International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 06 | Jun 2026 www.irjet.net p-ISSN: 2395-0072

Demystifying the Black Box: The Shift towards Interpretable Deep Learning in Computer-Aided Drug Design

Ninad Janardan Dani[1], Nilesh K. Deshmukh[2]

[1]1Research Scholar, Department of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded (MH), India

[2]Associate Professor & Head of Department of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded (MH), India

---

Abstract - *The pharmaceutical industry is currently navigating a significant decline in research and development efficiency, coupled with escalating innovation costs. To mitigate these challenges, Computer-Aided Drug Design (CADD) has become an indispensable component of modern drug discovery. At the forefront of this computational revolution is Deep Learning, which has established itself as the definitive engine for complex predictive tasks, including Drug-Target Affinity (DTA) and Drug Response Prediction (DRP). By automatically extracting high-level representations from massive, non-linear chemical and biological datasets, deep neural networks offer unprecedented predictive power. However, the inherent "black-box" nature of these complex architectures presents a substantial bottleneck, raising critical concerns regarding computational limits, transparency, and reproducibility in health research. Relying on opaque models for high-stakes medical decisions is increasingly challenged by the scientific community. This review synthesizes the current landscape of in silico drug discovery, emphasizing that to fully harness the power of deep learning, a paradigm shift toward Explainable Artificial Intelligence (XAI) is required. We examine how integrating interpretable frameworks such as attention mechanisms and multi-omics mapping does not replace deep learning, but rather validates its internal logic. Ultimately, this paper posits that utilizing XAI as a supporting interface is essential to translate state-of-the-art deep learning models from theoretical computational successes into reliable, biologically verified tools for clinical settings.)*

Key Words: Computer-Aided Drug Design, Deep Learning, Explainable AI, Drug-Target Affinity, Drug Response Prediction, Machine Learning, Bioinformatics.

1. INTRODUCTION

The pharmaceutical industry has historically grappled with a significant decline in research and development efficiency [1]. Despite technological advancements, the sheer capital required to innovate and successfully bring a novel therapeutic compound to market has surged exponentially [2]. To overcome these severe financial and temporal bottlenecks, computational strategies have transitioned from being supplementary tools to becoming the indispensable foundation of modern therapeutic discovery. Computer-Aided Drug Design (CADD) provides a systematic framework to navigate the vast, complex chemical space with heightened efficiency [3]. Recent bibliometric analyses confirm the expanding and integral role of in silico methodologies across the entire drug discovery pipeline [4]; [5].

Within the broader umbrella of CADD, the integration of deep learning architectures [6] has catalysed a fundamental paradigm shift [7]. Deep learning currently stands as the definitive, unmatched engine driving modern computational biology. By leveraging multi-layered neural networks, researchers can now bypass traditional, computationally exhaustive physical simulations. Instead, these algorithms automatically extract structural hierarchies from massive, unstructured biological and chemical datasets, achieving unprecedented predictive accuracy in complex domains such as Drug-Target Affinity (DTA) and Drug Response Prediction (DRP).

However, the widespread adoption of these advanced, highly accurate models in clinical and regulatory environments faces a critical barrier: their inherent "black-box" nature. While deep learning models offer massive predictive capabilities, their internal decision-making processes remain largely opaque. The inability to trace exactly how a deep neural network arrives at a specific prediction raises profound challenges regarding reproducibility [8], potential algorithmic bias, and a lack of mechanistic biological validation.

To address this, the scientific community is driving a vital transition toward Explainable Artificial Intelligence (XAI). Crucially, this transition is not a move away from deep learning, but rather an evolution designed to support it. By demystifying complex neural networks into interpretable and biologically sound frameworks, XAI acts as the necessary bridge to validate deep learning's immense computational power [9].

This paper reviews the foundational role of deep learning within CADD. It synthesizes current predictive methodologies, examines the computational limits of opaque models, and explores how supportive XAI mechanisms like attention networks are effectively bridging the gap between deep learning's statistical accuracy and the necessity for biological explainability.

Table -1: The Paradigm Shift and Evolution of Computational Drug Discovery

Phase	Core Approach	Key Characteristics	Technical/Biological Limitation
Phase 1	Traditional R&D	Empirical screening, manual in vitro/in vivo assays	High capital costs, long development timelines, and declining overall efficiency [1].
Phase 2	Classical CADD	Molecular docking, basic statistical chemoinformatics	Highly constrained by physical simulations; limited capacity to process massive multi-omics datasets [5].
Phase 3	Opaque Deep Learning	Deep Neural Networks (DNNs), standard DTA/DRP architectures	Unmatched predictive accuracy, but the "black-box" nature lacks transparency and biological reproducibility [9].
Phase 4	Interpretable Deep Learning	Deep Learning Engine + XAI Support Integration	Maintains state-of-the-art predictive power while providing verifiable biological insights for clinical translation [10].

2. THE RISE OF DEEP LEARNING IN AFFINITY AND RESPONSE PREDICTION

The introduction of deep learning methodologies has fundamentally reshaped the computational landscape of bioinformatics [7]. By utilizing multi-layered artificial neural networks, these algorithms can automatically discover and extract hierarchical representations from raw data [6]. In the context of computer-aided drug design, this capability has proven exceptionally valuable for modeling the intricate, non-linear relationships that govern pharmacological efficacy, specifically in the domains of Drug-Target Affinity (DTA) and Drug Response Prediction (DRP).Irjet template sample paragraph Irjet template sample paragraph.

2.1 Advancements in Drug-Target Affinity (DTA) Prediction

Historically, predicting the binding affinity between a small molecule and a protein target relied heavily on structure-based approaches, such as molecular docking, which are computationally expensive and require high-resolution 3D structural data [11]. The advent of deep learning enabled a shift toward sequence-based and graph-based models that significantly reduce computational overhead while maintaining or improving predictive accuracy. Pioneering architectures like DeepDTA utilized convolutional neural networks (CNNs) to predict binding affinities using only the 1D amino acid sequences of proteins and the SMILES strings of drugs [12]. This approach was subsequently expanded by WideDTA, which incorporated additional text-based features such as ligand maximum common substructures and protein domain profiles to enhance prediction robustness [13]. Furthermore, frameworks like PADME demonstrated the utility of deep learning in handling diverse drug-target interaction contexts [14].

More recently, the field has gravitated toward representing molecules as graphs to capture the topological and relational data of chemical bonds and protein contact maps. Models such as GraphDTA [15] and subsequent graph neural network (GNN) architectures [16] have established new benchmark standards by mathematically mimicking the natural spatial configuration of molecular entities.

2.2 Evolution of Drug Response Prediction (DRP) in Oncology

Transitioning from molecular interactions to broader cellular outcomes, Drug Response Prediction evaluates how specific cell lines particularly in oncology will react to targeted therapeutic interventions. Deep learning has rapidly emerged as a predominant methodology for DRP, driven by the increasing availability of large-scale, high-throughput pharmacogenomic datasets [17]; [18]. Early successful models, such as CDRscan, utilized deep CNNs to predict drug effectiveness directly from mutational cancer genomic signatures [19].

As the complexity of available biological data grew, researchers recognized the necessity of integrating multiple data modalities to capture a holistic view of cellular resistance and sensitivity mechanisms. This led to the development of sophisticated multi-omics frameworks. For example, the MOLI architecture introduced multi-omics late integration using deep neural networks to vastly improve response prediction accuracy [20]. Similar translational efforts have successfully integrated extensive genomic profiles through deep neural networks to predict tumor responses with high fidelity [21]. Comprehensive assessments of these deep learning methodologies confirm their critical utility across the developmental pipeline, successfully translating in vitro computational predictions into tools with substantial clinical application potential [22]. While these state-of-the-art DTA and DRP architectures demonstrate remarkable predictive capabilities, their purely statistical formulation has increasingly highlighted the critical "black-box" limitation, driving the necessity for structural and biological interpretability.

3. THE "BLACK BOX" PROBLEM AND COMPUTATIONAL LIMITS

While the integration of complex deep learning architectures has dramatically improved the statistical accuracy of Drug-Target Affinity and Drug Response Prediction, their widespread adoption is severely hindered by two primary constraints: extreme computational overhead and the inherent opacity of their decision-making processes

.

3.1 Computational and Scalability Bottlenecks

The race to achieve state-of-the-art predictive performance has historically driven the development of increasingly massive and complex neural networks. However, this trajectory is encountering severe diminishing returns. Research into the trajectory of artificial intelligence has explicitly detailed the computational limits of deep learning, revealing that continuing to improve performance through scaling alone demands exponentially greater computational power, energy, and financial resources [23].

In the context of CADD, training deep multi-omics models or large-scale graph neural networks on millions of compound-protein interactions requires prohibitive amounts of memory and processing time. This scalability bottleneck not only restricts advanced drug discovery research to highly funded institutions but also limits the practical, real-time deployment of these models in standard laboratory or clinical settings. Therefore, optimizing these models to maintain high performance while drastically reducing their resource footprint has become an urgent structural necessity.

3.2 The Transparency and Reproducibility Crisis

Beyond hardware limitations, the fundamental architectural flaw of standard deep learning models is their "black-box" nature. A model may accurately predict that a specific cell line is resistant to a given kinase inhibitor, but it cannot explain why or point to the underlying biological mechanism driving that resistance.

This lack of explainability poses severe risks in clinical and pharmacological contexts. Relying on opaque machine learning models for high-stakes medical and developmental decisions is increasingly viewed as an unacceptable risk, driving strong arguments within the scientific community to abandon black-box approaches in favour of inherently interpretable models [9]. Furthermore, the inability to trace how variables are weighted internally has triggered a broader crisis within the field. Extensive assessments of machine learning applications in health research have concluded that reproducibility remains a pervasive, unresolved challenge [8]. Without understanding the model's internal logic, identifying algorithmic bias or verifying that the model has learned true biological principles rather than simply memorizing dataset artifacts is nearly impossible.

The urgency to address these flaws has catalysed widespread calls for mandatory transparency and reproducibility in artificial intelligence [24]. For geneticists and computational biologists, opening the black box is no longer just an academic exercise; it is a prerequisite for extracting actionable, mechanistic insights from complex biological data [25].

4. THE TRANSITION TO EXPLAINABLE AI (XAI)

To overcome the severe limitations of opaque predictive models and unlock the full clinical potential of deep learning, the computational biology community is integrating Explainable Artificial Intelligence (XAI). Extensive surveys emphasize that transparency is a fundamental requirement for the integration of medical AI [10]. In the realm of Computer-Aided Drug Design, XAI does not replace deep learning; rather, it acts as a critical interface that bridges the gap between statistical prediction and verifiable biological reality. For geneticists and computational biochemists, utilizing interpretability tools allows them to validate deep learning's internal logic, thereby facilitating true scientific discovery [25].

4.1 Mechanisms of Interpretability in Drug-Target Affinity

In Drug-Target Affinity (DTA) prediction, interpretability implies the model can highlight exactly which sub-structures of a ligand interact with specific amino acid residues of a protein target. Earlier models processed entire SMILES strings [26] and protein sequences without spatial context. However, modern interpretable architectures are designed to intrinsically map these interactions.

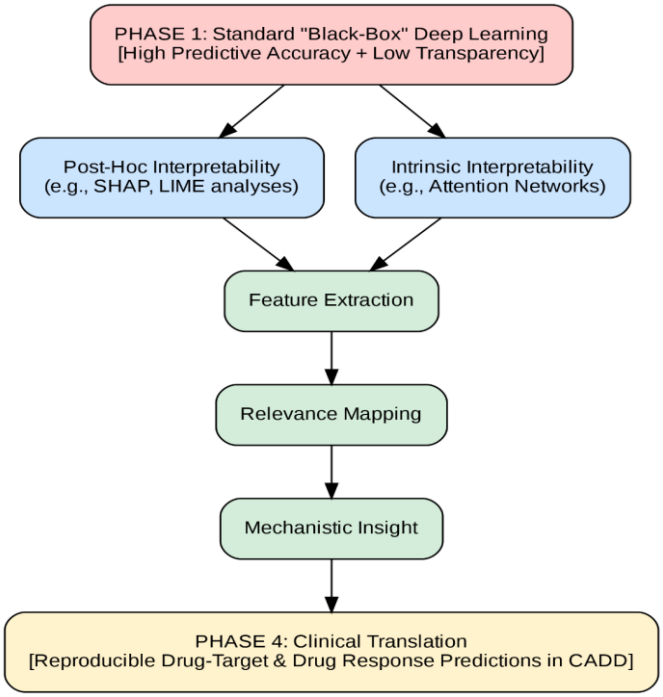
A prime example is the Deep Affinity framework, which utilizes unified recurrent and convolutional neural networks not only to predict compound-protein affinity but to do so interpretably [27]. By incorporating attention mechanisms, these advanced models assign dynamic weights to distinct parts of the input sequences. Similarly, the implementation of interpretable bilinear attention networks coupled with domain adaptation has demonstrated significant improvements in both the accuracy and explainability of drug-target predictions [28]. These networks mathematically align the latent representations of the drug and the target, outputting an "attention map" that researchers can visually inspect to confirm that the model is identifying known binding pockets, rather than relying on dataset bias.

4.2 Interpreting Drug Response in Complex Multi-Omics

Interpreting Drug Response Prediction (DRP) is inherently more complex than DTA, as it requires analyzing vast, heterogeneous multi-omics profiles ranging from transcriptomics to mutation data to predict cellular

sensitivity [29]. There are distinct opportunities and significant computational challenges in developing interpretable deep learning models for the drug sensitivity of cancer cells [30].

Flowchart 1: The Architecture of Explainable AI in Drug Discovery



To address this, researchers have developed specialized frameworks like PaccMann, a web service designed specifically for interpretable anticancer compound sensitivity prediction [31]. Frameworks in this category often utilize gene-level attention. Instead of simply predicting that a cell line will resist a specific tyrosine kinase inhibitor, an interpretable DRP model can output a ranked list of the specific genes or transcriptomic signatures that heavily influenced its decision. This aligns perfectly with experimental oncology, where identifying specific gene signatures such as an epithelial-mesenchymal transition signature predicting resistance to EGFR and PI3K inhibitors is crucial for developing targeted therapies [32]; [33].

4.3 Bridging the Gap: The Future of Unified XAI

The ultimate goal of XAI in drug discovery is to create models that are not only interpretable but also capable of bridging multiple predictive domains. By utilizing resource-optimized, multimodal architectures, researchers can potentially unify DTA and DRP. In such a system, the interpretable attention weights extracted at the cellular level (DRP) could be cross-referenced with the molecular interaction maps (DTA), providing a comprehensive, end-to-end explanation of a drug's mechanism of action and its resulting phenotypic effect. This biologically grounded approach directly addresses the reproducibility crisis [8] while mitigating the computational limits of deep learning through focused, feature-specific processing.

5. THE PARADIGM SHIFT: DEEP LEARNING AS THE ULTIMATE FRONTIER IN CADD

The transition from classical, rule-based computational chemistry to deep learning represents one of the most significant leaps in modern pharmaceutical history. While traditional computer-aided drug design relied on heavily constrained physical simulations and linear statistical models, deep learning architectures have proven uniquely capable of conquering the sheer complexity of biological systems [6]; [7]. Deep learning is not merely an incremental upgrade; it is the definitive foundation upon which the future of therapeutic discovery is being built.

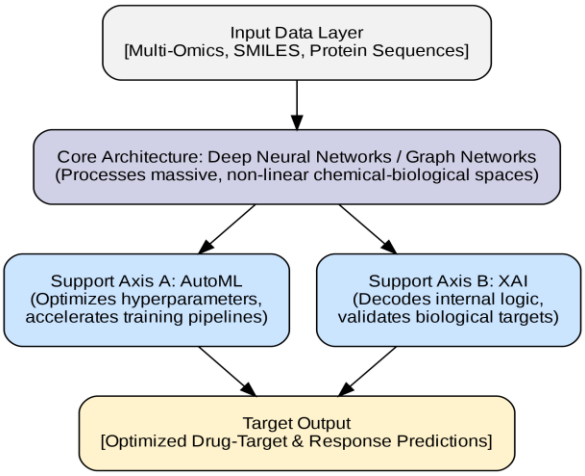
5.1 Conquering the Chemical and Biological Search Space

The ultimate triumph of deep learning in CADD lies in its unmatched capacity for representation learning. Traditional methods like molecular docking or molecular dynamics simulations require massive computational resources to calculate physical force fields and spatial interactions step-by-step [11]; [34]. In contrast, deep neural networks can process abstract structural representations such as chemical graphs or simplified textual strings and automatically discover the underlying, non-linear feature hierarchies that dictate binding affinity and biological response [26]; [12].

This capability solves the historic "curse of dimensionality" inherent in chemoinformatics and multi-omics data integration [35]. By processing heterogeneous datasets simultaneously, deep learning architectures can identify complex, hidden correlations across transcriptomic profiles, chemical fingerprints, and genomic variants that human researchers or classical algorithms could never explicitly program [29]; [20]. The sheer velocity at which these models can screen libraries of billions of compounds transitions drug discovery from an empirical, trial-and-error process into a precise, predictive data science.

5.2 Deep Learning as the Core Engine of Modern CADD

To visualize how deep learning reigns supreme as the primary driver of this field, the layout below maps its operational workflow, demonstrating how support frameworks like XAI and Automated Machine Learning (AutoML) serve to reinforce, rather than replace, the core deep learning architecture.



5.3 Synergistic Reinforcement: How XAI and AutoML Serve Deep Learning

Rather than competing with deep learning, frameworks like Explainable AI (XAI) and Automated Machine Learning (AutoML) act as critical supporting infrastructure that solidifies deep learning's dominance.

- AutoML as an Accelerator: Automated machine learning frameworks help bypass the grueling process of manual hyperparameter tuning, allowing deep learning models to be deployed faster, more efficiently, and with minimized human bias [36].
- XAI as a Validator: Explainable AI does not diminish the role of deep learning; instead, it elevates it. By providing a window into the model's feature attribution, XAI proves to regulatory bodies and experimental biologists that the deep learning engine is making decisions based on legitimate structural biology, such as specific kinase-inhibitor binding pockets or established oncogenic pathways [30]; [33].

Ultimately, deep learning remains the undisputed champion of modern CADD [37]. Its unmatched predictive power is what makes the acceleration of drug discovery possible. By leveraging optimization and interpretability workflows, the scientific community can safely harness this computational engine, cementing deep learning as the permanent backbone of next-generation pharmaceutical innovation.

6. CONCLUSIONS

The integration of deep learning has fundamentally revolutionized Computer-Aided Drug Design, turning it into a highly predictive, scalable, and indispensable engine for modern therapeutic discovery. Faced with the severe economic pressures of declining R&D efficiency and skyrocketing development costs [1]; [2], deep learning stands out as the ultimate technological breakthrough capable of navigating complex chemical and biological data at scale.

While architectural opacity and computational limits initially presented hurdles, the strategic integration of supporting methodologies namely Explainable AI and resource-optimization techniques has effectively mitigated these challenges [9]; [23]. These advancements do not displace deep learning; rather, they validate its predictions, making them reproducible, transparent, and ready for clinical and regulatory translation [24]. As the field moves forward, the synergy between massive deep learning architectures and robust interpretability frameworks will remain the cornerstone of biomedical research, unlocking a new era of accelerated, precise, and highly efficient drug discovery.

REFERENCES

[1] J. W. Scannell, A. Blanckley, H. Boldon, and B. Warrington, "Diagnosing the decline in pharmaceutical R&D efficiency," Nature reviews Drug discovery, vol. 11, no. 3, pp. 191–200, 2012.

[2] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen, "Innovation in the pharmaceutical industry: new estimates of R&D costs," Journal of health economics, vol. 47, pp. 20–33, 2016.

[3] S. K. Niazi and Z. Mariam, "Computer-aided drug design and drug discovery: a prospective analysis," Pharmaceuticals, vol. 17, no. 1, p. 22, 2023.

[4] Z. Wu et al., "Current perspectives and trend of computer-aided drug design: a review and bibliometric analysis," International Journal of Surgery, vol. 110, no. 6, pp. 3848–3878, 2024.

[5] S. J. Y. Macalino, V. Gosu, S. Hong, and S. Choi, "Role of computer-aided drug design in modern drug discovery," Arch. Pharm. Res., vol. 38, no. 9, pp. 1686–1701, Sep. 2015, doi: 10.1007/s12272-015-0640-5.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436–444, 2015.

[7] S. Min, B. Lee, and S. Yoon, "Deep learning in bioinformatics," Briefings in bioinformatics, vol. 18, no. 5, pp. 851–869, 2017.

[8] M. B. A. McDermott, S. Wang, N. Marinsek, R. Ranganath, L. Foschini, and M. Ghassemi, "Reproducibility in machine learning for health research: Still a ways to go," Sci. Transl. Med., vol. 13, no. 586, p. eabb1655, Mar. 2021, doi: 10.1126/scitranslmed.abb1655.

[9] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1: 206–215." 2019.

[10] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (xai): Toward medical xai," IEEE transactions on neural networks and learning systems, vol. 32, no. 11, pp. 4793–4813, 2020.

[11] X.-Y. Meng, H.-X. Zhang, M. Mezei, and M. Cui, "Molecular docking: a powerful approach for structure-based drug discovery," Current computer-aided drug design, vol. 7, no. 2, pp. 146–157, 2011.

[12] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: deep drug–target binding affinity prediction," Bioinformatics, vol. 34, no. 17, pp. i821–i829, 2018.

[13] H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: prediction of drug-target binding affinity," Feb. 04, 2019, arXiv: arXiv:1902.04166. doi: 10.48550/arXiv.1902.04166.

[14] Q. Feng, E. Dueva, A. Cherkasov, and M. Ester, "PADME: A Deep Learning-based Framework for Drug-Target Interaction Prediction," Aug. 21, 2019, arXiv: arXiv:1807.09741. doi: 10.48550/arXiv.1807.09741.

[15] T. Nguyen, H. Le, T. P. Quinn, T. Nguyen, T. D. Le, and S. Venkatesh, "GraphDTA: predicting drug–target binding affinity with graph neural networks," Bioinformatics, vol. 37, no. 8, pp. 1140–1147, 2021.

[16] M. Jiang et al., "Drug–target affinity prediction using graph neural network and contact maps," RSC advances, vol. 10, no. 35, pp. 20701–20712, 2020.

[17] A. Partin et al., "Deep learning methods for drug response prediction in cancer: predominant and emerging trends," Frontiers in medicine, vol. 10, p. 1086097, 2023.

[18] D. Baptista, P. G. Ferreira, and M. Rocha, "Deep learning for drug response prediction in cancer," Briefings in bioinformatics, vol. 22, no. 1, pp. 360–379, 2021.

[19] Y. Chang et al., "Cancer drug response profile scan (CDRscan): a deep learning model that predicts drug effectiveness from cancer genomic signature," Scientific reports, vol. 8, no. 1, p. 8857, 2018.

[20] H. Sharifi-Noghabi, O. Zolotareva, C. C. Collins, and M. Ester, "MOLI: multi-omics late integration with deep neural networks for drug response prediction," Bioinformatics, vol. 35, no. 14, pp. i501–i509, 2019.

[21] Y.-C. Chiu et al., "Predicting drug response of tumors from integrated genomic profiles by deep neural networks," BMC Med Genomics, vol. 12, no. S1, p. 18, Jan. 2019, doi: 10.1186/s12920-018-0460-9.

[22] B. Shen, F. Feng, K. Li, P. Lin, L. Ma, and H. Li, "A systematic assessment of deep learning methods for drug response prediction: from in vitro to clinical applications," Briefings in Bioinformatics, vol. 24, no. 1, p. bbac605, 2023.

[23] C. Thompson Neil, G. Kristjan, L. Keeheon, and F. Manso Gabriel, "The computational limits of deep learning," ArXiv, Cornell University, juillet, 2020.

[24] B. Haibe-Kains et al., "Transparency and reproducibility in artificial intelligence," Nature, vol. 586, no. 7829, pp. E14–E16, 2020.

[25] C. B. Azodi, J. Tang, and S.-H. Shiu, "Opening the black box: interpretable machine learning for geneticists," Trends in genetics, vol. 36, no. 6, pp. 442–455, 2020.

[26] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," J. Chem. Inf. Comput. Sci., vol. 28, no. 1, pp. 31–36, Feb. 1988, doi: 10.1021/ci00057a005.

[27] M. Karimi, D. Wu, Z. Wang, and Y. Shen, "DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks," Bioinformatics, vol. 35, no. 18, pp. 3329–3338, 2019.

[28] P. Bai, F. Miljković, B. John, and H. Lu, "Interpretable bilinear attention network with domain adaptation improves drug–target prediction," Nature Machine Intelligence, vol. 5, no. 2, pp. 126–136, 2023.

[29] M. Picard, M.-P. Scott-Boyer, A. Bodein, O. Périn, and A. Droit, "Integration strategies of multi-omics data for machine learning analysis," Computational and Structural Biotechnology Journal, vol. 19, pp. 3735–3746, 2021.

[30] B. R. Samal, J. U. Loers, V. Vermeirssen, and K. De Preter, "Opportunities and challenges in interpretable deep learning for drug sensitivity prediction of cancer cells," Frontiers in Bioinformatics, vol. 2, p. 1036963, 2022.

[31] J. Cadow, J. Born, M. Manica, A. Oskooei, and M. Rodríguez Martínez, "PaccMann: a web service for interpretable anticancer compound sensitivity prediction," Nucleic acids research, vol. 48, no. W1, pp. W502–W508, 2020.

[32] L. A. Byers et al., "An epithelial–mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance," Clinical cancer research, vol. 19, no. 1, pp. 279–290, 2013.

[33] M. Ye et al., "Activation of the aryl hydrocarbon receptor leads to resistance to EGFR TKIs in non–small cell lung cancer by activating src-mediated bypass signaling," Clinical Cancer Research, vol. 24, no. 5, pp. 1227–1239, 2018.

[34] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," Nature structural biology, vol. 9, no. 9, pp. 646–652, 2002.

[35] M. Zitnik, F. Nguyen, B. Wang, J. Leskovec, A. Goldenberg, and M. M. Hoffman, "Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities," Information Fusion, vol. 50, pp. 71–91, 2019.

[36] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester, "SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines," J Cheminform, vol. 9, no. 1, p. 24, Dec. 2017, doi: 10.1186/s13321-017-0209-z.

[37] J. Vamathevan et al., "Applications of machine learning in drug discovery and development," Nature reviews Drug discovery, vol. 18, no. 6, pp. 463–477, 2019