EXPLAINABLE AI
Assignment-5

Name: K.SAI TEJA
HTNO:2303A52325
Batch:35
Instructor:Dr.vairachilai Shenbagavel

Lung Cancer Risk Prediction Report

1. Key Findings:

• The dataset consists of X rows and Y columns with demographic and lifestyle-related features (e.g., Age, Gender, Smoking, Alcohol, Fatigue, Wheezing, etc.).
• Target variable indicates lung cancer risk (Yes/No).
• EDA showed strong correlation between Smoking, Age, Alcohol, and the target.
• Class imbalance was present but corrected using SMOTE during preprocessing.
• Both Machine Learning (ML) and Deep Learning (DL) models were implemented and compared.

2. Comparison Table of ML vs DL Results

Machine Learning Results

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	67%	0.48%	0.68%	0.56%
Decision Tree	0.63%	0.42%	0.52%	0.47%
Random Forest	0.69%	0.50%	0.56%	0.53%

KNN	0.63%	0.44%	0.60%	0.51%
SVM (optional)	0.67%	0.48%	0.73%	0.58%
XGBoost (optional)	0.70%	0.52%	0.60%	0.56%

Deep Learning Results

Model	Test Accuracy
MLP	0.7153%
CNN	0.7127%
LSTM	0.7089%

3. Insights from XAI Visualizations

• SHAP Analysis showed that features such as Smoking, Age, and Alcohol consumption have the highest impact on lung cancer prediction.
• LIME Explanation highlighted how combinations of lifestyle habits (e.g., Smoking = Yes + Fatigue = Yes) significantly increase the probability of being classified as Cancer Risk.
• These techniques improved trust and interpretability of the models, especially important in healthcare.

4. Final Recommendation

• Random Forest achieved the best performance among ML models with high F1-score and interpretability.
• CNN achieved the highest accuracy among DL models but at the cost of reduced explainability.
• For real-world medical applications:
  - ML (Random Forest + SHAP/LIME) is preferred for its interpretability and balanced accuracy.
  - DL models may be explored further if higher accuracy is prioritized over explainability.