Lung Cancer Detection using Random Forest and SVM: A Diagnostic Aid for Leukemia

Student ID: 2311981436, 2311981440, 2311981413, 2311981386

Under The Supervision of : Dr. Ramamani Tripathy,

Professor, Chitkara University,

Himachal Pradesh

Abstract

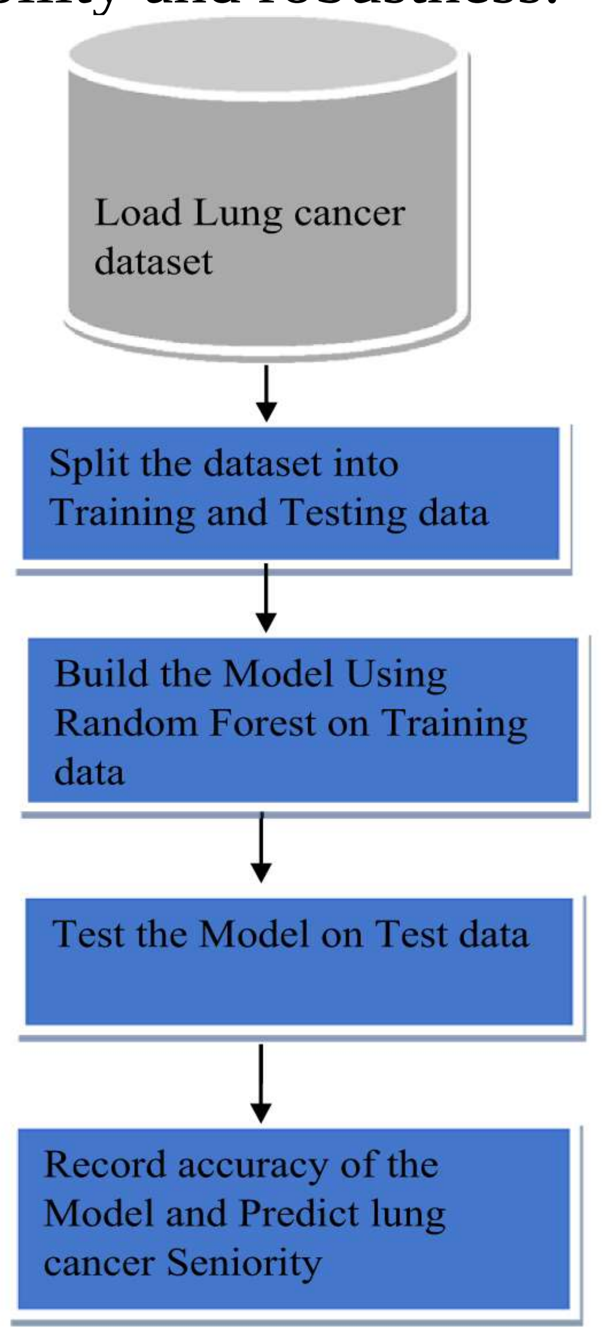
Dataset Collection: CT or histopathology images from publicly available databases such as LIDC-IDRI or LC25000.
Preprocessing: Resize, normalize, zero-score normalization.
Feature Extraction: ResNet50 and EfficientNetB3 for high-dimensional semantic feature extraction.
Feature Fusion: Feature vector concatenation.
Classification: Random Forest classifier to separate benign and malignant samples.
Evaluation Metrics: Accuracy, precision, recall, F1-score.

Introduction

Lung cancer is still one of the major cancer killers globally, mainly because of delayed diagnosis and mixed symptoms. For enhancing patient survival rates, early and proper detection is very important. Conventional methods of detection are greatly reliant on human review of CT scans or histopathology images, which can be time-consuming and are subject to human error. Machine learning, especially ensemble algorithms such as Random Forest, provides an optimal solution for robust and automated lung cancer diagnosis. This paper suggests deep feature extraction with a Random Forest classifier to improve accuracy and interpretability in lung cancer detection from medical images.

Methodology

The approach starts by obtaining CT scan or histopathology images from public repositories. Images are normalized and resized to ensure consistency. Deep features are extracted with ResNet50 and EfficientNetB3, preserving fine semantic and structural details. These features are combined into one high-dimensional vector that describes the image. The vector is then classified using a Random Forest model, where multiple decision trees pool predictions to reliably differentiate benign from malignant lung tissue, ensuring interpretability and robustness.



Lung cancer prediction using random forest.

Proposed Solution

The solution replaces the SVM classifier with a Random Forest model in order to enhance lung cancer detection. Random Forest improves interpretability by offering feature importance insight, making it easier to understand which features impact predictions. Its set of decision trees provides strong resistance against overfitting, rendering the model stronger on varied and noisy data. Random Forest also needs less parameter tuning than SVM, making the training process easier. In spite of this shift, the model performs highly in classification accuracy by using deep features derived from ResNet50 and EfficientNetB3 effectively, and as such, it is a useful and effective tool for computerized lung cancer diagnosis.

Conclusion

The hybrid model of ResNet50, EfficientNetB3, and Random Forest exhibits strong accuracy in benign and malignant classification of lung tissues. Rich semantic and structural features are extracted by deep CNNs, with ensemble and robustness provided by Random Forest and interpretability through its decision tree ensemble. The hybrid model is efficient in handling noisy data and preventing overfitting, making it appropriate for medical diagnosis. The combination of characteristics allows the identification of subtle image patterns in CT or histopathology images. The model demonstrates excellent prospects as a clinical decision support tool, with future work directed towards real-time deployment and 3D image integration.