RN SHETTY TRUST®
RNS INSTITUTE OF TECHNOLOGY
Autonomous Institution Affiliated to VTU, Recognized by GOK, Approved by AICTE
(NAAC 'A+ Grade' Accredited, NBA Accredited (UG - CSE, ECE, ISE, EIE and EEE)
Channasandra, Dr. Vishnuvardhan Road, Bengaluru - 560 098
Ph:(080)28611880,28611881 URL: www.rnsit.ac.in
Department of CSE (Data Science)

1. AI-Powered Prediction of Drug-Protein Binding Affinity Using Machine Learning

2. Abstract / Purpose

The aim of this project is to develop a machine learning model to predict the binding affinity between drug-like compounds and proteins. Using public datasets and established deep learning architectures, the model will help virtually screen drug candidates, accelerating early-stage drug discovery and reducing the need for expensive lab experiments.

3. Introduction / Background

Discovering new drugs requires identifying molecules that bind effectively to disease-related proteins, a process that is slow and costly. Advances in AI now allow researchers to predict drug-protein interactions efficiently using computational models. This project explores applying machine learning to address this challenge.

4. Objectives

- To build a model that predicts binding strengths of drugs to proteins.
- To use public datasets for training and evaluation.
- To validate AI's utility for drug discovery.

5. Scope of the Project

- Computational research only; no lab work involved.
- Focus on model development and evaluation.
- Uses existing ML architectures and open datasets.

6. Literature Review / Existing System

- Artificial intelligence (AI) and deep learning have transformed drug discovery workflows, making virtual screening and molecular design faster and more accurate.
- AlphaFold and its successors have achieved protein structure predictions at near-experimental quality, radically improving structural biology.
- Graph Neural Networks (GNNs) allow for direct modeling of molecular graphs, improving drug-target interaction predictions.
- Deep learning models such as convolutional neural networks and transformers enable improved protein model quality and accurate tertiary structure prediction.
- AI-driven models enable high-throughput screening of millions of compounds, significantly reducing costs and timelines for early-stage drug discovery.
- Generative AI models (e.g., DiffDock) have revolutionized molecular docking and "de novo" drug design methodologies.
- AI-based tools enhance ADMET and toxicity prediction but still require further validation using in vivo data.
- Multiomics and NLP approaches contribute to target identification and mapping complex biological pathways.
- Challenges in AI drug discovery include data validity, bias, and lack of comprehensive patient data, highlighting the need for robust validation frameworks.
- Template-based modeling (TBM) is accurate with close homologs but struggles with distant evolutionary relationships; DL methods have improved this.
- Classical ML methods offer useful foundations, but modern deep neural networks often outperform them for binding affinity tasks.
- Knowledge graphs, biomedical graphs, and heterogeneous graph neural networks are integrated to optimize drug discovery and prediction of DTIs.
- Drug repositioning and polypharmacology benefit greatly from GNNs and interpretable deep models.
- Explainable AI is increasingly important for mechanism prediction and understanding drug action.
- Accessible, open-source databases like DrugBank, PubChem, PDB, and ZINC provide the data backbone for modern predictive systems

7. Proposed System / Methodology

- Collect and preprocess data from DrugBank, PubChem, PDB.
- Represent molecules as graphs/sequences for ML input.
- Implement and train a deep learning model (e.g., Graph Neural Network) for binding prediction.
- Validate results on benchmark datasets.

8. Software & Hardware Requirements

Software: Python 3.x, PyTorch/TensorFlow, RDKit, Pandas, NumPy

Hardware: Standard desktop with GPU preferred (for training); cloud services optional for larger experiments.

9. Work Plan / Time Schedule

Data collection & review: 2 weeks

Model setup & initial coding: 3 weeks

Training, testing, and optimization: 5 weeks

Analysis and reporting: 2 weeks

10. Expected Outcome

The project is expected to produce a trained machine learning model that accurately predicts how strongly drug compounds bind to target proteins. The results will show how AI methods can improve the speed and effectiveness of early drug screening, making the discovery process more efficient for research and development.

11. Conclusion

This work demonstrates the practical value of AI in drug discovery by focusing on binding affinity prediction. The approach highlights how even student research using existing models and public datasets can make a meaningful contribution to pharmaceutical research and computational biology.

12. References

[1] Blanco-González, A. et al., "The Role of AI in Drug Discovery: Challenges and Opportunities." PMC, 2023.

[2] ACS Omega, "AI-Driven Drug Discovery: A Comprehensive Review," 2025.

[3] Fu, C. et al., "The future of pharmaceuticals: Artificial intelligence in drug discovery," 2025.

[4] Arxiv, "Artificial intelligence in drug discovery: A comprehensive literature review," 2025.

[5] Klaus, B. et al., "AI Models and Drug Discovery Within Pharmaceutical Research," 2023.

[6] Meng, Y. et al., "Protein structure prediction via deep learning: an in-depth review," 2025.

[7] Yao, R. et al., "Knowledge mapping of graph neural networks for drug discovery," 2024.

[8] Wiley, "The Role of AI in Drug Discovery," 2023.

[9] Jumper, J. et al., "Highly accurate protein structure prediction with AlphaFold," Nature, 2021.

[10] Zhang, O. et al., "Graph Neural Networks in Modern AI-aided Drug Discovery," Arxiv, 2025.

[11] Patel, V. et al., "Artificial intelligence and machine learning in drug discovery and development," 2022.

[12] Huang, B. et al., "Protein Structure Prediction: Challenges, Advances, and Future Directions," 2023.

[13] Wang, C. et al., "Drug discovery and mechanism prediction with explainable AI," Nature, 2025.

[14] Gangwal, A. et al., "Unleashing the power of generative AI in drug discovery," 2024.

[15] Bertoline, L.M.F. et al., "Before and after AlphaFold2: An overview of protein structure prediction," Frontiers in Bioinformatics, 2023.

Signature of Subject Teacher **Signature of Students**

Prof.Navyashree K S