GOVT SKSJ TECHNOLOGICAL INSTITUTE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Mini Project presentation on

"DRUG RESPONSE PREDICTION USING GENOMIC DATA"

Under the guidance of

Prof. Asha V

Assistant Professor, Dept of CSE

Presented by,

HANUMAGOUDA MALIPATIL(1SK23CS017)

TEJAS C(1SK23CS050)

TILAK KH (1SK23CS052)

URVASHI TANWAR(1SK23CS055)

OVERVIEW

- Introduction
- Problem Statement
- Objective
- Features
- Tools and Technology
- Process Flow
- User Interface Preview
- Flow Chart
- Literature Survey
- Expected Outcome
- References

INTRODUCTION

Personalized medicine is an emerging approach in healthcare that **customizes** medical treatment based on individual **patient profiles**. Advances in genomics and machine learning provide opportunities to analyse how genetic mutations, gene expressions, and clinical features like BMI influence drug efficacy.

This project proposes a system that integrates patient genomic data, clinical parameters, and drug molecular features to predict drug responses, providing actionable insights for healthcare providers.

PROBLEM STATEMENT

Traditional drug prescriptions often fail to account for genetic variability and individual patient differences, leading to ineffective treatments or adverse drug reactions. Current systems lack reliable predictive frameworks that integrate genomic and clinical data guide drug selection. There is a need for a robust AI/ML-based system that can analyze patient-specific features and predict drug responses with accuracy and interpretability.

OBJECTIVES

OBJECTIVE 1

To develop an AI model that predicts drug response using patient genomic and clinical data.

OBJECTIVE 2

To validate the model for accuracy in forecasting efficacy and adverse effects.

OBJECTIVE 3

To support personalized medicine by guiding safer and more effective prescriptions

FEATURES

1. Integration of patient genomic and clinical data for analysis.
2. Drug feature encoding using **molecular fingerprints.**
3. Supervised learning models for drug response prediction.
4. **Dual outputs:** IC50 value prediction (regression) or responder/non-responder classification.
5. Interpretability layer using **SHAP or LIME** for explainable AI results.
6. Potential extension into clinical decision-support systems.

TOOLS AND TECHNOLOGY

1. Front-End (User Interface) :

React.js, HTML, CSS, JavaScript

Features on UI:

input patient genomic and clinical details, upload drug information (SMILES string or structure), display prediction results (responder or non-responder), and show visualizations such as graphs and SHAP plots to understand .

2. Backend:

Programming Language: Python Frameworks: Scikit-learn, TensorFlow/PyTorch, XGBoost

3. Dataset:

GDSC for drug response, CCLE for cell line, Index database

4. AI/ML Layer

Languages and Frameworks: Python, scikit-learn (for models like RandomForest and XGBoost),

LightGBM, TensorFlow,RDKit (for drug feature extraction), and SHAP (for interpretability)

PROCESS FLOW

1

Data Collection: Collection of patient data from genomic and clinical databases such as GDSC, TCGA, or CCLE.

2

Encode genomic features as binary and drug structures into numeric vectors, merging them into a single feature matrix.

3

Model training: Using supervised learning algorithm (e.g,XGBoost, RandomForest, LR, GBM).

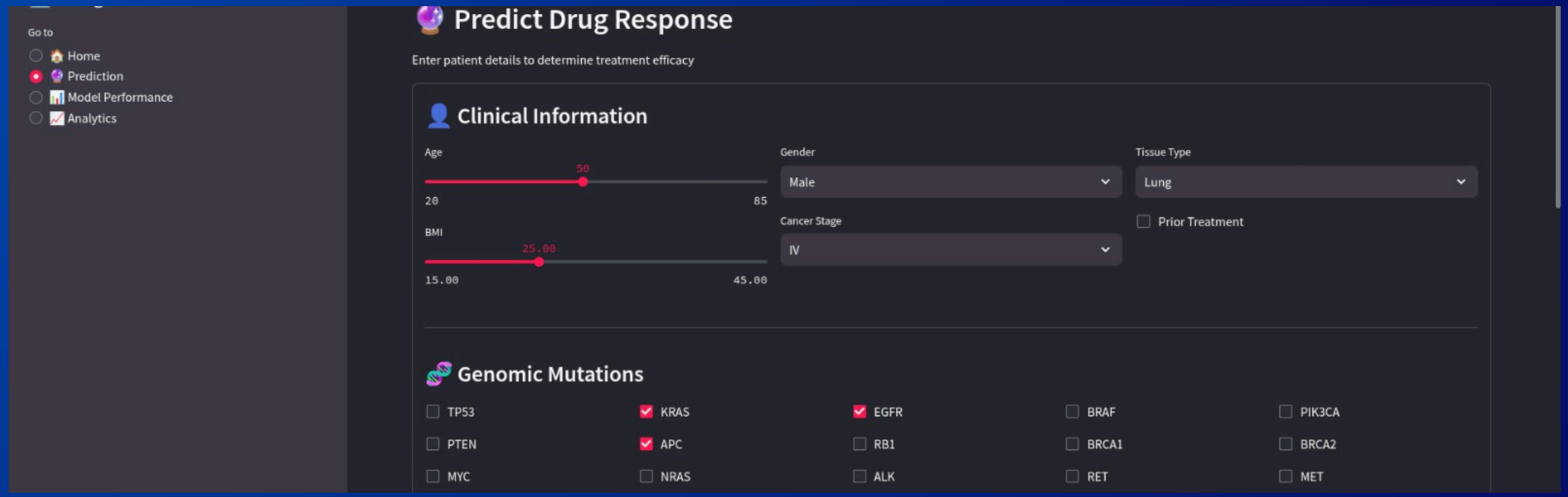
4

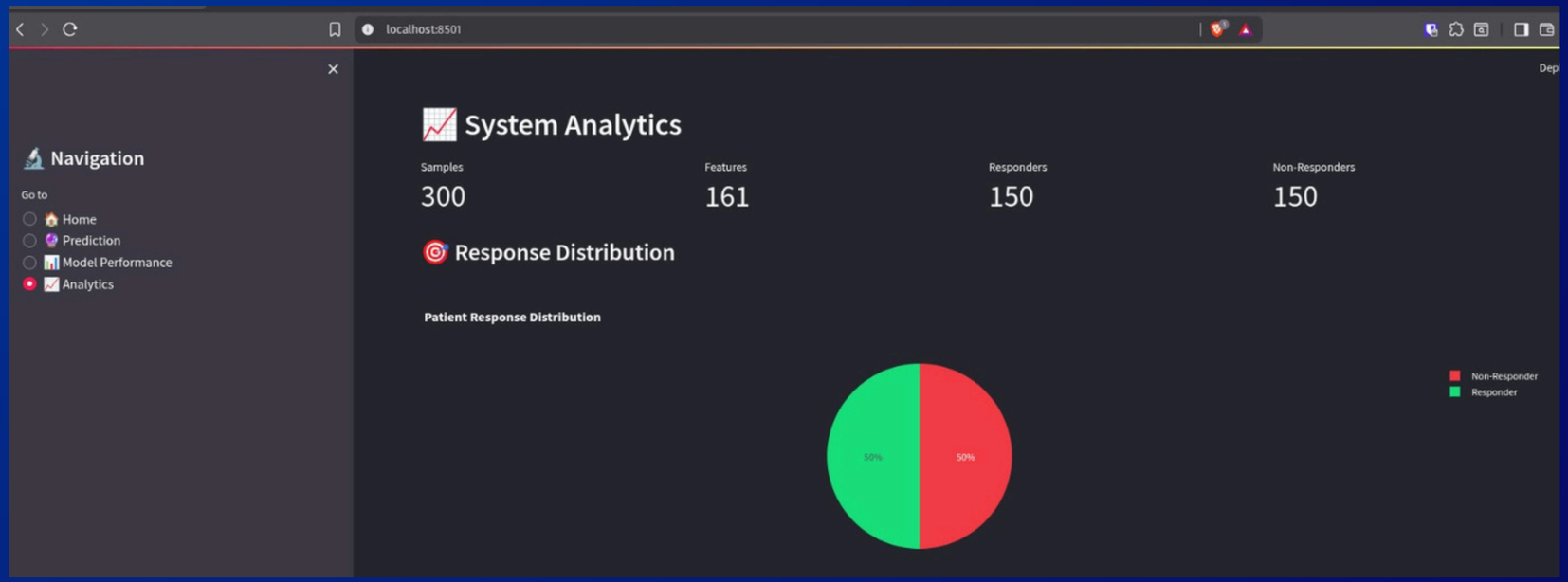
Model has to learn mapping: f(X) gives Y where y=1(responder) or 0(nonresponder).

5

Genomic Data + Clinical Data + Drug response= Predicted Response

USER INTERFACE PREVIEW





LITERATURE SURVEY

Sl. No.	Title / Source	Author(s) / Year	Methods / Algorithms Used	Key Findings	Drawbacks / Limitations
1	Machine learning models for predicting drug response based on pharmacogenomic data – PMC6656435	Costello et al., 2019	Random Forest, Support Vector Machine (SVM), Elastic Net Regression	Machine learning algorithms were applied to genomic and drug features to predict patient drug response with good accuracy.	Model interpretability was limited and dataset imbalance affected generalization.
2	Deep learning-based drug response prediction using multi-omics data – PMC10950891	Zhang et al., 2023	Deep Neural Networks (DNN), Autoencoders	Integrating genomic and transcriptomic data improved overall drug response prediction accuracy.	High computational cost and limited biological interpretability.
3	PharmGKB: Pharmacogenomics Knowledge Base	Whirl-Carrillo et al., 2021	Knowledge-based data curation and association mapping	Provides comprehensive data linking genetic variations with observed drug responses.	Serves as a data source; does not perform predictive modeling.
4	GDSC and CCLE datasets for cancer drug sensitivity prediction	Yang et al., 2013	Correlation-based analysis, regression, clustering	Developed large-scale genomic datasets for modeling drug sensitivity in cancer cell lines.	Heterogeneity and inconsistency between datasets limited performance.
5	Explainable AI for drug response prediction	Chen et al., 2022	XGBoost with SHAP interpretability framework	Demonstrated interpretable ML models that identify genomic features most relevant to drug response	Scalability remains limited when applied to multiple drugs or large datasets.

REFERENCES

[1] Y. Wang, L. Chen, and H. Zhang, “An explainable graph neural network for predicting cancer drug sensitivity (XGraphCDS),” IEEE Transactions on Computational Biology and Bioinformatics, vol. 21, no. 3, pp. 874–885, Mar. 2024.

[2] Y. Zhou, M. Liu, and Q. Zhao, “PharmaFormer: Predicting clinical drug responses through transformer-based architectures,” IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 22, no. 1, pp. 45–56, Jan. 2025.

[3] Y. A. Qadri, “Explainable AI: A perspective on drug response prediction,” IEEE Access, vol. 13, pp. 122345–122359, 2025.

ANY QUESTIONS?