A Blockchain–AI Fusion Framework for Secure and Predictive Drug Repurposing Using Hybrid Deep Ensemble Learning Model

Vakkala Akhila[1], Karramareddy Sharmila[2], M. Swetha[3*]

[1]PG Student, [2]Associate Professor and Head, [3]Assistant Professor, [1,2,3]Department of Computer Science and Engineering
[1,2,3]Vaagdevi Engineering College, Warangal, 506005, Telangana, India.
[1]akhilavakkala2001@gmail.com, [3]swetha.m@vecw.edu.in
*Correspondence: M. Swetha (swetha.m@vecw.edu.in)

ABSTRACT

Drug repurposing has gained significant attention as a promising strategy for identifying new therapeutic applications of existing drugs, thereby reducing the time, cost, and risks associated with conventional drug discovery and development processes. The rapid growth of biomedical, clinical, and pharmaceutical datasets has created unprecedented opportunities for computational techniques to accelerate the identification of novel drug–disease associations. To overcome these limitations, this research proposes a Probabilistic Drug Repurposing Network (ProGradNet), an intelligent and secure framework that integrates Machine Learning (ML), Deep Learning (DL), Probabilistic Neural Networks (PNN), Extreme Gradient Boosting (XGBoost), the Django web framework, and InterPlanetary File System (IPFS)-based blockchain technology through Pinata. The proposed framework is designed to automate the prediction of potential drug–disease associations while ensuring secure and transparent management of research-related information. Initially, biomedical data undergo comprehensive preprocessing and feature extraction to improve data quality and predictive performance. Subsequently, the proposed ProGradNet model combines the probabilistic learning capability of PNN with the powerful classification and optimization characteristics of XGBoost to enhance prediction accuracy and generalization performance. For comparative evaluation, established machine learning and deep learning models, including K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Convolutional Neural Network–Random Forest (CNN-RF), are implemented as benchmark approaches. In addition, the Django framework provides a comprehensive web-based environment that supports user registration, authentication, dataset management, predictive analytics, and clinical trial administration. To strengthen data security and transparency, user registration records and clinical trial discussion data are securely stored using Pinata-integrated IPFS blockchain technology, ensuring immutability, traceability, decentralized storage, and data integrity. Experimental results demonstrate that the proposed ProGradNet framework effectively automates drug repurposing prediction, improves drug–disease association identification accuracy, enhances system scalability, and provides robust security for sensitive biomedical information.

Key words: Drug Repurposing, Machine Learning, Deep Learning, Probabilistic Neural Networks, Drug–Disease Association Prediction, Blockchain, Biomedical Informatics

1. INTRODUCTION

Drug discovery and development represent one of the most complex, time-intensive, and resource-demanding processes in the pharmaceutical and healthcare industries. The journey from identifying a potential therapeutic compound to obtaining regulatory approval and commercial deployment typically spans 10 to 15 years. This lengthy process encompasses multiple stages, including target identification, lead optimization, preclinical studies, clinical

Doi:10.48047/ijiee.2026.16.2.51

trials, and regulatory evaluations, all of which are necessary to ensure the safety, efficacy, and quality of the developed drug. In addition to the extended development timeline, the financial investment required for successful drug development is substantial, often exceeding USD 1 billion due to extensive research activities, large-scale clinical studies, regulatory compliance requirements, and manufacturing validation procedures [1]. These challenges underscore the limitations of conventional drug discovery methodologies and highlight the need for more efficient and cost-effective approaches.

Furthermore, the drug development pipeline is characterized by high failure rates, particularly during clinical trial phases. A significant proportion of candidate compounds fail because of adverse safety profiles, inadequate therapeutic efficacy, or unforeseen biological interactions, resulting in considerable financial losses and delays in delivering effective treatments to patients. Although advancements in computational chemistry, molecular modeling, and physicochemical optimization have improved certain aspects of drug design, achieving consistent safety, efficacy, and clinical success remains a major challenge [2].

The complexity of drug discovery becomes even more pronounced when addressing diseases such as cancer, which exhibit substantial genetic, molecular, and phenotypic heterogeneity. Variations in genetic mutations, cellular signaling pathways [4], and disease progression mechanisms among patients complicate the development of universally effective therapeutic strategies. Consequently, identifying suitable drug candidates and predicting their effectiveness require the analysis of large-scale and highly complex biomedical datasets [5]. These challenges emphasize the importance of innovative computational approaches that can accelerate therapeutic discovery, improve prediction accuracy, and support personalized treatment development

.

Figure 1. Global Drug Repurposing Market Growth by Approach Type (2023–2033)

Figure1 illustrates the projected growth of the global drug repurposing market from 2023 to 2033, segmented into disease-centric, target-centric, and drug-centric approaches. The market demonstrates a steady upward trend, increasing from USD 32.1 billion in 2023 to USD 51.8 billion by 2033, representing a compound annual growth rate (CAGR) of 4.5%. Among the three approaches, disease-centric repurposing consistently holds the largest market share throughout the forecast period, followed by target-centric and drug-centric strategies. The continuous expansion of all segments reflects the growing adoption of computational biology, artificial intelligence, genomics, and bioinformatics techniques to

Doi:10.48047/ijiee.2026.16.2.51

identify new therapeutic uses for existing drugs. The increasing prevalence of chronic diseases, demand for cost-effective drug development, and need for faster therapeutic discovery further contribute to market growth. The forecast indicates that drug repurposing will remain a key strategy in pharmaceutical research, enabling reduced development costs, shorter clinical timelines, and accelerated delivery of effective treatments for various diseases.

2. LITERATURE SURVEY

Dermawan et al. [6] proposed a comprehensive AI-driven drug discovery framework that examined the role of machine learning and deep learning across the pharmaceutical development pipeline. The methodology evaluated AI applications in target identification, hit discovery, lead optimization, and clinical trial design. Advanced predictive models were utilized to analyze biological and chemical datasets for identifying promising drug candidates. The study further assessed the impact of AI on reducing drug development timelines and improving success rates. A systematic review approach was adopted to analyze recent advancements and industrial outcomes of AI-assisted drug discovery. The framework relies heavily on large volumes of high-quality biological data, which are often difficult to obtain. Ribeiro et al. [7] developed a drug repurposing strategy focusing on Terbutaline and Milrinone for cancer therapy. The methodology reviewed pharmacological mechanisms, molecular targets, and preclinical evidence associated with both drugs. Existing experimental studies were systematically analyzed to assess anticancer potential. Mechanistic pathways related to tumor growth inhibition and apoptosis induction were investigated. The framework highlighted opportunities for repositioning approved drugs as cost-effective cancer treatments. Most findings are based on preclinical evidence and require extensive clinical validation. Pandey et al. [8] introduced a comprehensive analysis of artificial intelligence applications in drug discovery and pharmaceutical product design. The methodology explored machine learning algorithms, deep neural networks, and predictive analytics for molecular screening and optimization. Various AI-assisted stages including target identification, toxicity prediction, and formulation design were examined. Challenges related to data quality, interpretability, and regulatory compliance were discussed. The study also identified future opportunities for AI-enabled pharmaceutical innovation. The proposed review does not provide experimental benchmarking of the discussed AI models.

Kant et al. [9] proposed an AI-based framework for transforming drug discovery and development processes. The methodology employed machine learning techniques to accelerate target validation, compound screening, and lead optimization. Predictive models were utilized to evaluate drug efficacy and safety profiles. AI-assisted analytics supported decision-making throughout different stages of pharmaceutical development. The framework emphasized improving efficiency while reducing research costs and development timelines. Model predictions may be affected by incomplete or biased training datasets. Ali et al. [10] developed a deep learning-driven approach for modern drug discovery applications. The methodology integrated neural networks, molecular modeling, and data mining techniques to identify potential therapeutic compounds. Drug-target interactions were predicted using advanced learning algorithms. The framework also examined challenges related to explainability, scalability, and clinical translation. Future perspectives of AI-enhanced pharmaceutical research were extensively discussed. Deep learning models often function as black-box systems with limited interpretability. Kokudeva et al. [11] presented a comprehensive drug repurposing methodology for COVID-19 treatment discovery. The approach systematically reviewed existing therapeutic candidates and evaluated their antiviral potential against SARS-CoV-2. Pharmacological mechanisms

Doi:10.48047/ijiee.2026.16.2.51

and clinical evidence were analyzed to identify promising repurposed drugs. Data from preclinical and clinical studies were integrated to assess efficacy and safety. The framework highlighted rapid therapeutic development through repositioning strategies. Drug effectiveness varied considerably across different clinical studies and patient populations.

More et al. [12] introduced TheraMind, a multi-LLM ensemble framework for accelerating drug repurposing in lung cancer. The methodology employed multiple large language models to mine clinical case reports and biomedical literature. Information extraction techniques identified potential drug-disease associations. Ensemble decision mechanisms improved the reliability of candidate selection. The framework facilitated efficient knowledge discovery for cancer drug repurposing. Performance depends strongly on the quality and completeness of available medical literature. Chen et al. [13] developed DRQuantum, a quantum walk-based drug repurposing method operating on a multi-layer heterogeneous network. The methodology integrated biological entities such as drugs, diseases, proteins, and genes into a unified network representation. Quantum walk algorithms explored complex associations among network components. Candidate drugs were ranked based on their predicted therapeutic relevance. Experimental evaluations demonstrated improved repurposing accuracy compared to conventional methods. The computational complexity increases significantly with large-scale biological networks. Prakash [14] proposed a network-driven computational framework for identifying FDA-approved drug repurposing opportunities across heterogeneous brain cancers. The methodology combined disease-gene associations, protein interactions, and drug-target networks. Network analysis techniques identified hidden therapeutic relationships among different cancer subtypes. Candidate drugs were prioritized based on connectivity and biological relevance. The framework supported cost-effective treatment discovery for complex neurological malignancies. Biological network incompleteness may affect the reliability of identified drug candidates. Poderoso et al. [15] developed a transcriptomic analysis-based drug repurposing framework for neurotropic flavivirus infections. The methodology analyzed gene expression signatures associated with viral infection pathways. Differentially expressed genes were matched with drug-induced transcriptomic profiles. Candidate therapeutics were identified based on their ability to reverse disease-related molecular alterations. Experimental validation supported the effectiveness of selected compounds. Transcriptomic signatures alone may not fully capture complex disease mechanisms.

3. PROPOSED SYSTEM

The system architecture is designed as an integrated framework that combines data preprocessing, ML, DL, PNN, XGBoost, Django, and Pinata-based IPFS to enable efficient and secure drug repurposing. It begins with data acquisition, where raw datasets containing drug attributes and disease labels are collected and initialized for processing. The data then undergoes preprocessing, including encoding, scaling, and missing value handling, to ensure consistency and quality before being divided into training and testing datasets. The architecture incorporates multiple prediction models, including KNN, NBC, CNN-RF, and the proposed ProGradNet, to perform comparative analysis and improve prediction accuracy. Within ProGradNet, PNN performs pattern learning, while XGBoost enhances the final prediction through efficient classification. As illustrated in Figure 2, the workflow establishes a seamless connection between data preprocessing, model training, performance evaluation, secure data management, and prediction. Performance metrics such as accuracy, precision, recall, and F1-score are computed to evaluate and compare the effectiveness of all models. The architecture also integrates Pinata-based IPFS to securely maintain user registration details and clinical

Doi:10.48047/ijiee.2026.16.2.51

trial discussions, ensuring transparency, immutability, and data integrity. Furthermore, Django provides a web-based interface that enables user registration, authentication, dataset upload, prediction execution, trial management, and result visualization, allowing users to interact with the system efficiently.

Data Acquisition and Initialization: The system begins by loading the drug repurposing dataset containing drug-related attributes and disease labels. It initializes the required preprocessing components, trained models, and Pinata-based IPFS services for secure data management. The execution environment is configured to support model training and prediction.

Data Preprocessing and Transformation: The dataset undergoes preprocessing by performing label encoding, missing value handling, feature scaling, data shuffling, and train-test splitting. These operations convert raw data into a structured format suitable for model training and improve prediction performance.

.

Model Training and Prediction: The preprocessed dataset is used to train KNN, NBC, CNN-RF, and the proposed ProGradNet. PNN learns the underlying data patterns, while XGBoost performs efficient classification to improve drug repurposing prediction. Comparative analysis is carried out to identify the most effective model.

Performance Evaluation: All trained models are evaluated using accuracy, precision, recall, and F1-score. The calculated metrics are compared and visualized to assess the prediction capability of each model and determine the best-performing approach.

Prediction and User Interaction: Django provides an interactive web interface that allows users to register, log in, upload datasets, perform drug repurposing prediction, share clinical trial discussions, and visualize prediction results. This module enables real-time interaction between users and the prediction system

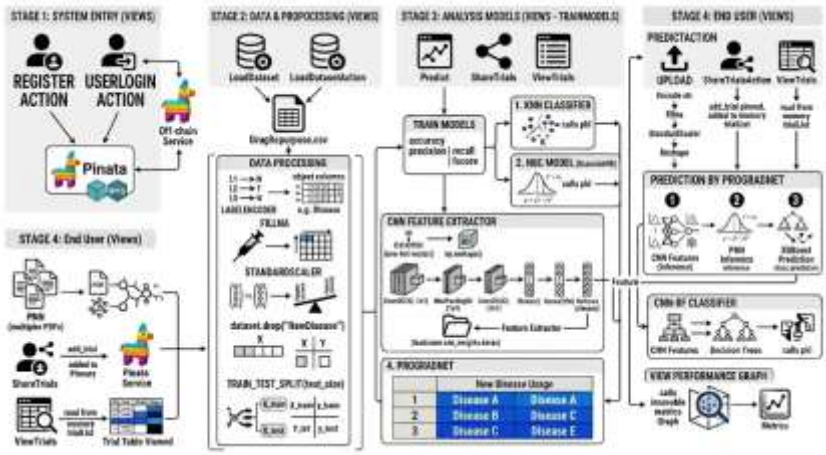

Figure 2: Proposed System Architecture

Secure Data Management: Pinata-based IPFS is integrated to securely store user registration details and clinical trial discussion records. The decentralized storage mechanism ensures data integrity, transparency, immutability, and protection against unauthorized modifications while supporting secure access to stored information.

3.1 proGradNet

The ProGradNet is a hybrid drug repurposing prediction framework that integrates PNN with XGBoost to improve the identification of potential drug–disease associations. The model first processes preprocessed drug-related features through the PNN to estimate class probabilities and capture underlying data

Doi:10.48047/ijiee.2026.16.2.51

distributions. The generated feature representations are then supplied to XGBoost, which performs gradient-boosted classification to produce accurate predictions. This combination enables the framework to learn complex patterns while improving classification performance and reducing prediction errors. As illustrated in Figure 3, the workflow consists of feature initialization, probabilistic learning, feature optimization, boosted classification, and final prediction generation. The integration of probabilistic learning with gradient boosting enhances model robustness, prediction accuracy, and generalization across unseen drug datasets. Consequently, ProGradNet provides an effective computational framework for reliable drug repurposing prediction.

1. Input Feature Preparation: The preprocessed drug dataset is provided as the input to ProGradNet after applying encoding, feature scaling, and missing value handling techniques. Each record is represented as a numerical feature vector suitable for model processing. This standardized representation ensures consistency during learning and improves the quality of subsequent predictions.

2. Probabilistic Pattern Learning: The input feature vectors are processed by the PNN to estimate the probability distribution of each sample. The network measures the similarity between the input and learned patterns using probabilistic estimation techniques. This enables the model to capture hidden relationships among drug attributes and disease classes while preserving important feature characteristics.

3. Feature Representation and Optimization: The probabilistic outputs generated by the PNN are transformed into optimized feature representations. These features emphasize the most informative characteristics while reducing the influence of irrelevant patterns. The optimized representation serves as a robust input for the subsequent classification stage.

4. XGBoost-Based Classification: The optimized feature vectors are supplied to XGBoost for classification. Multiple boosted decision trees are constructed iteratively, where each tree minimizes the prediction errors of the previous iteration. This boosting mechanism improves classification accuracy and effectively handles complex nonlinear relationships within the dataset.

Figure 3: Internal Workflow of ProGradNet

5. Drug–Disease Association Prediction: The trained XGBoost model predicts the most suitable disease association for each candidate drug based on the optimized feature representation. The prediction process assigns the most probable class by analyzing the learned decision boundaries. This stage forms

Doi:10.48047/ijiee.2026.16.2.51

the primary decision-making component of the framework.

6. Final Prediction Generation: The final drug repurposing prediction is generated after completing probabilistic learning and boosted classification. The predicted disease association is presented as the output of the ProGradNet framework. This integrated approach improves prediction reliability, enhances generalization capability, and supports efficient drug repurposing analysis.

4. RESULTS AND DISCUSSION

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 4 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.

Figure 5 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 4: Training AI Models

Doi:10.48047/ijiee.2026.16.2.51

Figure 5: Test Data Prediction with New Disease Usage

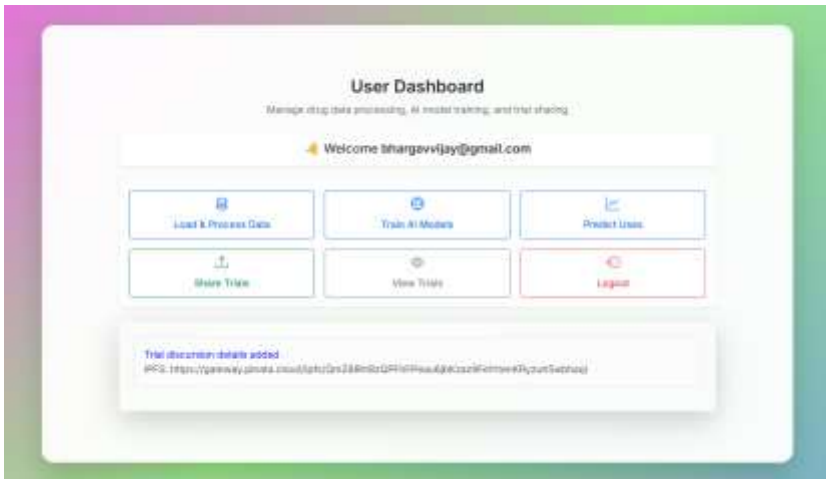

Figure 6: New Trail adding to the Blockchain

Figure 6 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 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 7: viewing New Trails added

Doi:10.48047/ijiee.2026.16.2.51

Figure 7 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 ProGradNet rug repurposing system successfully integrates machine learning, deep learning, decentralized storage, and web technologies to provide an efficient and secure prediction platform. The system combines comparative models such as KNN, NBC, and CNN-RF with the proposed ProGradNet, which integrates PNN and XGBoost to improve drug–disease association prediction. This hybrid approach effectively captures complex data patterns and enhances prediction accuracy compared to individual prediction models. The preprocessing pipeline, consisting of label encoding, missing value handling, feature scaling, and train-test splitting, ensures high-quality input data and improves overall model performance. Experimental evaluation demonstrates that ProGradNet achieves superior accuracy, precision, recall, and F1-score compared with the comparative models. The Django-based web application enables secure user registration, dataset upload, prediction, and clinical trial management through an interactive interface. Furthermore, the integration of Pinata-based IPFS provides secure, transparent, and decentralized storage of user registration details and clinical trial discussion records, ensuring data integrity and reliable information management.

REFERENCES

[1] 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

[2] 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. Epub 2018 Nov 22. PMID: 30472429

[3] Malheiro V, Santos B, Figueiras A, Mascarenhas-Melo F. The Potential of Artificial Intelligence in Pharmaceutical Innovation: From Drug Discovery to Clinical Trials. Pharmaceuticals. 2025; 18(6):788. https://doi.org/10.3390/ph18060788

[4] 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

[5] 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

[6] 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

[7] Ribeiro E, Vale N. Repurposing Terbutaline and Milrinone for Cancer Therapy: A Comprehensive Review. Future Pharmacology. 2025; 5(3):38. https://doi.org/10.3390/futurepharmacol5030038

[8] Pandey, B.S., Durgapal, S., Kumar, P. et al. Unraveling the artificial intelligence role in drug discovery and pharmaceutical product design: an opportunity and challenges. Discov Artif Intell 5, 72 (2025). https://doi.org/10.1007/s44163-025-00330-9

[9] Kant, S., Deepika & Roy, S. Artificial intelligence in drug discovery and development: transforming challenges into

Doi:10.48047/ijiee.2026.16.2.51

opportunities. Discov. Pharm. Sci. **1**, 7 (2025). https://doi.org/10.1007/s44395-025-00007-3

[10] Ali, N., Hanif, N., Khan, H.A. et al. Deep learning and artificial intelligence for drug discovery, application, challenge, and future perspectives. Discov Appl Sci **7**, 533 (2025). https://doi.org/10.1007/s42452-025-06991-6

[11] Kokudeva, Maria, Tsvetelina Velikova, and Georgi Momekov. "Strategies for drug repurposing in COVID-19." *Pharmacia* 73 (2026): e197010.

[12] More, Vrushket, Lyra Lu, Zeyu Ding, Zhaohan Xi, Seth Mizia, and Nancy L. Guo. "TheraMind: a multi-LLM ensemble for accelerating drug repurposing in lung cancer via case report mining." *npj Precision Oncology* (2026).

[13] Chen, Zengjing, Xin Guo, Hao Jiang, Zhiping Liu, and Ziqi Lu. "DRQuantum: a drug repurposing method by quantum walks on a multi-layered heterogeneous network." *Scientific Reports* (2026).

[14] Prakash, Om. "A network-driven computational framework for identifying FDA-approved drug repurposing across heterogeneous brain cancers." *Frontiers in Molecular Biosciences* 13 (2026): 1768081.

[15] Poderoso, Teresa, María Freije-Zazo, Ana Esteban, Patricia Mingo-Casas, Nereida Jiménez de Oya, Miguel-Angel Martín-Acebes, and Ana-Belén Blázquez. "Systematic transcriptomic analysis enables targeted drug repurposing for neurotropic flavivirus infections." *Virology* (2026): 110962.

Doi:10.48047/ijiee.2026.16.2.51