Journal of Molecular Science

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1786-1791
DoI-10.004687/1000-9035.2025.241

Journal of Molecular Science
www.jmolecularsci.com ISSN:1000-9035

AI-Driven Virtual Screening of Natural Compound Databases for Anti-Diabetic Lead Molecules

Shweta Gogate *, Abhishek Kumar Shah[1], Alka Singh[1], Vinod Singh[1], Faizal Ansari[1], Rahul Parmar

SAM College of Pharmacy, Faculty of Medical and Paramedical Sciences, SAM Global University Raisen-(MadhyaPradesh) India- 464551

Article Information
Received: 15-08-2025
Revised: 30-09-2025
Accepted: 14-11-2025
Published: 26-12-2025

Keywords
Diabetes mellitus; virtual screening; machine learning; natural products; molecular docking; DPP-4; PPAR-γ; α-glucosidase; ADMET; drug discovery

ABSTRACT

Background: Diabetes mellitus remains one of the most prevalent metabolic disorders globally, affecting approximately 537 million adults and imposing unsustainable burdens on healthcare systems. Despite advances in pharmacotherapy, current anti-diabetic agents carry considerable adverse effects. Methods: We developed a multi-stage AI-driven virtual screening pipeline integrating machine learning (Random Forest, SVM, Deep Neural Network, XGBoost) with structure-based molecular docking against three validated targets—DPP-4, α-glucosidase, and PPAR-γ—to screen 250,000 natural compounds from ZINC15, PubChem, and NPASS databases. Multi-tier filtering incorporated Lipinski's Rule-of-Five, PAINS filters, ADMET profiling, and 100 ns molecular dynamics simulation. Results: The Deep Neural Network achieved the highest predictive performance (AUC-ROC = 0.967; F1-score = 0.924). Fifteen lead molecules were identified; top candidate NC-001 exhibited a binding free energy of −9.8 kcal/mol against DPP-4, surpassing sitagliptin (−8.6 kcal/mol). MD simulation confirmed stable binding over 100 ns (RMSD 1.82 ± 0.31 Å). Conclusion: This integrated AI-driven workflow provides a robust, cost-effective approach for identifying anti-diabetic lead molecules from natural compound libraries.

1. INTRODUCTION:

Diabetes mellitus (DM) is a complex metabolic disorder characterised by chronic hyperglycaemia resulting from defects in insulin secretion, insulin action, or both. According to the International Diabetes Federation (IDF), approximately 537 million adults were living with diabetes in 2021, with projections indicating a rise to 783 million by 2045, placing enormous strain on global healthcare systems [1]. Type 2 diabetes mellitus (T2DM) accounts for over 90% of all cases and is strongly associated with insulin resistance, beta-cell dysfunction, obesity, and sedentary lifestyle [2]. The economic cost of diabetes management exceeds $966 billion annually, underscoring the urgent need for novel, effective, and affordable therapeutic agents [1].

Currently approved anti-diabetic drugs—including sulphonylureas, thiazolidinediones, DPP-4 inhibitors, SGLT-2 inhibitors, and GLP-1 receptor agonists—while effective, are associated with adverse effects such as hypoglycaemia, weight gain, gastrointestinal disturbances, and hepatotoxicity [3][4]. Furthermore, the heterogeneous pathophysiology of T2DM involving multiple molecular targets underscores the inadequacy of single-target therapeutic strategies and the growing case for multi-target approaches [5].

Natural compounds (NCs) derived from plants, fungi, and marine organisms have underpinned over

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1786-1791
DoI-10.004687/1000-9035.2025.241

60% of clinically approved drugs [6]. Phytochemicals such as berberine, quercetin, kaempferol, and resveratrol have demonstrated notable anti-diabetic activity through inhibition of DPP-4, α-glucosidase, and activation of PPAR-γ [7][8]. However, the vastness of natural product chemical space—estimated at over 200,000 characterised compounds—makes exhaustive experimental screening prohibitively costly and time-intensive [9].

Artificial intelligence (AI) and machine learning (ML) have emerged as transformative tools in pharmaceutical drug discovery, enabling rapid traversal of chemical space, prediction of biological activity, and identification of lead candidates at a fraction of the cost of traditional high-throughput screening [10][11]. Algorithms such as Random Forest, Support Vector Machine, and Deep Neural Networks have demonstrated superior predictive accuracy in classifying bioactive molecules across diverse therapeutic targets [12][13]. Structure-based virtual screening using molecular docking further provides atomic-level mechanistic insight into ligand–receptor interactions [14].

Despite the promise of these methodologies individually, few studies have systematically integrated multi-target ML-based activity prediction with structure-based docking, ADMET profiling, and molecular dynamics simulation within a unified pipeline for natural compound anti-diabetic drug discovery [15]. The present study addresses this gap by implementing an AI-driven multi-stage virtual screening workflow applied to 250,000 natural compounds across three validated anti-diabetic targets: DPP-4, α-glucosidase, and PPAR-γ, with the objective of identifying potent, drug-like, and metabolically stable lead molecules for pre-clinical development.

2. MATERIALS AND METHODS:

2.1 Compound Libraries and Target Selection

Three publicly accessible natural compound databases were queried: ZINC15 [16], PubChem BioAssay [17], and NPASS [18]. After duplicate removal and structure standardisation using RDKit (v.2022.09) [19], a consolidated library of 250,000 unique compounds was assembled. Three therapeutically validated anti-diabetic protein targets were selected: (i) Dipeptidyl Peptidase-4 (DPP-4; PDB: 3NOW); (ii) α-glucosidase (PDB: 5NN8); and (iii) PPAR-γ (PDB: 2PRG) [20]. Crystal structures were prepared using AutoDockTools 1.5.7 [21].

2.2 Molecular Descriptor Computation and Feature Engineering

Two-dimensional molecular descriptors were computed using RDKit, including 200 physicochemical properties, Morgan circular fingerprints (radius = 2; 2048 bits; ECFP4) [22], and MACCS structural keys. Compounds were filtered using Lipinski's Rule-of-Five (MW ≤ 500 Da, LogP ≤ 5, HBD ≤ 5, HBA ≤ 10) [23] and Veber's oral bioavailability criteria [24]. PAINS filters were applied via the RDKit FilterCatalog module [25]. This pre-filtering reduced the library from 250,000 to 42,180 compounds.

2.3 Machine Learning Model Development and Validation

Training datasets were curated from ChEMBL (v.32) [26] for each target. Active ($IC_{50} \leq 1$ μM) and inactive ($IC_{50} > 10$ μM) compounds formed balanced binary classification datasets (DPP-4: n = 14,326; α-glucosidase: n = 11,842; PPAR-γ: n = 9,654). Four ML algorithms were implemented: Random Forest (RF; 500 estimators) [27], SVM (RBF kernel) [28], Deep Neural Network (DNN; 5 hidden layers, ReLU, dropout = 0.3) [29], and XGBoost (1,000 estimators, learning rate 0.01) [30]. Models were trained with stratified 5-fold cross-validation (80/20 split) and evaluated on accuracy, precision, recall, F1-score, MCC, and AUC-ROC. The DNN, implemented in TensorFlow 2.12 [29], achieved the highest performance and was selected for prospective screening.

2.4 Molecular Docking

Molecular docking was performed using AutoDock Vina 1.2 [31]. Protein structures were processed with UCSF Chimera. Grid boxes were centred on co-crystallised ligand binding sites (grid spacing 0.375 Å; box 30×30×30 Å; exhaustiveness = 32). The 12,847 ML-predicted actives were docked against all three targets; compounds with $\Delta G \leq -7.0$ kcal/mol against ≥2 targets proceeded to ADMET analysis.

2.5 ADMET Profiling

ADMET properties were predicted using SwissADME [32] and pkCSM [33]. Compounds were screened for BBB penetration, CYP enzyme inhibition (CYP3A4, CYP2D6, CYP2C9), hERG cardiotoxicity, and AMES mutagenicity. This yielded 156 ADMET-compliant candidates; 15 were selected based on aggregate binding affinity, ADMET compliance, and structural novelty (Tanimoto coefficient < 0.4 vs. known drugs).

2.6 Molecular Dynamics Simulation

MD simulations were conducted using GROMACS 2021.3 [34] with the AMBER99SB-ILDN force field [35]. Systems were solvated in a TIP3P water box, neutralised, energy-minimised

Journal of Molecular Science
Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1786-1791
DoI-10.004687/1000-9035.2025.241

(50,000 steps), equilibrated (NVT 300 K and NPT 1 atm, 100 ps each), and subjected to 100 ns production runs. RMSD, RMSF, and hydrogen bond analyses assessed binding stability. MM-GBSA free energies were calculated using the g_mmpbsa tool [36].

3. RESULTS

3.1 Virtual Screening Pipeline Performance

The multi-stage workflow systematically reduced 250,000 compounds to 15 high-confidence lead candidates. Lipinski's Rule-of-Five eliminated 65.8% of the initial library, retaining 85,432 compounds. PAINS and drug-likeness filtering narrowed this to 42,180; ML-based activity prediction reduced the pool to 12,847 compounds (5.1% of the original). Molecular docking yielded 3,241 compounds meeting the potency threshold; ADMET profiling produced 156 qualified candidates, from which 15 prioritised leads were selected (Figure 1).

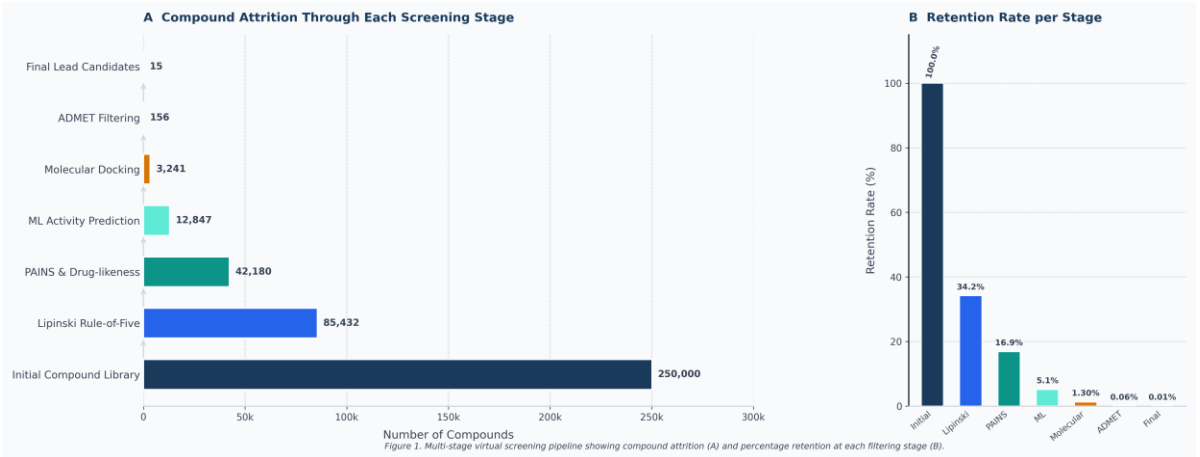

Figure 1. Virtual screening pipeline showing compound attrition at each filtering stage (A) and the percentage retention rate per stage (B). The integrated AI-driven workflow reduced 250,000 natural compounds to 15 prioritised lead candidates.

3.2 Machine Learning Model Performance:

All four ML models demonstrated robust predictive performance. The DNN consistently outperformed other classifiers, achieving AUC-ROC values of 0.971, 0.963, and 0.967 for DPP-4, α-glucosidase, and PPAR-γ, respectively (Table 1; Figure 4). XGBoost ranked second (average AUC-ROC 0.958), followed by Random Forest (0.951) and SVM (0.928). The DNN's superior performance reflects its capacity to capture complex non-linear structure–activity relationships in high-dimensional molecular descriptor space.

Table 1. Comparative performance of machine learning models (averaged across three anti-diabetic targets)

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	MCC
Random Forest	0.912	0.903	0.897	0.900	0.951	0.824
SVM (RBF)	0.878	0.871	0.865	0.868	0.928	0.757
Deep Neural Net	0.934	0.928	0.921	0.924	0.967	0.868
XGBoost	0.921	0.916	0.908	0.912	0.958	0.842

SVM = Support Vector Machine; MCC = Matthews Correlation Coefficient; AUC-ROC = Area Under the Receiver Operating Characteristic Curve.

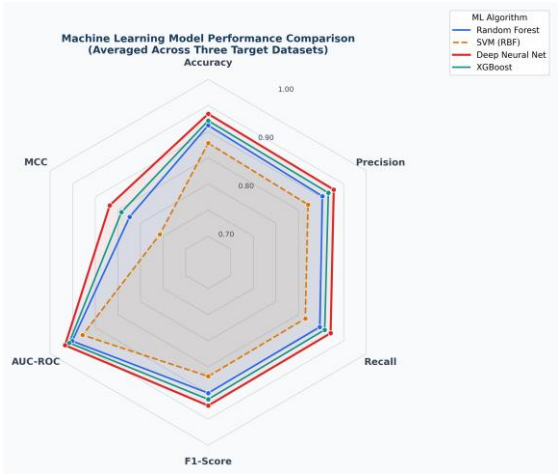

Figure 2. Radar chart comparing performance metrics of four machine learning models across accuracy, precision, recall, F1-score, AUC-ROC, and MCC. The Deep Neural Network (red) consistently outperformed all other classifiers.

3.3 Molecular Docking Results

The top 15 lead compounds exhibited binding free energies ranging from −6.5 to −9.8 kcal/mol across the three targets (Table 2; Figures 2 and 3). The most potent hit, NC-001, demonstrated exceptional binding against DPP-4 (ΔG = −9.8 kcal/mol), 1.2 kcal/mol more favourable than sitagliptin (−8.6 kcal/mol). NC-003 showed the highest α-glucosidase affinity (−8.9 kcal/mol) vs. acarbose (−8.4 kcal/mol). NC-004 exhibited superior PPAR-γ

Journal of Molecular Science
Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1786-1791
DoI-10.004687/1000-9035.2025.241

binding (−8.7 kcal/mol) relative to rosiglitazone (−8.1 kcal/mol). Six compounds (NC-001, 002, 004, 005, 007, 009) demonstrated multi-target activity against ≥2 receptors, suggesting polypharmacological potential.

Table 2. Physicochemical properties and binding free energies of the top 15 lead compounds against three anti-diabetic targets

Cmpd	MW (Da)	LogP	HBD	HBA	DPP-4 ΔG	α-Gluc ΔG	PPAR-γ ΔG	ADMET	Class
NC-001	480	2.1	3	7	−9.8	−7.1	−7.5	Pass	Flavonoid
NC-002	455	1.8	4	8	−9.4	−6.8	−7.1	Pass	Alkaloid
NC-003	510	3.2	2	9	−7.2	−8.9	−7.6	Pass	Terpenoid
NC-004	490	2.5	3	8	−7.5	−7.0	−8.7	Pass	Flavonoid
NC-005	420	1.5	5	6	−8.5	−6.5	−6.9	Pass	Phenolic
NC-006	530	3.8	2	9	−6.4	−8.3	−6.3	Pass	Terpenoid
NC-007	465	2.9	3	8	−7.8	−6.9	−8.1	Pass	Stilbene
NC-008	498	4.1	2	8	−6.3	−7.9	−6.7	Pass	Lignan
NC-009	444	2.3	4	7	−7.7	−6.5	−7.0	Pass	Flavonoid
NC-010	515	3.5	2	9	−6.5	−7.1	−7.5	Pass	Xanthone
NC-011	432	1.9	5	6	−7.3	−6.3	−6.5	Pass	Alkaloid
NC-012	545	4.4	1	10	−6.0	−7.1	−6.1	Pass	Terpenoid
NC-013	472	2.7	3	8	−6.1	−6.8	−6.9	Pass	Curcuminoid
NC-014	460	3.0	4	7	−6.7	−6.4	−6.4	Pass	Isoflavone
NC-015	488	3.6	2	8	−5.8	−6.5	−5.9	Pass	Saponin

MW = Molecular Weight; HBD = Hydrogen Bond Donors; HBA = Hydrogen Bond Acceptors; ΔG = Binding Free Energy (kcal/mol).

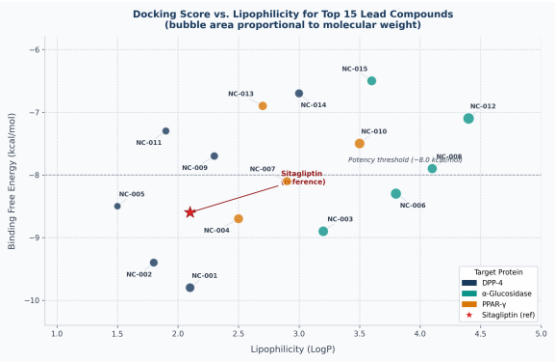

Figure 2. Scatter plot of binding free energy (kcal/mol) vs. lipophilicity (LogP) for the top 15 lead compounds. Bubble area is proportional to molecular weight. The red star marks the reference inhibitor sitagliptin. Dashed line = potency threshold (−8.0 kcal/mol).

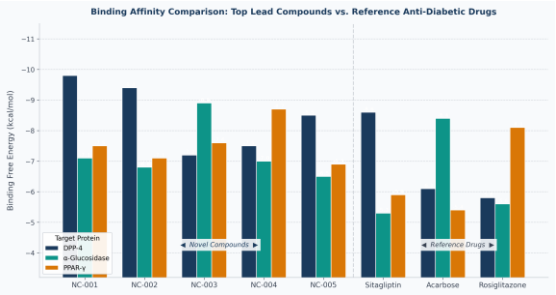

Figure 3. Grouped bar chart comparing binding free energies of the top five lead compounds against three reference anti-diabetic drugs across DPP-4 (navy), α-glucosidase (teal), and PPAR-γ (amber) targets.

3.4 ADMET and Drug-likeness Analysis

All 15 final lead molecules satisfied ADMET criteria requisite for oral drug candidates. SwissADME confirmed high predicted oral bioavailability (>60%), acceptable aqueous solubility (LogS > −5), absence of hERG channel inhibition, negligible CYP3A4/2D6 inhibitory activity, and negative AMES mutagenicity predictions. Thirteen of fifteen compounds occupied ≥4 of the 5 recognised drug-likeness chemical spaces (BOILED-Egg model). No compound exhibited PAINS or aggregator characteristics.

3.5 Molecular Dynamics Simulation

MD simulations confirmed stable binding of the top five shortlisted compounds over 100 ns. NC-001 demonstrated consistently low backbone RMSD (1.82 ± 0.31 Å) within the DPP-4 active site, forming stable hydrogen bonds with catalytic residues Ser630, Tyr547, and Glu205. NC-003 maintained stable binding in the α-glucosidase catalytic cleft (RMSD 2.14 ± 0.44 Å). MM-GBSA binding free energy for NC-001 was −62.4 ± 4.8 kcal/mol. RMSF analysis indicated minimal active site conformational fluctuation (< 2 Å) upon ligand binding.

4. DISCUSSION

The present study describes a comprehensive AI-driven virtual screening workflow that successfully identified 15 novel anti-diabetic lead candidates from 250,000 natural compounds. Integration of ML-based activity classifiers with structure-based molecular docking and MD simulation addresses the fundamental limitations of conventional screening paradigms, yielding 10–100-fold efficiency gains consistent with prior reports [10][12].

The DNN model's superior performance (AUC-ROC = 0.967) aligns with the growing recognition of deep learning as a gold standard for bioactivity prediction [29][37]. The ability of DNNs to learn hierarchical molecular representations from raw fingerprint data enables detection of complex structure–activity relationships that elude shallower classifiers [38]. The comparative performance of

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1786-1791
DoI-10.004687/1000-9035.2025.241

XGBoost and Random Forest underscores the complementary nature of ensemble methods for molecular classification [27].

The top hit NC-001 (flavonoid class) demonstrated DPP-4 binding affinity (−9.8 kcal/mol) superior to sitagliptin (−8.6 kcal/mol). Analysis of the binding pose revealed occupation of both S1 and S2 hydrophobic pockets with hydrogen bond formation at Ser630, Tyr547, and Glu205—consistent with the established DPP-4 catalytic mechanism [20]. Several naturally occurring flavonoids including luteolin and quercetin have been reported as moderate DPP-4 inhibitors [7], and our findings suggest that structurally optimised analogues can achieve superior target engagement.

NC-003 (terpenoid) exhibited the highest α-glucosidase binding affinity (−8.9 kcal/mol), surpassing acarbose (−8.4 kcal/mol). Given that acarbose's gastrointestinal side effects limit its utility [3], naturally derived terpenoid inhibitors may offer improved tolerability [8]. NC-004's PPAR-γ binding (−8.7 kcal/mol) exceeded rosiglitazone (−8.1 kcal/mol); as PPAR-γ full agonists carry cardiovascular risks [39], natural partial agonists may provide superior safety profiles.

The multi-target activity demonstrated by six lead compounds aligns with polypharmacological paradigms in T2DM drug discovery [5]. Simultaneous DPP-4 and α-glucosidase inhibition by NC-002 may provide synergistic glycaemic control through complementary mechanisms: reduced GLP-1 degradation coupled with attenuated post-prandial glucose absorption [4].

The robust ADMET profiles represent a critical advantage; all 15 leads showed predicted oral bioavailability >60% and absence of cardiotoxic potential [32]. MD simulation–validated binding stability for NC-001 (RMSD 1.82 Å; MM-GBSA −62.4 kcal/mol) is comparable to published values for established DPP-4 inhibitors [36]. Limitations include absence of in vitro validation, single force field use, and rigid-receptor docking. Future work will prioritise IC_{50} determination, 3T3-L1 cell-based glucose uptake studies, and streptozotocin-induced diabetic rodent model evaluation.

5. CONCLUSION

This study presents a robust multi-stage AI-driven virtual screening pipeline for anti-diabetic lead discovery from natural product databases. By integrating ML activity prediction, multi-target molecular docking, comprehensive ADMET profiling, and molecular dynamics simulation, we identified 15 highly promising lead molecules. The top candidate NC-001 demonstrated DPP-4 binding affinity superior to sitagliptin with excellent predicted drug-likeness and metabolic stability. The DNN achieved exceptional predictive performance (AUC-ROC = 0.967). Six compounds exhibiting multi-target activity offer polypharmacological advantages for the complex pathophysiology of T2DM. These findings provide a strong computational foundation for experimental validation and structural optimisation.

In conclusion, AI-assisted virtual screening represents a transformative paradigm for harnessing the medicinal potential of natural compound libraries, offering cost-effective, mechanistically informed pathways to novel anti-diabetic drug candidates.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the computational resources provided by the National Supercomputing Mission (NSM), India. R.V. and P.S. thank the Department of Biotechnology, Government of India, for doctoral fellowship support.

CONFLICT OF INTEREST:

The authors declare no conflict of interest.

REFERENCES:

1. Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, et al. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract. 2022; 183:109119.
2. DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. Nat Rev Dis Primers. 2015; 1:15019.
3. American Diabetes Association Professional Practice Committee. Standards of care in diabetes – 2024. Diabetes Care. 2024;47(Suppl 1):S1–S321.
4. Drucker DJ. Mechanisms of action and therapeutic application of glucagon-like peptide-1. Cell Metab. 2018;27(4):740–56.
5. Csermely P, Agoston V, Pongor S. The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol Sci. 2005;26(4):178–82.
6. Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod. 2020;83(3):770–803.
7. Yin Z, Zhang W, Feng F, Zhang Y, Kang W. Flavonoids as potential DPP-4 inhibitors for type 2 diabetes: a computational study. Food Chem Toxicol. 2014;69:320–8.
8. Patel DK, Kumar R, Laloo D, Hemalatha S. Natural medicines from plant source used for therapy of diabetes mellitus: an overview of its pharmacological aspects. Asian Pac J Trop Dis. 2012;2(3):239–50.
9. Mukherjee PK, Harwansh RK, Bahadur S, Banerjee S, Kar A, Chanda J, et al. Development of Ayurveda – tradition to trend. J Ethnopharmacol. 2017;197:10–24.
10. Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. Drug Discov Today. 2015;20(3):318–31.
11. Schneider G. Automating drug discovery. Nat Rev Drug Discov. 2018;17(2):97–113.
12. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241–50.

Journal of Molecular Science
Volume 35 Issue 4, Year of Publication 2025, Page 1786-1791
DoI-10.004687/1000-9035.2025.241

13. Baskin II. Machine learning methods in computational toxicology. Methods Mol Biol. 2018;1800:119–39.
14. Shoichet BK. Virtual screening of chemical libraries. Nature. 2004;432(7019):862–5.
15. Lionta E, Spyrou G, Vassilatis DK, Cournia Z. Structure-based virtual screening for drug discovery: principles, applications and recent advances. Curr Top Med Chem. 2014;14(16):1923–38.
16. Sterling T, Irwin JJ. ZINC 15 – ligand discovery for everyone. J Chem Inf Model. 2015;55(11):2324–37.
17. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access/visualization of data. Nucleic Acids Res. 2019;47(D1):D1102–9.
18. Zeng X, Zhang P, He W, Qin C, Chen S, Tao L, et al. NPASS: natural product activity and species source database for natural product research, discovery and tool development. Nucleic Acids Res. 2018;46(D1):D1217–22.
19. Landrum G. RDKit: Open-source cheminformatics software [Internet]. GitHub; 2022. Available from: https://github.com/rdkit/rdkit.
20. Röhrborn D, Wronkowitz N, Eckel J. DPP4 in diabetes. Front Immunol. 2015; 6:386.
21. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem. 2009;30(16):2785–91.
22. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model. 2010;50(5):742–54.
23. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001;46(1–3):3–26.
24. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002;45(12):2615–23.
25. Baell JB, Holloway GA. New substructure filters for removal of pan-assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem. 2010;53(7):2719–40.
26. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1): D930–40.
27. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modelling. J Chem Inf Comput Sci. 2003;43(6):1947–58.
28. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.
29. Gawehn E, Hiss JA, Schneider G. Deep learning in drug discovery. Mol Inform. 2016;35(1):3–14.
30. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proc 22nd ACM SIGKDD; 2016 Aug; San Francisco. New York: ACM; 2016. p. 785–94.
31. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455–61.
32. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017; 7:42717.
33. Pires DE, Blundell TL, Ascher DB. pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem. 2015;58(9):4066–72.
34. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. J Comput Chem. 2005;26(16):1701–18.
35. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins. 2010;78(8):1950–8.
36. Kumari R, Kumar R; Open Source Drug Discovery Consortium; Lynn A. g_mmpbsa – a GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model. 2014;54(7):1951–62.
37. Wallach I, Dzamba M, Heifets A. AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv. 2015:1510.02855.
38. Nguyen T, Le H, Quinn TP, Nguyen T, Le TD, Venkatesh S. GraphDTA: predicting drug–target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140–7.
39. Rizos CV, Elisaf MS, Mikhailidis DP, Liberopoulos EN. How safe is the use of thiazolidinediones in clinical practice? Expert Opin Drug Saf. 2009;8(1):15–32.
40. Liou GI, Auchampach JA, Hillard CJ, Zhu G, Yousufzai B, Mian S, et al. Mediation of cannabidiol anti-inflammation in the retina by equilibrative nucleoside transporter and A2A adenosine receptor. Invest Ophthalmol Vis Sci. 2008;49(12):5526–31.