GOVT SKSJ TECHNOLOGICAL INSTITUTE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Mini Project presentation on

"DRUG RESPONSE PREDICTION USING GENOMIC & CLINICAL 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

1. Introduction
2. Problem Statement
3. Literature summary
4. System Architecture
5. Input/Output Flow
6. Methodology
7. Screenshots
8. Future Scope
9. Key Takeaways
10. References

INTRODUCTION

Precision medicine is revolutionizing healthcare by tailoring treatment strategies to individual patient characteristics. Leveraging breakthroughs in genomics and artificial intelligence, we can now decipher the complex interplay between genetic mutations, clinical factors, and drug responses.

This project presents an AI-driven system that seamlessly integrates genomic profiles, clinical data, and molecular fingerprints to predict personalized treatment outcomes, empowering healthcare providers with data-driven insights to optimize therapeutic decisions and improve patient care.

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 integrating genomic and clinical data to guide drug selection, highlighting the need for a robust AI/ML-based system that analyzes patient-specific features to predict drug responses with accuracy and interpretability.

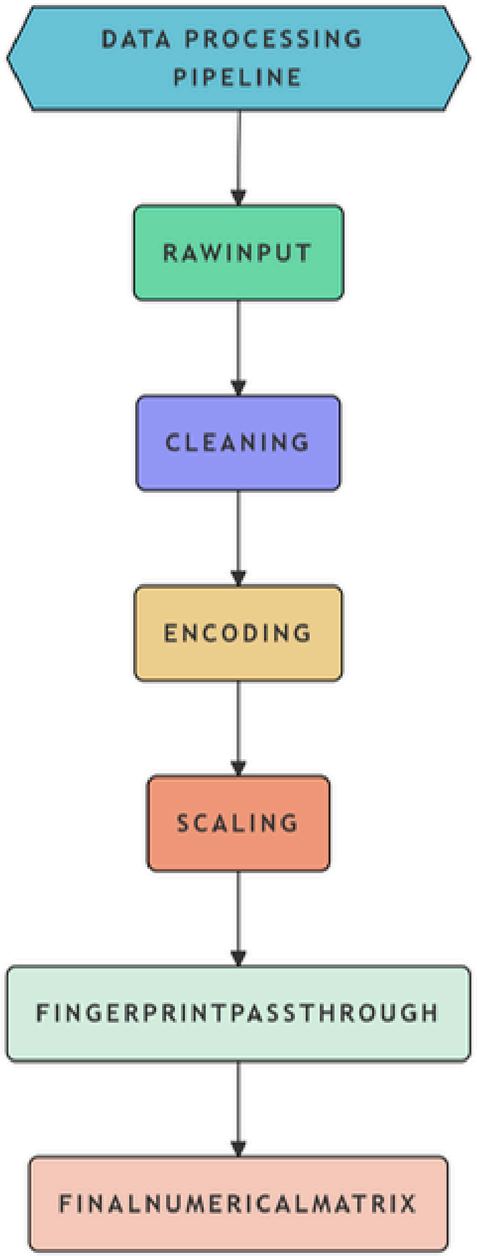
LITERATURE SUMMARY

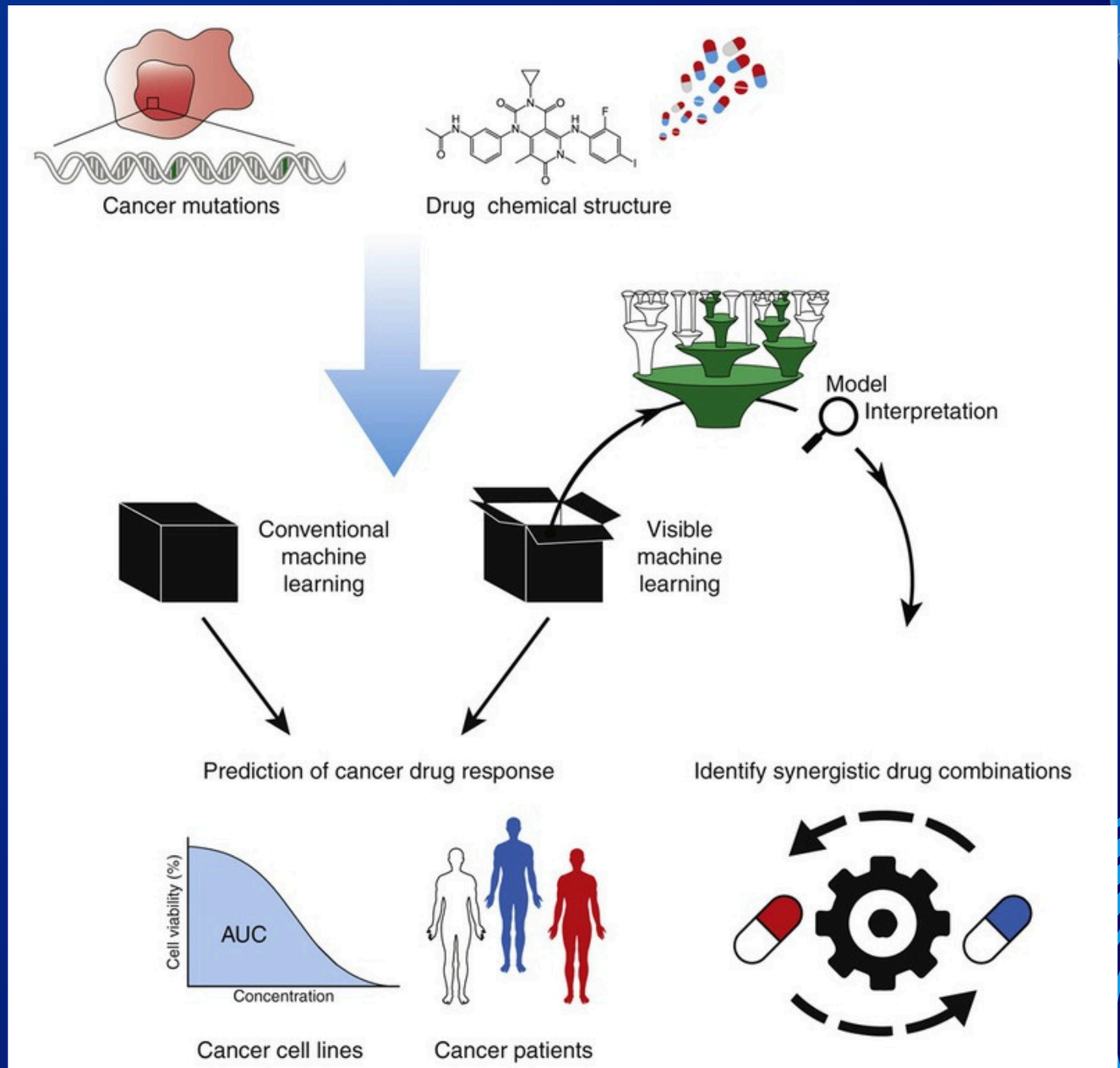
1. **Explainable GNN models (Wang et al., 2024)**
- Use graph neural networks to model gene–drug interactions.
- Provide pathway-level interpretability but are computationally intensive.
2. **MMDRP framework (Taj et al., 2024)**
- Integrates biomarker discovery with drug response prediction in a single pipeline.
- Risk of overfitting on small datasets; mixed external validation results.
3. **PharmaFormer (Zhou et al., 2025)**
- Transformer-based model trained on RNA-seq tumor data.
- Improves patient-level prediction but limited by scarce clinical datasets.
4. **PASO multimodal transformer (Wu et al., 2025)**
- Fuses multi-omics and drug features using multi-scale attention.
- High performance, but requires large compute and careful tuning.
5. **Explainable AI review (Qadri, 2025)**
- Highlights importance of interpretability for clinical and regulatory acceptance.
- Notes lack of empirical benchmarks for XAI in drug response prediction

SYSTEM ARCHITECTURE

1. Frontend / User Interface:
   - Python scripts with CLI-based interaction
   - Visualizations and plots using matplotlib and seaborn
   - Model interpretability via SHAP
2. Backend / ML Layer:
   - Python as main programming language
   - Machine Learning Libraries: scikit-learn, XGBoost, RandomForest, LightBGM
   - Data Handling: pandas, numpy
   - Drug Feature Extraction: RDKit
3. Datasets:
   - GDSC – Genomics of Drug Sensitivity in Cancer
   - CCLE – Cancer Cell Line Encyclopedia
   - Index Database – Additional reference datasets

INPUT & OUTPUT FLOW





METHODOLOGY

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

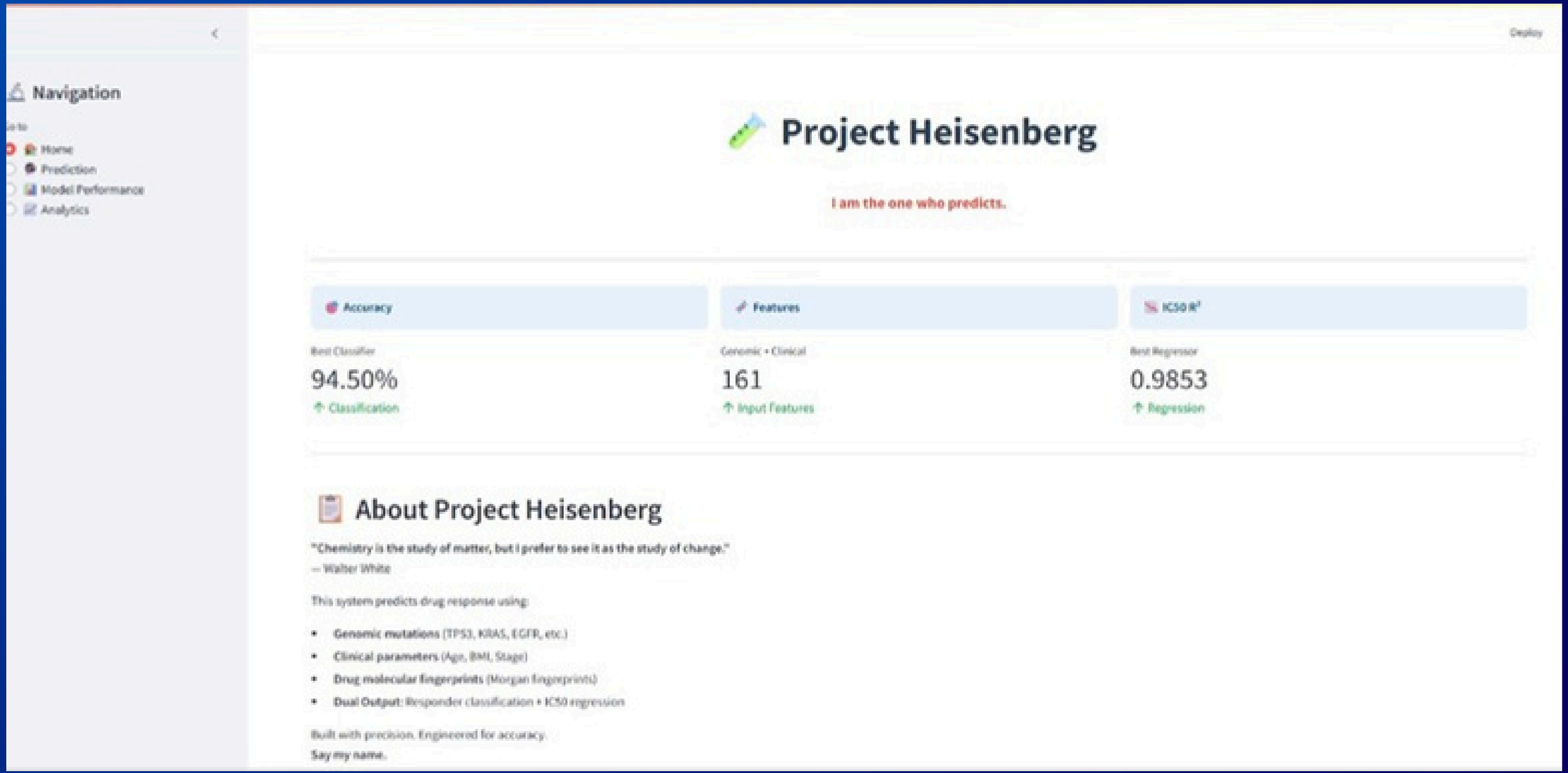
2 Encode genomic features as binary and drug structures into numeric vectors, merging them into a single feature matrix.

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

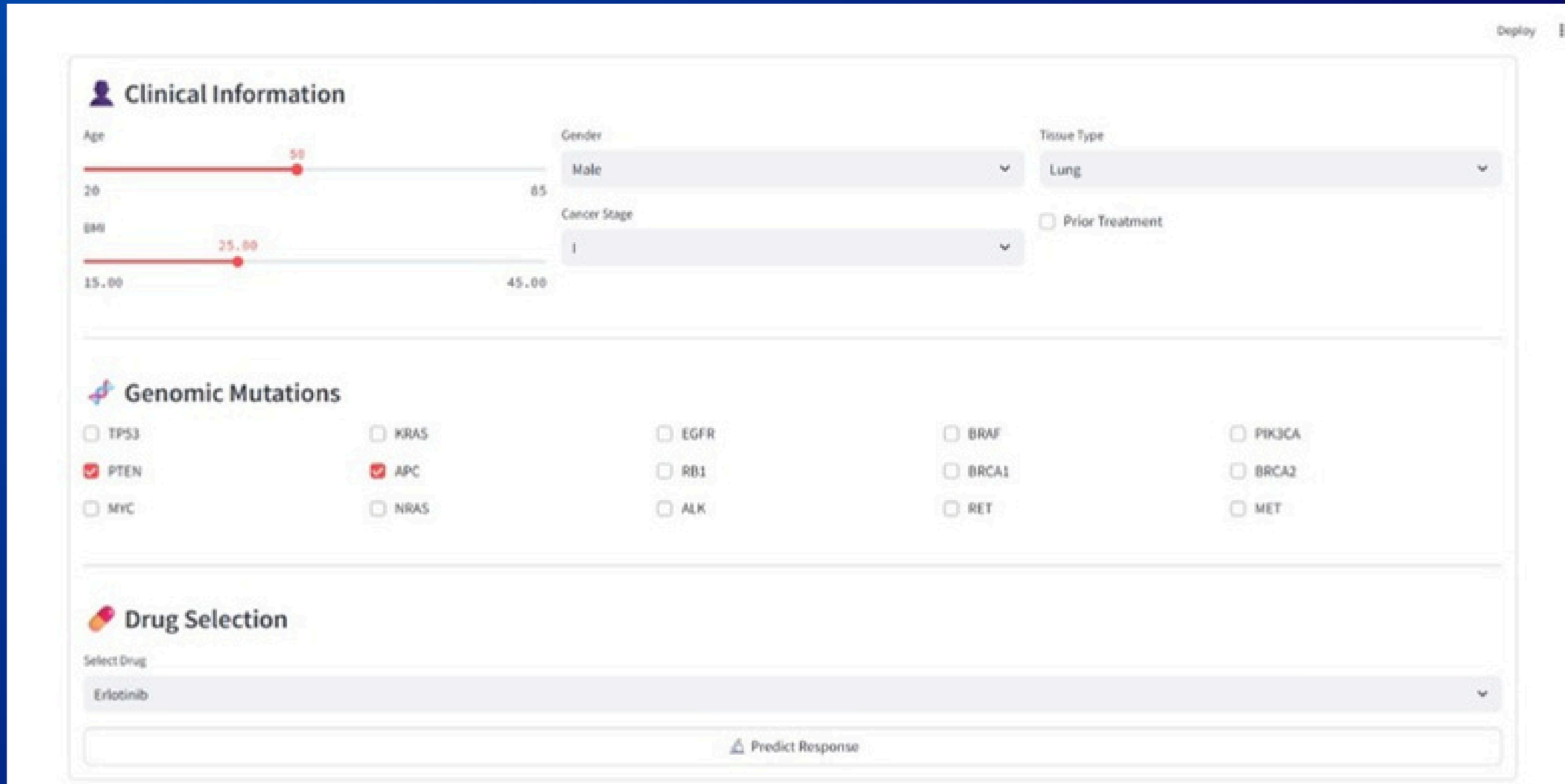
4 Model has to learn mapping: f(X) gives Y where y=1(responder) or 0(nonresponder).

5 Genomic Data + Clinical Data + Drug structure= Predicted Response

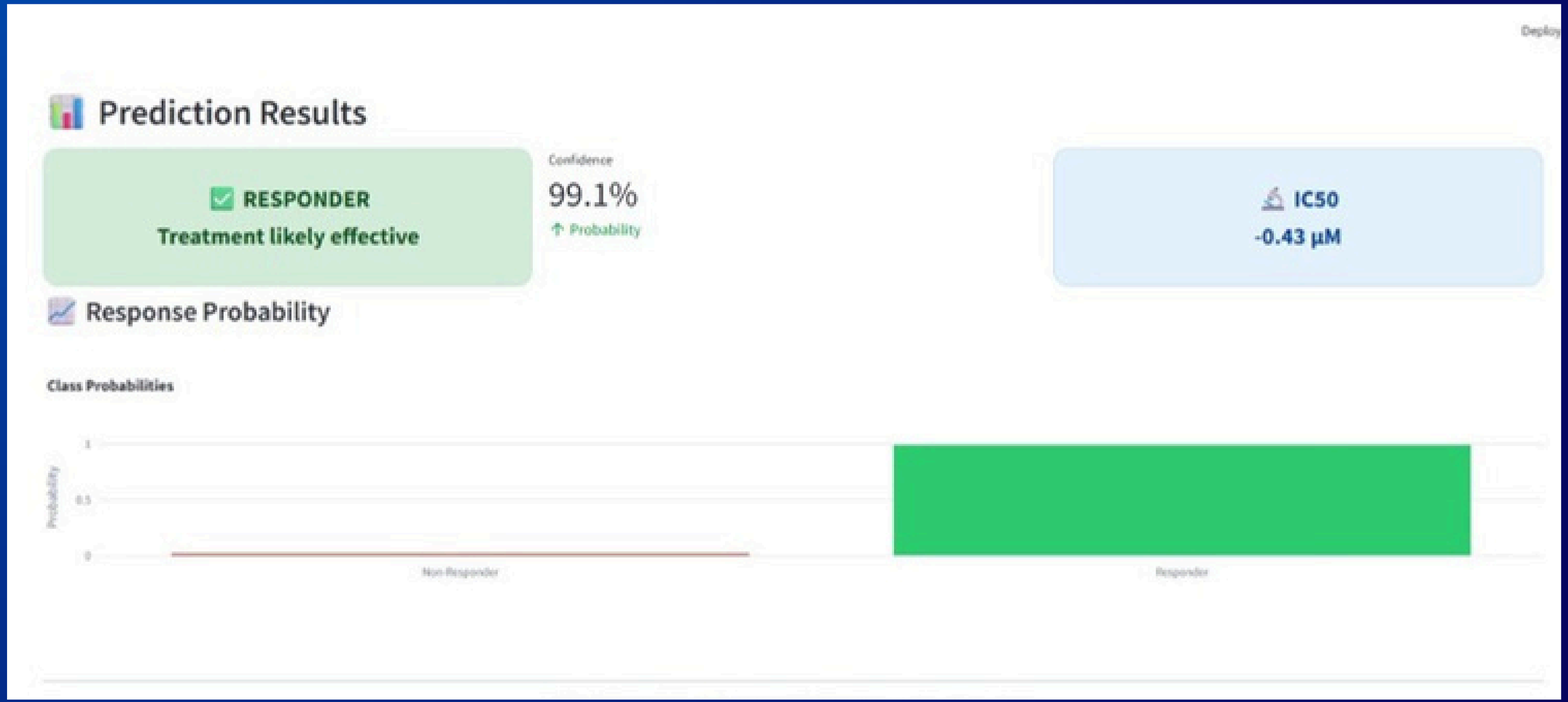
SCREENSHOTS



SCREENSHOTS



SCREENSHOTS



SCREENSHOTS

Model Performance – Precision Engineered

Classification Models

	Model	Accuracy	ROC-AUC	Time(s)
0	XGBoost_Class	0.9450	0.9876	2.08
1	LightGBM_Class	0.9390	0.9871	0.51
2	RandomForest_Class	0.8860	0.9651	0.43

Regression Models (IC50 Prediction)

	Model	R2	MSE	Time(s)
0	XGBoost_Reg	0.9853	0.1846	0.44
1	LightGBM_Reg	0.9826	0.2189	0.29
2	RandomForest_Reg	0.9815	0.2326	0.74

"I did it for me. I liked it. I was good at it. And I was really... I was alive."

— Walter White, Breaking Bad

SCREENSHOTS

Project Heisenberg

I am the one who predicts.

System Analytics

Samples	Features	Responders	Non-Responders
1000	161	500	500

Response Distribution

Patient Response Distribution

FUTURE SCOPE

1. **Multi-drug combination** response prediction
2. Real-time clinical decision support system
3. Enhanced **explainability** (SHAP, LIME, gene–drug interactions)
4. Inclusion of **multi-omics** data
5. Cloud deployment for scalability
6. **Continuous learning** from real patient outcomes
7. Integration with drug discovery pipelines
8. Movement toward **clinical trials** & regulatory approvals

KEY TAKEAWAYS

This project shows how machine learning can be used to predict drug response based on genomic and molecular data, enabling progress toward personalized cancer treatment. Using datasets like GDSC and CCLE, the system extracts meaningful features, trains models such as XGBoost and neural networks, and provides interpretable results using tools like SHAP. The final interface allows users to input data and view predictions clearly, demonstrating a practical step toward AI-assisted decision making in precision medicine.

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?