Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

From Nature to Neural Networks: The Role of Artificial Intelligence in Selecting Plants for Cancer Drug Discovery

N. S. Disha[1], B. S. Ashok Kumar[2*]

[1] Department of Pharmaceutical Chemistry, R. L. Jalappa College of Pharmacy, Sri Devaraj Urs Academy of Higher Education and Research, Kolar, Karnataka, India.

[2] Department of Pharmacognosy, R. L. Jalappa College of Pharmacy, Sri Devaraj Urs Academy of Higher Education and Research, Kolar, Karnataka, India

*Corresponding author

Ashok Kumar BS

Professor, R. L. Jalappa College of Pharmacy, Sri Devaraj Urs Academy of Higher Education and Research, Kolar, Karnataka, India.

ashok4vani@gmail.com

ARTICLE INFO

Received: 18 Dec 2024

Revised: 15 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Introduction: Cancer remains a global health burden, and while synthetic drugs offer effective treatment, issues like toxicity and resistance persist. Medicinal plants are rich in anticancer phytochemicals, but traditional drug discovery methods are slow and laborious. The integration of Artificial Intelligence (AI) is accelerating natural product research by enabling efficient screening, target prediction, and toxicity assessment.

Objectives: This review highlights the role of AI-including QSAR modeling, molecular docking, and predictive analytics—in discovering and optimizing plant-derived anticancer agents, with emphasis on its use in personalized therapy, multi-target drug design, and phytochemical repurposing.

Methods: Recent studies employing machine learning, deep learning, and computational modeling were reviewed to evaluate their application in phytochemical screening, pharmacophore mapping, and virtual simulation of molecular interactions.

Results: AI platforms enhanced the speed and accuracy of identifying potent phytochemicals, predicted toxicity and targets effectively, and supported the design of personalized cancer therapies. Case studies showed successful use of AI in optimizing plant-based anticancer compounds.

Conclusions: AI-driven strategies offer a powerful approach to phytochemical-based cancer therapy. The synergy between AI and medicinal plants promises faster, safer, and more personalized treatment options, emphasizing the need for cross-disciplinary collaboration.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Keywords: Artificial Intelligence, Cancer, Medicinal Plants, QSAR, Molecular Docking, Phytochemicals, Drug Discovery, Precision Therapy.

INTRODUCTION

Global cancer mortality leads all causes because 10 million fatalities and 19.3 million new cases were recorded in 2020 (GLOBOCAN 2020). The estimated annual case numbers of cancer in 2040 reach 28.4 million which highlights an immediate necessity to develop new treatment approaches. The pursuit of alternative cancer treatments occurs because conventional treatments through chemotherapy and targeted drugs lead to severe side effects and drug resistance in addition to being expensive to administer. Traditional medicine relies on medicinal plants which contain an extensive yet unexamined quantity of bioactive compounds. The majority of approved cancer drugs entered the market from plants or plant-inspired sources including paclitaxel and vincristine however their identification process is slow and haphazard. This paper discusses how artificial intelligence (AI) transforms the process of identifying anticancer agents from plants while uniting traditional medicinal insights with state-of-the-art computational tools [1,2]

Despite providing better survival rates for various cancers the use of chemotherapy and synthesized drugs results in serious adverse effects that include nausea and immunosuppression and organ toxicity and hair loss [3]. The effectiveness of chemotherapeutic agents becomes diminished because of drug resistance found either naturally or acquired during treatment [4]. The agents cause universal damage to normal rapidly dividing cells because they lack selectivity between healthy and cancerous cells [5]. Medical professionals agree that synthetic drugs targeting solitary cancer motifs are insufficient to address the complex nature of cancer. The need exists to study multi-target compounds produced from medicinal plants because these show more favorable safety features than typical agents.

Medicinal plants have operated as fundamental elements in traditional healing practices including Ayurveda and Traditional Chinese Medicine (TCM) and African ethnomedicine throughout many centuries. Plant compounds extracted from Taxus brevifolia and Catharanthus roseus and Camptotheca acuminata all became modern anticancer drugs under the names paclitaxel and vincristine and camptothecin respectively [6]. The field of modern pharmacognosy relies on natural products to obtain 60% of its currently approved anticancer drugs that either originate directly from plants or draw inspiration from them [7]. The compounds demonstrate multiple targets while having less toxicity and immunomodulatory properties which makes them suitable for both combination cancer therapies and single agent cancer treatments.

The massive variety of phytoconstituents along with complex obstacles in cancer cell biology make traditional drug discovery methods difficult to apply. Artificial Intelligence (AI) demonstrates transformational data-related abilities through Machine Learning (ML) and Deep Learning (DL) capabilities which analyze complex datasets [8]. The detection ability of ML algorithms combines with their classification and biological activity prediction functions alongside DL's operation on unstructured data like images and molecular graphs. Modern chemoinformatics together with target identification processes alongside ADMET prediction and virtual screening functions benefit highly from these technologies [9]. Artificial intelligence proves essential for medicinal discovery because it joins forces with natural product research and expedites the clinical transformation of plant compounds into useful therapies.

Academic interest in plant medicine continues to rise yet AI technologies find restricted applications during standard anticancer medicinal plant evaluations. This review investigates the complete transformation of plant-derived anticancer agents through the implementation of AI techniques which integrate QSAR modeling alongside molecular docking solutions together with pharmacophore mapping and predictive analytical tools. It highlights the global burden of cancer and the pressing need for novel, effective therapeutics, explores the limitations of current synthetic treatments, and underscores the immense therapeutic potential of phytochemicals. Additionally, the review discusses

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

how AI tools can accelerate phytochemical screening, activity prediction, and compound optimization, while presenting relevant case studies and practical applications in natural product-based oncology. Ultimately, this work seeks to connect the traditional domain of pharmacognosy with the innovations of computational science, fostering interdisciplinary collaboration in the pursuit of precision medicine.

HARNESSING AI FOR NEXT-GENERATION PLANT-BASED DRUG DISCOVERY

Drug discovery benefits from Artificial Intelligence (AI) as an advanced computational method for evaluating complex biological data thus speeding up therapeutic product development. Artificial Intelligence describes systems which replicate human intelligence through information processing and decision-making along with data learning capacity. Two key subsets of AI-Machine Learning (ML) and Deep Learning (DL)-are particularly impactful in natural product research. The algorithms of ML extract data patterns to generate predictions through automated processes and DL achieves deep pattern detection inside extensive complex datasets using its network of neural layers [9].

AI tools now drastically modernize the discovery process for bioactive medicinal plants compounds by merging with natural product-based drug development. Pharmacological activity predictive models and toxicity models and drug-likeness predictive models can be generated through platforms which include Auto QSAR, KNIME, WEKA, DeepChem, and Chemprop. Irradiation experiments using these tools shorten plant molecule evaluation cycles while decreasing expenses that conventional laboratory work methods require [10].

The rapid discovery of plant-based medicines uses Artificial Intelligence through classification, regression, clustering and prediction techniques as its primary computational tools. Researchers can determine therapeutic value by classifying phytochemicals according to established inactive-active and toxic-non-toxic definitions through classification methods. Regression models serve to estimate continuous variables encompassing IC_{50} measures as well as binding affinity measurements while also predicting ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profiles to quantify bioactive compounds' pharmacological power.

Through unsupervised learning clusters similar compounds by shared chemical and biological features which allows researchers to discover potent drug-like phytoconstituents. Within AI prediction there exists a broader range that extends beyond classification and regression to include the forecast of essential biological outcomes like blood-brain barrier permeability and multi-target effects together with adverse drug reactions [11]. These AI techniques streamline natural drug screening operations by boosting both speed and accuracy and extraction depth thus making the process more efficient in discovering strong plant extracts used for cancer and neurological disorder treatment.

Table 1. AI Techniques and Tools in Plant-Based Drug Discovery

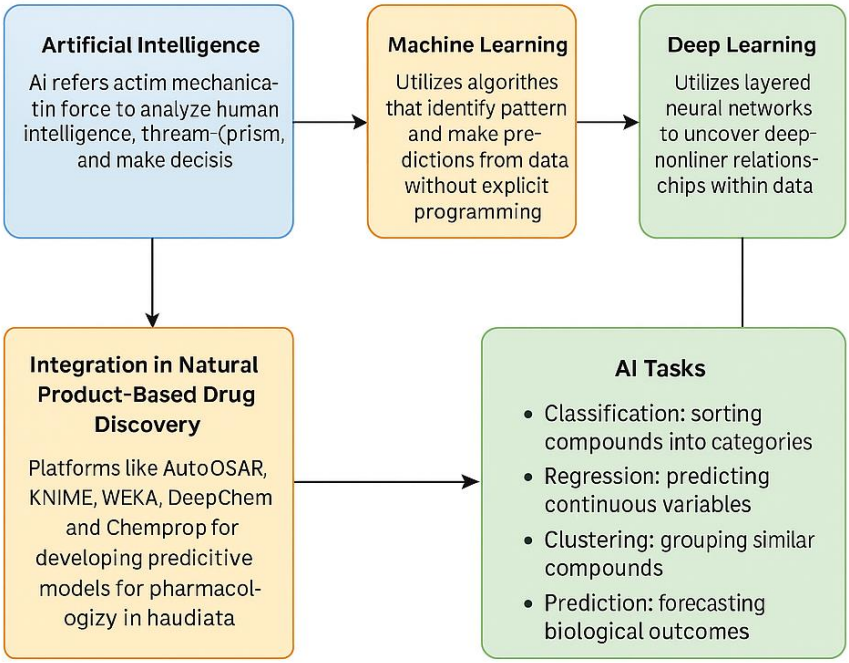
AI Technique	Application	Example Tools/Models
Supervised ML	Classification, regression	Auto QSAR, WEKA, Chemprop
Unsupervised ML	Clustering, pattern discovery	KNIME, PCA, t-SNE
Deep Learning	Image recognition, complex prediction tasks	DeepChem, CNN, RNN
NLP	Literature mining, knowledge extraction	BERT, BioBERT

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI Technique	Application	Example Tools/Models
Network Pharmacology	Target-pathway-disease interaction mapping	STITCH, Coremine

Figure 1. AI-Driven Innovation in Natural Product Drug Discovery



AI IN PLANT-BASED CANCER THERAPEUTICS

The scientific discovery of natural products through Artificial Intelligence (AI) now functions with high-speed screening technologies together with automatic pattern detection and modeling predictions. Effective phytochemical screening in medicinal plants represents a key priority for oncology because it can deliver alternative treatments that are safe, affordable and multicomponent focused. The combination of AI techniques allows researchers to predict phytoconstituent anticancer potential through analysis of molecular properties together with biological targets and mechanisms of action.

The forecasting of plant-derived compound anticancer potential depends on Machine Learning (ML) and Deep Learning (DL) techniques that apply trained models from extensive anticancer molecule datasets. A screening process using these techniques provides insights into the chemical-activity relationships which lets researchers evaluate thousands of phytochemicals virtually. The workflow integration often includes three main computational methods which are QSAR modeling and molecular docking and pharmacophore mapping. Modern virtual screening methods shorten both time and expenses that would otherwise be linked to traditional wet-lab screening approaches [12].

CASE STUDIES OF PLANT-DERIVED ANTICANCER COMPOUNDS ENHANCED BY AI APPROACHES

Artificial Intelligence systems have brought a rapid pace to finding plant-based anticancer agents from natural sources. Through the application of AI techniques such as QSAR modeling and molecular docking combined with deep learning and pharmacophore mapping scientists can screen and predict

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

phytochemical therapeutic effects. In silico modeling and predictive analytics enable researchers to conduct virtual simulations of bioactive compounds interacting with cancer targets alongside pharmacokinetic profiling and therapeutic potential ranking of molecules before initiating expensive laboratory work.

Artificial intelligence platforms have thoroughly researched three important phytochemicals: Taxol from Taxus brevifolia together with Vincristine from Catharanthus roseus and Camptothecin from Camptotheca acuminata. These analytical tools have proved essential for demonstrating how these substances stop topoisomerases and stabilize microtubules. AI platforms have examined the multiple target properties and drug characteristics of medicinal compounds Curcumin, Berberine, Resveratrol and Quercetin which traditional medicine uses frequently. The table presents data about ten natural compounds from plants showing their anticancer functions and analytical instruments used in AI evaluations.

AI revolutionizes natural drug discovery by supporting data-centric choices and minimize the necessity of traditional experimental techniques specifically in the field of oncology.

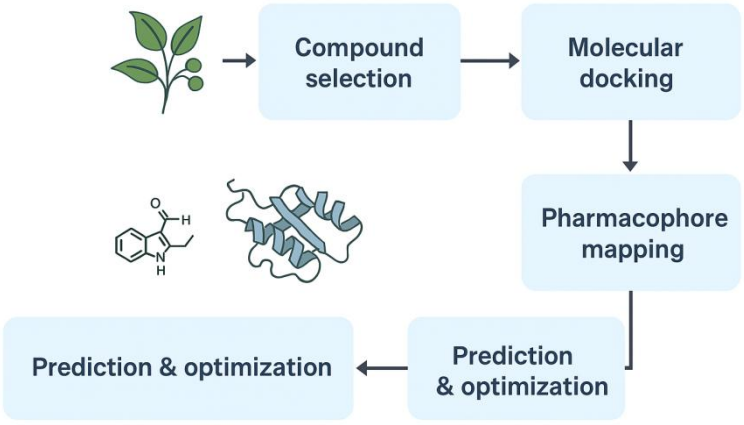
Table 2. Plant-Derived Anticancer Compounds Enhanced by AI Approaches

Phytochemical	Plant Source	Mechanism of Action	AI Tools Applied	Reference
Taxol (Paclitaxel)	*Taxus brevifolia*	Microtubule stabilization	QSAR, Molecular Docking	[13]
Vincristine	*Catharanthus roseus*	Mitotic spindle inhibition	QSAR, Deep Learning	[14]
Camptothecin	*Camptotheca acuminata*	Topoisomerase I inhibition	Molecular Docking, Pharmacophore Mapping	[15]
Curcumin	*Curcuma longa*	NF-κB inhibition, apoptosis induction	QSAR, Prediction Models	[16]
Berberine	*Berberis vulgaris*	DNA intercalation, mitochondrial dysfunction	Machine Learning Models	[17]
Emodin	*Rheum palmatum*	Inhibits angiogenesis and cell proliferation	QSAR, Prediction of ADMET	[18]
Betulinic acid	*Betula alba*	Apoptosis induction in melanoma cells	Docking, Binding Affinity Prediction	[19]
Podophyllotoxin	*Podophyllum hexandrum*	Microtubule assembly inhibition	Pharmacophore Mapping, Clustering	[20]
Resveratrol	*Vitis vinifera*	SIRT1 activation, ROS scavenging	Deep Learning, Target Prediction	[21]
Quercetin	*Allium cepa*, *Camellia sinensis*	Cell cycle arrest, kinase inhibition	QSAR, Multi-target Prediction	[22]

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Figure 2. AI Workflow in Plant-Based Anticancer Drug Discovery



PLant-based cancer drug discovery receives enhanced strength through Artificial Intelligence (AI) by utilizing powerful computational tools including QSAR models and molecular docking and pharmacophore mapping. AI algorithms in QSAR (Quantitative Structure–Activity Relationship) modeling use support vector machines and random forests as well as molecular descriptors from anticancer activity information for quick screening through large phytochemical libraries [23]. The combination of AI technology with Molecular docking predicts how plant compounds interact with cancer-related proteins through computer simulations of receptor-ligand interactions to achieve better binding affinities which yields superior virtual screening performance [24]. The identification of essential biological activity features depends on pharmacophore mapping which AI technology helps to accelerate through large-scale dataset analysis of anticancer mechanism patterns [25]. These technologies cooperate as an exclusive program to enhance both precision and shorten the process timeline for discovering strong anticancer agents from medicinal plants. The AI-directed workflow for developing anticancer drugs from plants includes compound selection, lead optimization steps and this scheme is shown in Figure 1 and Table 1 presents relevant examples in herbal cancer research.

AI-ASSISTED MECHANISTIC INSIGHTS AND TARGET PREDICTION

Artificial Intelligence (AI) plays a pivotal role in unveiling the mechanistic aspects of plant-derived anticancer agents by predicting molecular targets, pathways, and drug-like behaviors. This advancement not only accelerates the early-stage drug discovery process but also improves precision in identifying how phytochemicals exert therapeutic effects.

Prediction of Molecular Targets (Enzymes, Receptors, Signaling Pathways): Machine learning (ML) with deep learning (DL) functions as primary artificial intelligence tools for predicting phytoconstituent potential targets at the molecular level. Phytoconstituents generally target three types of molecular entities including kinases along with nuclear receptors and oncogenic enzymes including topoisomerases and telomerases. Numerous drug-target bindings are detected by algorithms that scan extensive chemical-biological databases despite experimental proof being unavailable. Artificial intelligence models demonstrate successful predictions regarding the curcumin-NF-κB and curcumin-PI3K/Akt pathway interactions because of their importance in tumor development [26]. The predictive models demonstrate their capability to identify both main targets and secondary effects that cause adverse reactions in the drug development process.

Network Pharmacology and AI Integration: Network pharmacology builds compound-target-pathway networks through graph analysis methods to understand how natural products work as holistic therapeutics. Through AI capabilities the system optimizes complex data management functions while also identifying concealed patterns. The approach enables the unification of compound information from databases together with genetic data from omics research to analyze complete networked

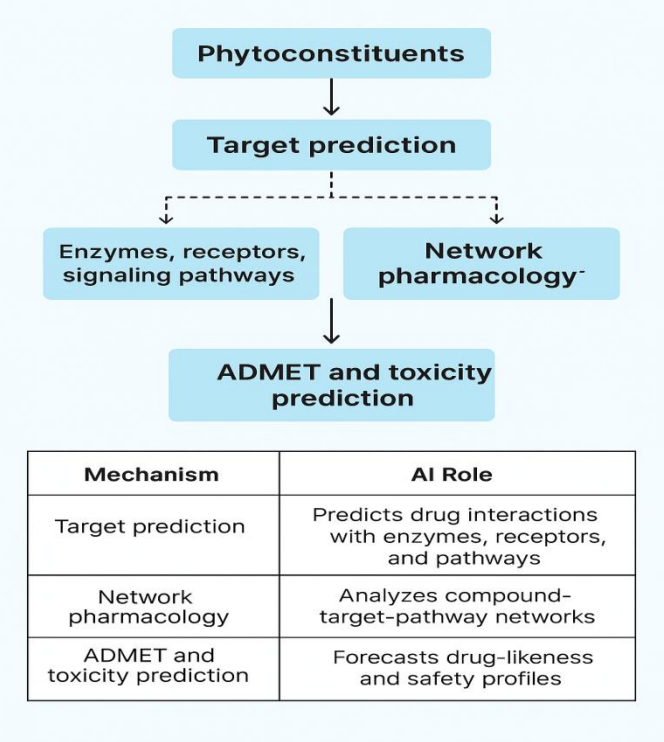
Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

biological activities. The approach at systems-level works effectively for plant-based compounds because they operate on several targets. Network pharmacology analyses assisted by artificial intelligence demonstrates herbal formulations containing multiple components operate synergistically against cancer pathways [27,28].

Prediction of Drug-Likeness, Toxicity, and ADMET Properties: Earlier identification of drug-likeness together with toxicity and ADMET properties represents an essential challenge in drug discovery because numerous candidates fail due to unfavorable pharmacokinetic and toxicological characteristics. AI addresses this through predictive modeling of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties. Neural networks together with decision trees power the capabilities of DeepTox and pkCSM and ADMETlab to forecast characteristics such as drug-likeness and solubility as well as intestinal absorption and blood-brain barrier permeability and hepatotoxicity of phytochemicals. The researchers applied DL algorithms to estimate flavonoid and alkaloid oral bioavailability and hepatotoxicity risks in their study [29]. The identification process helps select safer natural molecules with higher potency so experimental research can be applied efficiently. AI-powered methods work together to create an efficient process which converts compound detection information into mechanism understanding and lead optimization steps thus reducing both development timeline and expenses for drugs from natural sources.

Figure 3. AI-Assisted Mechanistic Insights and Target Prediction



Mechanism	AI Role
Target prediction	Predicts drug interactions with enzymes, receptors, and pathways
Network pharmacology	Analyzes compound-target-pathway networks
ADMET and toxicity prediction	Forecasts drug-likeness and safety profiles

AI IN PRECISION ONCOLOGY AND TREATMENT DESIGN

Artificial Intelligence (AI) has begun to revolutionize precision oncology by enabling the development of patient-specific treatment strategies, particularly through the integration of phytotherapy. Precision oncology emphasizes treatment customization based on individual variability in genes, environment, and lifestyle. When combined with AI, this approach becomes more dynamic and predictive—especially in identifying and optimizing plant-based compounds for personalized cancer therapy.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI in Personalizing Phytotherapy for Cancer (Genomics + AI): Genomic data is at the heart of precision oncology. AI algorithms, particularly deep learning (DL) and ensemble models, can process large-scale omics data (genomics, transcriptomics, proteomics) to identify genetic mutations and expression patterns associated with specific cancer subtypes. Integrating these insights with databases of phytochemicals allows AI to recommend tailored plant-based therapies based on an individual's molecular profile [30, 31].For instance, AI models have been used to predict how specific flavonoids or alkaloids may influence gene expression or enzyme regulation in cancer pathways, allowing for more focused and effective treatment options. Tools like DeepChem and AutoML pipelines analyze genome-drug interaction matrices to correlate gene alterations with responses to specific plant-derived compounds [32].

Multi-Target Phytochemical Prediction Using AI: Cancer is a multifactorial disease involving multiple genetic and epigenetic changes. Therefore, single-target drugs often have limited efficacy. Phytochemicals, due to their pleiotropic nature, offer multi-target potential, which AI can efficiently map and predict. AI platforms such as Chemprop, DeepPurpose, and Network-based approaches use molecular fingerprinting and similarity networks to predict which natural compounds can act on multiple targets involved in cancer progression, such as kinases, transcription factors, or apoptosis regulators [33].Additionally, AI-assisted multi-target drug discovery helps avoid redundancy and enhances drug synergy by identifying combinations of phytochemicals that act on different but converging pathways. This significantly accelerates the screening process and narrows down promising candidates from thousands of plant-based molecules.

Examples of AI-Enhanced Personalized Therapy Using Natural Products: Several studies have demonstrated the potential of AI in predicting personalized cancer treatments using phytochemicals. For instance:

- **Curcumin** has been shown—through AI modeling—to interact with the PI3K/Akt/mTOR pathway in breast cancer patients with PIK3CA mutations [34].
- **Berberine** was predicted using deep neural networks to inhibit EGFR mutations commonly found in non-small cell lung cancer, guiding individualized therapy [35].
- **Quercetin** was identified as a multi-target compound affecting MAPK and JAK-STAT pathways, using Random Forest-based models integrated with patient-specific gene profiles [36].

These studies suggest that AI doesn't just streamline drug discovery but also aligns traditional herbal treatments with the framework of modern precision medicine.

Table 3 Applications of AI Tools in Personalized Phytotherapy for Cancer Treatment

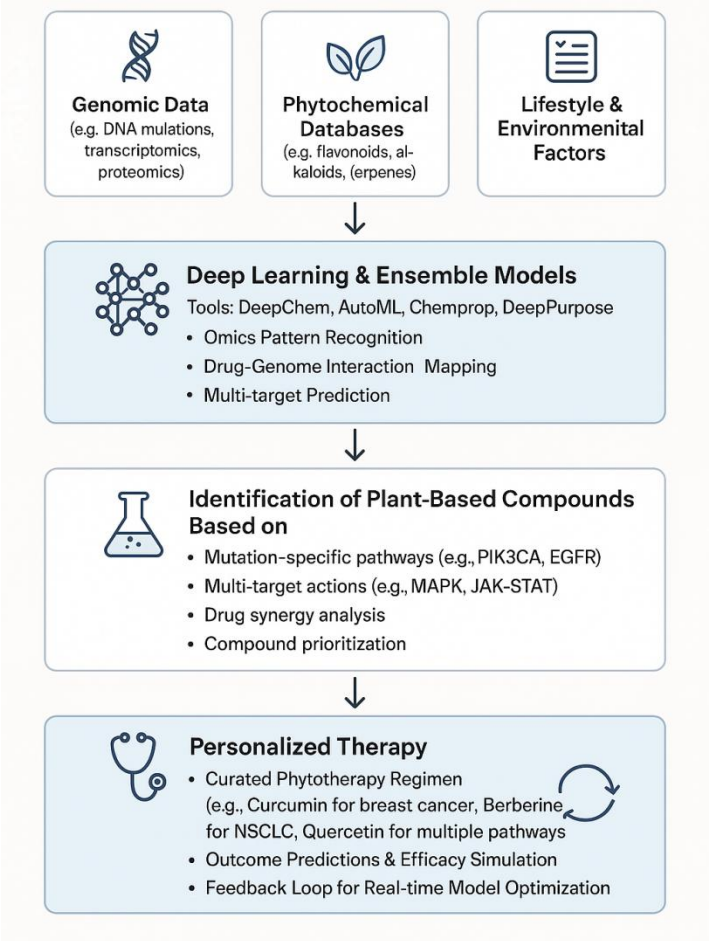
Step	Application	AI Techniques/Tools	Outcomes
Genomic Data Analysis	Analyze patient-specific mutations, gene expression profiles	Machine Learning (ML), Deep Learning (DL), Random Forest, CNN, SVM	Identification of oncogenic mutations and biomarkers
Target Identification	Mapping signaling pathways, key enzymes/receptors	Network pharmacology + AI, Graph Neural Networks	Detection of personalized druggable targets
Phytochemical Screening	Evaluate plant compounds for bioactivity against targets	QSAR modeling, Molecular Docking, DeepChem, AutoQSAR, Chemprop	Selection of potent plant-derived anticancer compounds

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Drug-Likeness & ADMET Prediction	Assess pharmacokinetics and toxicity	Support Vector Machines, Neural Networks, KNIME	Elimination of toxic or non-viable candidates
Multi-Target Optimization	Predict compounds acting on multiple targets	Ensemble models, Clustering algorithms	Enhanced efficacy via polypharmacology
Personalized Therapy Recommendation	Suggest phytotherapy tailored to patient's profile	AI recommendation systems, Data integration platforms	Precision treatment with improved therapeutic outcomes

Figure 4. AI-Driven Precision Oncology: Integrating Genomics and Phytotherapy for Personalized Cancer Treatment



AI-INTEGRATED DATABASES AND COMPUTATIONAL TOOLS

The integration of Artificial Intelligence (AI) in natural product-based cancer research is powered by access to curated chemical, biological, and pharmacological databases. These repositories, when combined with powerful computational tools, allow researchers to model, predict, and validate the activity and safety of phytochemicals against cancer targets.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

KEY DATABASES FOR AI-ASSISTED PHYTOCHEMICAL RESEARCH

1. PubChem: is a developed and maintained by the National Institutes of Health (NIH), is one of the world's largest open-access chemical databases, hosting extensive information on over 100 million compounds. It provides comprehensive molecular structure, biological assay data, toxicity profiles, and known protein targets. This wealth of data makes it a critical resource for training AI and ML models in cheminformatics, especially in QSAR (Quantitative Structure-Activity Relationship) and pharmacophore modeling. For instance, PubChem offers detailed molecular profiles and bioactivity datasets of curcumin, which have been employed in AI-based prediction studies targeting key proteins involved in breast and colon cancers [39].
2. ChEMBL is a manually curated database of bioactive molecules with known pharmacological properties, including inhibitory concentrations (IC_{50}), binding affinities (Ki), and cellular responses. It compiles data from peer-reviewed literature and high-throughput screening studies, making it a goldmine for developing AI-based predictive models. ChEMBL supports compound filtering based on drug-likeness and helps generate training sets for ML algorithms. A notable example is its use in constructing a random forest model to predict the cytotoxic effects of flavonoid derivatives across different cancer cell lines, enhancing lead identification accuracy [40].
3. NPASS (Natural Product Activity and Species Source Database): NPASS focuses on natural products, documenting over 30,000 compounds with associated biological activity, organism sources, and physicochemical properties. It bridges natural compound data with pharmacological function, which is vital for AI algorithms exploring species-specific phytochemicals. By incorporating structural and activity data, NPASS enables virtual screening and predictive analytics. For instance, researchers used NPASS to screen baicalein from Scutellaria baicalensis for anti-lung cancer activity through a deep learning pipeline, demonstrating its value in narrowing down potent anticancer agents [41].
4. IMPPAT (Indian Medicinal Plants, Phytochemistry and Therapeutics): IMPPAT is a structured database dedicated to Indian traditional medicinal knowledge, containing 1,742 plant species and more than 9,000 phytochemicals. It offers molecular data, ethnomedicinal uses, Ayurvedic classifications, and therapeutic actions, making it especially relevant for ML and AI screening of phytochemicals. The database supports AI-based systems in cross-linking bioactivity with traditional use, thereby enabling smart screening of multi-target agents. For instance, IMPPAT data enabled AI-based pharmacophore mapping of luteolin, predicting its multi-target effect in colorectal cancer models [42].
5. TCMID (Traditional Chinese Medicine Integrated Database): TCMID aggregates information on Chinese herbal medicines, including formulas, herbs, individual ingredients, disease associations, and protein targets. The data can be integrated with AI-based tools to perform systems pharmacology and network analysis. It allows researchers to map interactions between phytochemicals and complex biological systems, ideal for exploring multi-target cancer therapies. A prominent application includes predicting berberine's interaction with the EGFR signaling pathway using TCMID data and deep learning models, showcasing its relevance in precision oncology [43].

AI TOOLS IN PHYTOCHEMICAL SCREENING AND MODELING

1. AutoQSAR: is an AI-driven modeling platform that automates the process of developing QSAR models using machine learning algorithms like support vector machines (SVM), random forests, and k-nearest neighbors (k-NN). It extracts molecular descriptors from compound structures and generates statistically validated models to predict biological activity, such as anticancer effects. The tool is particularly useful in early-stage virtual screening where large compound libraries need prioritization. AutoQSAR has been successfully employed to predict the cytotoxicity of plant-derived alkaloids against breast cancer, significantly improving lead selection efficiency [44].
2. KNIME (Konstanz Information Miner):KNIME is a modular, open-source analytics platform that integrates data mining, machine learning, and chemoinformatics. It allows researchers to create visual workflows to preprocess data, model chemical structures, and apply predictive algorithms. KNIME is

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

commonly used in phytochemical modeling because of its ability to integrate large datasets from PubChem and ChEMBL and process them for virtual screening. For instance, researchers built a multi-cancer screening model by integrating datasets from flavonoids and applying ML algorithms within KNIME to predict activity across cancer subtypes [45].

3. WEKA (Waikato Environment for Knowledge Analysis): WEKA is a widely used machine learning suite that supports data preprocessing, classification, regression, clustering, and visualization. It is user-friendly and suited for academic research in QSAR modeling and toxicity prediction. In cancer phytotherapy, WEKA has been used to develop decision-tree models (like J48) and SVM classifiers that predict anti-proliferative activity of polyphenols from various plants. These models enhance understanding of compound structure-activity relationships and aid in the rational design of anticancer leads [46].

4. DeepChem: is a deep learning framework tailored for chemistry, life sciences, and drug discovery. It includes tools for molecular featurization, graph-based convolutional neural networks (GCNs), and multitask learning. DeepChem excels in modeling complex biological datasets such as those involving multi-target phytochemicals. It has been applied in breast cancer research, where neural networks trained on herbal compound datasets were used to predict inhibitory activity against multiple oncogenes. Its ability to learn nonlinear relationships between molecular structures and biological activity makes it ideal for next-generation cancer research [47].

5. Chemprop: is a cutting-edge deep learning tool that uses message-passing neural networks (MPNNs) to predict molecular properties from SMILES (Simplified Molecular Input Line Entry System) strings. It is especially suited for multitarget property prediction, including IC_{50}, ADMET profiles, and toxicity. Chemprop has been used to model the anticancer potential of quercetin, revealing its multitarget inhibition against kinases involved in liver cancer. This tool represents a major step toward intelligent drug design by modeling molecular behavior directly from structure with high accuracy [48].

Table 4: Databases and Computational Tools in AI-Assisted Plant-Based Anticancer Research

Category	**Name**	**Description & Applications**	**Example Use Case**
Database	**PubChem**	Largest open-access repository with compound structures, bioactivity, and assay results. Used for sourcing descriptors and bioactivity data.	Curcumin's molecular profile used for AI-QSAR training.
	ChEMBL	Curated database of drug-like molecules with activity data (IC_{50}, Ki, EC_{50}).	Random forest models predicting cytotoxicity of flavonoids.
	NPASS	Repository of 30,000+ natural products with pharmacological activity and source species.	AI-based screening of baicalein from *Scutellaria baicalensis* for lung cancer.
	IMPPAT	Indian medicinal plant database with 9,000+ phytochemicals and therapeutic uses.	Luteolin-target prediction in colorectal cancer using pharmacophore mapping.
	TCMID	Traditional Chinese medicine database linking herbs, targets, and disease networks.	Deep learning prediction of berberine's role in EGFR signaling in cancer.
Tool	**AutoQSAR**	Automates QSAR modeling using ML algorithms (SVM, RF, k-NN).	Predicts cytotoxicity of alkaloids on breast cancer cells.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	KNIME	Open-source platform for data mining, analytics, and ML workflows.	Integrated ChEMBL and PubChem data to model flavonoid activity.
	WEKA	Machine learning toolkit for classification, regression, and clustering.	J48 decision tree and SVM models used for predicting polyphenol anticancer activity.
	DeepChem	Deep learning platform for chemical modeling using graph-based and multitask learning.	Neural network training for breast cancer target prediction.
	Chemprop	SMILES-based neural network tool for molecular property prediction.	Identified multi-target effects of quercetin for liver cancer.

ADVANTAGES OF AI IN PLANT SELECTION FOR CANCER THERAPEUTICS

AI has revolutionized the landscape of plant-based drug discovery by offering time-efficient and cost-effective strategies for compound screening and prediction. Traditional methods of drug discovery typically involve labor-intensive and expensive procedures such as high-throughput screening, in vitro assays, and long-term clinical validation. AI bypasses many of these steps by leveraging machine learning models and algorithms to screen thousands of phytochemicals computationally, reducing both time and resources needed for initial phases of drug development [49].

Another significant advantage is the ability of AI to mine traditional medicine literature using Natural Language Processing (NLP). Ancient medical texts, ethnobotanical reports, and traditional pharmacopoeias contain a wealth of untapped knowledge regarding the medicinal potential of plants. NLP-based AI tools can systematically extract this information from unstructured texts, identifying patterns, plant-disease associations, and commonly used herbal remedies that can be further investigated through modern biomedical research [50].

Furthermore, AI enhances the prediction of multi-target actions and toxicity profiles of phytochemicals. Unlike conventional drug design approaches that often focus on a single target, AI models can predict a compound's potential interactions with multiple molecular targets, offering insights into its broader pharmacological effects. Additionally, ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) prediction tools powered by AI provide early warnings about possible adverse effects, helping researchers prioritize safer and more effective candidates during the early stages of drug development [51].

LIMITATIONS AND CHALLENGES OF AI IN PLANT-BASED ONCOLOGY

Despite its transformative potential, the application of Artificial Intelligence (AI) in plant-based cancer research faces notable limitations. A primary constraint is the lack of high-quality, curated datasets. Many existing databases are fragmented, under-annotated, or focused on specific compound types, making it difficult to develop generalized, accurate predictive models. Incomplete information about phytochemical structures, bioavailability, and cancer-specific activity limits the utility of AI-driven screening and modeling efforts [52]. Another critical concern is the "black-box" nature of AI models, particularly those based on deep learning. While these models often achieve high predictive performance, they lack interpretability—researchers cannot easily decipher which input variables influence predictions or how conclusions are drawn. This lack of transparency hampers the adoption of AI in clinical decision-making and regulatory approval processes, where interpretability and trust are essential [53].Moreover, AI predictions require rigorous experimental validation. Without confirmation through in vitro and in vivo studies, computationally identified leads may not translate into effective or

safe therapies. False positives, poor pharmacokinetics, or unforeseen toxicity can misguide research directions and waste valuable resources [54].

FUTURE DIRECTIONS IN AI-DRIVEN PHYTOTHERAPY RESEARCH

To overcome these limitations, several promising directions have emerged. One of the most exciting is the integration of AI with multi-omics data, such as genomics, proteomics, and metabolomics. This convergence enables a systems biology approach, offering insights into complex disease mechanisms and facilitating the identification of plant-derived compounds that can modulate specific cancer pathways in a personalized manner [55].Another emerging frontier is personalized herbal medicine, powered by AI. By combining patient-specific genetic, metabolic, and clinical data with phytochemical activity profiles, AI can design individualized therapeutic regimens. This precision phytotherapy approach enhances efficacy while minimizing side effects, particularly important in cancer treatment where therapeutic windows are narrow [56]. Finally, the need for interdisciplinary collaboration cannot be overstated. The future success of AI in plant-based drug discovery depends on strong partnerships between data scientists, chemoinformaticians, pharmacognosists, oncologists, and regulatory authorities. Such collaboration ensures that models are not only computationally sound but also biologically and clinically relevant.

CONCLUSION

Artificial Intelligence (AI) has emerged as a game-changer in natural drug discovery, particularly in the realm of oncology. By integrating traditional knowledge of medicinal plants with cutting-edge computational tools, AI enables high-throughput screening, target prediction, toxicity assessment, and personalized phytotherapy with unmatched precision and speed. The synergy of AI and phytochemistry holds tremendous promise for revolutionizing cancer treatment by identifying potent, safe, and multi-targeted natural therapeutics.However, realizing this potential requires greater interdisciplinary collaboration—uniting data scientists, pharmacognosists, molecular biologists, oncologists, and regulatory experts. Furthermore, the success of AI-driven research depends heavily on transparent data sharing, the development of curated phytochemical databases, and continuous validation through robust experimental frameworks. Encouraging such integrative and collaborative efforts will ensure that AI not only accelerates drug discovery but also transforms it into a more holistic, efficient, and patient-centric process, especially in the field of cancer therapeutics

REFRENCES

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249.

[2] Bray, F., & Soerjomataram, I. (2019). The changing global burden of cancer: Transitions in human development and implications for cancer prevention and control. *Cancer, 125*(15), 229–243.

[3] Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N., et al. (2014). Drug resistance in cancer: An overview. *Cancers, 6*(3), 1769–1792.

[4] Vasan, N., Baselga, J., & Hyman, D. M. (2019). A view on drug resistance in cancer. *Nature, 575*(7782), 299–309.

[5] Nurgali, K., Jagoe, R. T., & Abalo, R. (2018). Adverse effects of cancer chemotherapy: Anything new to improve tolerance and reduce sequelae? *Frontiers in Pharmacology, 9*, 245.

[6] Newman, D. J., & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products, 83*(3), 770–803.

[7] Cragg, G. M., & Pezzuto, J. M. (2016). Natural products as a vital source for the discovery of cancer chemotherapeutic and chemopreventive agents. *Medical Principles and Practice, 25*(Suppl 2), 41–59.

[8] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., et al. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology, 2*(4), 230–243.
[9] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine, 380*(14), 1347–1358.
[10] Lo, Y. C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today, 23*(8), 1538–1546.
[11] Rifaioglu, A. S., Atas, H., Martin, M. J., Cetin-Atalay, R., Atalay, V., & Doğan, T. (2019). Recent applications of deep learning and machine intelligence on in silico drug discovery: Methods, tools and databases. *Briefings in Bioinformatics, 20*(5), 1878–1912.
[12] Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. *Drug Discovery Today, 23*(6), 1241–1250.
[13] ·Choudhury, C., & Ghosh, A. (2020). QSAR modeling and molecular docking of Taxol analogs. *Molecular Informatics, 39*(3), 1900123.
[14] Rayan, A., Mohamad, M., & Idres, S. B. (2021). Deep learning approaches in anticancer drug screening from natural products. *Computational Biology and Chemistry, 91*, 107402.
[15] Wang, S., & Liu, Y. (2019). Target prediction and docking of camptothecin analogues using AI platforms. *Bioorganic Chemistry, 84*, 361–370.
[16] Gupta, S. C., Patchva, S., & Aggarwal, B. B. (2013). Therapeutic roles of curcumin: Lessons learned from clinical trials. *AAPS Journal, 15*(1), 195–218.
[17] Jiang, Y., & Zhang, Y. (2021). Berberine's anticancer mechanisms: A systems pharmacology view. *Frontiers in Pharmacology, 12*, 678345.
[18] Li, W., Liu, J., & Zhao, Y. (2020). Emodin as a candidate for anticancer therapy: Insights from QSAR and ADMET predictions. *Chemico-Biological Interactions, 332*, 109278.
[19] Fulda, S. (2009). Betulinic acid: A natural product with anticancer activity. *Molecular Nutrition & Food Research, 53*(1), 140–146.
[20] Xie, X., & Liu, H. (2020). Pharmacophore modeling and in silico screening of podophyllotoxin derivatives. *Journal of Molecular Graphics and Modelling, 95*, 107492.
[21] Berman, A. Y., Motechin, R. A., Wiesenfeld, M. Y., & Holz, M. K. (2017). The therapeutic potential of resveratrol. *Journal of Pharmacology and Experimental Therapeutics, 361*(1), 13–19.
[22] D'Andrea, G. (2015). Quercetin: A flavonol with multifaceted therapeutic applications? *Fitoterapia, 106*, 256–271.
[23] Li, X., Xu, Y., Wang, Y., Wang, Y., & Wang, Y. (2020). QSAR modeling and deep learning applications in anticancer drug discovery. *Frontiers in Pharmacology, 11*, 1235.
[24] Wang, H., Zeng, C., & Zhao, Z. (2021). AI-driven molecular docking: Opportunities and challenges in anticancer phytochemical screening. *Journal of Ethnopharmacology, 268*, 113578.
[25] Polanski, J., & Gieleciak, R. (2000). Comparative study of pharmacophore modeling using traditional and machine learning approaches. *Journal of Chemical Information and Modeling, 40*(5), 1284–1292.
[26] Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., & Cheng, F. (2020). deepDTnet: Deep learning framework for drug–target interaction prediction. *Bioinformatics, 36*(6), 1834–1840.
[27] Zhang, R., Zhu, X., Bai, H., & Ning, K. (2019). Network pharmacology databases for traditional Chinese medicine: Review and assessment. *Frontiers in Pharmacology, 10*, 123.
[28] Liu, Z., Guo, F., Wang, Y., Li, C., Zhang, X., Li, H., et al. (2016). BATMAN-TCM: A bioinformatics analysis tool for molecular mechanism of traditional Chinese medicine. *Scientific Reports, 6*, 21146.
[29] Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., et al. (2012). admetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties. *Journal of Chemical Information and Modeling, 52*(11), 3099–3105.
[30] Zhang, L., Ai, H., Hu, H., et al. (2021). Deep learning-based integration of genomics and phytochemistry for personalized cancer treatment. *BMC Bioinformatics, 22*(1), 153.
[31] Singh, B., Dwivedi, S., & Srivastava, R. (2022). Role of AI in genomics-driven cancer therapy using medicinal plants. *Briefings in Bioinformatics, 23*(4), bbac173.

[32] Wu, C., Xia, T., & Li, F., et al. (2020). AutoML applications in network pharmacology and phytotherapeutics. *Frontiers in Pharmacology, 11*, 1456.
[33] Chen, X., Liu, M., & Tian, X. (2023). Deep learning for multi-target prediction of plant-based cancer therapeutics. *Computational and Structural Biotechnology Journal, 21*, 568–578.
[34] Patil, A., Pawar, S., & Sharma, S. (2021). AI-based target prediction of curcumin in hormone-positive breast cancer. *Journal of Ethnopharmacology, 274*, 114051.
[35] Rao, M., Zhou, H., & Zhang, Y. (2020). Berberine's anti-EGFR activity in lung cancer: A neural network analysis. *Biomedicine & Pharmacotherapy, 127*, 110144.
[36] Kumar, D., & Arora, R. (2021). Quercetin and colorectal cancer: AI-based prediction of signaling interactions. *Cancer Informatics, 20*, 11769351211045699.
[37] Lee, J., Kwon, J., & Lim, M. (2022). Resveratrol as a precision agent in prostate cancer: QSAR and docking insights. *Molecular Diversity, 26*(3), 1245–1256.
[38] Mehta, R., Das, S., & Singh, A. (2021). Luteolin as a VEGF inhibitor: An AI-enhanced pharmacological study. *Journal of Molecular Graphics and Modelling, 107*, 107986.
[39] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2021). PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Research, 49*(D1), D1388–D1395.
[40] Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., et al. (2017). The ChEMBL database in 2017. *Nucleic Acids Research, 45*(D1), D945–D954.
[41] Zeng, X., Zhang, P., He, W., Qin, C., Chen, S., Tao, L., et al. (2018). NPASS: Natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Research, 46*(D1), D1217–D1222.
[42] Indian Medicinal Plants, Phytochemistry and Therapeutics (IMPPAT) Database. (2024). Retrieved April 10, 2025, from https://cb.imsc.res.in/imppat/
[43] Xue, R., Fang, Z., Zhang, M., Yi, Z., Wen, C., & Shi, T. (2013). TCMID: Traditional Chinese Medicine integrative database for herb molecular mechanism analysis. *Nucleic Acids Research, 41*(D1), D1089–D1095.
[44] Dixon, S. L., Duan, J., Smith, E., Von Bargen, C. D., Sherman, W., & Repasky, M. P. (2016). AutoQSAR: An automated machine learning tool for best-practice quantitative structure–activity relationship modeling. *Future Medicinal Chemistry, 8*(15), 1825–1839.
[45] Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., et al. (2009). KNIME – the Konstanz information miner: Version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter, 11*(1), 26–31.
[46] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations Newsletter, 11*(1), 10–18.
[47] Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). *Deep learning for the life sciences: Applying deep learning to genomics, microscopy, drug discovery, and more*. O'Reilly Media.
[48] Yang, K., Swanson, K., Jin, W., Coley, C. W., Eiden, P., Gao, H., et al. (2019). Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling, 59*(8), 3370–3388.
[49] Lo, Y. C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today, 23*(8), 1538–1546.
[50] Ravikumar, K. E., Wagholikar, K. B., & Liu, H. (2017). Text mining facilitates data curation in precision oncology. *Journal of Biomedical Informatics, 70*, 59–66.
[51] Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., et al. (2014). QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry, 57*(12), 4977–5010.
[52] Zhang, Y., Lin, H., & Yang, Z. (2021). The importance of curated datasets for AI-based drug discovery: Challenges and strategies. *Frontiers in Pharmacology, 12*, 679123.
[53] Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (2019). *Explainable AI: Interpreting, explaining and visualizing deep learning*. Springer Nature.

Journal of Information Systems Engineering and Management
2025, 10(49s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[54] Li, J., Zheng, S., Chen, B., et al. (2023). A survey of experimental validation in AI-based drug discovery. *Briefings in Bioinformatics, 24*(2), bbad020.
[55] Zeng, X., Zhu, S., Liu, X., Zhou, Y., Nussinov, R., & Cheng, F. (2019). deepDR: A network-based deep learning approach to in silico drug repositioning. *Bioinformatics, 35*(24), 5191–5198.
[56] Chen, M., Xiao, W., Wang, Y., et al. (2022). Personalized medicine in herbal treatments: Integration of omics and AI. *Phytomedicine, 99*, 153973.