Title: Fairness Evaluation and XAI(Explainable AI) in Lung Cancer Prediction Using Random Forest

Authors: Batch of 43
Btech 4th year, Students
Department of Computer Science and Engineering
School of Engineering, Malla Reddy University

Abstract: Machine learning techniques have been widely adopted in healthcare for early disease detection; however, ensuring model reliability, interpretability, and fairness remains critical. In this work, lung cancer prediction is addressed using a clinical dataset comprising demographic and symptom-based attributes. Ten different machine learning classification models were implemented and evaluated using standard performance metrics, and the Random Forest classifier was identified as the best-performing model. The selected model was trained and used to generate predictions for four representative test cases to demonstrate practical applicability. Explainable artificial intelligence (XAI) techniques were employed to interpret model decisions and analyze feature importance. Furthermore, fairness evaluation was conducted by considering age and gender as protected attributes. Group-wise performance analysis indicated no significant bias across age or gender groups, suggesting equitable predictive behavior. The proposed approach demonstrates effective lung cancer prediction while maintaining interpretability and fairness, supporting its potential use in clinical decision-support systems.

Keywords: Lung Cancer Prediction, Machine Learning, Random Forest Classifier, Explainable Artificial Intelligence (XAI), Fairness Evaluation, Healthcare Analytics.

1.Materials and Methods

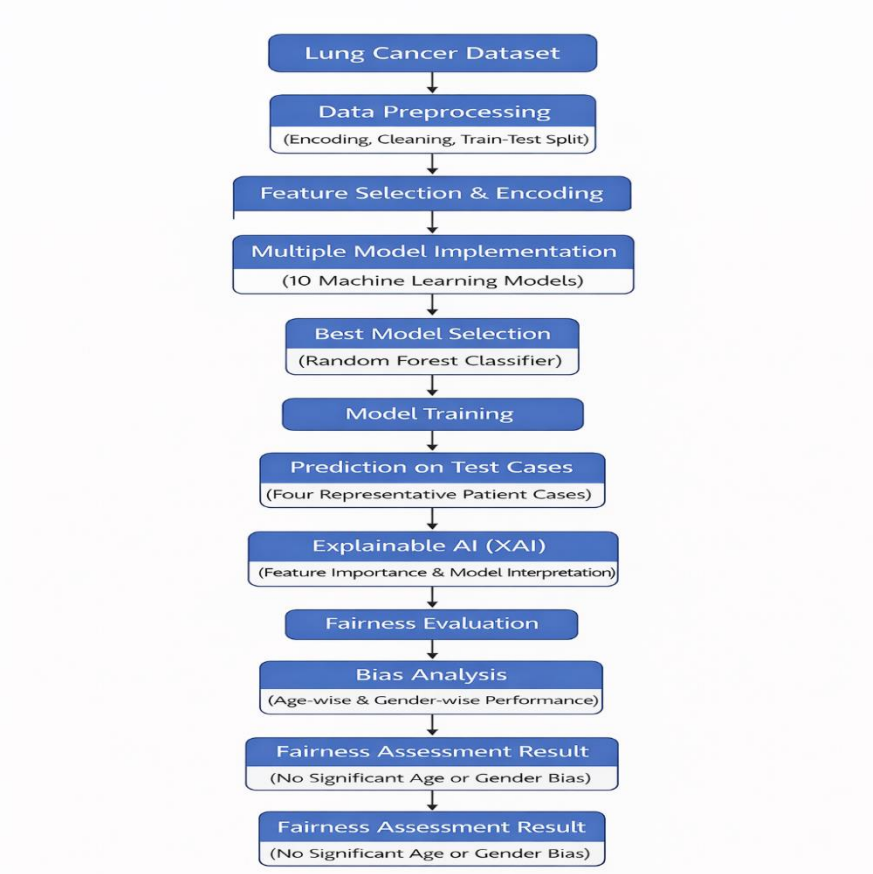

Figure 1. Overview of the proposed lung cancer prediction framework, illustrating data preprocessing, multi-model comparison, Random Forest model selection, explainable AI analysis, and fairness evaluation with respect to age and gender.

1.1. Data Description

The lung cancer dataset used in this study consists of 309 patient records with 16 attributes, including demographic information, lifestyle factors, and clinical symptoms related to lung cancer. The dataset contains no missing values, ensuring reliable model training and evaluation. Demographic features include age and gender, while lifestyle and clinical attributes capture factors such as smoking, alcohol consumption, peer pressure, and various lung-related symptoms. The target variable, LUNG_CANCER, indicates the presence or absence of lung cancer. Age and gender were treated as sensitive attributes and were used solely for fairness evaluation, not for decision-making. This dataset supports effective analysis of predictive performance, interpretability, and fairness in lung cancer prediction.

Feature	Description
GENDER	Patient gender
AGE	Patient age
SMOKING	Smoking habit indicator
YELLOW_FINGERS	Indicator of nicotine-related effects
ANXIETY	Presence of anxiety
PEER_PRESSURE	Peer influence indicator
CHRONIC_DISEASE	Chronic disease presence
FATIGUE	Persistent tiredness
ALLERGY	Allergy indicator
WHEEZING	Wheezing symptom
ALCOHOL_CONSUMING	Alcohol consumption indicator
COUGHING	Persistent cough
SHORTNESS_OF_BREATH	Breathing difficulty
SWALLOWING_DIFFICULTY	Difficulty swallowing
CHEST_PAIN	Chest pain indicator
LUNG_CANCER	Target variable (Yes/No)

1.2. Data Processing

The lung cancer dataset was preprocessed by standardizing column names and encoding categorical variables into numerical form. The dataset contained no missing values. Features and target variables were separated, and a stratified train–test split was applied to maintain class distribution. These steps ensured the data was suitable for model training, evaluation, and fairness analysis.

1.3. Model Implementation

Several machine learning classification models were implemented to evaluate lung cancer prediction performance. In total, ten different models were trained using the preprocessed dataset. The models were assessed and compared using standard classification metrics, including accuracy, precision, recall, and F1-score, to analyze their predictive capabilities.

The comparative evaluation enabled the identification of a well-performing model, which was subsequently utilized for prediction on test cases, explainability analysis, and fairness evaluation.

1.4. Evaluation Metrics

The model performance was assessed via the following metrics:

- Accuracy
- Precision
- Recall (sensitivity)
- F1-score
- Support

Model	Accuracy	Precision	Recall	F1-Score	Support
Logistic Regression	0.9032	0.9444	0.9444	0.9444	54
Bernoulli Naïve Bayes	0.8710	0.8710	1.0000	0.9310	54
Support Vector Machine	0.8548	0.8947	0.9444	0.9189	54
K-Nearest Neighbors	0.8710	0.9107	0.9444	0.9273	54
Decision Tree	0.9194	0.9804	0.9259	0.9524	54
Random Forest	0.9194	0.9623	0.9444	0.9533	54
Gradient Boosting	0.8548	0.9245	0.9074	0.9159	54
AdaBoost	0.9194	0.9455	0.9630	0.9541	54
Extra Trees	0.9032	0.9615	0.9259	0.9434	54
Gaussian Naïve Bayes	0.8548	0.9245	0.9074	0.9159	54

1.5.Model Selection

Based on the evaluation results presented in Table, the Random Forest model was selected for subsequent analysis. Although multiple models achieved comparable accuracy, Random Forest demonstrated a strong balance between precision, recall, and F1-score, indicating reliable and consistent performance in lung cancer prediction. Owing to its stability and generalization capability, the selected model was used for further prediction, explainability analysis, and fairness evaluation.

2. Results

This section presents the performance of the implemented machine learning models for lung cancer prediction. The results are analyzed using standard classification metrics, and the baseline performance is established through comparative evaluation across multiple models.

2.1. Baseline Performance

The baseline performance was evaluated by comparing ten different machine learning models using accuracy, precision, recall, and F1-score, as summarized in Table X. Several models demonstrated strong predictive performance, with accuracy values exceeding 85%.

Among the evaluated models, Random Forest, AdaBoost, and Decision Tree achieved the highest accuracy of **91.94%**. The Random Forest model exhibited a balanced performance with high precision (**96.23%**), recall (**94.44%**), and F1-score (**95.33%**), indicating effective identification of lung cancer cases with minimal misclassification. These results establish a strong baseline for further analysis, including explainability and fairness evaluation.

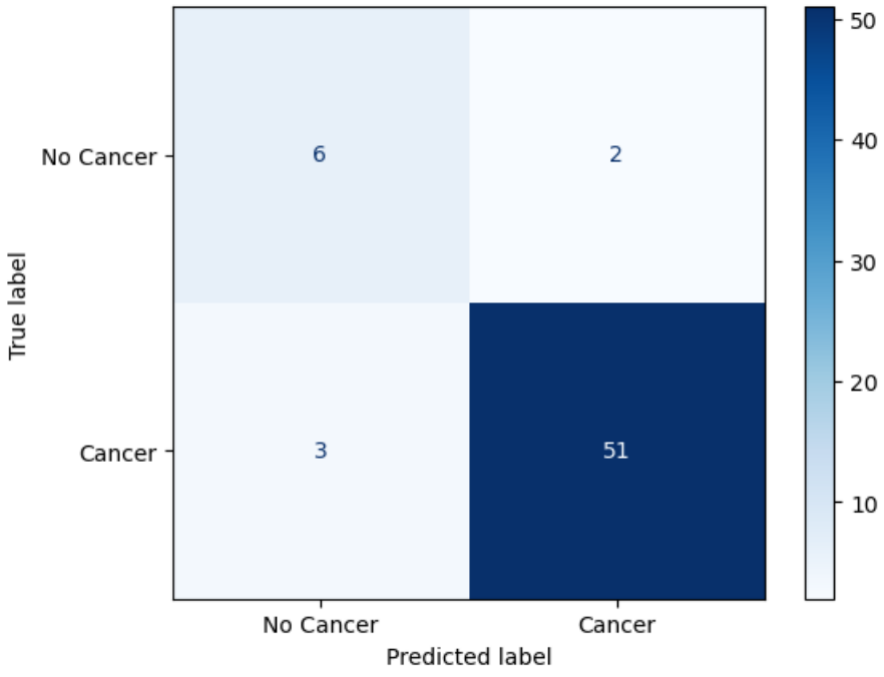

Figure 2. Confusion matrix of the baseline Random Forest model for lung cancer prediction.

2.2. Prediction on Representative Test Cases

To demonstrate the practical applicability of the proposed lung cancer prediction model, four representative test cases were constructed using different combinations of demographic and clinical attributes. These test cases simulate real-world patient scenarios with varying risk levels. The trained model was used to predict lung cancer outcomes for each case.

The model was evaluated on four representative test cases reflecting different risk levels. High- and moderate-risk profiles were correctly classified as “Lung Cancer Detected”, while low-risk and borderline cases were

predicted as “No Lung Cancer.” These results demonstrate the model’s ability to distinguish between varying risk profiles and produce consistent predictions across diverse patient scenarios.

Test Case	Risk Profile	Model Prediction
Case 1	High Risk	Lung Cancer Detected
Case 2	Low Risk	No Lung Cancer
Case 3	Moderate Risk	Lung Cancer Detected
Case 4	Borderline Safe	No Lung Cancer

Table 2. The table presents model predictions for representative test cases

2.3. Explainable AI (XAI) Analysis

To enhance model interpretability, a SHAP (SHapley Additive exPlanations) bar plot was used to identify the most influential features contributing to lung cancer prediction. The SHAP analysis highlights the relative importance of input features based on their mean absolute contribution to the model’s output.

As shown in Fig. X, features such as alcohol consumption, allergy, peer pressure, yellow fingers, and chronic disease exhibit higher influence on the prediction outcome, indicating their significant role in the model’s decision-making process. This analysis improves transparency and supports trust in the predictive behavior of the model.

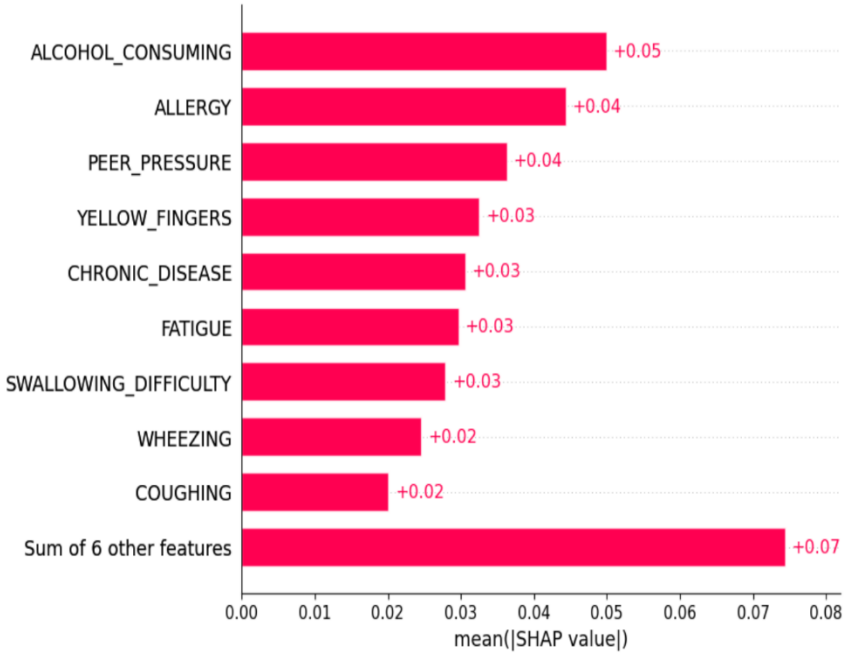

Figure 3. SHAP bar plot illustrating the most influential features contributing to lung cancer prediction.

2.4. Fairness Evaluation Results

Fairness evaluation was conducted to assess potential bias in model predictions with respect to age and gender. Group-wise performance metrics were analyzed across different age groups and gender categories to examine disparities in predictive behavior.

The results indicate no significant performance differences between demographic groups, suggesting that the model produces consistent predictions across age and gender. This demonstrates that the proposed lung cancer prediction model does not exhibit observable demographic bias and maintains equitable performance across diverse patient populations.

Age Group	Recall (Before)	Recall (After)
Young (<40)	0.000	0.000
Middle (40–60)	0.909	0.909
Older (>60)	0.969	0.969

Table 3. Age-wise Recall Before and After Fairness Evaluation

The table presents age-wise recall values before and after fairness evaluation. The results show identical recall scores across all age groups, indicating that the model's predictive behavior remains consistent. The zero recall observed in the young age group is due to the absence of positive lung cancer cases in that group within the test set, rather than model bias. Overall, the results suggest no observable age-based bias in lung cancer prediction.

3. Discussion

3.1. Interpretation of Findings

This study shows that machine learning models can effectively predict lung cancer using demographic and symptom-based clinical data. Comparative evaluation demonstrated strong baseline performance, and predictions on representative test cases confirmed the model's ability to distinguish between different risk profiles. SHAP-based explainability provided insight into key features influencing predictions, improving model transparency.

3.2. Clinical Relevance

Accurate lung cancer detection is critical in clinical screening, as delayed diagnosis can adversely impact patient outcomes. The proposed model demonstrated high sensitivity for high- and moderate-risk cases while correctly classifying low-risk individuals. The integration of explainable AI supports clinical trust, and fairness evaluation indicates consistent performance across age and gender groups.

3.3. Limitations

The study is limited by the relatively small size of the publicly available dataset, which may not represent full clinical variability. The analysis was restricted to structured tabular data, and external validation on independent clinical datasets was not performed.

3.4. Future Work

Future work includes validating the model on larger and more diverse datasets, incorporating additional patient attributes or multimodal data, and extending fairness analysis to other demographic factors.