Drug Repurposing and Adverse Effect Prediction Using Knowledge Graphs and Clinical Trial Data

Project Summary:

This project aims to develop a comprehensive framework for drug repurposing and adverse effect prediction by leveraging knowledge graphs and clinical trial data. The project will integrate diverse biomedical data sources, including drug-target interactions, gene-disease associations, and clinical trial outcomes, to build a knowledge graph. This graph will be used to identify potential drug repurposing opportunities and predict adverse effects associated with existing drugs.

Objectives:

1. Data Integration and Knowledge Graph Construction:

- Integrate data from multiple sources, including drug databases (e.g., DrugBank), gene-disease associations (e.g., DisGeNET), protein-protein interactions, and clinical trial databases (e.g., ClinicalTrials.gov).
- Construct a knowledge graph that captures the complex relationships between drugs, targets, diseases, and adverse effects.

2. Drug Repurposing:

- Develop algorithms to traverse the knowledge graph and identify potential drug repurposing candidates for specific diseases.
- Utilize machine learning techniques to rank the repurposing candidates based on the strength of evidence from the knowledge graph and clinical trial data.

3. Adverse Effect Prediction:

- Analyze historical clinical trial data to identify patterns and relationships between drugs and adverse effects.
- Use the knowledge graph to predict potential adverse effects for repurposed drugs by identifying similar patterns in the graph.

4. Validation:

- Validate the predicted drug repurposing candidates and adverse effects using in silico methods and available experimental data.
- Collaborate with clinical researchers to design experiments for further validation of the top-ranked drug candidates.

5. Development of a User-Friendly Platform:

- Develop a web-based platform that allows researchers to explore knowledge graphs, search for drug repurposing opportunities, and predict adverse effects.

Methodology:

1. Data Collection:

- Collect drug-related data from databases like DrugBank, PubChem, and ChemBL.
- Gather gene-disease association data from resources like DisGeNET and OMIM.
- Extract clinical trial data, focusing on outcomes and reported adverse effects, from ClinicalTrials.gov.

2. Knowledge Graph Construction:

- Use graph databases (e.g., Neo4j) to store and manage the knowledge graph.
- Nodes will represent entities such as drugs, genes, proteins, diseases, and clinical trials, while edges will represent relationships such as drug-target interactions, gene-disease associations, and drug-adverse effect correlations.

3. Algorithm Development:

- Implement graph traversal algorithms to explore the knowledge graph and identify drug repurposing candidates.
- Develop machine learning models to predict adverse effects based on patterns observed in the graph.

4. Validation and Testing:

- Use cross-validation and other statistical methods to assess the accuracy of the repurposing predictions.
- Compare the predicted adverse effects with known adverse events from clinical trial data to evaluate the model's performance.

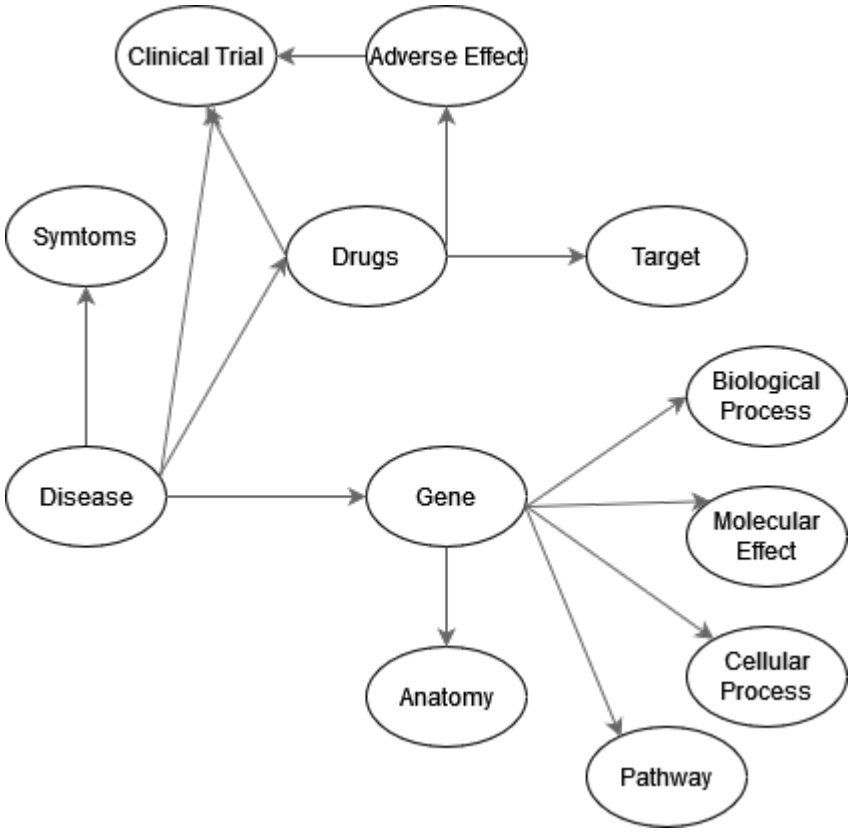
5. Platform Development:

- Use web technologies (e.g., Django, React) to build an interactive platform that provides access to the knowledge graph.
- Implement search and visualization tools to help users explore the graph and retrieve relevant information.

Expected Outcomes:

- A comprehensive knowledge graph that integrates diverse biomedical data sources.
- A list of potential drug repurposing candidates for various diseases, prioritized based on evidence strength.
- Predictions of adverse effects for repurposed drugs, validated using clinical data.

- A user-friendly platform for researchers to explore drug repurposing opportunities and predict adverse effects.



The toy diagram of the proposed knowledge graph established during the project

Collaborators:

- **Bioinformaticians:** For data integration and algorithm development.
- **Clinical Researchers:** For validation of drug repurposing candidates.
- **Software Developers:** For platform development and deployment.

Potential Impact:

This project has the potential to accelerate drug repurposing efforts, reduce the cost of drug development, and improve patient safety by predicting adverse effects before clinical application. The platform will serve as a valuable resource for researchers in academia and the pharmaceutical industry, fostering collaboration and innovation in drug discovery and safety assessment.

References:

1. Himmelstein DS, Zietz M, Rubinetti V, Kloster K, Heil BJ, Alquaddoomi F, Hu D, Nicholson DN, Hao Y, Sullivan BD, Nagle MW, Greene CS. Hetnet connectivity search provides rapid insights into how biomedical entities are related. Gigascience. 2022.
2. Boudin, M., Diallo, G., Drancé, M. et al. The OREGANO knowledge graph for computational drug repurposing. Sci Data 10, 871 (2023).
3. Bang, D., Lim, S., Lee, S. et al. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. Nat Commun 14, 3570 (2023).
4. Jiménez A, Merino MJ, Parras J, Zazo S. Explainable drug repurposing via path based knowledge graph completion. Sci Rep. 2024 Jul 18;14(1):16587.
5. Knox, C., Wilson, M., Klinger, C. M. & et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Res.52, 1265–1275 (2024).
6. Ritchie, M. E., Phipson, B., Wu, D. & et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, 1–13 (2015).
7. Szklarczyk, D. & et.al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res. 39, D561–D568 (2011).
8. Shih, Y.-K. & Parthasarathy, S. Identifying functional modules in interaction networks through overlapping markov clustering. Bioinformatics 28, i473–i479 (2012).
9. Fitch, A. M. & Jones, M. B. Shortest path analysis using partial correlations for classifying gene functions from gene expression data. Bioinformatics 25, 42–47 (2009).
10. Huang, L., Luo, H., Li, S., Wu, F.-X. & Wang, J. Drug–drug similarity measure and its applications. Briefings Bioinforma.22, bbaa265 (2021).

Dr. Ashmita Dey, DST INSPIRE Fellow, JSPS Fellow
Research Associate III
Indian Statistical Institute, Kolkata