WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Computational and AI-Based Approaches in Modern Drug Discovery

Miss. Patel Sayama Munir[1], Dr.Dharashive V.M.[2],Dr.Malpani S.G.[3], Miss. Katke Rudrani G[4],Miss Swami Dhanshree M[5], Mr .Dharashive Pratik B.[6],Mr.Badgire Vivek R[7],Miss. Jadhav Asmita[8],Miss. Kale Sakshi[9]

[2-9] Co-author-
Shivlengeshwar college of pharmacy Almala

ABSTRACT :

The drug discovery landscape has undergone a dramatic shift with the integration of computational methods and artificial intelligence (AI). These approaches promise to reduce cost and time, improve hit rates, and open new modalities by combining physics-based modeling, cheminformatics, machine learning (ML), and generative AI [1]. This review summarizes major computational approaches across the discovery pipeline—target identification, hit finding, lead optimization, ADMET prediction, and clinical translation—highlights recent breakthroughs (notably protein structure prediction and generative molecular design), discusses challenges and limitations, and outlines near-term opportunities and best practices for researchers and industry [2].

Keywords: Drug discovery, Artificial intelligence (AI),Machine learning (ML),Deep learning,Structure-based drug design (SBDD),Ligand-based drug design (LBDD),Molecular docking,Molecular dynamics (MD),Quantitative structure–activity relationship (QSAR)

1. Introduction

Traditional small-molecule drug discovery is both costly and time-consuming, with the average development timeline extending 10–15 years and costs surpassing US$1 billion when accounting for failed programs [3]. Attrition rates are high, particularly in the transition from preclinical to clinical stages, where safety and efficacy issues often derail promising candidates [4]. Computational methods and AI have emerged as transformative tools, enabling researchers to rapidly screen compounds, predict binding affinities, model ADMET properties, and generate novel molecules with desirable profiles [5]. Over the last decade, and especially since 2020, the convergence of large-scale data resources, improved algorithms, and accessible computing power has positioned computational drug discovery as a central component of pharmaceutical research [6].

2. Computational Approaches: An Overview

Drug discovery can broadly leverage two categories of computational methods. The first, physics-based approaches, rely on molecular simulations and structural data to model interactions. The second, data-driven approaches, employ statistical and machine learning methods to learn predictive models from experimental data [7]. Increasingly, hybrid systems are integrating these methodologies, allowing the accuracy of physical models to be combined with the generalizability and scalability of machine learning [8].

2.1 Physics-Based Methods

Physics-based methods rely on fundamental molecular principles. Structure-based drug design (SBDD) exploits target protein structures—obtained via X-ray crystallography, NMR, cryo-EM, or prediction algorithms—to guide ligand discovery [9]. Molecular docking estimates ligand poses within binding sites and scores them according to predicted affinity [10]. While docking is efficient, its accuracy is limited by approximations in scoring functions. Molecular dynamics (MD) simulations address this limitation by capturing the dynamic motion of biomolecules and solvent effects [11]. Free energy perturbation (FEP) techniques, though computationally demanding, provide highly accurate relative binding affinities, making them valuable for lead optimization [12].

2.2 Data-Driven and AI-Based Methods

The second category involves data-driven machine learning approaches, which rely on curated chemical and biological datasets. Traditional QSAR models correlate molecular descriptors with biological activity [13]. More recent advances use deep learning and graph neural networks (GNNs) to learn directly from chemical graph structures [14]. Transformers, initially designed for natural language processing, have been adapted to SMILES strings and protein sequences, achieving state-of-the-art results in activity prediction and de novo molecule design [15]. These models are particularly effective in

predicting drug–target interactions, ADMET properties, and toxicity risks [16].

3. Recent Breakthroughs

3.1 Protein Structure Prediction

One of the most transformative breakthroughs was the introduction of AlphaFold2 by DeepMind in 2020, which achieved unprecedented accuracy in protein structure prediction [17]. By 2021, the AlphaFold Protein Structure Database provided millions of predicted structures, including difficult-to-crystallize proteins [18]. This development revolutionized SBDD by making structural information accessible at proteome scale. More recently, AlphaFold3 extended predictions to protein complexes, enabling modeling of protein–protein and protein–ligand interactions [19]. Limitations remain, such as handling induced fit, post-translational modifications, and dynamic flexibility, but the impact on early-stage drug discovery has been profound [20].

3.2 Generative AI for Molecular Design

Generative AI models—variational autoencoders (VAEs), generative adversarial networks (GANs), and diffusion models—are now capable of proposing novel molecules optimized for activity, selectivity, and synthetic accessibility [21]. Reinforcement learning frameworks allow these models to be guided by project-specific objectives, such as potency and ADMET profiles [22]. More advanced retrieval-augmented generation models, like Rag2Mol, integrate 3D structural information from protein pockets, generating compounds tailored to binding sites [23]. These innovations reduce reliance on brute-force library screening and expand the accessible chemical space beyond traditional scaffolds [24].

3.3 Foundation Models and Multimodal Learning

Foundation models trained on massive chemical, structural, and biomedical text corpora are becoming general-purpose engines for drug discovery [25]. By learning representations across modalities—chemical structures, protein sequences, and literature—they enable transfer learning for diverse tasks such as drug repurposing, side-effect prediction, and mechanistic inference [26]. Pharmaceutical companies are forming data-sharing consortia to pool proprietary data and develop robust foundation models, though this raises issues around intellectual property and governance [27].

4. Pipeline Applications of AI

4.1 Target Identification and Validation

AI is increasingly used to mine multi-omic data and biomedical literature for target discovery. Network biology combined with ML allows researchers to identify disease-associated genes and prioritize druggable targets [28]. Predictive models integrate transcriptomic, proteomic, and pathway-level data to validate targets and identify biomarkers for patient stratification [29]. These approaches help reduce the risk of selecting undruggable or non-essential targets [30].

4.2 Hit Identification

Virtual screening is one of the earliest applications of computational methods. Modern AI models can screen ultra-large libraries (billions of compounds) using docking combined with ML-based scoring [31]. Phenotypic screening data can also be analyzed with ML to infer potential targets for active compounds [32]. Generative AI models can propose novel scaffolds where existing libraries fail to provide hits, shortening timelines from years to months [33].

4.3 Lead Optimization

AI-driven multi-objective optimization enables simultaneous tuning of potency, selectivity, and developability [34]. Bayesian optimization and reinforcement learning approaches propose analogs, while physics-based simulations refine affinity predictions [35]. Active learning further accelerates cycles by selecting compounds where experimental results are most informative for model improvement [36].

4.4 ADMET and Safety Prediction

Late-stage failures often result from poor pharmacokinetic or safety profiles. AI models trained on curated datasets can predict solubility, permeability, metabolic stability, and toxicity early in discovery [37]. Regulatory agencies increasingly accept computational ADMET evidence to reduce animal testing [38]. While predictions are not replacements for experiments, they serve as effective triage tools [39].

5. Validation and Benchmarking

Despite successes, the reliability of AI methods depends on rigorous benchmarking. Studies have shown that naive dataset splitting can inflate

performance, as test molecules may be too similar to training data [40]. Time-split and scaffold-split validation are recommended for realistic assessment [41]. Public benchmarks and community challenges, such as the DTI prediction contests, are vital for ensuring reproducibility and preventing overhyped claims [42].

6. Challenges and Limitations

6.1 Data Quality and Bias

Drug discovery datasets often contain experimental noise, batch effects, and chemical bias toward known scaffolds. Models trained on such data risk poor generalization [43].

6.2 Out-of-Distribution Generalization

AI models may fail when faced with novel chemical classes or targets not represented in training data. Out-of-distribution detection and uncertainty quantification are needed to mitigate risks [44].

6.3 Interpretability

While ML models excel at prediction, they often lack mechanistic interpretability. Hybrid models combining physics-based simulations with AI improve interpretability and trust [45].

6.4 Synthetic Accessibility

Generative models sometimes propose molecules that are chemically implausible or infeasible to synthesize. Integration with retrosynthetic planning tools is critical to ensure realistic outputs [46].

6.5 Ethical and Regulatory Concerns

Data pooling for foundation models raises issues around intellectual property, patient privacy, and fairness. Regulatory frameworks for AI in drug discovery are still evolving [47].

7. Case Studies

AI-driven approaches have already produced tangible outcomes. For instance, Exscientia advanced AI-designed drug candidates into clinical trials within a fraction of the traditional timeline [48]. Insilico Medicine identified a novel fibrosis target and advanced an AI-designed molecule to phase I trials in under three years [49]. Several partnerships between pharmaceutical companies and AI startups highlight the growing industrial confidence in computational methods [50].

8. Best Practices

Best practices for integrating AI in drug discovery include:

1. Define clear objectives for model use (hit discovery, ADMET triage, etc.) [51].
2. Rigorous data curation with careful train–test splitting [52].
3. Hybrid approaches that combine physics and ML for improved accuracy [53].
4. Uncertainty estimation to identify risky predictions [54].
5. Prospective experimental validation to close the feedback loop [55].
6. Transparent governance and documentation for regulatory compliance [56].
7. Future Directions

Advanced 3D Generative Models: Diffusion and retrieval-augmented approaches conditioned on protein pockets will enable more precise structure-guided design [57].
Multimodal Foundation Models: Unified models combining text, sequence, and structure data could accelerate hypothesis generation [58].
Reduced Animal Testing: Regulators are beginning to accept in silico ADMET predictions as partial replacements for animal models [59].
Human–AI Collaboration: AI will not replace scientists but will augment decision-making, enabling faster and more informed choices [60].

10. Conclusion

Computational and AI-based methods have moved from research novelties to essential components of modern drug discovery. Advances in protein structure prediction, generative molecular design, and predictive ADMET modeling are already reshaping pipelines. While limitations remain around data quality, interpretability, and regulatory acceptance, careful integration of AI with traditional methods offers the potential to significantly reduce

costs, timelines, and attrition rates in drug development. The future of drug discovery will likely be driven by a synergistic partnership between human expertise and artificial intelligence [61]

REFERENCES:

[1]. Ferreira FJN et al. AI-Driven Drug Discovery: A Comprehensive Review (2025).
[2] Su J. Artificial intelligence in drug discovery: A comprehensive review. arXiv (2025).
[3]Paul SM et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov (2010).
[4] DiMasi JA et al. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ (2016).
[5]Askr H. Deep learning in drug discovery: an integrative review and SLR (2023).
[6] Hughes JP et al. Principles of early drug discovery. Br J Pharmacol (2011).
[7] Özçelik R. Structure-Based Drug Discovery with Deep Learning (2023).
[8] Niazi SK. Recent Advances in Machine-Learning-Based Methods in Drug Discovery (2023).
[9] Kitchen DB et al. Docking and scoring in virtual screening for drug discovery. Nat Rev Drug Discoverv (2004).
[10] Morris GM et al. AutoDock4 and AutoDockTools4. J Comput Chem (2009).
[11] Hollingsworth SA, Dror RO. Molecular dynamics simulation for drug discovery. Annu Rev Biomed Data Sci (2018).
[12] Wang L et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by FEP. J Am Chem Soc (2015).
[13] Cherkasov A et al. QSAR modeling: where have you been? Where are you going to? J Med Chem (2014).
[14]Wu Z et al. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw (2021).
[15] Honda S et al. SMILES Transformer: Pre-trained molecular fingerprint for low data drug discovery. arXiv (2019).
[16] Xu Y et al. Deep learning for drug-induced liver injury prediction. Chem Res Toxicol (2020).
[17]Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature (2021).
[18] Varadi M et al. AlphaFold Protein Structure Database. Nucleic Acids Res (2022).
[19]DeepMind. AlphaFold3 Technical Report (2024).
[20] Desai D. Review of AlphaFold 3: Advances in Drug Design (2024).
[21]Gómez-Bombarelli R et al. Automatic chemical design using data-driven continuous representations of molecules. ACS Cent Sci (2018).
[22] Olivecrona M et al. Molecular de-novo design through deep reinforcement learning. J Cheminform (2017).
[23] Rag2Mol: Retrieval-augmented molecule generation framework (2025).
[24] Zavadskaya M. Integrating QSAR with reinforcement learning for inhibitor discovery. (2025).
[25] Walters WP. Big data in drug discovery and development: implications for preclinical research. Drug Discov Today (2022).
[26]Chen H et al. The rise of deep learning in drug discovery. Drug Discov Today (2018).
[27] Reuters. Pharma consortia building foundation models for drug discovery (2025).
[28]Barabási A-L et al. Network medicine: a network-based approach to human disease. Nat Rev Genet (2011).
[29] Li Y et al. Machine learning for biomarker discovery. Brief Bioinform (2020).
[30] Schadt EE et al. Molecular networks and network medicine. Nat Rev Genet (2009).
[31] Lyu J et al. Ultra-large library docking for discovering new chemotypes. Nature (2019).
[32] Corsello SM et al. The Drug Repurposing Hub: a next-generation drug library. Nat Med (2017).
[33]Walters WP, Murcko MA. Assessing the impact of generative AI on medicinal chemistry. J Med Chem (2020).
[34] Zhavoronkov A et al. Deep generative models for multi-objective drug design. Mol Pharm (2019).
[35]Abel R et al. Accurate computation of free energy differences in drug design. Nat Rev Drug Discoverv (2020).
[36]Schneider G. Automating drug discovery. Nat Rev Drug Discov (2018).
[37] Wang Y et al. ADMET prediction with deep learning. Front Pharmacol (2020).
[38] US FDA. Modernization Act for Reduced Animal Testing (2023).
[39]Ekins S. Industrializing AI in drug discovery. Expert Opin Drug Discov (2021).
[40] Sheridan RP. Time-split cross-validation as a method for estimating the goodness of prospective prediction. J Chem Inf Model (2013).
[41] Yang K et al. Are learned molecular representations ready for real-world drug discovery? Chem Sci (2019).
[42]Tatonetti NP. Data challenges in predictive drug discovery. Annu Rev Pharmacol Toxicol (2021).
[43] Chen L et al. Bias in chemical datasets for machine learning. J Chem Inf Model (2022).
[44]Zhou J et al. Out-of-distribution detection in deep learning for drug discovery. Brief Bioinform (2021).
[45] Jiménez-Luna J et al. Drug discovery with explainable AI. Nat Mach Intell (2020).
[46] Segler MHS et al. Planning chemical syntheses with deep neural networks and symbolic AI. Nature (2018).
[47]OECD. Principles on AI in drug discovery and healthcare. (2021).
[48] Exscientia Ltd. Press release: First AI-designed drug enters clinical trials (2020).
[49] Insilico Medicine. Press release: AI-discovered fibrosis drug in clinical trial (2021).
[50] Nature Biotech. Pharma–AI startup partnerships surge (2024).
[51]Walters WP. What do we want from AI in drug discovery? J Med Chem (2019).
[52] Fourches D et al. Trust, but verify: on the importance of chemical structure curation. J Chem Inf Model (2010).
[53]Zhavoronkov A. Deep learning enables rapid drug design. Mol Pharm (2019).
[54] Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: uncertainty in deep learning. ICML (2016).
[55] Schneider G. Automating drug discovery: closing the loop with active learning. Nat Rev Drug Discov (2018).

[56] EMA. Guidelines for use of AI in pharmaceutical R&D. (2023).
[57] Nigam A et al. Diffusion models for molecule generation. NeurIPS (2022).
[58] Su J et al. Multi-modal learning in drug discovery. arXiv (2025).
[59]Reuters. FDA considers AI models as replacements for animal testing. (2025).
[60] Jiménez-Luna J et al. The human–AI collaboration in medicinal chemistry. J Med Chem (2022).
[61] Ferreira FJN et al. AI-Driven Drug Discovery: A Comprehensive Review (2025).