AMRITA VISHWA VIDYAPEETHAM
DEEMED TO BE UNIVERSITY

Lung Cancer and Drug Response Prediction

Molecular biology & basic cellular physiology
Ethics, innovative research, businesses & IPR

Guru Jaya Surya Yadav CB.AI.U4AIM24101.
J. Tej Krishna Sai CB.AI.U4AIM24117.
P. Teja Prakash Royal CB.AI.U4AIM24136.
S. Ankith CB.AI.U4AIM24147.

Guided by:
Prof . Kelath Murali Manoj
Dr. Neelesh Ashok
Reshma Sanal mam

INTRODUCTION

Lung cancer is one of the leading causes of cancer-related deaths worldwide.

Early detection and personalized treatment are crucial for improving survival rates.

Machine Learning (ML) and Deep Learning (DL) can assist in both prediction and treatment response.

Objective: Predict lung cancer risk and drug response using ML/DL, and deploy via a user-friendly UI.

PROBLEM STATEMENT

Early detection of lung cancer remains a challenge.

Patients show varied responses to chemotherapy and other drugs.

Manual analysis of large patient data is time-consuming.

AI can bridge this gap by offering rapid, accurate predictions.

BIOLOGICAL SIGNIFICANCE

- Lung Cancer Risk Factors: Smoking, yellow fingers, anxiety, chronic disease, etc.
- Drug Response: Varies with patient genetics and health history.
- AI models help identify these risks and responses early.
- Supports personalized treatment like precision medicine.

OBJECTIVES

Built ML Model: Trained multiple models to predict drug response.

Data Analysis & Visualization: Explored key features and patterns.

Model Comparison & Selection: To identify the best model for drug response.

Building a model for lung cancer prediction from patients data

DATASET USED

Lung Cancer Dataset:

- Features: Age, Smoking, Yellow Fingers, Anxiety, Chronic Disease, etc.

- Label: Lung cancer presence

Drug Response Dataset:

- Features: Patient features, drug name

- Label: Response (Sensitive/Resistant)

Preprocessing:

- Missing value handling, encoding, normalization

ML/DL MODELS APPLIED

Lung Cancer Prediction:

- Logistic Regression, Random Forest, XGBoost, DNN

Drug Response Prediction:

- Decision Tree, SVM, XGBoost, CNN

Used grid search and cross-validation to optimize models.

MODEL EVALUATION

Evaluation Metrics:

Model Comparison:

	Model	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	0.685	0.642105	0.677778	0.659459
1	Random Forest	0.880	0.823529	0.933333	0.875000
2	SVM	0.900	0.850000	0.944444	0.894737
3	KNN	0.900	0.864583	0.922222	0.892473
4	Gradient Boosting	0.850	0.819149	0.855556	0.836957

Drug Response model

Comparing ML Models:

	Model	Accuracy	Precision	Recall	F1 Score
0	Random Forest	0.967742	0.983333	0.983333	0.983333
1	Logistic Regression	0.967742	0.983333	0.983333	0.983333
2	SVM	0.967742	0.967742	1.000000	0.983607
3	Decision Tree	0.967742	0.983333	0.983333	0.983333
4	Naive Bayes	0.951613	0.983051	0.966667	0.974790

Lung cancer model

Best Models:

Lung Cancer: Random Forest (Accuracy: 98%)

Drug Response: KNN (Accuracy: 90%)

UI DEVELOPMENT

- Built using Gradio (Python UI Library)
- Inputs: Age, Smoking, Drug Name, etc.
- Outputs: Lung Cancer: Risk Prediction (Yes/No)
- Drug Response: Predicted Response (Sensitive/Resistant)
- Integrated the best-performing models.

SYSTEM ARCHITECTURE

Input → Preprocessing → Trained Model (ML/DL) → Prediction → UI Output

Saved trained models using joblib

Connected models to Gradio app for real-time prediction

RESULTS & OBSERVATIONS

- Best-performing models delivered over 90% accuracy
- Smoking and age were key predictors for lung cancer
- Personalized prediction possible via trained drug model
- Real-time app for usage by clinicians

CONCLUSION & FUTURE WORK

Achieved dual-prediction system for diagnosis and treatment.

Built a working prototype UI using Gradio.

Future Plans:

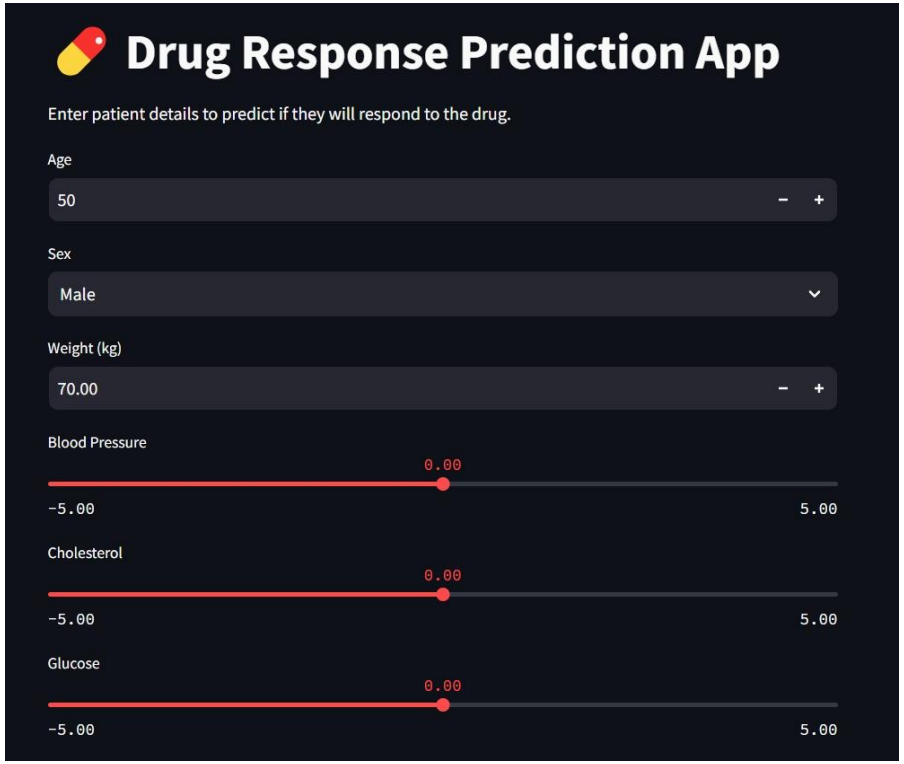
- Increase dataset size

- Use real-time clinical data via APIs

- Deploy on web for public use

UI SCREENSHOT

Streamlit app snapshot



ETHICAL & IPR CONSIDERATIONS

- Patient Data Privacy: All data anonymized for ethical use.
- AI for Assistance: Our model supports doctors, doesn't replace them.
- Fairness: Trained on diverse data to avoid bias.
- IPR: Potential for licensing or patenting the app in future.

CASE STUDY

REFERENCES

Kaggle Datasets: Lung Cancer & Drug Response

scikit-learn, XGBoost, TensorFlow

Gradio Documentation

Research articles on ML in healthcare