Mini-Review Article

Computational Approaches for Predicting the Biological Activities of Phytochemicals: Advances and Challenges

Naveena Sakthivel and Gnanendra Shanmugam*

Department of Biotechnology, Vivekanandha College of arts and Sciences for Women (Autonomous), Elayampalyam, Tiruchengode, Tamilnadu, India.

*Correspondence: gnani@vicas.org (GS)

Abstract : Phytochemicals, a diverse class of bioactive compounds derived from plants, play a pivotal role in modern drug discovery and natural product research. The ability to accurately predict their biological activities based on chemical structure is essential for enhancing screening efficiency and minimizing experimental costs. This review presents an overview of state-of-the-art computational strategies employed to forecast the pharmacological potential of phytochemicals. Key methodologies discussed include quantitative structure–activity relationship (QSAR) modeling, molecular docking, cheminformatics tools, and machine learning algorithms. Emphasis is placed on the use of molecular descriptors and structural fingerprints for functional classification of phytochemicals into categories such as antioxidant, anticancer, antimicrobial, and anti-inflammatory agents. The review also addresses current challenges, including limitations in data availability, issues of model interpretability, and the critical need for experimental validation. With ongoing advancements in artificial intelligence and big data analytics, predictive modeling continues to evolve, offering transformative opportunities for the identification and development of plant-derived therapeutics. Integrating computational predictions with empirical research holds significant promise for accelerating the discovery of novel bioactive compounds.

Keywords: phytochemicals, biological activities, QSAR, machine learning, drug discovery.

Citation: Naveena Sakthivel and Gnanendra S. Computational Approaches for Predicting the Biological Activities of Phytochemicals: Advances and Challenge. Int J Adv Interdis Res 2025, e005.

Received: 05 May 2025
Revised: 10 May 2025
Accepted: 12 May 2025
Published: 15 May 2025

Introduction

Medicinal plants, commonly referred to as medicinal herbs, have been used in traditional healing systems since prehistoric times, including Indian Ayurveda, Traditional Chinese Medicine, and African herbal practices [1]. Their therapeutic effects are primarily attributed to phytochemicals—biologically active compounds synthesized and stored as secondary metabolites [2]. These metabolites, which accumulate in different plant parts such as roots, leaves, stems, flowers, and bark, serve ecological functions such as protection against pathogens, abiotic stress, and ultraviolet radiation, without impeding the plant's growth or reproduction [5]. Due to their evolutionary refinement, these compounds often display high structural diversity and specific biological activities, rendering them attractive for medicinal research [3,4].

However, phytochemicals are generally less accessible in pure form compared to synthetic alternatives. As a result, systematic strategies have been developed to investigate the phytochemical composition of medicinal plants. These include the classification of chemical groups and identification of

active constituents through additive model analysis and other computational techniques [6]. Many phytochemicals have demonstrated a broad spectrum of bioactivities, including antioxidant, anti-inflammatory, anticancer, and antibacterial properties, further underscoring their pharmaceutical relevance [6]. Advancements in computational tools have accelerated the exploration of structure–activity relationships among natural products. A growing body of research demonstrates clear associations between the chemical structure of natural molecules—such as the configuration of the carbon backbone, the position and type of functional groups, and the characteristics of side chains—and their biological activities. These relationships are increasingly modeled using in silico techniques such as Quantitative Structure–Activity Relationship (QSAR) analysis, which quantitatively links chemical structure to pharmacological effect [7,8]. To support these computational endeavors, several phytochemical databases have been developed to consolidate chemical, biological, and experimental information, enhancing the accessibility and reproducibility of data for drug discovery efforts [7].

This mini-review provides an overview of key computational methods for predicting the biological activity of phytochemicals, focusing on QSAR modeling, molecular descriptors, cheminformatics, and machine learning. It also discusses current limitations, including data quality, model interpretability, and the necessity for experimental validation, while highlighting emerging opportunities driven by artificial intelligence and big data analytics.

Biological Activities

Anticancer Activity

Medicinal plants have long been recognized as a valuable source of bioactive compounds with anticancer potential. Beyond their antitumor efficacy, these phytochemicals offer several advantages, including low toxicity, affordability, and widespread availability [9]. Cancer is a multifactorial disease characterized by a complex interplay of physical, environmental, metabolic, chemical, and genetic factors that contribute to its initiation and progression. A wide range of polyphenolic compounds—such as flavonoids, phenolic acids, anthocyanidins, and tannins—has demonstrated substantial pharmacological activity, particularly in modulating cancer-related pathways [10]. Recent studies underscore the therapeutic potential of bioactive compounds present in commonly used herbs, noting their ability to exert antioxidant, anti-inflammatory, antimicrobial, and anticancer effects when consumed in appropriate doses. Empirical research and traditional medical practices both support the efficacy of medicinal plants in the development of anticancer agents. Several phytochemicals have been successfully integrated into clinical applications, either in their natural forms or as chemically modified derivatives. Notable examples include vinblastine, vincristine, podophyllotoxin, paclitaxel (Taxol), and camptothecin—compounds derived from plants that have become standard components in modern chemotherapeutic regimens [11]. This review

explores these compounds' therapeutic effects and mechanisms of action while also addressing the challenges and limitations associated with their use, such as bioavailability, standardization, and regulatory hurdles [12].

Antiviral Activity

An increasing body of evidence supports the antiviral properties of various plant-derived compounds. For instance, rutin, a flavonoid glycoside found in numerous plant species, has demonstrated significant efficacy against avian influenza viruses [13]. Quercetin, a metabolite of rutin, is widely distributed in plants and has shown inhibitory activity against a range of viruses, including rhinovirus, dengue virus type-2, herpes simplex virus type 1 (HSV-1), poliovirus, adenovirus, and highly pathogenic influenza strains [14]. Comparative studies of several flavonoids—including epigallocatechin gallate (EGCG), epicatechin gallate (ECG), quercetin, daidzein, fisetin, baicalein, kaempferol, biochanin A, theaflavin, and digallate—have demonstrated notable antiviral activity, particularly against RNA viruses such as murine norovirus (MNV) and feline calicivirus (FCV) [15]. These findings reinforce previous observations of the broad-spectrum antiviral potential of flavonoids, especially in targeting RNA virus replication mechanisms [16].

Anti-inflammatory Activity

Inflammation is a complex biological response to infection, injury, or stress, often triggered by pathogenic microorganisms such as bacteria, viruses, and fungi. Medicinal plants have played a central role in traditional healthcare systems for managing inflammatory conditions [17]. In their response to environmental stressors and microbial attacks, plants produce a range of secondary metabolites with diverse biological effects, including pronounced anti-inflammatory activity [18]. The anti-inflammatory efficacy of several flavonoid aglycones—kaempferol, quercetin, apigenin, chrysin, diosmetin, luteolin, daidzein, genistein, and hesperetin—has been extensively studied. Among these, luteolin was found to exert the most potent inhibitory effects on inflammatory mediators such as nitric oxide (NO) and tumor necrosis factor-alpha (TNF-α) [19,20]. The increasing application of phytomedicine in managing chronic inflammatory disorders such as rheumatoid arthritis (RA) and inflammatory bowel disease (IBD) highlights the growing recognition of plant-based therapies as accessible, cost-effective alternatives to conventional drugs [21,22]. These natural products represent a promising resource for the development of next-generation anti-inflammatory agents.

Prediction Tools

Quantitative Structure–Activity Relationship (QSAR)

Quantitative Structure–Activity Relationships (QSARs) are mathematical models that establish correlations between the chemical structure of a compound and its biological activity [23]. These models rely on regression analysis, classification algorithms, and pattern recognition techniques to

quantitatively predict pharmacological outcomes. QSAR workflows typically involve several stages: data collection, calculation of molecular descriptors, feature selection, model building, validation, and prediction. Descriptor calculation—capturing physicochemical, electronic, and topological properties—is commonly performed using tools such as PaDEL-Descriptor or Dragon [25]. Feature selection algorithms like principal component analysis (PCA) and random forest are employed to extract the most informative molecular descriptors, minimizing redundancy and enhancing model accuracy. Predictive modeling is then carried out using statistical and machine learning approaches, including multiple linear regression (MLR), support vector machines (SVM), and artificial neural networks (ANN) [24,26]. Model performance is validated through internal (e.g., cross-validation) and external (e.g., test set) methods to ensure predictive reliability.

SWISS Target Prediction

SwissTargetPrediction is a web-based tool designed to identify probable biological targets for small molecules, leveraging both 2D and 3D similarity to known ligands [27]. The tool provides prediction scores ranging from 0 to 1, indicating the likelihood of interaction with a given target. In practical applications, this platform has been used to assess large compound datasets—such as 250 phytochemical structures—by predicting their top-ranked molecular targets [28]. SwissTargetPrediction supports early-stage drug discovery by offering insights into the molecular targets of novel compounds, thus streamlining target validation and hit-to-lead optimization. The tool accepts input in the form of SMILES strings, returning ranked targets along with associated bioactivity data, and facilitating integration with experimental validation workflows [29].

PASS (Prediction of Activity Spectra for Substances)

PASS (Prediction of Activity Spectra for Substances) is another widely used web-based tool that forecasts the biological activity profile of organic compounds based solely on their structural formula [30]. The system draws on a training dataset of over one million biologically active molecules and predicts more than 4,000 pharmacological and biochemical activities [31,32]. It assigns a probability value (Pa) for each predicted activity, indicating the confidence level of the prediction. PASS has been used to explore diverse biological functions, including anticonvulsant effects, neurotransmitter modulation, anti-inflammatory properties, and activities targeting neurodegenerative conditions [33]. In antimicrobial research, PASS has helped to identify thymidine analogs with potent antibacterial activity exceeding their antifungal efficacy, thus demonstrating its value in guiding in vitro validation [34].

Machine Learning Model Development and Evaluation

Machine learning (ML) is increasingly integral to predictive phytochemical research, providing powerful methods for modeling complex relationships

between molecular features and biological outcomes. These approaches are particularly valuable for elucidating mechanisms of action (MoA) and identifying drug targets for phytochemicals whose effects are not yet fully understood [35]. In recent studies, supervised algorithms such as k-nearest neighbors (kNN), support vector machines (SVM), random forests (RF), and extreme gradient boosting (XGB) have been utilized to develop predictive models for the anticancer activity of plant-derived compounds, particularly against breast cancer [36]. Among these, the kNN algorithm is favored for its simplicity and effectiveness in classifying compounds based on chemical similarity [37]. Random forest, an ensemble technique, aggregates outputs from multiple decision trees to enhance model stability and performance in both classification and regression tasks [38]. These ML models often employ molecular fingerprints like MACCS and Morgan2 for structural representation. By mapping molecular features to biological activity, ML techniques offer new avenues for phytochemical drug design, enabling predictions of efficacy and patient-specific responses. These innovations are transforming natural product research by accelerating lead compound identification and optimizing therapeutic development pipelines [39].

Conclusion

This review underscores the critical role of computational approaches in advancing phytochemical research and drug discovery. The ability to predict the biological activities of phytochemicals based on their chemical structures has become a cornerstone of modern pharmacognosy and natural product chemistry. Advances in cheminformatics, molecular docking, and machine learning have significantly improved the identification of bioactive compounds, elucidation of mechanisms of action, and optimization of therapeutic potential. By integrating in silico predictions with experimental validation, researchers can streamline the drug discovery pipeline, reduce associated costs, and accelerate the development of plant-derived therapeutics. Despite these advances, key challenges remain—particularly in ensuring data quality, enhancing model interpretability, and capturing the complexity of biological systems. Moving forward, emphasis should be placed on refining predictive algorithms, improving the transparency of machine learning models, and expanding access to comprehensive, well-curated phytochemical and bioactivity datasets.

References

[1] Ingle SG, Gade AK, Hedawoo GB. Systematic Review on Phytochemicals Structure and Activity Databases. Phytomedicine Plus. 2024 Sep 30:100644.

[2] Vitale S, Colanero S, Placidi M, Di Emidio G, Tatone C, Amicarelli F, D'Alessandro AM. Phytochemistry and biological activity of medicinal plants in wound healing: an overview of current research. Molecules. 2022 Jun 1;27(11):3566.

[3] Koza NA, Adedayo AA, Babalola OO, Kappo AP. Microorganisms in plant growth and development: roles in abiotic stress tolerance and secondary metabolites secretion. Microorganisms. 2022 Jul 28;10(8):1528.

[4] Hussein RA, El-Anssary AA. Plants secondary metabolites: the key drivers of the pharmacological actions of medicinal plants. Herbal medicine. 2019 Jan 30;1(3):11-30.

[5] Al-Khayri JM, Rashmi R, Toppo V, Chole PB, Banadka A, Sudheer WN, Nagella P, Shehata WF, Al-Mssallem MQ, Alessa FM, Almaghasla MI. Plant secondary metabolites: The weapons for biotic stress management. Metabolites. 2023 May 31;13(6):716.

[6] Jakubczyk A, Złotek U, Szymanowska U, Rybczyńska-Tkaczyk K, Jęderka K, Lewicki S. In vitro antioxidant, anti-inflammatory, anti-metabolic syndrome, antimicrobial, and anticancer effect of phenolic acids isolated from fresh Lovage Leaves [Levisticum officinale Koch] Elicited with Jasmonic Acid and Yeast Extract. Antioxidants. 2020 Jun 25;9(6):554.

[7] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, Li B, Madabhushi A, Shah P, Spitzer M, Zhao S. Applications of machine learning in drug discovery and development. Nature reviews Drug discovery. 2019 Jun;18(6):463-77.

[8] Nahar L, Sarker SD. Computational phytochemistry: An overview. Computational Phytochemistry. 2024 Jan 1:1-58.

[9] Khan T, Ali M, Khan A, Nisar P, Jan SA, Afridi S, Shinwari ZK. Anticancer plants: A review of the active phytochemicals, applications in animal models, and regulatory aspects. Biomolecules. 2019 Dec 27;10(1):47.

[10] Teodor ED, Ungureanu O, Gatea F, Radu GL. The potential of flavonoids and tannins from medicinal plants as anticancer agents. Anti-Cancer Agents in Medicinal Chemistry-Anti-Cancer Agents). 2020 Dec 1;20(18):2216-27.

[11] Pathak K, Pathak MP, Saikia R, Gogoi U, Sahariah JJ, Zothantluanga JH, Samanta A, Das A. Cancer chemotherapy via natural bioactive compounds. Current Drug Discovery Technologies. 2022 Jul 1;19(4):4-23.

[12] Chintada V, Golla N. Exploring the Therapeutic Potential of Bioactive Compounds from Plant Sources. InBiotechnological Intervention in Production of Bioactive Compounds: Biosynthesis, Characterization and Applications 2025 Feb 26 (pp. 229-247). Cham: Springer Nature Switzerland.

[13] Behl T, Rocchetti G, Chadha S, Zengin G, Bungau S, Kumar A, Mehta V, Uddin MS, Khullar G, Setia D, Arora S. Phytochemicals from plant foods as potential source of antiviral agents: An overview. Pharmaceuticals. 2021 Apr 19;14(4):381.

[14] Ross IA. Plant-Based Therapeutics, Volume 1: Cannabis Sativa. Springer Nature; 2023 Aug 8.

[15] Lalani S, Poh CL. Flavonoids as antiviral agents for enterovirus A71 (EV-A71). Viruses. 2020 Feb 6;12(2):184.

[16] Choudhury A. From plants to animals: A study of phytopathogenic infections. InViral, Parasitic, Bacterial, and Fungal Infections 2023 Jan 1 (pp. 29-36). Academic Press.

[17] Bhatla SC, Lal MA. Secondary metabolites. InPlant physiology, development and metabolism 2023 Dec 5 (pp. 765-808). Singapore: Springer Nature Singapore.

[18] Elshafie HS, Camele I, Mohamed AA. A comprehensive review on the biological, agricultural and pharmaceutical properties of secondary metabolites based-plant origin. International journal of molecular sciences. 2023 Feb 7;24(4):3266.

[19] Virshette SJ, Patil MK, Somkuwar AP. A review on medicinal plants used as anti-inflammatory agents. J. Pharmacogn. Phytochem. 2019 Apr;8(4):1641-6.

[20] Gautam RK, Roy K, Thapa G, Arora D, Parashar S, Gurung B, Deb L. Perspective of plant medicine in therapy of rheumatoid arthritis. Indian J Pharm Sci. 2020

Oct 30;82(5):741-65.

[21] Sarker SD, Nahar L, editors. Computational phytochemistry. Elsevier; 2024 Mar 6.

[22] Huang T, Sun G, Zhao L, Zhang N, Zhong R, Peng Y. Quantitative structure-activity relationship (QSAR) studies on the toxic effects of nitroaromatic compounds (NACs): A systematic review. International Journal of Molecular Sciences. 2021 Aug 9;22(16):8557.

[23] De P, Kar S, Ambure P, Roy K. Prediction reliability of QSAR models: an overview of various validation tools. Archives of Toxicology. 2022 May;96(5):1279-95.

[24] Consonni V, Ballabio D, Todeschini R. Chemical space and molecular descriptors for QSAR studies. Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development. 2023 Jan 1:303-27.

[25] Abdalfattah S, Knorz C, Ayoobi A, Omer EA, Rosellini M, Riedl M, Meesters C, Efferth T. Identification of Antagonistic Action of Pyrrolizidine Alkaloids in Muscarinic Acetylcholine Receptor M1 by Computational Target Prediction Analysis. Pharmaceuticals. 2024 Jan 8;17(1):80.

[26] Abdulellah AK, Khudair WA, Jihad SM, Fadhil HR. Molecular Modeling, ADME and Density Functional Theory Studies of Novel Acylhydrazone Derivatives Targeted Epidermal Growth Factor Receptor.

[27] Alzahrani A. Identification of Therapeutic Agents Targeting Mitogen-Activated Protein Kinase 14 in the Treatment and Management of Asthma. Biosciences Biotechnology Research Asia. 2024 Jul 1;21(2):583-98.

[28] Khaiitova M. Computer-aided evaluation of targets and biological activity spectra for new piperidine derivatives. Journal of Clinical Medicine of Kazakhstan. 2023;20(4):60-7.

[29] Shamsuddin T, Hosen MA, Alam MS, Emran TB, Kawsar SM. Uridine derivatives: Antifungal, PASS outcomes, ADME/T, drug-likeliness, molecular docking and binding energy calculations. Med. Sci. Int. Med. J. 2021;10(4):1373-86.

[30] Rudik A, Dmitriev A, Lagunin A, Filimonov D, Poroikov V. MetaPASS: a web application for analyzing the biological activity spectrum of organic compounds taking into account their biotransformation. Molecular informatics. 2021 Apr;40(4):2000231.

[31] Muduli SK, Sharma I, Sharma N, Khurana N. A systemic review and PASS assisted prediction of phytoconstituents for epilepsy. InAIP Conference Proceedings 2023 Sep 8 (Vol. 2800, No. 1). AIP Publishing.

[32] Hosen MA, Munia NS, Al-Ghorbani M, Baashen M, Almalki FA, Hadda TB, Ali F, Mahmud S, Saleh MA, Laaroussi H, Kawsar SM. Synthesis, antimicrobial, molecular docking and molecular dynamics studies of lauroyl thymidine analogs against SARS-CoV-2: POM study and identification of the pharmacophore sites. Bioorganic Chemistry. 2022 Aug 1;125:105850.

[33] Chihomvu P, Ganesan A, Gibbons S, Woollard K, Hayes MA. Phytochemicals in drug discovery—A confluence of tradition and innovation. International journal of molecular sciences. 2024 Aug 13;25(16):8792.

[34] Das AP, Agarwal SM. PhyIndBC: Development of a machine learning tool for screening of potential breast cancer inhibitors from phytochemicals. Current Plant Biology. 2025 Jan 6:100435.

[35] Samad A, Ajmal A, Mahmood A, Khurshid B, Li P, Jan SM, Rehman AU, He P, Abdalla AN, Umair M, Hu J. Identification of novel inhibitors for SARS-CoV-2 as therapeutic options using machine learning-based virtual screening, molecular docking and MD simulation. Frontiers in Molecular Biosciences. 2023 Mar 7;10:1060076.

[36] Idowu SO, Fatokun AA. Artificial intelligence (AI) to the rescue: deploying machine

learning to bridge the biorelevance gap in antioxidant assays. SLAS TECHNOLOGY: Translating Life Sciences Innovation. 2021 Feb;26(1):16-25.

[37] Mihaylova S, Tsvetkova A, Georgieva E, Vankova D. Bioactive phyto-compounds with antimicrobial effects and AI: Results of a desk research study. Microorganisms. 2024 May 24;12(6):1055.

[38] Egbuna C, Rudrapal M, Tijjani H, editors. Phytochemistry, computational tools, and databases in drug discovery. Elsevier; 2022 Nov 30.