

Exploring New Drug Uses through Hybrid Learning and Blockchain-Supported Data Validation

M. Ganesh¹, Gaddam Richitha², B Sai Jagadeesh Goud², Gannarapu Ramani², Gudepu Bharath²

¹Assistant Professor, ²UG Student, ^{1,2}Department of Computer Science and Engineering

^{1,2}Vaagdevi Engineering College, Bollikunta, Warangal, 506005, Telangana, India.

ABSTRACT

Drug repurposing has gained significant attention as an efficient strategy for identifying new therapeutic applications of existing drugs, thereby reducing both development time and cost compared to traditional drug discovery processes. Current drug discovery approaches rely on experimental procedures, expert analysis, and extensive clinical trials, which are time-intensive and computationally inefficient when handling large-scale biomedical data. These methods often struggle to process complex and high-dimensional datasets, resulting in slower analysis and limited predictive capability. Additionally, these systems lack robust mechanisms for secure data management, making clinical records and trial discussions susceptible to inconsistencies and unauthorized modifications. To overcome these limitations, this work proposes an intelligent drug repurposing framework that integrates Machine Learning (ML), Deep Learning (DL), and blockchain technologies. The system utilizes baseline models such as K-Nearest Neighbors (KNN) and Gaussian Naive Bayes (GNB) for comparative analysis, along with a hybrid DrugNet model that combines Convolutional Neural Networks (CNN) for feature extraction and Random Forest (RF) for classification. This hybrid approach enhances the ability to capture complex patterns in drug-related data and improves prediction accuracy. Furthermore, blockchain integration using Web3 ensures secure storage of user data, clinical interactions, and trial information, providing transparency, immutability, and data integrity. The proposed framework enables automated prediction of potential drug-disease associations through a unified processing pipeline, supporting real-time analysis and decision-making. By combining advanced Artificial Intelligence (AI) techniques with decentralized data management, the system improves scalability, reliability, and efficiency in drug repurposing. This approach offers a practical and secure solution for accelerating pharmaceutical research and supporting data-driven medical innovation.

Key words: Drug repurposing, machine learning, deep learning, blockchain, convolutional neural networks, drug-disease prediction.

1.INTRODUCTION

The process of discovering and developing new drugs is inherently complex, time-consuming, and financially demanding. Bringing a single compound from initial identification to final market approval often requires a prolonged period, typically ranging between 10 and 15 years. This extended duration reflects the multiple stages of validation, regulatory approvals, and clinical evaluations required to ensure safety and effectiveness. In addition to time constraints, the financial burden associated with this process is substantial, frequently exceeding USD 1 billion due to extensive research activities, clinical trials, and compliance requirements [1]. These challenges highlight the limitations of traditional drug development approaches. Beyond economic and temporal constraints, the pipeline is further complicated by high failure rates during clinical trials. As shown as figure 1 many candidate drugs fail due to safety concerns or insufficient therapeutic effectiveness, leading to significant resource loss and delays.

Although advancements in optimizing physicochemical properties have improved certain aspects of drug design, achieving consistent safety and efficacy remains a major challenge [2]. This issue becomes even more critical when addressing complex diseases such as cancer. Cancer exhibits significant genetic and molecular heterogeneity, which complicates treatment development.

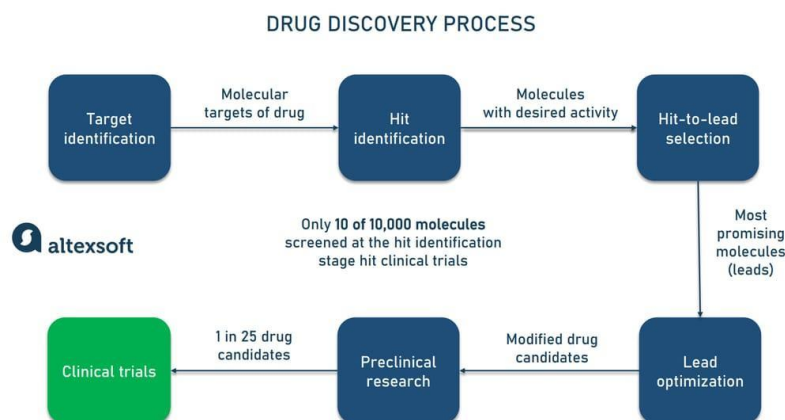


Figure 1: AI in Drug Discovery and Repurposing.

Variations in genetic profiles and signaling pathways across patients increase the difficulty of designing universally effective therapies [3]. Given these challenges, there is an urgent need to explore innovative strategies that can improve the efficiency of drug discovery. This need is particularly important in the case of cancer due to its high global prevalence and biological complexity. The integration of advanced computational approaches can support better analysis of complex biomedical data and enable more targeted therapy development [4].

These methods offer the potential to reduce development time while improving prediction accuracy. Artificial intelligence has emerged as a transformative solution in this domain. By leveraging techniques such as machine learning, deep learning, and sequence-based neural models, AI systems can analyze large-scale biomedical datasets and identify complex patterns [5]. These capabilities allow accurate prediction of drug–target interactions and optimization of molecular properties. As a result, AI-driven approaches significantly reduce both time and cost associated with drug development, while improving overall efficiency and precision.

2.LITERATURE SURVEY

Herraiz-Gil et al. [6] explored the application of AI-based methods for drug repurposing and natural product-inspired drug design in cancer, focusing on their potential to address the challenges and limitations of traditional drug discovery approaches. they delved into various AI-based approaches (machine learning, deep learning, and others) that are currently being employed for these purposes, and the role of experimental techniques in these approaches. By systematically reviewing the literature, they aimed to provide a comprehensive overview of the current state of AI-assisted cancer drug discovery workflows, highlighting AI’s contributions to accelerating drug development, reducing costs, and improving therapeutic outcomes. They also discussed the challenges and opportunities associated with the integration of AI into the drug discovery pipeline, such as data quality, interpretability, and ethical considerations. Mak et al. [7] presented the pharmaceutical industry is facing challenges in sustaining their drug development programmes because of increased R&D costs and reduced efficiency, they discussed the major causes of attrition rates in new drug approvals, the possible ways that AI can

improve the efficiency of the drug development process and collaboration of pharmaceutical industry giants with AI-powered drug discovery firms.

Malheiro et al. [8] developed, allowing machines to learn from large amounts of data and improve their performance over time. The pharmaceutical industry has greatly benefited from the development of this technology. AI has revolutionized drug discovery and development by enabling rapid and effective analysis of vast volumes of biological and chemical data during the identification of new therapeutic compounds. The algorithms developed can predict the efficacy, toxicity, and possible adverse effects of new drugs, optimize the steps involved in clinical trials, reduce associated time and costs, and facilitate the implementation of innovative drugs in the market, making it easier to develop precise therapies tailored to the individual genetic profile of patients. Despite significant advancements, there are still gaps in the application of AI, particularly due to the lack of comprehensive regulation. Wang et al. [9] proposed a novel NMF-based drug repositioning model that incorporates biological context (NMFIBC), which integrated drug and disease similarity networks through graph-regularized optimization to enhance predictive performance. This design enhanced both the robustness and interpretability of association prediction. Extensive benchmarking on multiple gold-standard datasets demonstrates that NMFIBC outperforms existing methods across a range of metrics, including AUC, precision, and F1-score. Moreover, case studies involving clinically relevant drugs validate the biological plausibility of the predicted associations using public databases such as DrugBank, CTD, and KEGG. They proposed framework provides a powerful, context-aware AI strategy for discovering actionable insights in drug repositioning.

Serrano et al. [10] transformed the pharmaceutical industry, revolutionizing everything from drug development and discovery to personalized medicine, including target identification and validation, selection of excipients, prediction of the synthetic route, supply chain optimization, monitoring during continuous manufacturing processes, or predictive maintenance, among others. While the integration of AI promises to enhance efficiency, reduce costs, and improve both medicines and patient health, it also raises important questions from a regulatory point of view. In this review article, they will present a comprehensive overview of AI's applications in the pharmaceutical industry, covering areas such as drug discovery, target optimization, personalized medicine, drug safety, and more. Dermawan et al. [11] categorized based on AI methods, clinical phase, and therapeutic area. The percentages of AI methods used, clinical phase stages, and the therapeutic regions were analyzed to identify trends. AI methods included machine learning (ML) at 40.9%, molecular modeling and simulation (MMS) at 20.7%, and deep learning (DL) at 10.3%. Oncology accounted for the majority of studies (72.8%), followed by dermatology (5.8%) and neurology (5.2%). In clinical phases, 39.3% of studies were in the preclinical stage, 23.1% in Clinical Phase I, and 11.0% in the transitional phase. Clinical outcome reporting was observed in 45% of studies, with 97% reporting industry partnerships.

Vora et al. [12] enabled the prioritization and optimization of lead compounds, reducing the need for extensive and costly animal testing. Personalized medicine approaches can be facilitated through AI algorithms that analyze real-world patient data, leading to more effective treatment outcomes and improved patient adherence. Baker et al. [13] Proposed repurposing existing drugs leverages established safety data and reduces development time and costs. Second, AI can broaden the search for effective therapies by identifying unexpected connections between drugs and potential new targets.

Garg et al. [14] enhanced computer-aided drug design (CADD), offering unprecedented precision in identifying potential anticancer compounds. Traditionally, cancer drug design has been a complex, resource-intensive process, but AI introduces new opportunities to accelerate discovery, reduce costs,

and optimize efficiency. This manuscript delved into the transformative applications of AI-driven methodologies in predicting and developing anticancer drugs, critically evaluating their potential to reshape the future of cancer therapeutics while addressing their challenges and limitations. Abbas et al. [15] evaluated AI enhancing various stages of drug development, including target identification, lead optimization, de novo drug design, and drug repurposing. AI tools, such as AlphaFold for protein structure prediction and AtomNet for structure-based drug design, have significantly accelerated the discovery process, improved efficiency and reduced costs. Success stories like Insilico Medicine's AI-designed molecule for idiopathic pulmonary fibrosis and BenevolentAI's identification of baricitinib for COVID-19 highlight AI's transformative potential.

3. PROPOSED SYSTEM

The system architecture is designed as an integrated framework that combines data processing, machine learning, deep learning, and blockchain technologies to enable efficient drug repurposing. It begins with data acquisition, where raw datasets containing drug attributes and disease labels are collected and initialized for processing. The data is then passed through a preprocessing pipeline that performs encoding, normalization, and cleaning to ensure consistency and quality. Following this, the dataset is partitioned into training and testing subsets to support reliable model development and validation. The architecture incorporates multiple learning models, including KNN, GNB, and the proposed hybrid DrugNet, to perform comparative analysis and improve prediction accuracy. The CNN component within DrugNet extracts deep feature representations, which are further processed by RF for robust classification. As illustrated in Figure 2, the workflow establishes a seamless connection between data preparation, model training, and evaluation modules. Performance metrics such as accuracy, precision, recall, and F1-score are computed to assess model effectiveness. In addition to predictive capabilities, the system integrates a blockchain layer to securely manage user information and clinical trial discussions. This ensures data immutability, transparency, and protection against unauthorized modifications. The architecture also includes user interface modules that allow dataset uploading, prediction execution, and result visualization. These interfaces provide an interactive environment for users to engage with the system efficiently.

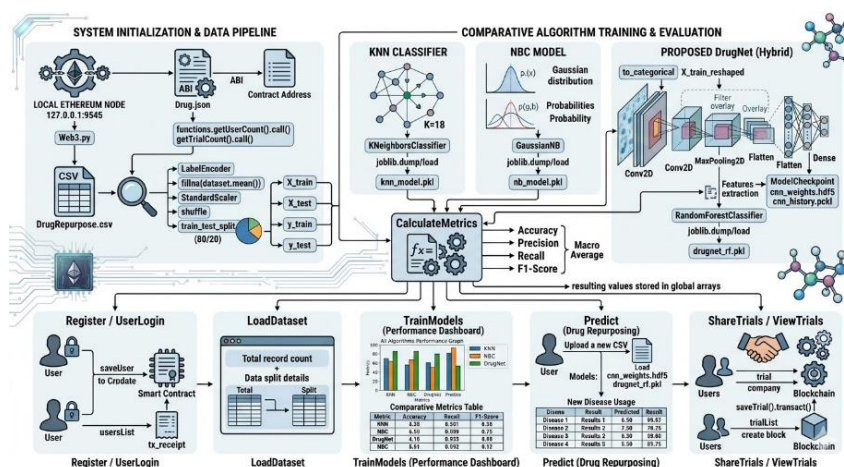


Figure 2: Proposed System Architecture

1. Data Acquisition and Initialization: The system begins by loading the dataset containing drug-related attributes and disease labels. It establishes connections with the blockchain network to retrieve user and trial information. Initial setup includes configuring system parameters and preparing the

environment for further processing. This step ensures that both data and infrastructure are ready for execution.

2. Data Preprocessing and Transformation: The raw dataset undergoes preprocessing where categorical features are encoded, missing values are handled, and feature scaling is applied. The data is then shuffled and split into training and testing subsets. These transformations ensure that the dataset is clean, normalized, and suitable for model training. It improves the efficiency and accuracy of subsequent algorithms.

3. Model Training and Feature Learning: Multiple models such as KNN, GNB, and the hybrid DrugNet are trained using the prepared dataset. The CNN component in DrugNet extracts deep features, which are further processed by RF for classification. Each model learns patterns from the data to perform predictions. This step forms the core intelligence of the system.

4. Performance Evaluation and Comparison: The trained models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score. These metrics are calculated to compare the effectiveness of different algorithms. The results are stored and visualized to identify the best-performing model. This ensures objective assessment of model performance.

5. Prediction and User Interaction: The system provides an interface for users to upload new data and obtain predictions for drug repurposing. The selected model processes the input and generates results, which are displayed in a structured format. This module enables real-time interaction and practical usage of the system. It bridges the gap between backend processing and end users.

6. Blockchain Integration and Data Management: The blockchain layer is used to securely store user details and clinical trial discussions. Transactions are recorded immutably, ensuring transparency and data integrity. Users can share and view trial information through the system interface. This component enhances trust and security within the overall architecture.

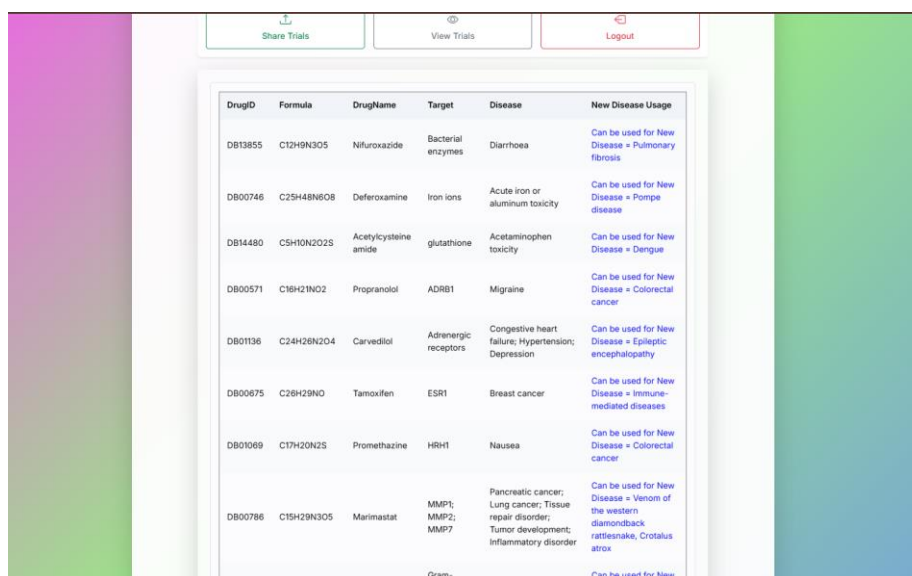
4. RESULTS ANALYSIS

The results demonstrate the effective functioning of the proposed drug repurposing system by integrating machine learning, deep learning, and blockchain technologies. The system provides an interactive interface where users can upload data, perform predictions, and visualize outcomes efficiently. Comparative analysis shows that the hybrid DrugNet model outperforms traditional models like KNN and GNB in terms of accuracy and reliability. The CNN component successfully extracts deep features, while Random Forest enhances classification performance. Additionally, blockchain integration ensures secure, transparent, and tamper-proof data management.



Figure 3: Training AI Models

Figure 3 illustrates the Training AI Models phase, where multiple algorithms are evaluated using quantitative performance metrics to determine their effectiveness in drug repurposing prediction. The figure depicts that the KNN Classifier achieves an accuracy of 75.365, with precision 72.378, recall 77.771, and F-score 71.473. It further shows that the NBC Model attains an improved accuracy of 85.386, along with precision 85.233, recall 87.805, and F-score 85.272, indicating better classification capability. The proposed DrugNet model demonstrates the highest performance with an accuracy of 93.111, precision 91.859, recall 93.768, and F-score 91.577. The comparison clearly represents that the proposed approach outperforms conventional models across all evaluation metrics.



DrugID	Formula	DrugName	Target	Disease	New Disease Usage
DB13855	C12H9N3O5	Nifuroxazide	Bacterial enzymes	Diarrhoea	Can be used for New Disease = Pulmonary fibrosis
DB00746	C25H48N6O8	Deferoxamine	Iron ions	Acute iron or aluminum toxicity	Can be used for New Disease = Pompe disease
DB14480	C5H10N2O2S	Acetylcysteine amide	glutathione	Acetaminophen toxicity	Can be used for New Disease = Dengue
DB00571	C16H21NO2	Propranolol	ADRB1	Migraine	Can be used for New Disease = Colorectal cancer
DB01136	C24H26N2O4	Carvedilol	Adrenergic receptors	Congestive heart failure; Hypertension; Depression	Can be used for New Disease = Epileptic encephalopathy
DB00675	C26H29NO	Tamoxifen	ESR1	Breast cancer	Can be used for New Disease = Immune-mediated diseases
DB01069	C17H20N2S	Promethazine	HRH1	Nausea	Can be used for New Disease = Colorectal cancer
DB00786	C15H29N3O5	Marimastat	MMP1; MMP2; MMP7	Pancreatic cancer; Lung cancer; Tissue repair disorder; Tumor development; Inflammatory disorder	Can be used for New Disease = Venom of the western diamondback rattlesnake, Crotalus atrox
			Gram-		Can be used for New Disease =

Figure 4: Test Data Prediction with New Disease Usage

Figure 4 illustrates the Test Data Prediction with New Disease Usage, where the system generates predictions for potential new applications of existing drugs. It depicts how trained models analyze test data to identify possible disease associations and repurposing opportunities. The figure highlights the

output generation process that links drugs with new therapeutic uses based on learned patterns. It represents the final stage of the workflow where actionable insights are produced.

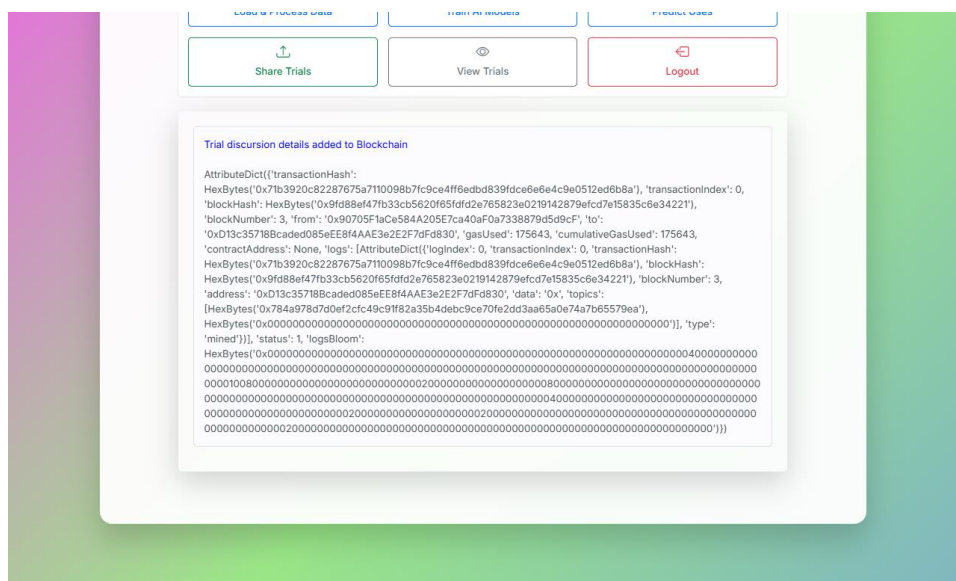


Figure 5: New Trail adding to the Blockchain

Figure 5 depicts the New Trail adding to the Blockchain, where the submitted trial data is securely recorded as a blockchain transaction. It illustrates the generation of transaction hash, block number, gas usage, and other cryptographic details ensuring immutability. The figure emphasizes how blockchain technology maintains transparency, traceability, and tamper-proof storage of trial records. It represents the successful conversion of user-submitted data into a permanent distributed ledger entry.

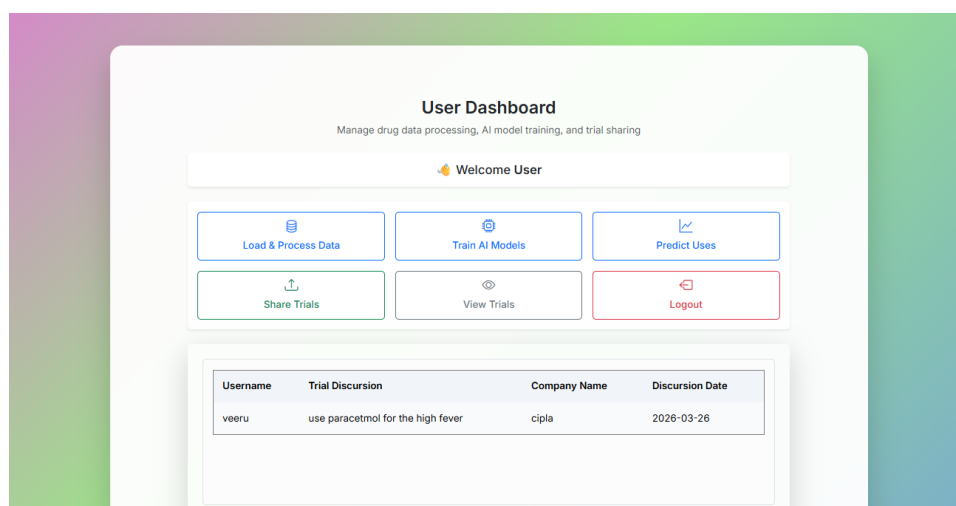


Figure 6: viewing New Trails added

Figure 6 illustrates the viewing New Trails added, where stored trial records can be retrieved and displayed for user verification and analysis. It depicts the structured presentation of key details such as username, trial discussion, company name, and date of submission. The figure highlights the accessibility of blockchain-stored data in a readable and organized format. It represents the retrieval mechanism that connects decentralized storage with user-level visibility.

5. CONCLUSION

The developed DrugNet-based prediction system successfully integrates machine learning, deep learning, and blockchain technologies to provide an efficient and secure prediction platform. The system combines baseline models such as KNN and GNB with the proposed DrugNet model, which utilizes CNN for feature extraction and Random Forest for classification. This hybrid approach improves the ability to capture complex data patterns and enhances prediction accuracy compared to individual models. The preprocessing pipeline ensures high-quality input data, which further contributes to improved model performance. The evaluation results demonstrate that the DrugNet model achieves better accuracy, precision, recall, and F1-score than traditional approaches. The integration of Flask enables real-time user interaction, making the system practical and user-friendly. Additionally, blockchain integration ensures secure and immutable storage of user and trial data, enhancing system reliability.

REFERENCES

- [1] Hughes, J.P.; Rees, S.; Kalindjian, S.B.; Philpott, K.L. Principles of early drug discovery. *Br. J. Pharmacol.* 2011, 162, 1239–1249.
- [2] Waring, M.J.; Arrowsmith, J.; Leach, A.R.; Leeson, P.D.; Mandrell, S.; Owen, R.M.; Pairaudeau, G.; Pennie, W.D.; Pickett, S.D.; Wang, J.; et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discov.* 2015, 14, 475–486.
- [3] Vasan, N.; Baselga, J.; Hyman, D.M. A view on drug resistance in cancer. *Nature* 2019, 575, 299–309.
- [4] Garg, P.; Singhal, G.; Kulkarni, P.; Horne, D.; Salgia, R.; Singhal, S.S. Artificial Intelligence-Driven Computational Approaches in the Development of Anticancer Drugs. *Cancers* 2024, 16, 3884.
- [5] Wang, L.; Song, Y.; Wang, H.; Zhang, X.; Wang, M.; He, J.; Li, S.; Zhang, L.; Li, K.; Cao, L. Advances of Artificial Intelligence in Anti-Cancer Drug Design: A Review of the Past Decade. *Pharmaceutics* 2023, 16, 253.
- [6] Herráiz-Gil S, Nygren-Jiménez E, Acosta-Alonso DN, León C, Guerrero-Aspizua S. Artificial Intelligence-Based Methods for Drug Repurposing and Development in Cancer. *Applied Sciences*. 2025; 15(5):2798. <https://doi.org/10.3390/app15052798>
- [7] Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. *Drug Discov Today*. 2019 Mar;24(3):773-780. doi: [10.1016/j.drudis.2018.11.014](https://doi.org/10.1016/j.drudis.2018.11.014). Epub 2018 Nov 22. PMID: 30472429
- [8] Malheiro V, Santos B, Figueiras A, Mascarenhas-Melo F. The Potential of Artificial Intelligence in Pharmaceutical Innovation: From Drug Discovery to Clinical Trials. *Pharmaceutics*. 2025; 18(6):788. <https://doi.org/10.3390/ph18060788>
- [9] Wang Y, Wang Y, Hu Y, Wang J. Biology-Informed Matrix Factorization: An AI-Driven Framework for Enhanced Drug Repositioning. *Biology*. 2025; 14(5):549. <https://doi.org/10.3390/biology14050549>
- [10] Serrano DR, Luciano FC, Anaya BJ, Ongoren B, Kara A, Molina G, Ramirez BI, Sánchez-Guirales SA, Simon JA, Tomietto G, et al. Artificial Intelligence (AI) Applications in Drug Discovery and Drug Delivery: Revolutionizing Personalized Medicine. *Pharmaceutics*. 2024; 16(10):1328. <https://doi.org/10.3390/pharmaceutics16101328>

-
- [11] Dermawan D, Alotaiq N. From Lab to Clinic: How Artificial Intelligence (AI) Is Reshaping Drug Discovery Timelines and Industry Outcomes. *Pharmaceuticals*. 2025; 18(7):981. <https://doi.org/10.3390/ph18070981>
- [12] Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial Intelligence in Pharmaceutical Technology and Drug Delivery Design. *Pharmaceutics*. 2023; 15(7):1916. <https://doi.org/10.3390/pharmaceutics15071916>
- [13] Baker, R.E., Mahmud, A.S., Miller, I.F. et al. Infectious disease in an era of global change. *Nat Rev Microbiol* 20, 193–205 (2022). <https://doi.org/10.1038/s41579-021-00639-z>
- [14] Garg P, Singhal G, Kulkarni P, Horne D, Salgia R, Singhal SS. Artificial Intelligence–Driven Computational Approaches in the Development of Anticancer Drugs. *Cancers*. 2024; 16(22):3884. <https://doi.org/10.3390/cancers16223884>
- [15] M. K. G. Abbas, Abrar Rassam, Fatima Karamshahi, Rehab Abunora, Maha Abouseada, First published: 12 May 2024, <https://doi.org/10.1002/cbic.202300816>