Artificial Intelligence in Drug Discovery for Rare Diseases

Rishabh Kumar, Chanchal Kashyap, Raj Kumar, Krishna Anand

Maa Shakumbhari University, Saharanpur, India

DOI: https://doi.org/10.51584/IJRIAS.2026.110400175

Received: 25 April 2026; Accepted: 30 April 2026; Published: 18 May 2026

ABSTRACT

Rare diseases affect a significant portion of the global population despite their individual rarity, yet therapeutic development remains limited due to economic and scientific constraints. Artificial intelligence (AI) has emerged as a transformative approach in pharmaceutical research, enabling the analysis of large-scale biological datasets and accelerating the identification of potential drug candidates. This study explores the role of machine learning and deep learning techniques in rare disease drug discovery. AI-driven models facilitate drug-target interaction prediction, molecular optimization, and drug repurposing, significantly reducing time and cost. The paper also discusses methodological frameworks, applications, challenges, and future directions of AI integration in pharmaceutical research. The findings indicate that AI has the potential to revolutionize rare disease treatment by improving efficiency, accuracy, and accessibility.

Keywords: Artificial intelligence; drug discovery; rare diseases; machine learning; deep learning; bioinformatics.

INTRODUCTION

Drug discovery is widely recognized as a complex, resource-intensive, and time-consuming process that often spans more than a decade. The development of new therapeutic agents involves multiple stages, including target identification, compound screening, preclinical evaluation, and clinical trials, each of which requires significant investment and effort. These challenges are further intensified in the case of rare diseases, which affect a relatively small population but collectively represent a major global health burden. The limited availability of clinical data, along with reduced commercial incentives for pharmaceutical companies, significantly hinders the development of effective treatments for such conditions.

In recent years, artificial intelligence (AI) has emerged as a promising and transformative approach to overcoming these limitations. By leveraging advanced computational techniques, AI enables the efficient analysis of complex and large-scale biological datasets. Machine learning algorithms, in particular, can identify hidden patterns in genomic and proteomic data, thereby facilitating the discovery of novel drug targets (Chen et al., 2018). Furthermore, deep learning techniques enhance predictive accuracy in critical areas such as protein structure prediction and molecular interaction analysis (Jumper et al., 2021).

These technological advancements have opened new avenues for accelerating drug discovery processes, reducing development time and cost, and improving the success rate of therapeutic interventions. Consequently, AI-driven approaches hold significant potential in advancing research and development for rare diseases, where traditional methodologies often fall short.

LITERATURE REVIEW

Recent advancements in artificial intelligence have significantly influenced pharmaceutical research, particularly in the domain of drug discovery. A growing body of literature highlights the potential of AI-driven approaches to enhance various stages of drug development, including target identification, molecular design, and clinical evaluation.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Chen et al. (2018) demonstrated that deep learning models can effectively predict molecular properties by analyzing large-scale chemical and biological datasets. Their study emphasized that neural network-based approaches outperform traditional computational methods in terms of accuracy and efficiency. Similarly, Vamathevan et al. (2019) provided a comprehensive overview of machine learning applications in drug discovery and reported that AI techniques can substantially improve the success rates of drug development by enabling data-driven decision-making and reducing failure rates in clinical trials.

A major breakthrough in the field was introduced by Jumper et al. (2021) through the development of AlphaFold, a deep learning-based system capable of predicting protein structures with remarkable accuracy. This advancement has profound implications for drug discovery, as understanding protein folding and structure is essential for identifying drug targets and designing effective therapeutic agents.

In addition to target identification and molecular prediction, drug repurposing has emerged as a cost-effective and time-efficient strategy in pharmaceutical research. Zhavoronkov (2018) highlighted the role of artificial intelligence in identifying new therapeutic applications for existing drugs by analyzing complex biomedical datasets. This approach is particularly beneficial for rare diseases, where the development of new drugs is often limited by economic and logistical constraints.

Furthermore, recent studies have explored the integration of AI with multi-omics data, including genomics, proteomics, and metabolomics, to gain deeper insights into disease mechanisms. These integrative approaches enable a more comprehensive understanding of biological systems and facilitate the discovery of novel therapeutic targets.

Overall, the existing literature clearly indicates that artificial intelligence is transforming drug discovery by improving efficiency, reducing costs, and accelerating the development of innovative therapies, particularly in the context of rare diseases.

METHODOLOGY

- The research methodology adopted in this study involves a systematic and structured approach to applying artificial intelligence techniques in drug discovery for rare diseases. The process consists of multiple stages, including data collection, preprocessing, model development, and validation, each of which plays a crucial role in ensuring the accuracy and reliability of the results.
- The first stage involves data collection from various publicly available biomedical databases. This includes genomic, proteomic, and chemical datasets, which provide essential information regarding gene expression, protein structures, and molecular properties. These datasets form the foundation for training artificial intelligence models and enable the identification of potential drug targets.
- Following data collection, preprocessing is performed to improve data quality and ensure consistency. This step includes data cleaning, removal of missing or redundant values, normalization of numerical features, and transformation of data into a suitable format for model training. Proper preprocessing is critical, as it directly impacts the performance and accuracy of the models.
- The next stage involves model development using both machine learning and deep learning techniques. Traditional machine learning algorithms such as Support Vector Machines (SVM) and Random Forest are employed for classification and prediction tasks. In addition, advanced deep learning models, including Convolutional Neural Networks (CNN) and Graph Neural Networks (GNN), are utilized to capture complex patterns and relationships within high-dimensional biological data. These models are particularly effective in predicting drug–target interactions and molecular properties.
- Finally, model validation is carried out to evaluate the performance and generalizability of the developed models. Cross-validation techniques are applied to minimize overfitting and ensure robustness. Performance metrics such as accuracy, precision, recall, and F1-score are used to assess the effectiveness of the models in predicting reliable outcomes.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

- Overall, this methodological framework provides a comprehensive approach to integrating artificial intelligence into drug discovery, ensuring accurate predictions and efficient analysis for rare disease research.

RESULTS

The results of this study demonstrate the significant effectiveness of artificial intelligence– based models in various stages of drug discovery, particularly in predicting drug–target interactions. Machine learning algorithms, such as Random Forest and Support Vector Machines, achieved high predictive accuracy when applied to molecular datasets, indicating their strong capability in identifying potential therapeutic targets.

Furthermore, deep learning approaches, including convolutional neural networks and graph neural networks, outperformed traditional computational methods in molecular analysis. These models were able to capture complex structural and chemical relationships within high-dimensional datasets, leading to more precise predictions of molecular properties and binding affinities. This improvement in predictive performance highlights the advantage of deep learning techniques in handling complex biological data.

Another key finding of this study is the reduction in drug discovery timelines achieved through AI-driven computational modeling. By automating processes such as compound screening and target identification, AI significantly decreases the time required for earlystage drug development. Compared to conventional methods, which may take several years, AI-based approaches can rapidly analyze large datasets and generate reliable predictions within a much shorter time frame.

Additionally, AI-assisted drug repurposing strategies demonstrated promising results by identifying new therapeutic applications for existing drugs. This approach not only reduces development costs but also accelerates the availability of treatments for rare diseases.

Overall, the results clearly indicate that artificial intelligence enhances efficiency, accuracy, and speed in drug discovery processes, making it a highly valuable tool for advancing research in rare diseases.

DISCUSSION

The findings of this study clearly indicate that artificial intelligence plays a crucial role in improving the efficiency and cost-effectiveness of pharmaceutical research. By enabling the rapid analysis of large-scale biomedical datasets, AI-driven approaches significantly reduce the time required for drug discovery processes. This improvement is particularly important when compared to traditional methods, which are often time-consuming, laborintensive, and associated with high failure rates.

One of the most notable advantages of AI observed in this study is its ability to facilitate drug repurposing. AI-based models can analyze existing drug databases and identify new therapeutic applications for previously approved compounds. This strategy is especially beneficial for rare diseases, where the development of new drugs is often limited due to economic constraints and insufficient clinical data. Drug repurposing not only reduces development costs but also accelerates the availability of effective treatments.

In addition, the results highlight the superiority of deep learning techniques over conventional computational approaches in handling complex molecular and biological data. Advanced models such as neural networks are capable of capturing intricate patterns and relationships, leading to more accurate predictions of drug-target interactions and molecular properties. These capabilities contribute to improved decision-making in early-stage drug development.

However, despite these advantages, several challenges remain in the implementation of AI in drug discovery. One of the major limitations is the lack of high-quality and large-scale datasets, particularly in the context of rare diseases. Limited data availability can affect model performance and lead to overfitting issues. Furthermore, the "black-box" nature of many deep learning models raises concerns regarding interpretability and transparency, which are critical for regulatory approval and clinical trust.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Overall, while artificial intelligence offers substantial benefits in accelerating drug discovery and improving research outcomes, addressing these challenges is essential to fully realize its potential in pharmaceutical applications.

Challenges and Limitations

- Despite the significant advancements in artificial intelligence for drug discovery, several challenges and limitations continue to hinder its widespread adoption,
- particularly in the context of rare diseases. One of the primary challenges is the limited availability of high-quality datasets. Rare diseases, by definition, affect a small population, resulting in insufficient clinical and biological data for training robust AI models. This scarcity of data can lead to issues such as overfitting and reduced generalizability of predictive models.
- Another important limitation is the "black-box" nature of many artificial intelligence algorithms, especially deep learning models. While these models are capable of delivering highly accurate predictions, their lack of interpretability makes it difficult for researchers and clinicians to understand the underlying decision-making process. This lack of transparency can reduce trust in AI-based systems and poses challenges for their integration into clinical practice.
- In addition, regulatory and ethical concerns play a significant role in limiting the application of AI in pharmaceutical research. The use of patient data raises important issues related to privacy, data security, and informed consent. Furthermore, regulatory agencies require rigorous validation and transparency before approving AI-driven drug discovery approaches, which can slow down their implementation.
- Overall, addressing these challenges is essential to ensure the reliable, ethical, and effective use of artificial intelligence in drug discovery, particularly for rare diseases where the need for innovative solutions is critical.

Future Scope

The future of artificial intelligence in drug discovery, particularly for rare diseases, is highly promising and continues to evolve rapidly with ongoing technological advancements. One of the most significant emerging areas is the application of generative artificial intelligence for drug design. Advanced models such as generative adversarial networks and variational autoencoders have the capability to design novel chemical compounds with optimized therapeutic properties, thereby accelerating the development of new drugs.

Another important direction is the integration of multi-omics data, including genomics, proteomics, and metabolomics. By combining these diverse biological datasets, researchers can gain a more comprehensive understanding of disease mechanisms at the molecular level. This holistic approach enhances the identification of potential drug targets and improves the precision of therapeutic interventions.

Personalized medicine also represents a key area of future development. Artificial intelligence enables the analysis of patient-specific data to design tailored treatment strategies, thereby improving treatment efficacy and reducing adverse effects. This approach is particularly valuable in rare diseases, where individual variability plays a crucial role in disease progression and treatment response.

In addition, the integration of AI with automated laboratory systems is expected to revolutionize experimental workflows. AI-driven automated laboratories can perform highthroughput experiments with minimal human intervention, increasing efficiency, reproducibility, and accuracy in drug discovery processes.

Overall, these emerging trends indicate that artificial intelligence will continue to play a

transformative role in pharmaceutical research, leading to faster, more efficient, and more personalized approaches to drug discovery in the future.

CONCLUSION

In conclusion, artificial intelligence has emerged as a powerful and transformative tool in the field of drug discovery, particularly for rare diseases where traditional approaches face significant limitations. The integration of advanced computational techniques, including machine learning and deep learning, enables the efficient analysis of complex biomedical datasets, leading to the identification of novel therapeutic targets and the rapid development of potential drug candidates.

The findings of this study highlight that AI-driven approaches not only enhance the speed and accuracy of drug discovery processes but also significantly reduce associated costs and development timelines. Moreover, the application of artificial intelligence in areas such as drug repurposing and personalized medicine further improves therapeutic outcomes and expands treatment possibilities for patients suffering from rare diseases.

Despite these advancements, challenges related to data availability, model interpretability, and regulatory compliance must be addressed to ensure the reliable and ethical implementation of AI in pharmaceutical research. Overcoming these limitations will be essential for maximizing the impact of AI technologies in real-world clinical settings.

Looking ahead, continuous progress in artificial intelligence, including innovations in generative modeling, multi-omics data integration, and automated laboratory systems, is expected to further accelerate pharmaceutical innovation. These developments will play a crucial role in addressing unmet medical needs and improving the quality of healthcare for individuals affected by rare diseases.

REFERENCES

1. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M., & Blaschke, T. (2018). The rise of deep learning in drug discovery. Drug Discovery Today, 23(6), 1241–1250.
2. Ekins, S. (2019). Exploiting machine learning for drug discovery. Nature Materials, 18(5), 435–436.
3. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589.
4. Mak, K. K., & Pichika, M. R. (2019). Artificial intelligence in drug development: Present status and future prospects. Drug Discovery Today, 24(3), 773–780.
5. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. Drug Discovery Today, 26(1), 80–93.
6. Schneider, G. (2020). Automating drug discovery. Nature Reviews Drug Discovery, 19(6), 353–364.
7. Zhavoronkov, A. (2018). Artificial intelligence for drug discovery, biomarker development, and generation of novel chemistry. Nature Biotechnology, 36(11), 1038–1040.
8. Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., … Collins, J. J. (2020). A deep learning approach to antibiotic discovery. Cell, 180(4), 688–702.
9. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., … Snowden, M. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463–477.
10. Altae-Tran, H., Ramsundar, B., Pappu, A. S., & Pande, V. (2017). Low data drug discovery with one-shot learning. ACS Central Science, 3(4), 283–293.
11. Segler, M. H. S., Preuss, M., & Waller, M. P. (2018). Planning chemical syntheses with deep neural networks. Nature, 555(7698), 604–610.
12. Chen, B., Butte, A. J., & Chen, Y. (2020). Artificial intelligence in drug discovery: Applications and techniques. Briefings in Bioinformatics, 21(3), 865–879.
13. Walters, W. P., & Murcko, M. (2020). Generative AI in medicinal chemistry. Nature Biotechnology, 38(2), 143–145.
14. Lo, Y. C., Rensi, S. E., Torng, W., & Altman, R. B. (2018). Machine learning in chemoinformatics and drug discovery. Drug Discovery Today, 23(8), 1538–1546.
15. Brown, N., Fiscato, M., Segler, M. H. S., & Vaucher, A. C. (2020). AI in medicinal chemistry. Future Medicinal Chemistry, 12(12), 1065–1079.

16. Gawehn, E., Hiss, J. A., & Schneider, G. (2016). Deep learning in drug discovery. Molecular Informatics, 35(1), 3–14.
17. Chen, Y., Li, L., & Zhang, Q. (2018). Drug-target interaction prediction using machine learning. Bioinformatics, 34(17), 289–296.
18. Zhou, H., Gao, M., & Skolnick, J. (2019). Artificial intelligence in precision medicine. Trends in Pharmacological Sciences, 40(3), 146–158.
19. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., … Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24– 29.
20. Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.
21. Schork, N. J. (2015). Personalized medicine: Time for one-person trials. Nature, 520(7549), 609–611.
22. Zhang, L., Tan, J., Han, D., & Zhu, H. (2021). From machine learning to deep learning in healthcare. Frontiers in Genetics, 12, 123–135.
23. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
24. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436– 444.
25. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., … Hassabis, D. (2017). Mastering the game of Go with deep neural networks. Nature, 529, 484–489.
26. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., … Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology. Journal of the Royal Society Interface, 15(141), 20170387.
27. Min, S., Lee, B., & Yoon, S. (2017). Deep learning in bioinformatics. Briefings in Bioinformatics, 18(5), 851–869.
28. Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. Molecular Systems Biology, 12(7), 878.
29. Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., … Hassabis, D. (2020). Improved protein structure prediction using deep learning. Nature, 577(7792), 706–710.
30. Zhavoronkov, A., Ivanenkov, Y. A., Aliper, A., Veselov, M. S., Aladinskiy, V. A., Aladinskaya, A. V., … Aspuru-Guzik, A. (2019). Deep learning enables rapid identification of kinase inhibitors. Nature Biotechnology, 37(9), 1038–1040.