Japan Journal of Research

Review Article

Deep Learning in Drug Discovery: Current Landscape and Future Prospects

Sensen Liu[1] and Yin-Shan Lin[2]

[1]Washington Univeristy in St. Louis, St.Louis, MO, USA
[2]Northeastern University, Boston, MA, USA

Abstract

Deep learning (DL) has emerged as a transformative technology in drug discovery, offering the potential to accelerate and optimize various stages of the drug development pipeline. While numerous reviews have summarized the broader landscape of machine learning (ML) in this field, this review focuses specifically on deep learning, highlighting its unique strengths and challenges. We examine the current state-of-the-art deep learning algorithms applied in drug discovery, categorizing them by their architectural designs and applications. We further identify emerging trends and potential areas for future research, emphasizing the need for continued exploration and innovation at the intersection of deep learning and drug discovery.

Correspondence

Sensen Liu

Washington Univeristy in St. Louis, St.Louis, MO, USA
E-mail: lius@ese.wustl.edu

- Received Date: 17 May 2024
- Accepted Date: 29 May 2024
- Publication Date: 31 May 2024

Keywords

Deep learning; Drug discovery; Neural networks; Artificial intelligence; Therapeutics

Introduction

The traditional drug discovery process is plagued by inherent complexities, exorbitant costs, and protracted timelines, often spanning a decade or more [1]. To address these formidable challenges, computational approaches, particularly machine learning (ML), have been increasingly integrated into the pharmaceutical research landscape. While previous reviews have explored the broader applications of ML in drug discovery [2], the distinct capabilities and rapid advancements of deep learning (DL) necessitate a focused and comprehensive evaluation.

Deep learning, characterized by its multi-layered neural networks, has demonstrated remarkable proficiency in extracting intricate patterns and representations from large-scale datasets [3]. In the realm of drug discovery, DL algorithms have been applied to a wide array of tasks, from target identification and validation to virtual screening, lead optimization, and drug repurposing [4]. This review systematically examines the current landscape of deep learning (DL) applications in drug discovery, assessing different DL architectures, their performance, and identifying areas for further research. Our goal is to provide a comprehensive overview of DL's current state and future potential in this transformative field.

Deep Learning Algorithms in Drug Discovery

Target Identification and Validation

DL models, including convolutional neural networks (CNNs) and graph neural networks (GNNs), have shown promise in target identification and validation. CNNs, leveraging principles of local feature extraction and hierarchical composition successful in image recognition, are well-suited for modeling biochemical interactions due to their inherent locality [5]. By identifying local patterns, such as hydrogen bond donors and acceptors, CNNs can construct complex representations of molecular binding [6]. GNNs, adept at modeling graph-structured data like molecular structures, have shown efficacy in predicting various molecular properties [7]. However, challenges such as model interpretability and generalizability remain [8], with potential solutions lying in integrating multi-omics data and explainable AI (XAI) techniques.

Virtual Screening and Lead Discovery

Generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs), are powerful tools for generating novel chemical structures with desired properties. VAEs learn the underlying distribution of molecular structures, enabling the generation of novel compounds with specific constraints like target binding affinity and synthetic accessibility [9]. By encoding molecules into a continuous latent space, VAEs can sample and generate diverse molecules with potentially improved properties, while incorporating conditional information allows for task-specific optimization [10]. Balancing exploration and exploitation, VAEs expand the chemical space while ensuring generated molecules possess desired characteristics [11]. GANs have been applied to molecular de novo design, generating drug-like molecules tailored to specific targets or properties [12]. However, challenges such as synthetic feasibility and novelty remain areas for further research.

Citation: Liu S, Lin Y-S. Deep Learning in Drug Discovery: Current Landscape and Future Prospects. Japan J Res. 2024;5(4):028

Lead Optimization and Property Prediction

DL models are employed in lead optimization to refine compound properties for safety and efficacy. Recurrent neural networks (RNNs), traditionally utilized in natural language processing (NLP), have found new applications in predicting ADMET (absorption, distribution, metabolism, excretion, toxicity) properties. RNNs can predict drug metabolism and toxicity based on time-series data, but data limitations and complex metabolic processes present challenges [13]. Transformer architectures, renowned for their attention mechanisms that weigh the importance of different input elements, have demonstrated exceptional performance in various NLP tasks. This attention mechanism has proven valuable in predicting molecular properties and DTIs, as it allows models to focus on relevant chemical features and interactions [14]. Transformer-based models have achieved state-of-the-art results in predicting protein-ligand binding affinities and identifying potential drug targets [15]. However, addressing data scarcity and model interpretability are crucial for advancing this area [16].

Drug Repurposing and Combination Therapy

DL models transform drug repurposing and combination therapy by mining large-scale datasets for hidden therapeutic potential [17]. In drug repurposing, natural language processing (NLP) models analyze drug-disease associations, gene expression profiles, and clinical outcomes to identify potential new indications for existing drugs [18]. DL models also facilitate combination therapy by predicting synergistic drug combinations based on molecular, pharmacological, and clinical data [19].

However, challenges persist. Data heterogeneity, varying in quality and format, complicates integration and analysis. Confounding factors, like patient demographics, can obscure the true effects of drug combinations. The need for interpretable models remains paramount as well.

Addressing these challenges through the development of novel DL architectures, incorporating attention mechanisms, holds the potential to further accelerate drug repurposing and combination therapy. This could ultimately transform drug discovery and significantly improve patient outcomes.

Future Directions

Future directions in DL-driven drug discovery include:

Explainable AI (XAI): Developing models that provide transparent and understandable explanations for their predictions.

Multi-Omics Integration: Combining diverse data sources for a holistic understanding of biological systems. Transfer Learning and Domain Adaptation: Leveraging pre-trained models to accelerate development and reduce the need for large training datasets.

Reinforcement Learning: Applying reinforcement learning techniques to drug design can enable the iterative optimization of molecules towards desired properties, potentially leading to the discovery of novel compounds with enhanced efficacy and safety profiles.

Collaboration and Data Sharing: Fostering collaboration between academia, industry, and regulatory bodies is crucial for sharing data, expertise, and resources, thereby accelerating progress in DL-driven drug discovery.

Conclusion

Deep learning has emerged as a powerful and versatile tool in drug discovery, demonstrating the potential to transform various stages of the drug development pipeline. While significant progress has been made, several challenges and opportunities remain. By addressing these challenges and embracing emerging trends, the field of DL-driven drug discovery is poised to make significant contributions to the development of safer, more effective, and affordable therapeutics.

References

1. Malerba F, Orsenigo L. The evolution of the pharmaceutical industry. Bus Hist. 2015;57(5):664-687.
2. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-477.
3. Liu S, Ching S. Recurrent information optimization with local, metaplastic synaptic dynamics. Neural Comput. 2017;29(9):2528-2552.
4. Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-1250.
5. Bissantz C, Kuhn B, Stahl M. A medicinal chemist's guide to molecular interactions. J Med Chem. 2010;53(14):5061-5084.
6. Wallach I, Dzamba M, Heifets A. Atomnet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery. arXiv preprint arXiv:1510.02855. 2015.
7. Wu Z, Ramsundar B, Feinberg EN, et al. Moleculenet: a benchmark for molecular machine learning. Chem Sci. 2018;9(2):513-530.
8. Jiang D, Wu Z, Hsieh CY, et al. Could graph neural networks learn better molecular representation for drug discovery? a comparison study of descriptor-based and graph-based models. J Cheminform. 2021;13:1-23.
9. Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268-276.
10. Pravalphruekul N, Piriyajitakonkij M, Phunchongharn P, Piyayotai S. De novo design of molecules with multiaction potential from differential gene expression using variational autoencoder. J Chem Inf Model. 2023;63(13):3999-4011.
11. Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. In: International Conference on Machine Learning. PMLR; 2018:2323-2332.
12. Sanchez-Lengeling B, Outeiral C, Guimaraes GL, Aspuru-Guzik A. Optimizing distributions over molecular space. an objective-reinforced generative adversarial network for inverse-design chemistry (organic). ChemRxiv. 2017.
13. Xu Y, Dai Z, Chen F, et al. Deep learning for drug-induced liver injury. J Chem Inf Model. 2015;55(10):2085-2093.
14. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with alphafold. Nature. 2021;596(7873):583-589.
15. Chen R, Liu X, Jin S, Lin J, Liu J. Machine learning for drug-target interaction prediction. Molecules. 2018;23(9):2208.

16. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-584.
17. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41-58.
18. Pan X, Lin X, Cao D, et al. Deep learning for drug repurposing: Methods, databases, and applications. Wiley Interdiscip Rev Comput Mol Sci. 2022;12(4):e1597.
19. Preuer K, Lewis RPI, Hochreiter S, Bender A, Bulusu KC, Klambauer G. Deepsynergy: predicting anti-cancer drug synergy with deep learning. Bioinformatics. 2018;34(9):1538-1546.