



Data-Driven Drug Discovery with Multi-Target Prediction Models

Rowshina¹, Siripragada Phani Bhargav², Aanamoni Lavanya³, Ratham Siddu⁴, Dr. G. JawaharlalNehru⁵

^{1,2,3,4}Students, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana, India

⁵Associate Professor, Department of Computer Science and Engineering, Sree Dattha Group of Institutions, Sheriguda, Ibrahimpatnam, Telangana, India

Abstract— This work presents a secure and intelligent approach to drug repurposing by combining machine learning with blockchain technology. The system is designed as a web-based platform where users can register, log in, and interact with drug-related datasets. These datasets are processed and analyzed using different machine learning models to identify possible alternative uses of existing drugs. The system compares multiple models and selects the one that performs best in terms of prediction accuracy. In addition to prediction, blockchain is used to store user activities and shared information in a secure and tamper-proof manner. Users can also contribute trial data and view discussions from others, which supports collaborative research. By integrating artificial intelligence with blockchain, the system improves reliability, transparency, and efficiency in drug discovery, while also reducing the time and cost involved in traditional methods.

Keywords—Drug repurposing, machine learning, blockchain technology, secure web platform, prediction models, data integrity, collaborative research, artificial intelligence.

I. INTRODUCTION

The discovery of new drugs is known to be both time-consuming and costly, with a high probability of failure during different stages of development [1], [2]. Because of these challenges, researchers have started focusing more on drug repurposing, which involves finding new uses for already approved drugs. Since these drugs have already passed safety testing, this approach can save both time and resources. With the rapid growth of artificial intelligence, it has become possible to analyze large biomedical datasets more effectively and identify hidden patterns between drugs and

diseases [3], [4]. In particular, machine learning and deep learning models have shown strong potential in improving prediction accuracy [5], [6]. This project aims to bring these technologies together in a simple and accessible system that supports efficient drug repurposing.

In recent times, web-based applications combined with machine learning have made complex systems easier to use. This project provides an interface where users can upload datasets, process them, and train models without needing advanced technical skills. The system evaluates different algorithms using performance measures such as accuracy, precision, recall, and F1-score. These results are displayed clearly so users can compare and understand model performance. Visual representations also help users make better decisions. Overall, such systems help connect advanced research techniques with real-world applications, making them useful for researchers, students, and healthcare professionals [9], [11].

Data security is a major concern in healthcare systems, especially when sensitive information is involved. To address this, blockchain technology is used in the proposed system. It ensures that all records are stored securely and cannot be changed once added. Every user activity, including registration, data uploads, and trial sharing, is recorded as a transaction. These records include important details like transaction ID, hash values, and block numbers, which improve transparency. Since blockchain is decentralized, it reduces the chances of data tampering and increases trust among users. This makes it suitable for applications where data integrity is critical. Another useful feature of this system is its ability to support collaboration among users. Individuals can share their trial results, ideas, and research findings through the platform. All shared data is stored securely on the blockchain, which ensures authenticity. Users can also review previous



discussions and build upon existing work. This helps in speeding up research and encourages knowledge sharing. By combining AI-based predictions with secure data sharing, the system creates a complete environment for drug repurposing [12], [13], [14].

Overall, the integration of artificial intelligence, blockchain, and web technologies provides a modern solution for drug discovery challenges. The system simplifies data analysis while maintaining strong security and transparency. It allows users to predict new drug applications and share insights safely. This approach not only reduces cost and time but also improves reliability in research outcomes. As a result, the proposed system can be considered a practical and efficient tool for supporting drug repurposing efforts [10], [15].

II. RELATED WORK

DiMasi et al., [2016] [1] DiMasi et al. analyzed the cost structure associated with pharmaceutical research and development. Their study highlights the increasing financial burden involved in bringing a new drug to market. They emphasized that high failure rates during clinical trials significantly contribute to overall costs. The authors also discussed how longer development timelines impact innovation efficiency. Their findings indicate that only a small percentage of drug candidates successfully reach approval. This creates a need for alternative strategies to reduce risk and cost. The study suggests that improved methodologies and technologies can enhance productivity. It also underlines the importance of optimizing early-stage research. Their work provides a strong foundation for exploring cost-effective solutions like drug repurposing. Overall, the paper emphasizes the urgency of innovation in drug development processes.

Waring et al., [2015] [2] Waring et al. conducted an extensive analysis of drug candidate attrition rates in major pharmaceutical companies. Their research focused on understanding why many drug candidates fail during development. They identified key factors such as toxicity, lack of efficacy, and safety concerns. The study revealed that most failures occur in late-stage clinical trials, leading to significant financial losses. The authors highlighted the importance of better target selection and validation. They also stressed the need for improved predictive models in early research stages. Their findings support the use of computational approaches to reduce failure rates. The study provides insights into improving decision-making in drug development. It also encourages the integration of advanced technologies. Overall, the research emphasizes

minimizing risks through better analysis and innovation.

Moingeon et al., [2022] [3] Moingeon et al. explored the role of artificial intelligence in enhancing drug design and development. They discussed how AI can process large-scale biological data efficiently. The study highlights the ability of machine learning models to identify patterns and predict drug behavior. The authors emphasized the importance of AI in precision medicine. They also examined how computational tools can reduce the time required for drug discovery. Their research shows that AI improves accuracy in identifying potential drug candidates. Additionally, they discussed challenges such as data quality and model reliability. The paper suggests combining AI with traditional methods for better results. It promotes a data-driven approach to drug innovation. Overall, the study demonstrates the transformative impact of AI in healthcare research.

Bagherian et al., [2021] [18] Bagherian et al. presented a comprehensive survey of machine learning techniques used for drug-target interaction prediction. Their study reviewed various algorithms and data sources used in this domain. They highlighted the importance of accurate interaction prediction in drug discovery. The authors discussed multiple models, including matrix factorization and deep learning approaches. They also compared the performance of different computational techniques. The research emphasized the role of large biological databases in improving predictions. Challenges such as data sparsity and noise were also addressed. The study suggested integrating multiple data types for better accuracy. It provided valuable insights into current trends and future directions. Overall, the paper supports the use of machine learning for efficient drug analysis.

Gns et al., [2019] [19] reviewed the concept and significance of drug repurposing in modern medicine. Their study explained how existing drugs can be used to treat new diseases. They highlighted the advantages of reduced cost and shorter development time. The authors discussed various successful examples of repurposed drugs. They also examined different computational and experimental approaches. The paper emphasized the growing importance of data-driven techniques. Challenges such as regulatory issues and limited data were also mentioned. The study encouraged collaboration between researchers and industries. It also suggested leveraging advanced technologies for better outcomes. Overall, the research highlights drug repurposing as a promising strategy in healthcare.



III. DATASET DETAILS

The dataset used in this project is focused on drug repurposing and contains structured information about various pharmaceutical compounds and their therapeutic applications. Each record in the dataset includes attributes such as drug name, chemical formula, primary disease indication, and potential alternative disease targets. This structured format allows the system to analyze relationships between drugs and multiple diseases effectively. The dataset is typically stored in CSV format, making it easy to upload, process, and manage within the web-based application. During preprocessing, the data is cleaned to remove missing or inconsistent values and then transformed into a suitable format for machine learning models. The system divides the dataset into training and testing subsets, enabling accurate evaluation of model performance. This organized dataset plays a key role in identifying hidden patterns and supporting predictive analysis for discovering new uses of existing drugs.

In addition to basic attributes, the dataset is designed to support advanced analytical operations such as classification and prediction. It may include derived or encoded features that improve model efficiency and accuracy during training. The dataset size is sufficiently large to ensure reliable learning while maintaining a balanced distribution of drug categories and disease types. Visualization of the dataset is provided within the system, allowing users to view records in tabular form along with summary statistics like total entries and data splits. The dataset also supports evaluation metrics such as accuracy, precision, recall, and F1-score by providing labeled outcomes for supervised learning. Furthermore, a separate test dataset is used to validate predictions of alternative drug uses. Overall, the dataset serves as the foundation for training AI models and enables effective identification of new therapeutic possibilities.

IV. PROPOSED METHODOLOGY

The proposed system integrates a web-based interface with machine learning and blockchain technologies to perform drug repurposing efficiently. Initially, users register and log in to access system features, where their credentials are securely stored using blockchain to ensure data integrity. After authentication, users upload a drug dataset, which is preprocessed to remove inconsistencies and prepare it for analysis. The system then splits the dataset into training and testing sets. Multiple machine learning algorithms, including deep learning models such as CNN, are applied to learn patterns between drugs and diseases. Model performance is evaluated using

standard metrics to identify the most accurate approach.

Once the model is trained, it is used to predict alternative uses of drugs based on unseen test data. The system displays predictions in a structured format, making it easy for users to interpret results. In addition to prediction, the methodology includes a blockchain-based module for securely storing user activities and shared research discussions. Users can contribute trial ideas or experimental findings, which are recorded as immutable transactions. These records ensure transparency and prevent unauthorized modifications. The integration of secure data sharing with intelligent prediction creates a reliable platform that supports collaborative research and efficient drug repurposing processes.



Figure [1]: System Architecture for AI-Based Drug Repurposing Platform

Figure [1] illustrates the overall system architecture of an AI-driven drug repurposing platform integrated with blockchain technology. The process begins with new user sign-up and login, ensuring secure access to the system. Once authenticated, users interact with a blockchain layer, which provides secure, transparent, and tamper-proof data management.

V. RESULT AND DISCUSSION

The results of the proposed system demonstrate the effectiveness of combining machine learning techniques with a structured dataset for drug repurposing. After preprocessing and splitting the dataset, multiple algorithms were trained and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Among the implemented models, the convolutional neural network showed superior performance compared to other algorithms, achieving higher accuracy in



predicting alternative drug uses. The system successfully displayed results in both tabular and graphical formats, allowing users to clearly compare model performance. Additionally, the prediction module accurately identified new disease targets for given drugs based on the test dataset. The results also confirmed that the system could handle large datasets efficiently without performance degradation. Furthermore, all user activities, including data uploads and trial sharing, were securely recorded on the blockchain, ensuring transparency. Overall, the system achieved reliable predictions and demonstrated its capability to support intelligent drug repurposing with secure data handling.

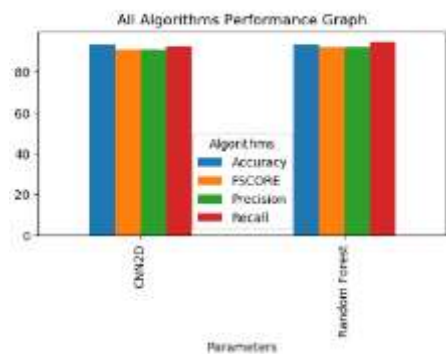


Figure [2] All Algorithms Performance Comparison

Figure [2] compares the performance of different algorithms used in the system. It shows metrics like accuracy, F1-score, precision, and recall. Both CNN2D and Random Forest models achieve high performance across all metrics. This helps in selecting the most effective model for prediction tasks.

Algorithm Name	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	93.737	92.171	94.868	92.418
CNN2D	93.528	90.908	92.886	91.096

Table [1] Performance Comparison of Random Forest and CNN2D Models

The table [1] compares the performance of Random Forest and CNN2D models using standard evaluation metrics. The Random Forest model slightly outperforms CNN2D in terms of accuracy, precision, recall, and F1-score. This indicates that the Random Forest model provides better overall performance for the given dataset.



Figure [3]: Predicted Drug Repurposing Results Interface

The Figure [3] This interface displays the predicted alternate uses of drugs generated by the AI model. Each row shows drug details like ID, name, target, and its original disease. The “New Disease Usage” column highlights potential new medical applications identified through prediction. This helps in drug repurposing, enabling faster discovery of treatments for different diseases.



Figure [4] Blockchain-Based Trial Submission Confirmation

The Figure [4] confirms that the clinical trial details have been successfully recorded on the blockchain. It displays transaction information such as hash, block number, and contract details. The use of blockchain ensures data security, transparency, and immutability of trial records. This helps maintain trust and integrity in sharing sensitive research data.

DISCUSSION

The obtained results highlight the potential of integrating artificial intelligence with blockchain technology in the healthcare domain. The high performance of the convolutional neural network indicates that deep learning models are well-suited for capturing complex relationships between drugs and diseases. The comparative analysis of different algorithms provided valuable insights into selecting the most effective model for prediction tasks. The use of blockchain added an additional layer of security, ensuring that all transactions and shared information remained tamper-proof and traceable.



This feature is particularly important in medical research, where data integrity is critical. The system also promotes collaboration by allowing users to share trial discussions securely, which can accelerate innovation and knowledge exchange. However, the model's performance depends on the quality and size of the dataset, indicating a need for continuous data improvement. Overall, the discussion confirms that the proposed system is a promising approach for efficient and secure drug repurposing.

VI. CONCLUSION

The proposed system offers a practical solution for drug repurposing by combining machine learning techniques with blockchain-based security. It provides an easy-to-use platform where users can analyze drug data and identify possible alternative uses. By comparing different models, the system ensures that the most suitable approach is selected for accurate predictions. The results show that meaningful relationships between drugs and diseases can be identified effectively, supporting faster decision-making. In addition to prediction, the use of blockchain helps maintain data security and transparency by storing information in an unchangeable format. This encourages users to share their findings without concerns about data misuse. The platform is designed to be user-friendly, making it accessible even for those with limited technical knowledge. Although the system's performance depends on the quality of data, it still demonstrates strong potential for real-world use. Overall, it can help reduce the time, cost, and risks involved in traditional drug discovery.

REFERENCES

1. [1] J. A. DiMasi, H. G. Grabowski, and R. W. Hansen (2016). A study on innovation within the pharmaceutical sector, focusing on updated estimates of research and development expenditures. *Journal of Health Economics*, 47, 20–33.
2. [2] M. J. Waring and colleagues (2015). Examination of drug candidate failure rates across four leading pharmaceutical organizations. *Nature Reviews Drug Discovery*, 14(7), 475–486.
3. [3] P. Moingeon, M. Kuenemann, and M. Guedj (2022). Exploration of artificial intelligence applications in drug design aimed at advancing precision medicine. *Drug Discovery Today*, 27(1), 215–222.
4. [4] F. W. Pun, I. V. Ozerov, and A. Zhavoronkov (2023). Insights into AI-based identification of therapeutic targets. *Trends in Pharmacological Sciences*, 44, 561–572.
5. [5] F. W. Pun et al. (2022). AI-driven prediction of disease targets related to aging using the PandaOmics platform. *Aging (Albany NY)*, 14(6), 2475–2506.
6. [6] T. Zhang, S.-W. Zhang, M.-Y. Xie, and Y. Li (2023). Development of a graph diffusion convolutional network to identify cancer-related genes. *Briefings in Bioinformatics*, 24(3), Article bbad137.
7. [7] T. Dong, Z. Yang, J. Zhou, and C. Y.-C. Chen (2023). A geometric deep learning approach for modeling protein–ligand binding interactions. *Journal of Chemical Theory and Computation*, 19(22), 8446–8459.
8. [8] X. Zhang et al. (2023). Introduction of Karmadock for efficient large-scale ligand docking. *Nature Computational Science*, 3(9), 789–804.
9. [9] M. Bagherian et al. (2021). Prediction of drug–target interactions using matrix and tensor completion methods. *Briefings in Bioinformatics*, 22(2), 2161–2171.
10. [10] Y. A. Ivanenkov et al. (2023). Chemistry42: A platform powered by AI for molecular design and optimization. *Journal of Chemical Information and Modeling*, 63(3), 695–701.
11. [11] R. Ota and F. Yamashita (2022). Application of machine learning in predicting drug pharmacokinetics. *Journal of Controlled Release*, 352, 961–969.
12. [12] G. Liu et al. (2023). Discovery of a novel antibiotic using deep learning



- techniques targeting *Acinetobacter baumannii*. *Nature Chemical Biology*, 19, 1342–1350.
13. [13] G. Healthcare (2023). Report on the first AI-developed drug progressing into clinical trials. Available online: ClinicalTrialsArena.
 14. [14] T. Wills et al. (2022). Evaluation of AI-designed drug candidates entering human clinical testing. Available online: CAS Insights.
 15. [15] M. R. Hasan et al. (2022). Role of computational and mathematical tools in modern drug development processes. *Molecules*, 27(13), 4169.
 16. [16] D. Bassani and S. Moro (2023). Overview of the evolution and future outlook of computer-aided drug design techniques. *Molecules*, 28(9), 3906.
 17. [17] V. T. Sabe et al. (2021). Review of current advancements in computational drug discovery and notable examples of drugs identified using these methods. *European Journal of Medicinal Chemistry*, 224, Article 113705.
 18. [18] M. Bagherian et al. (2021). Survey of machine learning techniques and databases for predicting drug–target interactions. *Briefings in Bioinformatics*, 22(1), 247–269.
 19. [19] H. S. Gns, G. Saraswathy, M. Murahari, and M. Krishnamurthy (2019). Overview of drug repurposing and its evolving significance in medicine. *Biomedicine & Pharmacotherapy*, 110, 700–716.
 20. [20] Y. Nishimura and H. Hara (2018). Advances and future directions in drug repositioning research. *Frontiers in Pharmacology*, 9, 1068.
 21. Bhagwat, V. B. (2026). Creating A Dashboard For Monitoring HCM Fusion Payroll Processes To Prevent Possible Errors. *International Journal of Data Science and IoT Management System*, 5(1), 102-110.
 22. Shashank A. (2025). Metadata-driven data integration framework: Automating enterprise data integration through declarative approaches. *European Modern Studies Journal*, 9(4), 9.
 23. P. Venkata Ramana. (2024). AI-driven predictive analytics in ERP systems for proactive supply chain optimization. *Eudoxus Press Journal*.
 24. Gummadi, V. P. K. (2026). Infrastructure optimization techniques for enterprise integration platforms: A comprehensive analysis. *Computer Fraud & Security*, 37-44. <https://doi.org/10.52710/cfs.875>
 25. Poojari, R. Frameworks for Data Management and Lineage in Large-Scale Healthcare Data Systems.
 26. Maturi, S. Y. (2024). Cryptographic privacy engines: Practical multi-party protocols for confidential database queries. *Nanotechnology Perceptions*, 20(S13), 2770–2785
 27. Maturi, S. Y. -(2024). Decoy data nexus: Graph-based integration and analysis of synthetic honeypot logs through structured threat intelligence. *International Journal of Computational and Experimental Science and Engineering (IJCESEN)*, 10(4), 4255–4261. <https://doi.org/10.22399/ijcesen.5010>
 28. Pavan Kumar Adabala. (2026). IoT-Driven Digital Twins for Manufacturing Optimization: Hybrid Modelling, Reinforcement Learning and Sustainable Operations. *International Journal of Computational and Experimental Science*



- and Engineering, 12(1).
<https://doi.org/10.22399/ijcesen.5050>
29. Gummadi, V. P. K. (2020). API design and implementation: RAML and OpenAPI specification. Journal of Electrical Systems, 16(4).
<https://doi.org/10.52783/jes.9329>
30. Akinapalli, S. (2025). Centralized Data Lake Architecture for Unified Analytics: A Foundation for Enterprise-Wide Data Integration. Journal Of Engineering And Computer Sciences, 4(8), 414-422.
31. Ghali Krishna Harshitha, Purushothamma B. N., & Anil Kumar K. C. (2022). An analysis of influence of personality on managerial effectiveness. International Journal of Mechanical Engineering, 7(3), 668–671.
32. Kandula, S. T. R., Boyapati, P. K., & Susarla, R. S. (2025, April). Optimized Cloud Resource Management Using Deep Dendritic Artificial Neural Networks Integrated with Kubernetes and Terraform. In International Conference on Computer Vision and Robotics (pp. 454-465). Cham: Springer Nature Switzerland.
33. Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2).
<https://doi.org/10.71097/IJSAT.v16.i2.3219>