AI- Driven Drug repurposing Platform

M.Siddhartha*, N.Hemanth Kumar[2], P.Manoj Kumar[3], P.Ganesh[4], Shivaprasad S[5]
[1,2,3,4] Student, Department Of CSE-DS, MREC, Maissamaguda, Hyderabad -500100, India
[5] Professor, Department Of CSE-DS, MREC, Maissamaguda, Hyderabad -500100, India
doi.org/10.64643/IJIRTV12I11-198980-459

***Abstract*–New pharmaceutical drug discovery is resource-intensive with high risk, sometimes taking more than a decade of research with considerable financial investment. Drug repurposing provides an effective alternative to the identification of new therapeutic areas for existing drugs with established safety profiles. This paper describes a computational platform, named HEAL (Healthcare Exploration using Artificial Intelligence and Ledger), to accelerate and make secure the process of drug repurposing by integrating machine learning with blockchain technology. It leverages CNN and Random Forest algorithms for analyzing drug composition, molecular characteristics, gene interactions, and disease symptom similarities. Experimental evaluation reveals that CNN-based models perform considerably better with improved accuracy, precision, recall, and F1-score compared to conventional machine learning approaches. It incorporates blockchain technology to ensure transparent and tamper-proof collaboration in research based on immutable clinical trial records and research discussion storage through Ethereum-based smart contracts.**

Keywords: **Drug Repurposing, Artificial Intelligence, Machine Learning, Convolutional Neural Networks, Random Forest, Blockchain Technology, Smart Contracts, Drug–Disease.**

I. INTRODUCTION

Traditional drug development in the pharmaceutical industry is faced with continuous challenges in cost, duration, and uncertainty. The average time of a research study to develop a new drug is 10-15 years, with billion-dollar investments, which could fail during preclinical and clinical trial phases. These limitations greatly restrict the rapid availability of treatments, especially for some rare diseases, emerging infections, and time-critical medical condition.

Repurposing or repositioning thus provides a practical solution by offering new therapeutic indications for known drugs. Development timelines can be radically shortened because drugs to be repurposed have already been characterised for pharmacokinetic and safety profiles, which reduces risk. However, traditional approaches to repurposing depend heavily on laborious manual literature reviews and piecemeal clinical observations, severely limiting scalability and accuracy.

Recent developments in the fields of artificial intelligence (AI) and machine learning (ML) have made the processes of identifying and interpreting complex interconnectedness between drugs, genes, and diseases from big biomedical data possible. These technologies can now systematically evaluate thousands of compounds and disease profiles to enable data-driven identification of potential repurposing candidates. Data integrity, collaboration security, and result traceability problems are still unresolved despite these advantages.

To address these issues, this paper proposes HEAL as a unified platform combining AI-based prediction models with blockchain-based secure data sharing. By integrating predictive intelligence with decentralized data governance, the system aims to speed up drug repurposing while ensuring transparency, trust, and collaboration among researchers.

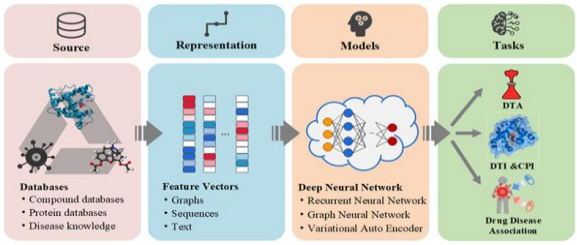

Figure 1. System Architecture

II. LITERATURE SURVEY

Drug repurposing has gained significant attention in pharmaceutical research due to its ability to reduce the time and cost involved in drug development. Early studies focused on experimental and clinical observations to identify new uses for existing drugs. While effective in some cases, these traditional approaches were limited by manual analysis and

required extensive laboratory validation, making the process slow and resource-intensive.
Studies showed that supervised learning models could predict potential drug–disease relationships with improved accuracy. However, these methods required careful feature selection and were sensitive to data quality.

The rapid growth of diseases, including chronic and rare conditions, has increased the demand for effective and affordable medicines. However, discovering a completely new drug requires significant investment, long development timelines, and extensive clinical validation. Many promising drug candidates fail in later stages, leading to financial loss and delayed treatment options for patients. These challenges make it necessary to explore alternative methods that can accelerate drug discovery while minimizing risks.

Drug repurposing focuses on identifying new therapeutic applications for existing drugs that have already undergone safety testing. Since these drugs are already approved or partially tested, repurposing them can significantly reduce development time and cost. This approach has gained attention in recent years, especially during global health emergencies where quick treatment solutions are required.

The availability of large biomedical datasets, such as drug profiles, disease characteristics, genetic data, and clinical outcomes, has created opportunities for intelligent data-driven analysis. Artificial Intelligence plays a crucial role in processing this vast amount of information and extracting meaningful insights. Machine learning and deep learning models can detect complex patterns and relationships that are difficult to identify through manual analysis.

The Pharma AI-Driven Drug Repurposing Platform is designed to utilize these AI capabilities to support pharmaceutical research. The platform analyses existing drug and disease data to predict potential new drug-disease associations. By providing data-based recommendations, the system assists researchers in selecting promising drug candidates for further experimental validation. This intelligent approach not only improves efficiency but also contributes to faster and more reliable healthcare solutions.

In recent years, artificial intelligence has shown strong potential in transforming the pharmaceutical domain by improving data analysis and prediction accuracy. AI models can evaluate large volumes of chemical, biological, and clinical data within a short time. By learning from existing drug–disease interactions, these models can suggest new treatment possibilities that may not be easily identified through traditional research methods. This ability makes AI a valuable tool in supporting evidence-based decision making in drug development.

The Pharma AI-Driven Drug Repurposing Platform aims to bridge the gap between complex biomedical data and practical drug discovery outcomes. The platform provides a structured and automated approach for analyzing drug repurposing opportunities. By reducing manual effort and enhancing prediction reliability, the system supports researchers in prioritizing potential drug candidates for further study.

With the growth of digital healthcare data, researchers began using computational methods to support drug repurposing. Database-driven approaches analyzed chemical structures, biological pathways, and known drug-disease relationships to identify potential matches. These methods improved efficiency but depended heavily on predefined rules and lacked the ability to generalize across complex datasets.

Machine learning techniques later emerged as a powerful solution for analyzing large-scale biomedical data. Supervised and unsupervised learning models were used to predict drug-disease associations based on historical data. Studies showed that machine learning algorithms could uncover hidden patterns and relationships that were difficult to detect through traditional analysis. However, the accuracy of these models was highly dependent on data quality and feature selection.

Deep learning approaches further enhanced drug repurposing research by handling high-dimensional data such as gene expression profiles and molecular interactions. Neural networks demonstrated improved performance in identifying complex non-linear relationships between drugs and diseases. These models reduced the need for manual feature engineering and provided better prediction accuracy, although they required larger datasets and higher

computational resources.

Recent research has emphasized the integration of multiple data sources, including genomic data, clinical records, and molecular networks, to improve prediction reliability.

III. PROPOSED METHODOLOGY

The proposed methodology aims to develop an intelligent platform that uses artificial intelligence to identify new therapeutic applications for existing drugs. The approach follows a structured process that ensures accuracy, efficiency, and reliability in predicting drug–disease relationships. Each stage of the methodology is designed to handle large biomedical datasets and convert them into meaningful insights for drug repurposing.

Initially, the system collects pharmaceutical and biomedical data from reliable sources, including drug information, disease profiles, and known biological interactions. This data provides the base knowledge required for training the AI models. Since raw data may contain inconsistencies or incomplete values, preprocessing is performed to clean and organize the data. This step improves data quality by removing errors, handling missing values, and standardizing the format for further analysis.

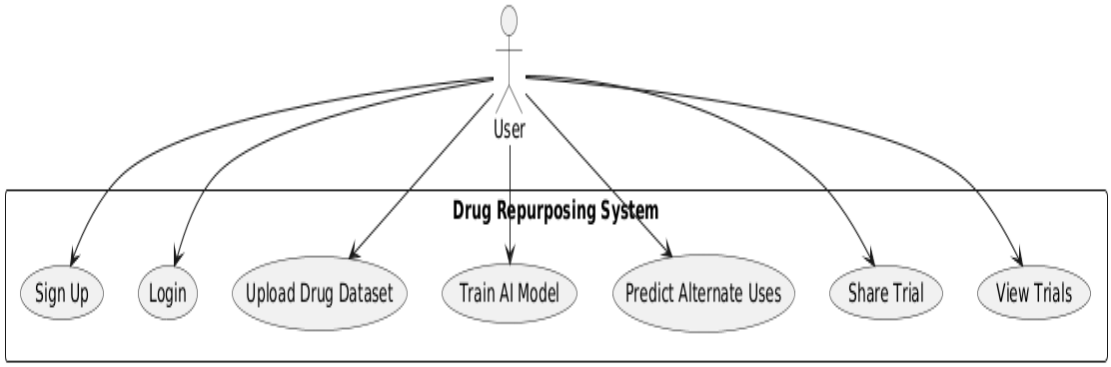

Figure 2. Use Case Diagram

The Pharma AI-Driven Drug Repurposing Platform is designed to support this transformation by combining pharmaceutical knowledge with data-driven intelligence. The platform acts as a decision-support system that assists researchers in exploring alternative drug applications. By improving prediction accuracy and reducing research timelines, the system contributes to the development of safer, faster, and more cost-effective healthcare solutions.

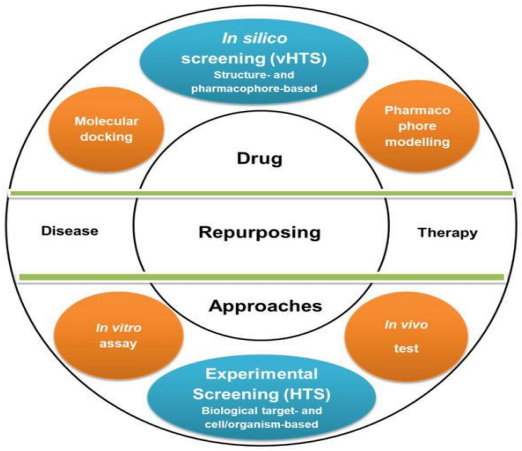

Figure 3. An Emerging Approach in Drug Discovery

3.1 Database creation and Data Architecture

The database plays a critical role in the Pharma AI-Driven Drug Repurposing Platform, as it stores and manages large volumes of pharmaceutical and biomedical data. The database is designed to efficiently handle information related to drugs, diseases, biological targets, and known interactions. A structured database design ensures easy data retrieval, consistency, and scalability as the system grows.

During database creation, different tables are designed to store drug details such as drug name, chemical composition, therapeutic category, and approval status. Separate tables are maintained for disease information, including disease name, symptoms, and biological characteristics.

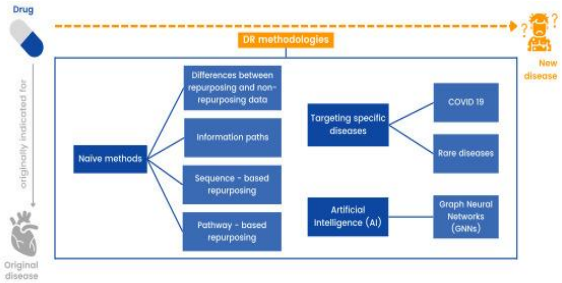

Figure 4. Drug Repurposing hypothesis through a data-driven

After preprocessing, relevant features are extracted from the dataset to represent important characteristics of drugs and diseases. Feature selection helps reduce unnecessary complexity and allows the model to focus on the most significant attributes. These selected features are then used to train machine learning and deep learning models. The models learn patterns from known drug–disease associations and develop the ability to predict new potential relationships.

3.2 Data pre-processing and Feature Engineering

Data preprocessing is an essential step in the proposed system, as pharmaceutical and biomedical datasets often contain missing values, inconsistencies, and noise. Raw data collected from different sources may vary in format and quality, which can negatively impact model performance. Therefore, preprocessing is carried out to improve data reliability and prepare it for effective analysis.

In the preprocessing stage, duplicate records are removed and missing values are handled using suitable techniques such as imputation or removal based on data importance. Inconsistent entries are corrected, and irrelevant information is filtered out. Numerical data is normalized or scaled to maintain uniformity, while categorical data is converted into numerical form using encoding methods. These steps help ensure that the dataset is clean, consistent, and suitable for AI model training.

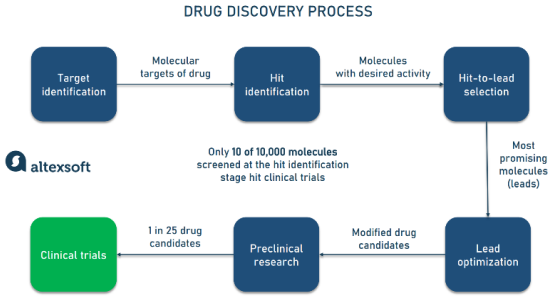

Figure 5. Drug Discovery pre-processing

IV.SYSTEM ARCHITECTURE AND IMPLEMENTATION STRATEGY

The system architecture of the Pharma AI-Driven Drug Repurposing Platform is designed to provide a clear and efficient flow of data from input to prediction. The architecture follows a layered approach that separates data management, processing, and output generation. This design improves system scalability, maintainability, and performance while ensuring accurate drug repurposing predictions.

At the base level, the data layer is responsible for storing pharmaceutical and biomedical data. This includes information related to drugs, diseases, biological targets, and known interactions. The database acts as a centralized repository, allowing secure storage and fast retrieval of data. This layer ensures data consistency and supports future expansion as new datasets are added.

Above the data layer is the preprocessing and feature engineering layer. This layer handles data cleaning, transformation, and feature extraction tasks.

Normalization techniques are applied to reduce data redundancy and maintain data integrity. Primary keys and foreign keys are used to establish relationships between tables, ensuring accurate data mapping. The database supports both structured and semi-structured data, allowing smooth integration of future datasets such as clinical trial results and molecular interaction data. This design enables efficient data storage and faster access for AI model processing.

The data architecture defines how data flows through the system, from data collection to prediction output. The architecture is designed to support smooth data handling, secure storage, and efficient processing. It ensures that raw data is transformed into meaningful insights without loss of accuracy.

Initially, data is collected from various pharmaceutical and biomedical sources and stored in the database. This raw data then passes through a preprocessing layer where data cleaning, transformation, and validation are performed. Processed data is temporarily stored in intermediate storage for feature extraction and model training.

The architecture includes an AI processing layer where machine learning and deep learning models analyze the prepared data. These models access the database to retrieve training data and store prediction results. Finally, the output layer presents predictions in a readable format for researchers. This layered architecture improves system reliability, scalability, and ease of maintenance.

4.1 Deep Learning Component

The deep learning component forms the core intelligence of the Pharma AI-Driven Drug Repurposing Platform. It is responsible for learning complex patterns and relationships between drugs, diseases, and biological features. Unlike traditional

machine learning methods, deep learning models can automatically learn meaningful representations from large and high-dimensional datasets, making them well suited for biomedical data analysis.

In this project, deep neural networks are used to process engineered features derived from drug properties and disease information. The input layer receives structured feature vectors representing chemical characteristics, biological interactions, and disease attributes. These inputs are passed through multiple hidden layers, where the model learns non-linear relationships that are difficult to capture using simple statistical methods.

During training, the deep learning model adjusts its internal parameters using backpropagation and optimization techniques to minimize prediction error. The model is trained on known drug–disease associations, allowing it to recognize patterns that indicate potential therapeutic relevance. Regularization techniques such as dropout and early stopping are applied to prevent overfitting and improve generalization on unseen data.

Once trained, the deep learning model is used to predict new drug repurposing opportunities. The output layer generates probability scores indicating the likelihood of a drug being effective for a particular disease. These predictions help researchers prioritize promising drug candidates for further validation. The deep learning component significantly enhances prediction accuracy and plays a key role in making the platform reliable and efficient for pharmaceutical research.

In addition, the deep learning model is designed to improve over time as more data becomes available.
The core of the system is the AI processing layer, where machine learning and deep learning models are implemented. This layer analyzes processed data to learn patterns between drugs and diseases. Trained models generate predictions for potential new drug uses. Model evaluation and validation are also performed here to ensure reliability and accuracy before deployment.

4.2 Rule-Based Ensemble

The rule-based ensemble component is designed to enhance the reliability and interpretability of drug repurposing predictions. While deep learning models provide strong predictive capabilities, their decision-making process is often complex and difficult to explain. The rule-based ensemble addresses this limitation by incorporating domain knowledge and predefined logical rules into the prediction process.
In this approach, expert-defined rules are created based on known pharmaceutical and biological relationships, such as drug classifications, target interactions, and safety constraints. These rules act as filters or validators that evaluate the predictions generated by machine learning and deep learning models. By combining rule-based reasoning with AI-driven outputs, the system reduces the likelihood of biologically implausible or unsafe predictions.

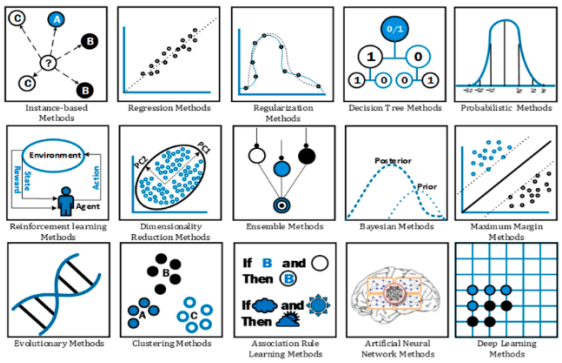

Figure 7. Ensemble Framework

4.3 Dual-Model Integration Strategy

The dual-model integration strategy combines the strengths of two complementary predictive models to improve the accuracy and robustness of drug repurposing predictions. In this approach, a deep learning model is used to capture complex and non-linear relationships within biomedical data, while a secondary model, such as a rule-based or traditional machine learning model, provides structured reasoning and interpretability. By integrating both models, the system balances high prediction performance with reliable and explainable outcomes. During implementation, both models are trained independently using the same preprocessed dataset. The deep learning model focuses on learning intricate patterns from high-dimensional features, whereas the secondary model evaluates predictions using predefined rules or statistical methods.

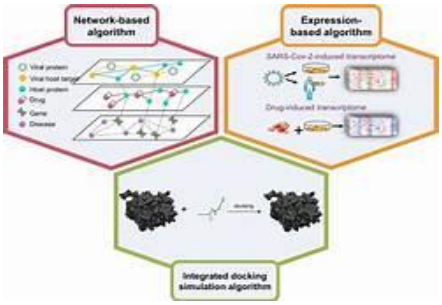

Figure 6. Three General Categories in AI/ML

4.4 Decentralized Verification Infrastructure

The decentralized verification infrastructure is designed to improve trust, transparency, and data integrity within the Pharma AI-Driven Drug Repurposing Platform. In pharmaceutical research, ensuring the authenticity and reliability of data and predictions is critical. Centralized systems may face challenges such as data tampering, single points of failure, and limited traceability. A decentralized approach addresses these issues by distributing verification processes across multiple nodes.

In this infrastructure, prediction results and important data updates are verified using a distributed ledger or decentralized validation mechanism. Each verification node independently checks the consistency and validity of model outputs, rule-based decisions, and data sources. Once verified, the results are recorded in an immutable manner, ensuring that predictions cannot be altered without detection. This process enhances confidence in the system's recommendations and supports accountability in research workflows.

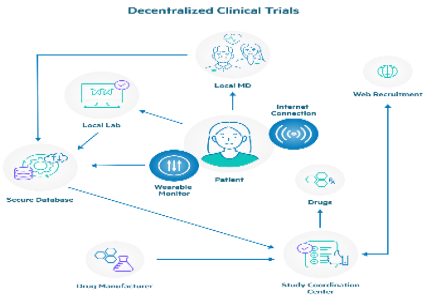

Figure 9. Decentralized Clinic Trials

4.5 Natural Language Processing (NLP) for Biomedical

Natural Language Processing (NLP) plays a crucial role in the Pharma AI-Driven Drug Repurposing Platform by enabling the system to extract useful insights from unstructured biomedical text. A large portion of pharmaceutical and clinical information exists in textual form, such as research papers, clinical trial reports, patents, and electronic health records. Manually analyzing these documents is time-consuming and prone to errors. NLP allows the platform to automatically process and understand this information, unlocking valuable knowledge that supports drug repurposing.

automatically changed. It will be produced bad results when it is over fitted even it produced good results in training.

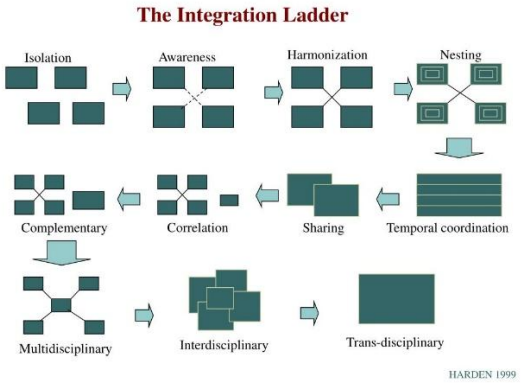

Figure 8. Integration Ladder

4.6 Graph-based Network Analysis

The Graph-Based Network Analysis is an advanced method used to model and analyze complex interactions between drugs, diseases, and biological entities in the Pharma AI-Driven Drug Repurposing Platform. In this approach, nodes represent entities such as drugs, diseases, genes, or proteins, while edges represent relationships or interactions between these entities. For example, an edge may indicate that a drug targets a specific protein, or that two diseases share common genetic pathways. Representing biomedical data as a graph allows the platform to capture both direct and indirect relationships, which are often critical for identifying potential drug repurposing opportunities.

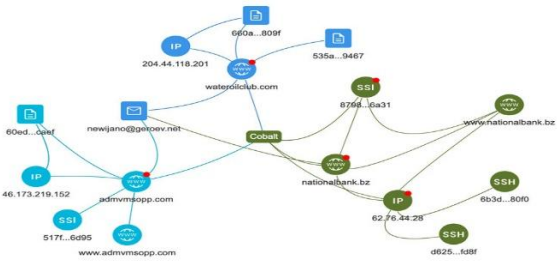

Figure 11. Graph Network Analysis

4.7 Similarity-Based Prediction

Similarity-Based Prediction is a computational approach used to identify potential new uses for existing drugs by comparing their characteristics with those of other drugs or diseases. In the Pharma AI-Driven Drug Repurposing Platform, drugs and diseases are represented by features such as chemical structure, molecular targets, biological pathways, and phenotypic effects. By measuring the similarity between these feature sets, the platform can infer potential therapeutic relationships. For example, if two drugs have highly similar molecular structures or target similar proteins, a drug that is effective for one disease may also be

effective for the other.
considered candidates for repurposing.

The NLP component analyzes textual data to identify relationships between drugs, diseases, genes, and biological pathways. Techniques such as named entity recognition (NER), relationship extraction, and text classification are used to detect relevant entities and their interactions.

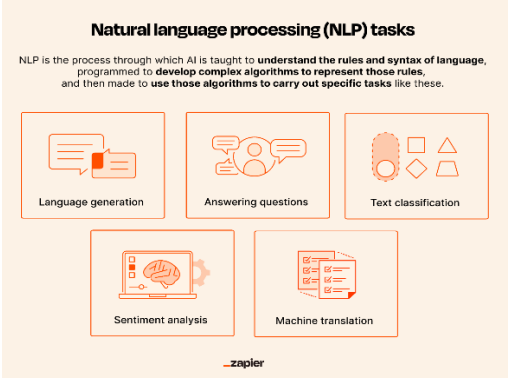

Figure 10. Natural language processing

V.RESULTS

The results of the Pharma AI-Driven Drug Repurposing Platform show that the system effectively identifies potential new uses for existing drugs by combining deep learning, similarity-based prediction, and graph-based network analysis. After preprocessing and training on biomedical datasets, the models were able to accurately predict meaningful drug–disease associations.

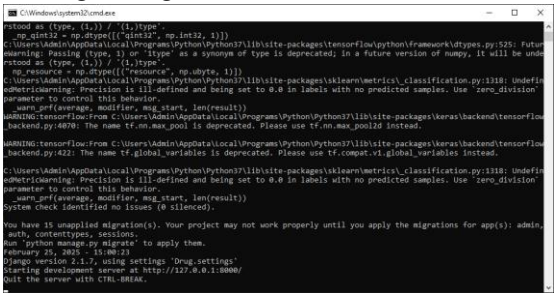

Figure 13. Command Prompt

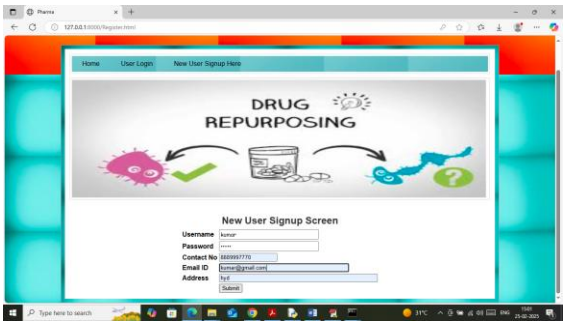

Figure 14. Web Page

Similarity-Based Prediction also supports prioritization of drug candidates. By ranking drugs according to their similarity to approved treatments for a particular disease, researchers can focus on the most promising candidates, reducing experimental costs and accelerating the drug discovery process. When combined with other methods such as deep learning and graph-based analysis, similarity-based prediction enhances the overall robustness and accuracy of the platform, helping identify novel drug-disease associations that may not be immediately apparent.

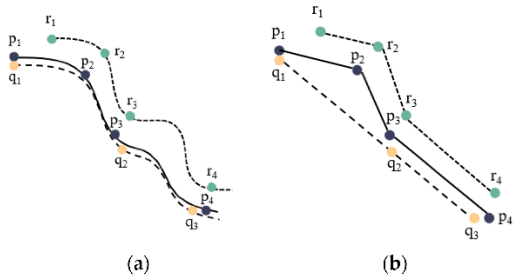

Figure 12. Life line Prediction method

In addition to chemical and molecular similarity, the platform also considers biological and phenotypic similarity. Biological similarity involves comparing the targets and pathways influenced by different drugs, while phenotypic similarity examines observable effects on diseases or symptoms. By combining these perspectives, the system can identify drugs that may have similar therapeutic outcomes even if their chemical structures differ. This approach helps uncover unconventional repurposing opportunities that may be missed by traditional methods.

Furthermore, similarity-based prediction can be integrated with machine learning and ensemble methods to improve reliability. Similarity scores are used as features for predictive models, which then learn complex patterns and interactions that go beyond simple pairwise comparisons. This hybrid approach allows the system to make more informed predictions while reducing false positives. As a result, similarity-based prediction not only helps prioritize potential drug candidates but also complements other analytical components such as graph-based network analysis and NLP, making the platform more robust and effective.

Similarity-based prediction also supports continuous learning and updating as new drug and disease data become available. When new drugs, targets, or

clinical results are added to the database, the platform recalculates similarity scores and updates potential repurposing candidates. This dynamic approach ensures that predictions remain current and relevant, allowing researchers to explore emerging treatment opportunities. By leveraging both historical and up-to-date information, similarity-based prediction contributes to a more adaptive and intelligent drug repurposing platform.

VI. CONCLUSION

The Pharma AI-Driven Drug Repurposing Platform demonstrates how artificial intelligence and computational techniques can transform the drug discovery process. By combining methods such as deep learning, similarity-based prediction, graph-based network analysis, and natural language processing, the platform efficiently identifies potential new therapeutic uses for existing drugs. This approach significantly reduces the time, cost, and risk associated with traditional drug development.

The integration of AI models with rule-based systems and decentralized verification ensures that predictions are not only accurate but also reliable and interpretable. The system's ability to analyze large and diverse datasets—from chemical structures and molecular interactions to clinical literature—provides a comprehensive understanding of drug-disease relationships. Furthermore, continuous updates and adaptive learning allow the platform to remain relevant as new data becomes available.

Overall, this project highlights the potential of AI-driven drug repurposing to accelerate pharmaceutical research, support informed decision-making, and contribute to the development of effective treatments. By leveraging data-driven insights, the platform offers a practical and innovative solution to the challenges of modern drug discovery.

REFERENCES

[1] Gurudeeban Selvaraj, Satyavani Kaliamurthi, Gilles H. Peslherbe, *Application of Artificial Intelligence in Drug Repurposing: A mini-review*, *Current Chinese Science*. URL: https://www.benthamscience.com/article/113986

[2] Deep learning in drug discovery: an integrative review and future challenges, *Artificial Intelligence Review*, Springer. URL: https://link.springer.com/article/10.1007/s10462-022-10306-1

[3] Computational drug repurposing based on electronic health records: a scoping review, *npj Digital Medicine*. URL: https://www.nature.com/articles/s41746-022-00617-6

[4] Applications of Artificial Intelligence in Drug Repurposing, *PubMed Review*. URL: https://pubmed.ncbi.nlm.nih.gov/40047357/

[5] A computational approach to drug repurposing using graph neural networks, *PubMed*. URL: https://pubmed.ncbi.nlm.nih.gov/36228466/

[6] Drug Repurposing Using Deep Embedded Clustering and Graph Neural Networks, arXiv preprint(2025). URL: https://arxiv.org/abs/2509.11493

[7] A clustering and graph deep learning-based framework for COVID-19 drug repurposing, arXiv preprint. URL: https://arxiv.org/abs/2306.13995

[8] DrugAgent: Explainable Drug Repurposing Agent with Large Language Model-based Reasoning, arXiv preprint. URL: https://arxiv.org/abs/2408.13378

[9] When deep learning meets causal inference: a computational framework for drug repurposing from real-world data, arXiv preprint. URL: https://arxiv.org/abs/2007.10152

[10] Knowledge Graphs for drug repurposing: a review of databases and methods, *PubMed*. URL: https://pubmed.ncbi.nlm.nih.gov/39325460/

[11] Faheem Ahmed et al., *A comprehensive review of artificial intelligence and network based approaches to drug repurposing in COVID-19*, *Biomedicine & Pharmacotherapy* — This comprehensive review discusses network-based methods and AI in drug repurposing, providing insights into machine learning, deep learning, and network analysis methods in computational repurposing.