International Journal of Computer Science Engineering Techniques (IJCSE) Volume 10 Issue 2 ,Mar-Apr 2026

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article (Review Article)*

Towards an Integrated AI Framework for Molecular Sciences: Connecting Structure Prediction, Drug Likeness Modeling and Molecular Design

Bharti B. Balande[1], Sujata Ambhore[2], Reema A. Lahane[3], Ramesh R. Manza[4],

Vanmathi C.[5], Sachin S. Bhusari[6]

[1]Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Maharashtra, India. [2]Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Maharashtra, India. [3]Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Maharashtra, India. [4]Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Maharashtra, India, Professor, Department of Software and Systems Engineering, Vellore Institute of Technology, Vellore Institute of Technology, Vandalur-Kelambakkam Road, Chennai-600127, Tamil Nadu, India., Chennai-600127, Tamil Nadu-India, Assistant Professor, Department of Chemical Technology, Dr. Babasaheb Ambedkar Marathwada University, University Campus, Near Soneri Mahal, Jaisingpura, Chhatrapati Sambhajinagar-431004, Maharashtra-India.

[1]*csit.bbb@bamu.ac.in*

Abstract - *Drug discovery remains costly and time-consuming; most of the costs could be explained by the high expenditure on the chemical space and the high rates of attrition observed during the later phases of development. AI has emerged as a potential solution to these issues that support the use of data to predict and create prospective candidates molecular entities. Recent advances in the pharmaceutical sector to date in the area of AI usage are critically reviewed in this review, as well as the conceptual backbone of prediction of protein structure, prediction of drug-likeness and ADMET, and de novo molecular design that are summarized within a unified framework of early-stage drug discovery. One of the literature review types that were undertaken was a narrative inquiry of peer-reviewed publications dated 2018 to 2025, based on major scientific databases. The results prove that AI-based approaches outperform or refine protein structure prediction, improve molecular property prediction, and provide easy passage of construction of chemical space. Moreover, it can be reported that integrative predictive-generative pipelines may reduce a false-positive rate and enhance selection of candidates at the early stage. In general, the further incorporation of AI methods with experimental validation is expected to put computational methods into an even greater perspective and speed up the curve of drug discoveries.*

Keywords - *Artificial Intelligence, Drug Discovery, ADMET modeling, generative models, molecular design*

1. Introduction

Drug discovery plays a central role in the development of new therapeutics, yet it remains one of the most complex and resource-intensive areas of biomedical research. Conventional drug development pipelines rely heavily on large-scale experimental screening followed by lengthy validation stages, often requiring 10–15 years and investments reaching billions of dollars1-3. Compounding this challenge is the vastness of chemical space, estimated to encompass more than 10^{60} possible molecules, making comprehensive exploration impractical using traditional approaches8.

In recent years, artificial intelligence (AI) has emerged as a promising alternative paradigm by enabling models to learn from existing molecular data and generate predictions about structural, chemical, and biological properties1-15. Breakthrough systems such as AlphaFold have demonstrated unprecedented accuracy in protein structure prediction, while

ISSN : 2455-135X
https://www.ijcsejournal.org/ *Type of Article* (Review Article)

broader deep learning frameworks have shown strong performance in modeling molecular interactions and binding behavior2-12-20. Alongside this, machine learning techniques are increasingly applied to tasks such as ADMET prediction, molecular docking, and compound prioritization, offering valuable decision support in early-stage screening22-32.

Despite this progress, most AI models are still developed and applied in isolation. Integrating structure prediction, property evaluation, and molecule generation into a cohesive and interoperable workflow remains an ongoing challenge13-14. Addressing this fragmentation is therefore an important research direction. Beyond drug discovery, AI-driven methodologies are also gaining traction in related domains such as protein engineering, molecular optimization, and synthetic biology, further highlighting their broad relevance to molecular sciences8-7-18.

Notably, AI-based approaches offer practical advantages by reducing experimental burden and accelerating the identification of promising candidate compounds5-6-10. Generative deep learning models, in particular, have enabled the design of novel molecules with tailored properties, including improved bioactivity, synthetic feasibility, and chemical novelty4-5-16. These techniques are now being progressively incorporated across multiple stages of the drug development pipeline, ranging from target identification and virtual screening to ADMET prediction and early clinical decision-making20-13-14-17.

A further advantage of generative approaches is their ability to support inverse molecular design, where compounds are constructed to meet predefined biological and pharmacological criteria. This capability provides a practical way to navigate the enormous complexity of chemical space more strategically8-17-30. Taken together, current evidence suggests that AI has the potential to substantially improve the efficiency, accuracy, and scalability of molecular research workflows21-22-24.

Figure [1] illustrates the role of artificial intelligence across three core stages of molecular discovery: (1) deep learning–based structure prediction, (2) rule-based or machine learning–driven assessment of drug-likeness22-23, and (3) generative AI models for molecular design24- 25. Rather than functioning as isolated components, these stages increasingly represent parts of an emerging interconnected workflow. The rapid evolution of AI systems, including conversational and automation-oriented tools, further underscores the potential for streamlining data handling and accelerating computational outputs within molecular research pipelines.

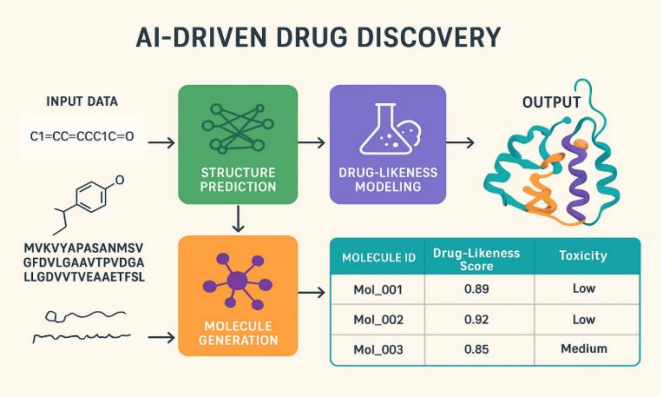

Figure 1. Integration of AI in molecular sciences explains the interaction of structure prediction, drug likeness measurement and molecule generation elements.

2. Problem Statement

Contemporary drug discovery efforts face multiple persistent challenges, including prolonged development timelines (often 10–15 years), extremely high financial costs reaching billions of US dollars, and substantial attrition rates during late-stage clinical evaluation. In addition, limitations in accurately predicting protein and molecular structures further reduce the efficiency of traditional discovery pipelines. The enormous size of chemical space, estimated to exceed 10^{60} potential molecules [4], presents an additional barrier, making exhaustive exploration practically impossible using conventional experimental and computational methods.

These limitations highlight the need for an integrated, AI-driven framework capable of reliably predicting molecular structures, evaluating drug-likeness and pharmacokinetic properties, and generating novel candidate compounds. Such a unified system has the potential to accelerate discovery processes while improving the overall quality and relevance of selected drug candidates.

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article* (Review Article)

3. Objectives

The present study is guided by the following objectives:

1. To apply artificial intelligence methods for accurate prediction of three-dimensional protein and molecular structures.
2. To develop predictive models for assessing drug-likeness characteristics and ADMET-related properties.
3. To design novel chemical entities using generative artificial intelligence approaches.
4. To integrate these components into a unified, cost-efficient framework for early-stage drug discovery.

4. Novelty and Contribution of the Present Study

Although many studies have independently explored AI-based structure prediction, molecular property assessment, and generative molecule design, these efforts are typically treated as separate research directions. The distinctive contribution of this review lies in its conceptual integration of these components into a coherent, closed-loop framework. By systematically synthesizing recent literature and organizing it within a unified model, this work demonstrates how predictive and generative AI techniques can complement each other to support more informed early-stage decisions, minimize false-positive candidates, and enhance the translational relevance of computational drug discovery.

5. Review Methodology

This study adopts a narrative review approach. Relevant literature was identified through systematic searches of major scientific databases, including PubMed, Scopus, Google Scholar, and Web of Science. Search queries combined key terms such as “artificial intelligence in drug discovery,” “protein structure prediction,” “AlphaFold,” “graph neural networks,” “ADMET prediction,” “generative models,” “variational autoencoders,” “generative adversarial networks,” and “molecular design.”

Peer-reviewed research articles, review papers, and influential reports published primarily between 2018 and 2025 were considered for inclusion. Studies were selected based on their relevance to AI applications in molecular sciences and their contributions to predictive modeling, generative methodologies, or integrative frameworks in drug discovery.

6. Materials and Methods

6.1 Review Methodology

The proposed AI-driven drug discovery framework consists of three interconnected components:

(i) structure prediction,

(ii) drug-likeness and ADMET evaluation, and

(iii) molecule generation with iterative refinement.

Rather than operating as isolated steps, these modules are designed to function as an integrated pipeline in which outputs from one component inform the next. The overall workflow of the framework is illustrated in **Figure 2**.

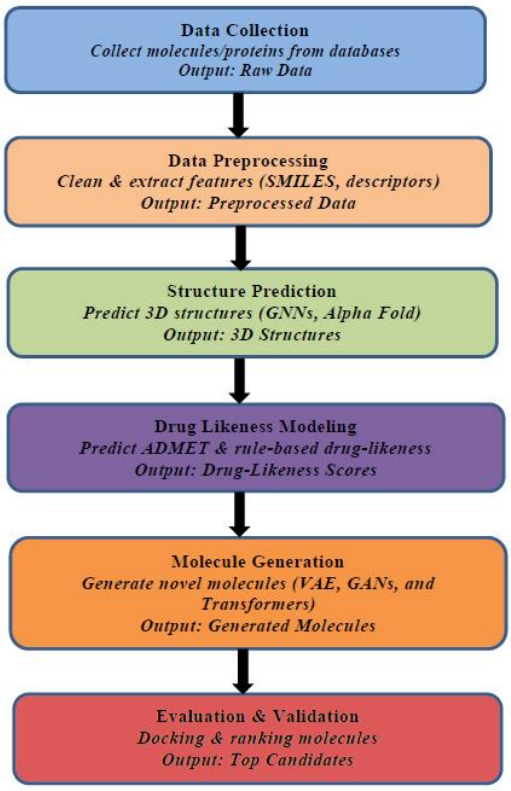

Figure 2. The framework depicts the use of artificial intelligence for the application of structure prediction, drug likeness and ADMET evaluation, generative modelling, and iterative refinement.

6.1 Review Methodology

Protein and molecular structures were modeled using deep learning-based approaches, primarily AlphaFold for

ISSN : 2455-135X
https://www.ijcsejournal.org/ *Type of Article* (Review Article)

protein structure prediction and graph neural networks for molecular representation learning. These models infer three-dimensional conformations from amino acid sequences and molecular graph structures, respectively. Accurate structural representations are essential for characterizing ligand-binding sites and understanding interaction geometries, which are critical for downstream screening and optimization.

6.3 Drug-Likeness and ADMET Modeling

Assessment of drug-likeness was performed using established rule-based criteria, including Lipinski's Rule of Five, alongside supervised machine learning classifiers trained on publicly available datasets such as ChEMBL and PubChem.

ADMET properties—covering absorption, distribution, metabolism, excretion, and toxicity—were predicted to identify and filter compounds with unfavorable pharmacokinetic profiles at an early stage. This step supports more informed prioritization before experimental validation.

6.4 Molecule Generation and Optimization

Generative modeling techniques, including variational autoencoders (VAEs) and generative adversarial networks (GANs), were employed to design novel chemical structures. Generated candidates were iteratively refined through a feedback mechanism that incorporates predictions from the structure modeling and ADMET evaluation modules. This iterative process supports continuous improvement of molecular candidates within a closed-loop optimization framework.

7. AI Methods in Molecular Sciences

7.1 Generative Models and Deep Learning

Several generative systems, such as REINVENT and GENTRL, have been developed to design molecules with targeted physicochemical properties while maintaining drug-like characteristics and synthetic feasibility. Recent studies also demonstrate that active learning strategies, which combine physics-based simulations with generative AI models, can further enhance molecular optimization and predictive performance26-31.

7.2 Molecular Property Prediction Using Machine Learning

Machine learning approaches are widely applied for predicting ADMET properties, bioactivity, and molecular docking outcomes28-29. Platforms such as ADMET-AI provide scalable and efficient prediction tools that support early-stage compound prioritization32. Benchmark datasets, including PharmaBench, have contributed to improving the robustness and comparability of predictive models across diverse chemical libraries29.

7.3 Molecular Optimization and Protein Engineering

Recent advances in AI have significantly enhanced the prediction of protein structures, ligand–protein interactions, and molecular optimization strategies2-12-20. Goal-directed molecular design systems integrated with automated synthesis platforms, such as Mol-AIR, have demonstrated the ability to generate candidate molecules guided by reward functions aligned with biological targets30.

8. Applications of Artificial Intelligence in Drug Discovery

AI-based technologies are increasingly influencing multiple stages of the drug discovery process, including target identification, compound library construction, and candidate prioritization11-14. For example, AI-guided approaches have reportedly reduced the discovery timeline for DDR1 kinase inhibitors from several months to only a few weeks5.

Both predictive and generative models are now used to explore novel chemical scaffolds while simultaneously optimizing bioactivity and drug-like properties in silico7-4-16. Emerging evidence also suggests that compounds discovered with AI assistance may demonstrate promising success rates in clinical development compared with traditional approaches24.

9. Expected Outcomes

The proposed framework is anticipated to offer several key benefits:

1. Reduction in time and cost associated with early-stage drug discovery.
2. Improved accuracy in structural and molecular property prediction.
3. Generation of novel candidate compounds using advanced computational methods.

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article* (Review Article)

4. Enhanced clinical success rates through effective pre-screening of compounds.
5. Support for personalized medicine and broader innovation in biotechnology.

10. Results

As this work focuses primarily on framework development and conceptual integration, results are expressed in terms of expected outcomes rather than experimental validation. The proposed pipeline is expected to improve accuracy in structure prediction, enhance screening efficiency, and prioritize biologically relevant compounds with favorable pharmacokinetic profiles. Integration of predictive and generative components is also anticipated to reduce false positives in early discovery and shorten the time required for candidate selection.

11. Discussion

Integrating AI capabilities across multiple stages of drug discovery represents a significant advancement beyond traditionally fragmented workflows. By combining structure prediction, molecular property evaluation, and generative design within a unified system, the framework supports more informed early-stage decision-making. Generative models expand exploration into less-sampled regions of chemical space, while predictive models help ensure that candidate compounds satisfy both structural and pharmacokinetic constraints.

11.1 Limitations

Despite its potential, the proposed framework has several limitations. First, the study is primarily conceptual, and the outcomes are based on anticipated performance rather than large-scale experimental validation. Model effectiveness is strongly influenced by the quality, diversity, and representativeness of available training datasets, which may introduce bias and limit generalizability across therapeutic domains.

Second, although AI-based predictions for ADMET properties and molecular interactions are improving, they cannot yet fully capture the complexity of biological systems and may still yield false positives or negatives. Third, some molecules generated by computational models may present challenges in practical synthesis or large-scale production. Finally, issues related to computational cost and model interpretability remain barriers to widespread clinical and industrial adoption. Future work should emphasize experimental validation, reinforcement learning strategies, explainable AI techniques, and stronger integration with wet-lab workflows.

12. Need and Significance

The application of artificial intelligence in molecular sciences offers meaningful opportunities to address global healthcare challenges by accelerating therapeutic discovery while reducing resource consumption. Early elimination of non–drug-like candidates can decrease late-stage clinical failure rates and improve overall productivity. In addition, such frameworks can strengthen national research capacity in computational molecular sciences and align with broader initiatives such as Aatmanirbhar Bharat and the Digital Health Mission.

13. Future Directions

Future studies should focus on incorporating explainable AI methodologies, expanding dataset diversity, enabling real-time experimental feedback, and strengthening collaboration between computational and experimental research teams. These developments are expected to enhance the reliability, transparency, and translational impact of AI-assisted drug discovery.

14. Conclusions

This study presents an integrated artificial intelligence framework for molecular sciences that encompasses structure prediction, drug-likeness and ADMET modeling, and molecular generation. The proposed approach offers a scalable and cost-effective strategy for early-stage drug discovery and holds considerable potential to improve the efficiency of translational research.

15. Data and Materials Availability

All data used in this study were obtained from publicly available databases, including ChEMBL, PubChem, and DrugBank.

ISSN : 2455-135X

https://www.ijcsejournal.org/ *Type of Article* (Review Article)

16. Declaration on Generative AI Use

AI-assisted tools were used solely to improve language clarity and readability. The authors carefully reviewed and edited all content and take full responsibility for the final manuscript.

References

[1] Gupta R, Srivastava D, Sahu M, Tiwari S, Ambasta RK, Kumar P. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Molecular diversity. 2021 Aug; 25(3):1315-60.

[2] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug 26; 596(7873):583-9.

[3] Lipinski CA. Lead-and drug-like compounds: the rule-of-five revolution. Drug discovery today: Technologies. 2004 Dec 1; 1(4):337-41.

[4] Segler MH, Kogej T, Tyrchan C, Waller MP. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS central science. 2018 Jan 24;4(1):120-31.

[5] Zhavoronkov A, Ivanenkov YA, Aliper A, Veselov MS, Aladinskiy VA, Aladinskaya AV, Terentiev VA, Polykovskiy DA, Kuznetsov MD, Asadulaev A, Volkov Y. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nature biotechnology. 2019 Sep; 37(9):1038-40.

[6] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug discovery today. 2018 Jun 1; 23(6):1241-50.

[7] Mente P. The use of functionalized polymers to make hollow carbon spheres to be used as supports in catalysis (Doctoral dissertation, University of the Witwatersrand, Johannesburg (South Africa)).

[8] Sanchez-Lengeling B, Aspuru-Guzik A. Inverse molecular design using machine learning: Generative models for matter engineering. Science. 2018 Jul 27; 361(6400):360-5.

[9] Rawlings CJ, Fox JP. Artificial intelligence in molecular biology: a review and assessment. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 1994 Jun 29; 344(1310):353-63.

[10] Dara S, Dhamercherla S, Jadav SS, Babu CM, Ahsan MJ. Machine learning in drug discovery: a review. Artificial intelligence review. 2022 Mar; 55(3):1947-99.

[11] Ferreira FJ, Carneiro AS. AI-Driven Drug Discovery: A Comprehensive Review. ACS omega. 2025 Jun 6.

[12] Kundaje A, Pollard KS, Ma J, Chang X, Chen M, Rohs R. Artificial intelligence in molecular biology. Molecular Cell. 2025 Jan 16; 85(2):193-8.

[13] Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N, Cedrón F, Novoa FJ, Carballal A, Maojo V, Pazos A, Fernandez-Lozano C. A review on machine learning approaches and trends in drug discovery. Computational and structural biotechnology journal. 2021 Jan 1; 19:4538-58.

[14] Ocana A, Pandiella A, Privat C, Bravo I, Luengo-Oroz M, Amir E, Gyorffy B. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. Biomarker Research. 2025 Mar 14; 13(1):45.

[15] Swargiary D, Saikia A, Basumatary D, Borah JC. Review on Advancement of AI in Cell Engineering and Molecular Biology. Artificial Intelligence (AI) in Cell and Genetic Engineering. 2025:169-92.

[16] Das U. Generative AI for drug discovery and protein design: the next frontier in AI-driven molecular science. Medicine in Drug Discovery. 2025 Jul 1:100213.

[17] Xia X, Zhang Y, Zeng X, Zhang X, Zheng C, Su Y. Artificial Intelligence in Molecular Optimization: Current Paradigms and Future Frontiers. International Journal of Molecular Sciences. 2025 May 19; 26(10):4878.

[18] Groff-Vindman CS, Trump BD, Cummings CL, Smith M, Titus AJ, Oye K, Prado V, Turmus E, Linkov I. The

ISSN : 2455-135X

convergence of AI and synthetic biology: the looming deluge. npj Biomedical Innovations. 2025 Jul 1; 2(1):20.

[19] Carr A, Cool J, Karaletsos T, Li D, Lowe AR, Otte S, Schmid SL. AI: A transformative opportunity in cell biology. Molecular Biology of the Cell. 2024 Dec 1;35(12):pe4.

[20] Luo M, Yang W, Bai L, Zhang L, Huang JW, Cao Y, Xie Y, Tong L, Zhang H, Yu L, Zhou LW. Artificial intelligence for life sciences: A comprehensive guide and future trends. The Innovation Life. 2024 Dec 6; 2(4):100105-.

[21] Bassey GE, Daniel EA, Okesina KB, Odetayo AF. Transformative Role of Artificial Intelligence in Drug Discovery and Translational Medicine: Innovations, Challenges, and Future Prospects. Drug Design, Development and Therapy. 2025 Dec 31:7493-502.

[22] Fan N, Chen J, Wang J, Chen ZS, Yang Y. Bridging data and drug development: machine learning approaches for next-generation ADMET prediction. Drug Discovery Today. 2025 Oct 4:104487.

[23] Tong X, Liu X, Tan X, Li X, Jiang J, Xiong Z, Xu T, Jiang H, Qiao N, Zheng M. Generative models for de novo drug design. Journal of Medicinal Chemistry. 2021 Sep 17; 64(19):14011-27.

[24] Jayatunga MK, Ayers M, Bruens L, Jayanth D, Meier C. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. Drug discovery today. 2024 Jun 1; 29(6):104009.

[25] Kneiding H, Balcells D. Augmenting genetic algorithms with machine learning for inverse molecular design. Chemical Science. 2024; 15(38):15522-39.

[26] Loeffler HH, He J, Tibo A, Janet JP, Voronov A, Mervin LH, Engkvist O. Reinvent 4: modern AI–driven generative molecule design. Journal of Cheminformatics. 2024 Feb 21; 16(1):20.

[27] Loeffler HH, Wan S, Klähn M, Bhati AP, Coveney PV. Optimal molecular design: Generative active learning combining REINVENT with precise binding free energy ranking simulations. Journal of Chemical Theory and Computation. 2024 Sep 3; 20(18):8308-28.

[28] Liu H, Zhu B, Nie S, Li H, Lin Y, Ma T, Shao X, Chen Q, Shen M, Zheng Y, Fan X. Advancing ADMET prediction through multiscale fragment-aware pretraining with MSformer-ADMET. Briefings in Bioinformatics. 2025 Sep; 26(5):bbaf506.

[29] Niu Z, Xiao X, Wu W, Cai Q, Jiang Y, Jin W, Wang M, Yang G, Kong L, Jin X, and Yang G. PharmaBench: Enhancing ADMET benchmarks with large language models. Scientific Data. 2024 Sep 10; 11(1):985.

[30] Park J, Ahn J, Choi J, Kim J. Mol-AIR: Molecular Reinforcement Learning with Adaptive Intrinsic Rewards for Goal-Directed Molecular Generation. Journal of Chemical Information and Modeling. 2025 Feb 24; 65(5):2283-96.

[31] Goryanin I, Goryanin I, Demin O. Revolutionizing drug discovery: Integrating artificial intelligence with quantitative systems pharmacology. Drug Discovery Today. 2025 Aug 6:104448.

[32] Swanson K, Walther P, Leitz J, Mukherjee S, Wu JC, Shivnaraine RV, Zou J. ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. Bioinformatics. 2024 Jul; 40(7):btae416.