https://doi.org/10.1093/nar/gkab897>
58. Liu, J., et al. (2024). AMDGT: Attention-aware multimodal fusion using a dual graph transformer for drug–disease association prediction. Knowledge-Based Systems, 284, Article 111329.
<https://doi.org/10.1016/j.knosys.2023.111329>
59. Yang, R., et al. (2024). GCNGAT: Drug–disease association prediction based on graph convolution neural network and graph attention network. Artificial Intelligence in Medicine, 150, Article 102805.
<https://doi.org/10.1016/j.artmed.2024.102805>
60. Jin, S., et al. (2021). HeTDR: Drug repositioning based on heterogeneous networks and text mining. Patterns, 2(8).
61. Hassanali Aragh, A., et al. (2024). MiRAGE: Mining relationships for advanced generative evaluation in drug repositioning. Briefings in Bioinformatics, 25(4).
<https://doi.org/10.1093/bib/bbae337>
62. Rodrigues, C. H., & Ascher, D. B. (2023). CSM-Potential2: A comprehensive deep learning platform for the analysis of protein

- interacting interfaces. *Proteins: Structure, Function, and Bioinformatics*. <https://doi.org/10.1002/prot.26615>
63. Venkatraman, V. (2021). FP-ADMET: A compendium of fingerprint-based ADMET prediction models. *Journal of Cheminformatics*, 13, Article 1–12. <https://doi.org/10.1186/s13321-021-00557-5>
64. Myung, Y., Pires, D. E., & Ascher, D. B. (2022). CSM-AB: Graph-based antibody–antigen binding affinity prediction and docking scoring function. *Bioinformatics*, 38(4), 1141–1143. <https://doi.org/10.1093/bioinformatics/btab762>
65. Rayka, M., & Firouzi, R. (2023). GB-score: Minimally designed machine learning scoring function based on distance-weighted interatomic contact features. *Molecular Informatics*, 42(3), Article 2200135. <https://doi.org/10.1002/minf.202200135>
66. Zhang, X., et al. (2023). TB-IECS: An accurate machine learning-based scoring function for virtual screening. *Journal of Cheminformatics*, 15(1), Article 63. <https://doi.org/10.1186/s13321-023-00731-x>
67. Hönig, S. M., et al. (2024). SpaceGrow: Efficient shape-based virtual screening of billion-sized combinatorial fragment spaces. *Journal of Computer-Aided Molecular Design*, 38(1), Article 13. <https://doi.org/10.1007/s10822-024-00551-7>
68. Loeffler, H. H., et al. (2024). REINVENT 4: Modern AI-driven generative molecule design. *Journal of Cheminformatics*, 16(1), Article 20. <https://doi.org/10.1186/s13321-024-00812-5>
69. Rifaioğlu, A. S., et al. (2021). MDeePred: Novel multi-channel protein featurization for deep learning-based binding affinity prediction in drug discovery. *Bioinformatics*, 37(5), 693–704. <https://doi.org/10.1093/bioinformatics/btaa858>
70. Zhang, L., et al. (2023). GPCNDTA: Prediction of drug-target binding affinity through cross-attention networks augmented with graph features and pharmacophores. *Computers in Biology and Medicine*, 166, Article 107512. <https://doi.org/10.1016/j.compbio.2023.107512>
71. Thafar, M. A., et al. (2022). Affinity2Vec: Drug-target binding affinity prediction through representation learning, graph mining, and machine learning. *Scientific Reports*, 12(1), Article 4751. <https://doi.org/10.1038/s41598-022-08787-9>
72. D’Souza, S., et al. (2023). Deep learning-based modeling of drug–target interaction prediction incorporating binding site information of proteins. *Interdisciplinary Sciences: Computational Life Sciences*, 15(2), 306–315. <https://doi.org/10.1007/s12539-023-00557-z>
73. Desai, N., Patel, D., & Vyas, N. (2022). Pharmacognosy with artificial intelligence: An interdisciplinary approach. In *Advances in data and information sciences*. Springer. https://doi.org/10.1007/978-981-16-5689-7_15
74. Duan, F.-L., et al. (2024). AI-driven drug discovery from natural products. *Advanced Agrochem*, 3(3), 185–187. <https://doi.org/10.1016/j.aac.2024.06.003>
75. Sarker, S. D., & Nahar, L. (2018). An introduction to computational phytochemistry. In S. D. Sarker & L. Nahar (Eds.), *Computational phytochemistry* (pp. 1–41). Elsevier. <https://doi.org/10.1016/B978-0-12-812364-5.00001-8>
76. Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J. J., Martin, N. I., Meijer, D., Terlouw, B. R., Biermann, F., Blin, K., Durairaj, J., Gorostiola González, M., Helfrich, E. J. N., Huber, F., Leopold-Messer, S., Rajan, K., de Rond, T., van Santen, J. A., Sorokina, M., ... Medema, M. H. (2023). Artificial intelligence for natural product drug discovery. *Nature Reviews Drug Discovery*, 22(11), 895–916. <https://doi.org/10.1038/s41573-023-00774-7>
77. Li, G., et al. (2022). Artificial intelligence-guided discovery of anticancer lead compounds

- from plants and associated microorganisms. *Trends in Cancer*, 8(1), 65–80. <https://doi.org/10.1016/j.trecan.2021.10.002>
- 78.Dias, T., Gaudêncio, S. P., & Pereira, F. (2018). A computer-driven approach to discover natural product leads for methicillin-resistant *Staphylococcus aureus* infection therapy. *Marine Drugs*, 17(1), Article 16. <https://doi.org/10.3390/md17010016>
- 79.Xiong, H.-S., et al. (2024). Application scenario design and prospect of generative artificial intelligence (AI) in intelligent manufacturing and supply chain of traditional Chinese medicine. *Zhongguo Zhong Yao Za Zhi*, 49(14), 3963–3970. <https://doi.org/10.19540/j.cnki.cjcmm.20240402.301>
- 80.Zeng, J., & Jia, X. (2024). Systems theory-driven framework for AI integration into the holistic material basis research of traditional Chinese medicine. *Engineering*. <https://doi.org/10.1016/j.eng.2024.04.009>
- 81.Zhong, Y., et al. (2022). An online, non-destructive method for simultaneously detecting chemical, biological, and physical properties of herbal injections using hyperspectral imaging with artificial intelligence. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 264, Article 120250. <https://doi.org/10.1016/j.saa.2021.120250>
- 82.Arora, S., et al. (2024). Artificial intelligence: A virtual chemist for natural product drug discovery. *Journal of Biomolecular Structure and Dynamics*, 42(7), 3826–3835. <https://doi.org/10.1080/07391102.2023.2216295>
- 83.Thanikkal, J. G., Dubey, A. K., & Thomas, M. (2023). An efficient mobile application for identification of immunity-boosting medicinal plants using shape descriptor algorithm. *Wireless Personal Communications*, 131(2), 1189–1205. <https://doi.org/10.1007/s11277-023-10476-3>
- 84.Merk, D., et al. (2018). Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators. *Communications Chemistry*, 1(1), Article 68. <https://doi.org/10.1038/s42004-018-0068-1>
- 85.Hesami, M., et al. (2020). Development of support vector machine-based model and comparative analysis with artificial neural network for modeling the plant tissue culture procedures: Effect of plant growth regulators on somatic embryogenesis of chrysanthemum, as a case study. *Plant Methods*, 16, Article 1–15. <https://doi.org/10.1186/s13007-020-00655-9>
- 86.Niazian, M., & Nalouisi, A. M. (2020). Artificial polyploidy induction for improvement of ornamental and medicinal plants. *Plant Cell, Tissue and Organ Culture*, 142(3), 447–469. <https://doi.org/10.1007/s11240-020-01888-1>
- 87.Sharafudeen, M., et al. (2024). Verified localization and pharmacognosy of herbal medicinal plants in a combined network framework. *Computers in Biology and Medicine*, 174, Article 108467. <https://doi.org/10.1016/j.compbiomed.2024.108467>
- 88.Das, P., Nayak, P. K., & Kesavan, R. K. (2024). Artificial neural networks (ANN)-genetic algorithm (GA) optimization on thermosonicated achocha juice: Kinetic and thermodynamic perspectives of retained phytocompounds. *Preparative Biochemistry & Biotechnology*. <https://doi.org/10.1080/10826068.2024.2378101>
- 89.Cirić, A., et al. (2020). Response surface methodology and artificial neural network approach for the optimization of ultrasound-assisted extraction of polyphenols from garlic. *Food and Chemical Toxicology*, 135, Article 110976. <https://doi.org/10.1016/j.fct.2019.110976>

90. Jha, A. K., & Sit, N. (2021). Comparison of response surface methodology (RSM) and artificial neural network (ANN) modelling for supercritical fluid extraction of phytochemicals from *Terminalia chebula* pulp and optimization using RSM coupled with desirability function (DF) and genetic algorithm (GA) and ANN with GA. *Industrial Crops and Products*, 170, Article 113769. <https://doi.org/10.1016/j.indcrop.2021.113769>
91. Ekpenyong, M. G., & Antai, S. P. (2024). Statistical versus neural network-embedded swarm intelligence optimization of a metallo-neutral-protease production: Activity kinetics and food industry applications. *Preparative Biochemistry & Biotechnology*. <https://doi.org/10.1080/10826068.2024.2328681>
92. Ray, A., et al. (2020). Application of artificial neural network (ANN) model for prediction and optimization of coronarin D content in *Hedychium coronarium*. *Industrial Crops and Products*, 146, Article 112186. <https://doi.org/10.1016/j.indcrop.2020.112186>
93. Liu, M., Li, Y., & Li, H. (2022). Deep learning to predict the biosynthetic gene clusters in bacterial genomes. *Journal of Molecular Biology*, 434(15), Article 167597. <https://doi.org/10.1016/j.jmb.2022.167597>
94. Klau, L. J., et al. (2022). The Natural Product Domain Seeker version 2 (NaPDoS2) webtool relates ketosynthase phylogeny to biosynthetic function. *Journal of Biological Chemistry*, 298(10), Article 102480. <https://doi.org/10.1016/j.jbc.2022.102480>
95. Aron, A. T., et al. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols*, 15(6), 1954–1991. <https://doi.org/10.1038/s41596-020-0317-5>
96. Sorokina, M., et al. (2021). COCONUT online: Collection of open natural products database. *Journal of Cheminformatics*, 13(1), Article 2. <https://doi.org/10.1186/s13321-020-00478-9>
97. Rutz, A., et al. (2021). The LOTUS initiative for open natural products research: Knowledge management through Wikidata. *bioRxiv*. <https://doi.org/10.1101/2021.02.28.433265>
98. Kim, H. W., et al. (2021). NPClassifier: A deep neural network-based structural classification tool for natural products. *Journal of Natural Products*, 84(11), 2795–2807. <https://doi.org/10.1021/acs.jnatprod.1c00399>
99. N. A. R., & G. K. R. (2024). A deep learning and docking simulation-based virtual screening strategy enables the rapid identification of HIF-1 α pathway activators from a marine natural product database. *Journal of Biomolecular Structure and Dynamics*, 42(2), 629–651. <https://doi.org/10.1080/07391102.2023.2194997>
100. Kim, H. W., et al. (2021). NPClassifier: A deep neural network-based structural classification tool for natural products. *Journal of Natural Products*, 84(11), 2795–2807. <https://doi.org/10.1021/acs.jnatprod.1c00399>
101. Chávez-Hernández, A. L., & Medina-Franco, J. L. (2023). Natural products subsets: Generation and characterization. *Artificial Intelligence in the Life Sciences*, 3, Article 100066. <https://doi.org/10.1016/j.ailsci.2023.100066>
102. Ye, L., et al. (2024). De novo genome assembly and functional insights of the first commercial pink *Auricularia cornea*. *Genomics*, 116(5), Article 110902. <https://doi.org/10.1016/j.ygeno.2024.110902>
103. An, T., et al. (2023). De novo biosynthesis of anticarcinogenic icariin in engineered yeast. *Metabolic Engineering*, 80, 207–215. <https://doi.org/10.1016/j.ymben.2023.10.003>
104. Wu, Y., & Dockendorff, C. (2018). Synthesis of a novel bicyclic scaffold inspired by the antifungal natural product sordarin. *Tetrahedron Letters*, 59(36), 3373–3376. <https://doi.org/10.1016/j.tetlet.2018.07.064>
105. Schuppe, A. W., et al. (2022). Unified total synthesis of the limonoid alkaloids:

- Strategies for the de novo synthesis of highly substituted pyridine scaffolds. *Chem*, 8(10), 2856–2887.
<https://doi.org/10.1016/j.chempr.2022.09.012>
106. Bung, N., et al. (2021). De novo design of new chemical entities for SARS-CoV-2 using artificial intelligence. *Future Medicinal Chemistry*, 13(6), 575–585.
<https://doi.org/10.4155/fmc-2020-0262>
 107. Wilson, D. M., et al. (2024). Targeted sampling of natural product space to identify bioactive natural product-like polyketide macrolides. *Nature Communications*, 15(1), Article 2534. <https://doi.org/10.1038/s41467-024-46721-x>
 108. Hartenfeller, M., et al. (2012). DOGS: Reaction-driven de novo design of bioactive compounds. *PLoS Computational Biology*, 8(2), Article e1002380.
<https://doi.org/10.1371/journal.pcbi.1002380>
 109. Merk, D., et al. (2018). Computer-assisted discovery of retinoid X receptor modulating natural products and isofunctional mimetics. *Journal of Medicinal Chemistry*, 61(12), 5442–5447.
<https://doi.org/10.1021/acs.jmedchem.8b00494>
 110. Grisoni, F., et al. (2019). Design of natural-product-inspired multitarget ligands by machine learning. *ChemMedChem*, 14(12), 1129–1134.
<https://doi.org/10.1002/cmdc.201900097>
 111. Sundarasekar, J., & Sahgal, G. (2024). Role of genomics and proteomics in drug discovery. In *Concepts in pharmaceutical biotechnology and drug development* (pp. 207–246). Springer. https://doi.org/10.1007/978-981-97-1148-2_11
 112. Wang, H., et al. (2020). Deep learning for plant genomics and crop improvement. *Current Opinion in Plant Biology*, 54, 34–41.
<https://doi.org/10.1016/j.pbi.2019.12.010>
 113. Singh, A., et al. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, 21(2), 110–124.
<https://doi.org/10.1016/j.tplants.2015.10.015>
 114. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7, Article 1419.
<https://doi.org/10.3389/fpls.2016.01419>
 115. Fu, P., et al. (2019). Hyperspectral leaf reflectance as proxy for photosynthetic capacities: An ensemble approach based on multiple machine learning algorithms. *Frontiers in Plant Science*, 10, Article 730.
<https://doi.org/10.3389/fpls.2019.00730>
 116. Mansoor, S., et al. (2024). Genomics, phenomics, and machine learning in transforming plant research: Advancements and challenges. *Horticultural Plant Journal*.
<https://doi.org/10.1016/j.hpj.2023.09.005>
 117. Teshome, F. T., et al. (2023). Unmanned aerial vehicle (UAV) imaging and machine learning applications for plant phenotyping. *Computers and Electronics in Agriculture*, 212, Article 108064.
<https://doi.org/10.1016/j.compag.2023.108064>
 118. ElManawy, A. I., et al. (2022). HSI-PP: A flexible open-source software for hyperspectral imaging-based plant phenotyping. *Computers and Electronics in Agriculture*, 200, Article 107248.
<https://doi.org/10.1016/j.compag.2022.107248>
 119. Hariharan, J., et al. (2023). An AI-based spectral data analysis process for recognizing unique plant biomarkers and disease features. *Computers and Electronics in Agriculture*, 204, Article 107574.
<https://doi.org/10.1016/j.compag.2022.107574>