LungXAI: An Explainable High-Performance Model for Multi-Class Lung Cancer Classification Using Deep Learning and RAG-Based Knowledge Retrieval

Saksham Mann (RA2211003011213)
School of Computing
Department of Computing Technologies
SRM Institute of Science and Technology
Chennai, India
saksham.mann@srmist.edu.in

Chirag Agrawal (RA2211003011252)
School of Computing
Department of Computing Technologies
SRM Institute of Science and Technology

Abstract— Lung cancer remains a critical global health challenge, being the third most common cancer worldwide with the highest mortality rate among all cancers. With 2.48 million new cases reported in 2022 and a 5-year survival rate of approximately 28.4%, early detection and accurate subtyping are essential for improved patient outcomes. This study introduces LungXAI, a clinically interpretable deep learning framework designed for multi-class classification of lung cancer subtypes from CT scan images. The proposed pipeline employs a Vision Transformer (ViT) / ResNet-50 architecture pretrained on ImageNet and fine-tuned for four-class classification (Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma, and Normal/Benign). The model integrates Gradient-weighted Class Activation Mapping (Grad-CAM) for visual explainability and a novel XAI-to-RAG bridge that automatically converts visual heatmap features to textual queries for Retrieval-Augmented Generation (RAG) based knowledge retrieval from curated medical sources. The framework demonstrates expected predictive performance with accuracy of 85–90%, Precision ≥86%, Recall ≥84%, F1-Score 85–88%, and AUC-ROC 0.90–0.93. To enhance interpretability and

I. INTRODUCTION

Lung cancer describes a malignant growth of abnormal cells in lung tissue, often resulting from prolonged exposure to carcinogens such as tobacco smoke. According to recent estimates, lung cancer is the third most common cancer worldwide while maintaining the highest mortality rate among all cancer types. Approximately 2.48 million new cases were reported in 2022 alone, with a concerning 5-year survival rate of only 28.4% [1]. Furthermore, 50–70% of cases are diagnosed at the metastatic stage, leading to poor treatment outcomes and limited therapeutic options.

Non-Small Cell Lung Cancer (NSCLC) accounts for approximately 85% of all lung cancer cases, comprising three major subtypes. Adenocarcinoma represents approximately 40% of cases and typically presents in peripheral lung regions. Squamous Cell Carcinoma is the second most common subtype, often found in central lung areas near major airways. Large Cell Carcinoma is rare but aggressive, characterized by rapid progression. Accurate subtyping is critical for targeted therapy and personalized

II. RELATED WORKS

A. Deep Learning for Lung Cancer Classification

Several studies have explored deep learning approaches for lung cancer classification using CT scan images. CNN-based models have achieved up to 93.06% accuracy on three-class classification tasks distinguishing Adenocarcinoma, Squamous Cell Carcinoma, and Large Cell Carcinoma [5]. DenseNet combined with AdaBoost fusion achieved 89.85% accuracy, demonstrating the effectiveness of ensemble learning in medical imaging [6]. Vision Transformers, particularly the Swin Transformer architecture, have achieved state-of-the-art performance with 97.14% accuracy and 0.993 AUC-ROC using localized self-attention mechanisms [7]. These results establish that the classification task is well-studied and high-performing models are achievable with current technology.

B. Explainable AI for Medical Imaging

III. SUMMARY OF THE DATASET

This study uses the CT Scan Images of Lung Cancer dataset available on Kaggle [11]. The dataset includes CT scan images categorized into five classes: Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma, Benign cases, and Normal cases. This dataset is designed for multi-class classification, enabling the development of models that can distinguish between different lung cancer subtypes and healthy tissue. Table I summarizes the dataset attributes.

TABLE I. DESCRIPTION OF DATASET CLASSES

Class	Description
Adenocarcinoma	Most common NSCLC subtype (~40%), typically peripheral presentation
Squamous Cell Carcinoma	Second most common, often central lung location

IV. DESCRIPTIVE METHODOLOGY

Understanding the clinical and imaging characteristics of the dataset provides valuable context for lung cancer classification. The preprocessing pipeline includes the following steps:

Image Preprocessing:

- Image resizing to 224×224 pixels to match model input requirements
- Pixel normalization using ImageNet statistics (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
- CLAHE (Contrast Limited Adaptive Histogram Equalization) for contrast enhancement
- Noise reduction techniques for improved image quality

Data Augmentation:

V. PROPOSED METHODOLOGY

The proposed framework, LungXAI, is a deep learning pipeline developed for multi-class lung cancer classification with integrated explainability and knowledge retrieval. It adopts a hybrid architecture combining Vision Transformers (ViT) or ResNet-50 with Grad-CAM explainability and RAG-based knowledge retrieval. The overall workflow includes data preprocessing, feature extraction, classification, XAI visualization, and RAG-based explanation generation, as depicted in the system architecture.

A. System Architecture

The LungXAI system consists of five integrated modules:

1. **Data Acquisition and Preprocessing:** CT scan images are collected and preprocessed with normalization, noise reduction, and CLAHE for contrast enhancement. Data augmentation including rotation, flipping, and scaling improves model robustness.

VI. CONCLUSION

Lung cancer remains a critical global health challenge requiring precise and interpretable diagnostic tools. This study introduced LungXAI, a deep learning framework designed to classify lung cancer subtypes from CT scan images while providing transparent, evidence-backed explanations. The model employs Vision Transformer/ResNet-50 architecture with Grad-CAM for visual explainability and integrates a novel XAI-to-RAG bridge that automatically connects visual evidence with medical knowledge retrieval. The expected performance demonstrates reliable classification across multiple cancer subtypes with accuracy of 85–90% and AUC-ROC of 0.90–0.93.

The proposed framework addresses the critical trust barrier in medical AI adoption by bridging the gap between deep learning accuracy and clinical reasoning. It offers not only predictions but also transparent, evidence-backed explanations that support clinical decision-making. This work directly supports SDG 3 (Good Health and Well-

REFERENCES

[1] World Health Organization. (2023). Global Cancer Statistics 2022: Lung Cancer Incidence and Mortality. WHO Cancer Report.

[2] Travis, W. D., et al. (2021). WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart. IARC Publications.

[3] Liu, Z., Lin, Y., Cao, Y., et al. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. Proceedings of IEEE/CVF ICCV, 10012–10022.

[4] Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nature Machine Intelligence, 1(5), 206–215.

[5] Zhang, L., et al. (2020). CNN-based Classification of Lung Cancer Subtypes from CT Images. IEEE Access, 8, 142365–142375.