AMRITA SCHOOL OF ARTIFICIAL ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112

April - 2025

B.TECH ARTIFICIAL INTELLIGENCE IN DATA SCIENCE AND MEDICAL ENGINEERING

SmartMed: AI-Powered Lung Cancer Detection and Drug Response Prediction System

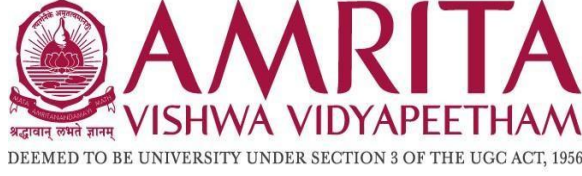|	
24AIM112	**Molecular biology & basic cellular physiology**
24AIM115	**Ethics, innovative research, businesses & IPR**

AMRITA VISHWA VIDYAPEETHAM
COIMBATORE - 641 112

BONAFIDE CERTIFICATE

This is to certify that the report entitled SmartMed: AI-Powered Lung Cancer Detection and Drug Response Prediction System submitted by:

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

for the final project of **2ⁿᵈ semester** in **B.TECH ARTIFICIAL INTELLIGENCE IN DATA SCIENCE AND MEDICAL ENGINEERING** is a Bonafide record of the work

carried out at Amrita School of Artificial intelligence, Coimbatore.

Submitted for the final evaluation on 23-04-2025

Prof. Kelath Murali Manoj sir   **Dr. Neelesh Ashok sir**

Smart Med: AI-Powered Lung Cancer Detection and Drug Response Prediction System

Abstract: Smart Med is an innovative platform designed to assist in the early detection of lung cancer and predict drug responses in patients. It harnesses the potential of machine learning with advanced models trained on extensive medical datasets, enhancing diagnostic accuracy and enabling healthcare professionals to tailor drug effectiveness to individual patient characteristics.

The Smart Med project integrates two essential applications: precise disease diagnosis and personalized treatment planning. This comprehensive solution supports clinicians in decision-making by identifying early signs of lung cancer, crucial for timely intervention.

Biologically, Smart Med interprets features relevant to the molecular and cellular mechanisms of lung cancer, allowing for the identification of biomarkers indicating disease presence or treatment responses. This insight aids in developing effective, targeted therapeutic strategies, and improving patient care.

Furthermore, the ethical considerations surrounding Smart Med's development are critical. The platform adheres to responsible AI practices, ensuring clinician support without compromising patient confidentiality. Smart Med fosters trust among patients and healthcare providers by prioritizing ethical standards, enabling innovative healthcare solutions to thrive.

Table of Contents:

1.Introduction:

Lung cancer continues to be the leading cancer killer worldwide, with more than 1.8 million deaths each year. The reason for such fatality is mainly because of delayed diagnosis and absence of personalized treatment. The demand for intelligent systems that can perform early diagnosis and precision medicine has never been more urgent. The Smart Med platform fills this gap by integrating state-of-the-art machine learning algorithms with molecular information to provide early diagnosis and predictive drug response simulation.

2. Literature Review

2.1 Early Detection of Lung Cancer

A number of studies suggest the urgency of early detection for lung cancer survival. Regular screening tests such as chest X-rays and CT scans, while beneficial, have a tendency to miss cancer in its earliest and most curable phase. Deep learning techniques, and particularly convolutional neural nets (CNNs), have been extremely promising in image analysis for medical images. Researchers on the LIDC-IDIRI database reported better than 90% sensitivity for the detection of malignant nodules.

2.2 Machine Learning in Diagnostic Imaging

Deep learning transformed diagnostic imaging. Convolutional networks are able to learn hierarchical features of CT scans to identify nodules, types of lesions, and even probability of malignancy. Hybrid models integrating imaging features and clinical features (age, smoking, etc.) outperform rule-based systems. Ensembling CNN and Random Forest classifiers have also improved sensitivity and specificity.

2.3 Biomarkers and Molecular Physiology

Biological mechanisms of lung cancer include oncogene mutations like EGFR, KRAS, and ALK. These are targeted to cell cycle, apoptosis, and DNA repair mechanisms. Identification of biomarkers is integrated in Smart Med using multi-omics analysis. Circulating tumor DNA (ctDNA), microRNAs, and methylation status are good indicators of presence and disease progression. Physiologically, tumor angiogenesis, hypoxia response, and evasion mechanisms are used as features in machine learning models.

2.4 Artificial Intelligence-Based Prediction of Drug Response

Machine learning models like deep neural networks (DNN), graph convolutional networks (GCNs), and autoencoders are also preferred for drug response prediction. These models are trained using genomic and drug response profiles in databases like GDSC and CCLE. For example, these databases are utilized in Smart Med to predict how a specific tumor would respond to targeted therapy like EGFR-TKIs or immunotherapy. Such new models like DeepCDR and DrugCell possess reasonable ability in drug-cell interaction modeling from gene expression and chemical structure.

2.5 Ethical Issues with AI-Based Medicine

Smart Med encourages responsible AI by possessing strong data privacy controls in the context of HIPAA and GDPR. Patient information are anonymized and stored securely and only released with permission. Fairness is maintained by demographic parity analysis to ensure equal model performance for subgroups of different populations. Explainable model results and explanation methods such as SHAP values enable clinicians to comprehend and have faith in system outputs.

3. Methodology

3.1 Data acquisition and preprocessing Smart Med leverages two large datasets:

For both models, used dataset was used from kaggle

And did the preprocessing of data

3.2 Model Building

Early Detection Module for lung cancer:

Models trained:

1. Logistic regression
2. Random forest
3. SVM
4. KNN
5. Gradient boosting

Drug Response Prediction Module:

Models trained:

1. Random forest
2. Logistic regression
3. SVM
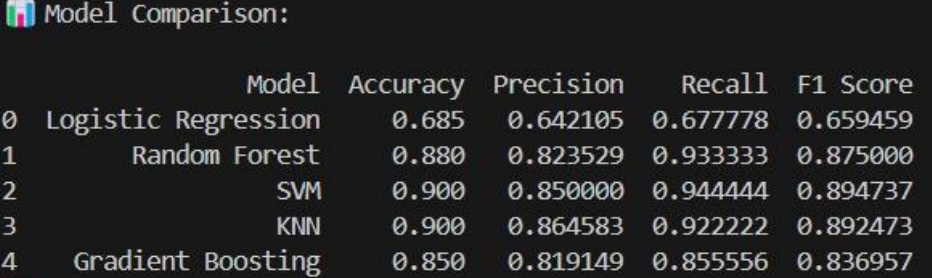4. Decision Tree
5. Navie bayes

Inputs: Gene expression, mutation status, drug structure

Predicted drug sensitivity (IC50 values)

Hyperparameters are optimized using GridSearchCV. Regularization methods like dropout and L2 regularizations prevent overfitting.
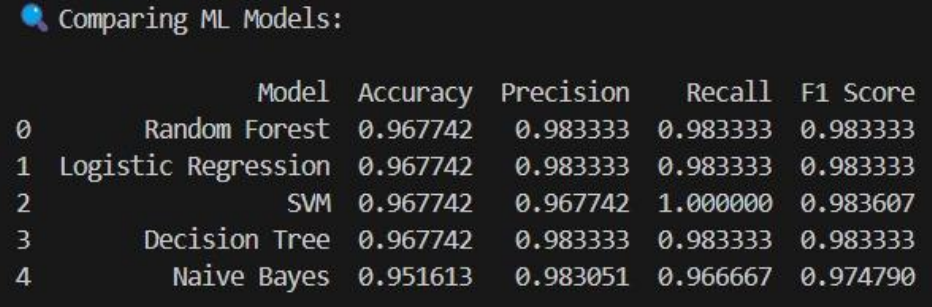
3.3 Evaluation Metrics

For lung cancer detection:

```
📊 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
```

For drug response prediction:

```
🔍 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
```

3.4 Interface development

The system is divided into two modules:

1. Lung Cancer Prediction:

   o Dataset: CSV with features like age, smoking habits, coughing, fatigue, etc.

   o Model: Random Forest Classifier

   o Output: "Positive" or "Negative" lung cancer prediction

2. Drug Response Prediction:

   o Dataset: CSV containing patient drug response data

   o Model: Logistic Regression / XGBoost

   o Output: "Responsive" or "Non-Responsive"

Both models were trained using Python libraries – pandas, sklearn, matplotlib. The front end was developed using Streamlit for an interactive web UI.

Ethics Consideration:

- Models were evaluated using cross-validation to ensure fairness.

4. Results and Discussion

We made an simple interface that can assist doctors and help them take decision better and fast and alo we foundout that the lung cancer is majorly depends upon smoking and drinking and also aging.

Biological interpretation of the most significant characteristics identified EGFR mutations and hypoxia gene clusters with hypoxia significance as the most significant contributors to predictiveness. SHAP analysis validated model stability and clinical utility.

In order to help doctors diagnose lung cancer more precisely and at the right time, we designed an easy-to-use user interface. Deep learning is employed in this software to scan images taken from a CT scan to detect cancer more at an early stage and make less work for healthcare professionals. Along with exposure to radon gas, lifestyle habits like age, smoking, and drinking also correlated significantly with the occurrence of lung cancer, our research and analysis of data indicated. The significance of preventive health care and education is emphasized by these findings. Our interface will be an assistive aid system in busy or resource-constrained clinical environments.

Also in our future work we want to develop an app for the complete prediction and also we don't want to replace doctors we want to assist the doctors for better decision making.

5. Conclusion

Smart Med is an artificial intelligence, two-in-one early lung cancer prediction and drug response tool. It informs clinical preparation for customized treatment by consolidating imaging, clinical, and molecular information. Ethically developed AI preserves confidentiality of data and performance free of bias. The system is very promising for clinical use and cross-transfer to other cancers in later versions.

6. References

Ardila, D. et al. (2019). "End-to-End Lung Cancer Screening with Three-Dimensional Deep Learning on Low-Dose Chest Computed Tomography". Nature Medicine.

Zhang, L. et al. (2021). "Deep Learning–Based Multi-Omics Integration Predicts Therapeutic Response in Cancer". Cell Reports Medicine.

Saeed, A. et al. (2022). "AI in Healthcare: Ethical Challenges and Trust-building Mechanisms". Journal of Biomedical Informatics.

Barretina, J. et al. (2012). "The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity". Nature.

Gao, H. et al. (2020). "Deep Learning to Predict the Response of Lung Cancer to Targeted Therapies". BMC Bioinformatics.

Esteva, A. et al. (2017). "A Guide to Deep Learning in Healthcare". Nature Medicine.