Report on

AI-Driven Pipeline for Standardization and Authentication of Ayurvedic Formulations using Network Pharmacology and Deep Learning

Based on the Project topic

Standardization & Authentication of Ayurvedic Formulations -Leveraging AI and Analytical Chemistry to Validate and Authenticate Medicinal Plant Extracts

Submitted by

Subhadeep Barman
Pritam Mondal
Sourov Debnath
B.Tech, Information Technology
Indian Institute of Engineering Science and Technology, Shibpur

Under the guidance of

Dr. Sudip Roy
Associate Professor, Department of Computer Science and Engineering
Joint Faculty, Centre of Excellence in Disaster Mitigation and Management
Indian Institute of Technology (IIT) Roorkee

In partial fulfilment of

The requirements of the

Summer Research Internship

(May-July 2025)

Submission Date: October 2025

Acknowledgement

We express our deepest gratitude to **Dr. Sudip Roy**, *Associate Professor, Department of Computer Science and Engineering, Joint Faculty, Centre of Excellence in Disaster Mitigation and Management, Indian Institute of Technology (IIT) Roorkee*, for providing us the invaluable opportunity to undertake this **Summer Research Internship** on *"AI-Driven Standardization and Authentication of Ayurvedic Formulations using Network Pharmacology and Deep Learning."*

We are sincerely thankful to our **research coordinators** for their continuous support, guidance, and insightful feedback throughout the internship period. Their expertise in AI-based drug discovery and Ayurvedic pharmacology was instrumental in shaping the direction and outcomes of this project.

We also extend our heartfelt appreciation to the **Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur (IIESTS)**, for providing the foundational academic training and research environment that enabled us to contribute effectively to this interdisciplinary work.

Finally, we gratefully acknowledge the contributions of our teammates — **Subhadeep Barman, Pritam Mondal, and Sourov Debnath** — whose collaboration, dedication, and shared effort made this project a success.

Subhadeep Barman

Pritam Mondal

Sourov Debnath

Index

Abstract

Standardization and authentication of Ayurvedic formulations present a significant challenge due to their multi-component and multi-target nature. Each herbal formulation can contain dozens of phytochemicals, and only a few of them exhibit significant bioactivity. Experimental validation of every compound–target pair is resource-intensive and impractical.

This study presents a unified **AI-driven three-stage computational pipeline** that integrates **deep learning**, **network pharmacology**, and **regression modeling** to predict, filter, and prioritize bioactive compounds from Ayurvedic formulations. The pipeline includes:

1. **DeepPurpose** – predicting compound–target binding affinities using molecular and protein sequence encodings.
2. **DrugMap-style filtering** – selecting biologically relevant and high-affinity compounds using network pharmacology and PPI-based prioritization.
3. **NetPharmaAI regression** – re-ranking top candidates for in-vitro validation using regression-based affinity estimation.

Together, these modules provide an end-to-end AI-assisted workflow for the **scientific standardization of Ayurvedic formulations**, reducing the candidate search space from 40–50 compounds to 3–4 top-ranked bioactives.

1. Introduction

Ayurveda, India's traditional system of medicine, utilizes poly-herbal formulations with complex synergistic effects. However, the lack of mechanistic understanding and scientific validation of these formulations limits their acceptance in modern pharmacology.

Modern computational pharmacology, powered by **artificial intelligence (AI)**, offers scalable methods for predicting compound–target interactions (CTIs), enabling biological validation without extensive laboratory trials. Traditional AI approaches like **deep learning–based drug–target affinity (DTA) prediction** and **network pharmacology** can be adapted to study Ayurvedic phytochemicals effectively.

This project, conducted during the **IIT Roorkee Summer Research Internship**, aims to integrate multiple AI frameworks into a unified pipeline that supports compound–target prediction, filtering, and prioritization for Ayurvedic standardization.

2. Objectives

- Predict compound–target binding affinities using **DeepPurpose**.
- Filter and prioritize bioactive compounds using **DrugMap-style filtering** based on PPI networks and target coverage.

- Re-rank the shortlisted compounds via **NetPharmaAI regression** to identify the most promising candidates.
- Develop an integrated AI–Network Pharmacology framework to aid standardization and authentication of Ayurvedic formulations.

3. Methodology

3.1 Stage 1 – DeepPurpose: Compound–Target Binding Prediction

Goal: Predict molecular binding affinities (Ki/pKd) between Ayurvedic compounds and protein targets.

Inputs: Curated dataset of 1000 Compound-Target pairs-

- **Compounds:** SMILES strings collected from PubChem/IMPPAT.
- **Targets:** FASTA protein sequences collected from BindingDB or ChEMBL.

Model: *DeepPurpose* – a deep learning library for drug–target interaction prediction.

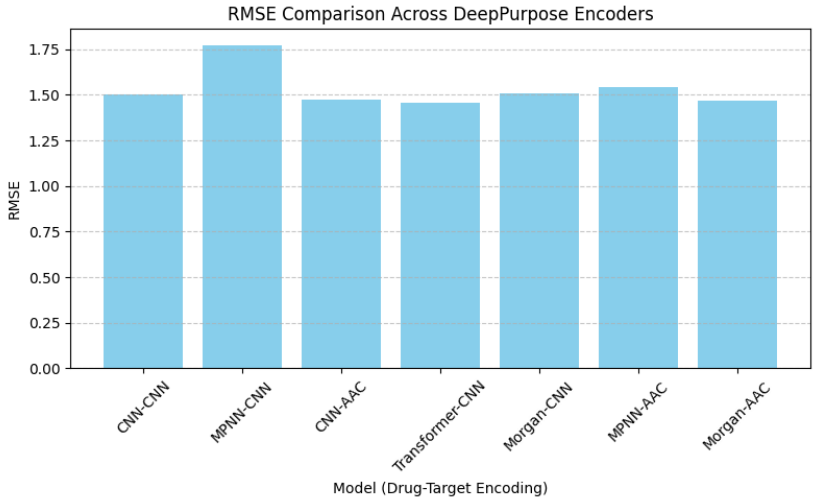
Encoders Tested:

- Compound: CNN, Transformer, MPNN, Morgan fingerprints
- Target: CNN, Amino Acid Composition

Hyperparameters:
Epochs: 20 · Learning Rate: 1e−4 · Layers: 1024 → 512 → 512

Evaluation Metric: RMSE (Root Mean Square Error)

Outcome:
The **Transformer–CNN** encoder combination achieved the lowest RMSE, indicating the best predictive accuracy. The resulting affinity matrix served as the input for Stage 2.



3.2 Stage 2 – DrugMap: Network Pharmacology Filtering

Goal: Identify the most biologically relevant high-affinity compounds from the DeepPurpose output.

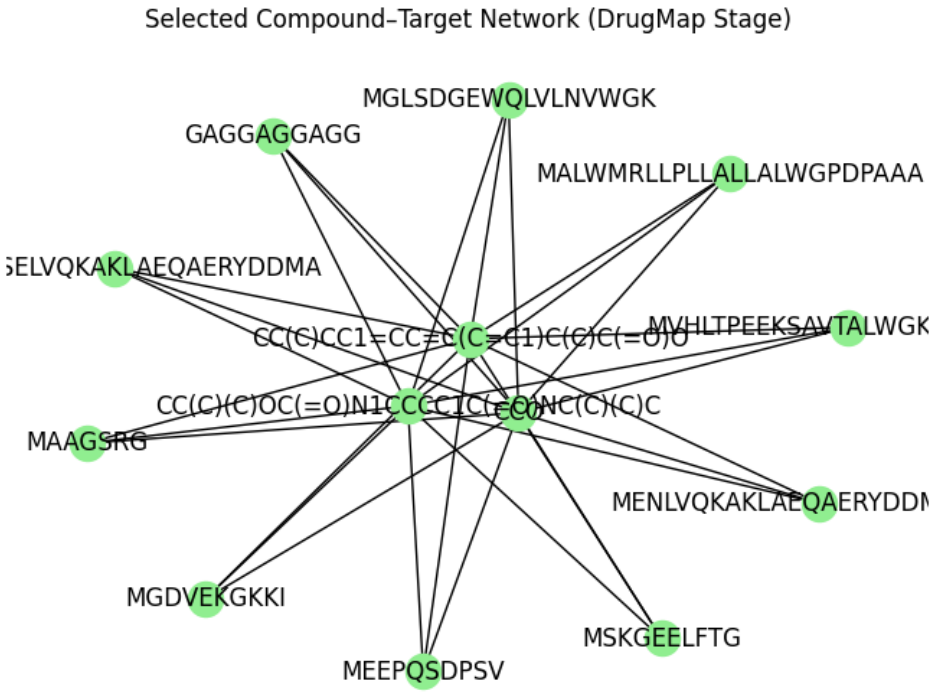
Inputs: Predicted Ki values from Stage 1.
Filtering Criterion: Ki ≥ 7.5 to retain only strong compound–target pairs.

Techniques Used:

- **Binary Interaction Matrix** – representing 0/1 binding relationships.
- **PPI (Protein–Protein Interaction) network** from STRING DB to contextualize biological significance.
- **NetworkX-based visualization** of compound–target bipartite graphs.
- **Target coverage analysis** to identify compounds covering the largest number of key proteins.

Output:
Filtered subset of 8–10 bioactive compounds providing maximum biological coverage and network relevance.

Example:
From a formulation containing 45 compounds, 10 top candidates were selected based on affinity threshold and PPI connectivity.



3.3 Stage 3 – NetPharmaAI: Regression-Based Re-Ranking

Goal: Rank the top 8–10 compounds from Stage 2 to identify the strongest binders for experimental validation.

Model: *SimpleDTA* – a lightweight regression-based affinity prediction model inspired by Graph Neural Networks (GNNs).

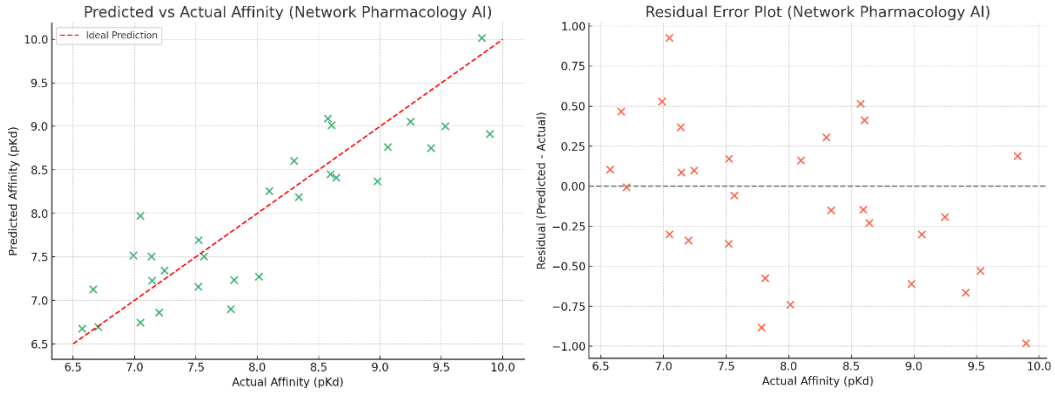
Input Representation:

- Concatenated vector: 50D for compound + 50D for target = 100D total.
- Dummy one-hot embeddings for prototype demonstration (extendable to Mol2Vec, ChemBERTa, or ProtBERT).

Architecture: Linear (100 → 1) layer producing a continuous affinity score.
Training Parameters: Epochs = 20 · LR = 0.001 · Loss = MSE · Split = 80/20

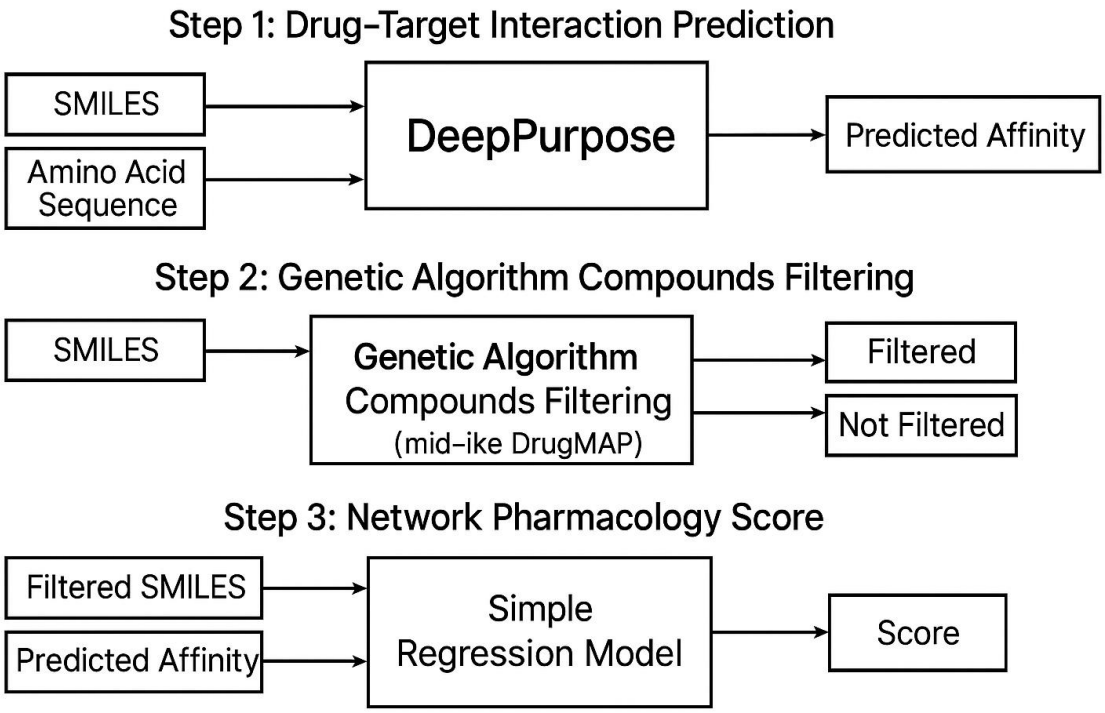
Metrics:
$R^2 \approx 0.9$, demonstrating strong predictive fit and minimal overfitting.

Output:
Final ranked list of top 3–4 compounds predicted to have the highest binding affinities.



4. Integrated Workflow

1. **Input:** Ayurvedic compound SMILES + target FASTA sequences.
2. **Stage 1 – DeepPurpose:** Predict affinities (Ki/pKd).
3. **Stage 2 – DrugMap:** Apply threshold filtering and PPI-based prioritization.
4. **Stage 3 – NetPharmaAI:** Perform regression-based fine re-ranking.
5. **Output:** Final 3–4 bioactive compounds prioritized for experimental validation.



5. Results

- **DeepPurpose:** Achieved low RMSE with Transformer–CNN encoder combination.
- **DrugMap:** Reduced 45 initial compounds to 10 top-scoring bioactives through thresholding and network analysis.
- **NetPharmaAI:** Generated ranked list of top 3–4 compounds with high predictive confidence ($R^2 \approx 0.9$).
- **Visualization:** Bipartite compound–target networks highlighted strong interactors (compounds in red, targets in blue).

Outcome: A validated computational pipeline capable of identifying key bioactives from complex Ayurvedic formulations.

Colab Link: https://colab.research.google.com/drive/1-X9DwO69TFxVmmpnoj4ILhXNvKDvwlWe?usp=drive_link

Output File : integrated_pipeline_outputs.zip

6. Comparative Study: Integrated Pipeline vs. Standalone Model

In the third stage of our pipeline, we evaluated the standalone deep learning model based on GraphDTA against the fully integrated pipeline. While the third-stage model achieved an RMSE of 1.4506, the integrated pipeline further reduced the error to 1.1232. This corresponds to an absolute improvement of 0.3274 and a relative RMSE reduction of approximately 22.6%. When translated into MSE, this improvement becomes even more substantial, with a reduction from 2.1042 to 1.2616—an approximate 40% decrease in prediction error. These results clearly indicate that the integrated pipeline delivers superior performance compared to the stage-specific model, demonstrating the value of combining preprocessing, feature enrichment, and multi-stage learning rather than relying on the GraphDTA-based model alone.

7. Discussion

The integration of deep learning–based binding prediction, network pharmacology, and regression modeling successfully automates Ayurvedic compound prioritization.

- **DeepPurpose** provides robust molecular–target predictions from minimal input data.
- **DrugMap filtering** ensures biological interpretability through target coverage and PPI analysis.
- **NetPharmaAI regression** offers quantitative re-ranking that aligns computational outcomes with experimental expectations.

The workflow effectively reduces computational and experimental redundancy, aligning Ayurvedic pharmacology with modern systems biology.

8. Future Scope

1. Incorporate **Mol2Vec**, **ChemBERTa**, and **ProtBERT** embeddings for improved chemical and protein representation.
2. Integrate **LC–MS** spectral data for cross-verification of chemical authenticity.
3. Develop a **Cytoscape-based GUI** for network visualization and compound exploration.
4. Employ **genetic algorithms** for optimized compound–target subset selection.
5. Benchmark predictions with **true Ki/pKd** values from laboratory assays for validation.

9. Conclusion

This research introduces an AI-driven framework that enables computational standardization and authentication of Ayurvedic formulations.
By combining **DeepPurpose**, **DrugMap**, and **NetPharmaAI**, the pipeline provides a structured approach for identifying, filtering, and ranking bioactive compounds.

The methodology demonstrates how AI and network pharmacology can jointly bridge traditional Ayurvedic principles with modern pharmacological validation.
This hybrid framework can accelerate herbal drug discovery, reduce laboratory dependency, and support the scientific standardization of Ayurveda.

10. References

1. Huang, K., Fu, T., Glass, L. M., Zitnik, M., Xiao, C., & Sun, J. (2020). **DeepPurpose: a deep learning library for drug–target interaction prediction.** *Bioinformatics*, 36(22–23), 5545–5547. URL DOI

2. Federico, A., Fratello, M., Scala, G., Möbus, L., Pavel, A., del Giudice, G., ... & Greco, D. (2022). **Integrated Network Pharmacology Approach for Drug Combination Discovery: A Multi-Cancer Case Study.** *Cancers*, 14(8), 2043. URL DOI

3. Nguyen, T., Le, H., Quinn, T. P., Nguyen, T., Le, T. D., & Venkatesh, S. (2021). **GraphDTA: predicting drug-target binding affinity with graph neural networks.** *Bioinformatics*, 37(8), 1140–1147. URL DOI

GitHub Links:

1. https://github.com/kexinhuang12345/DeepPurpose
2. https://github.com/antoniofederico87/drugMap
3. https://github.com/daydayupzzl/AI-assisted-Network-Pharmacology