# Multimodal Feature Fusion Pipeline for NSCLC Subtype Classification (ADC/SCC) Using ROI-Imputed CNN Embeddings, Radiomics, and Clinical Features with Explainable AI

NUSRAT FARZANA CHOUDHURY

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

## APPROVAL

This thesis titled on "**Multimodal Feature Fusion Pipeline for NSCLC Subtype Classification (ADC/SCC) Using ROI-Imputed CNN Embeddings, Radiomics, and Clinical Features with Explainable AI**", submitted by **Nusrat Farzana Choudhury (ID: 221-35-990)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

## BOARD OF EXAMINERS

**Chairman**

**Dr. S. M. Hasan Mahmud**
**Associate Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

**Internal Examiner 1**

**A.H.M Shahariar Parvez**
**Associate Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

**Internal Examiner 2**

**Tapushe Rabaya Toma**
**Assistant Professor**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

**Internal Examiner 3**

**Khalid Been Md. Badruzzaman Biplob**
**Lecturer (Senior Scale)**
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

**External Examiner**

**Dr. Md Sazzadur Rahman**
**Professor**
Institute of Information Technology
Jahangirnagar University, Bangladesh

**Department of Software Engineering**

**Faculty of Science and Information Technology**

**Supervisor's Approval Form**

| | | |
|---|---|---|
| Fall 2025 | B.Sc. In SWE | Campus: DSC |

| Student Name | Student ID |
|---|---|
| Nusrat Farzana Choudhury | 221-35-990 |

| Project/Thesis Information | |
|---|---|
| Project/Thesis Title | Multimodal Feature Fusion Pipeline for NSCLC Subtype Classification (ADC/SCC) Using ROI-Imputed CNN Embeddings, Radiomics, and Clinical Features with Explainable AI |
| Type of work | Interpretable Multimodal Imaging-Clinical Cancer Diagnosis |

| Supervisor's information | |
|---|---|
| Supervisor Name | Tapushe Rabaya Toma |
| Supervisor Initial | TRT |
| Completed Credit till now | 139 |
| How many credits are in this semeste | 6 |
| Amount (Due) | 0 BDT |
| Supervisor Consent | ☑ Yes ☐ No |

Supervisor Signature

## DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name : Nusrat Farzana Choudhury
Date of Birth : 20 September 2002
Title : Multimodal Feature Fusion Pipeline for NSCLC Subtype Classification (ADC/SCC) Using ROI-Imputed CNN Embeddings, Radiomics, and Clinical Features with Explainable AI
Academic Session : 2022-2025

I declare that this thesis is classified as:

| | |
|---|---|
| ☐ CONFIDENTIAL | (Contains confidential information under the Official Secret Act 1997)* |
| ☐ RESTRICTED | (Contains restricted information as specified by the organization where research was done)* |
| ☑ OPEN ACCESS | (I agree that my thesis will be published as online open access (Full Text) |

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

(Student's Signature)

(Supervisor's Signature)

Student's Name: Nusrat Farzana Choudhury
Student ID: 221-35-990
Date: 22 December 2025

Supervisor's Name: Tapushe Rabaya Toma
Designation: Assistant Professor,
Department of SWE
Date: 22 December 2025

## SUPERVISOR's DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

(Supervisor's Signature)

Full Name : Tapushe Rabaya Toma

Position : Assistant Professor

Date : 29 November 2025

## STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations, which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

(Student's Signature)

Full Name : Nusrat Farzana Choudhury

ID Number : 221-35-990

Date : 29 November 2025

# MULTIMODAL FEATURE FUSION PIPELINE FOR NSCLC SUBTYPE CLASSIFICATION (ADC/SCC) USING ROI-IMPUTED CNN EMBEDDINGS, RADIOMICS, AND CLINICAL FEATURES WITH EXPLAINABLE AI

NUSRAT FARZANA CHOUDHURY

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

December 2025

# ACKNOWLEDGEMENTS

I begin by expressing my deepest and unconditional gratitude to **the Almighty** for enabling me to complete this challenging journey of my thesis. It was only through His boundless mercy, constant guidance, and countless blessings that this extensive academic work became possible. His support remained with me at every stage, giving me strength, clarity, and patience throughout the process.

I would also like to convey my sincere respect and heartfelt gratitude to my supervisor, **Ms. Tapushe Rabaya Toma**, and my co-supervisor, **Mr. Musabbir Hasan Sammak**. Their unwavering support, thoughtful guidance, and academic integrity played a transformative role in shaping this research. Their willingness to correct my mistakes with care, their confidence in my abilities, and their openness in sharing knowledge greatly motivated me and strengthened my academic foundation. It is truly a privilege to have received their valuable time and expertise, which not only enhanced my technical skills but also instilled in me the essential values of sincerity, discipline, and humility that will continue to guide me in my future endeavors.

 

# DEDICATION

In the name of Allah, the Most Merciful, the Most Compassionate. This work is first dedicated to **the Almighty**, whose guidance sustained me through both clarity and confusion. When strength felt fragile and patience ran thin, His mercy never did. Every step, every lesson, and this completion itself exist only by His will. Alhamdulillah.

I dedicate this work to my **parents**, the silent force behind everything I am. Their sacrifices were endless, their prayers unwavering, and their belief in me unshakable. Long before I understood this journey, they were already carrying it with me. Whatever I have achieved stands firmly on their love and resilience. This is also for my **brothers**, my constant source of strength. Through words spoken and unspoken, they reminded me that I was never alone, turning pressure into courage and doubt into resolve.

I dedicate this to **all my teachers**, those who guided me for years and those who taught me for only a single day. Every lesson, correction, and moment of guidance shaped my thinking and discipline, leaving marks far deeper than time suggests. And finally, to my **friends** and silent **well-wishers**. To those who felt like home, who listened to my yapping, frustrations, and fears, who lifted me when I fell, and believed when I faltered. And to those who supported quietly from afar, unseen but deeply felt. This work carries pieces of you all.

This thesis is not mine alone. It is a shared journey of faith, sacrifice, and support. May Allah reward every heart behind it and make this effort a source of goodness, In sha Allah.

# ABSTRACT

Early and accurate identification of Non-Small Cell Lung Cancer (NSCLC) subtypes is critically important, as it enables reliable differentiation between Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SCC) and supports the adoption of truly personalized and targeted treatment strategies tailored to individual patients. Conventional biopsy-based diagnosis, however, is invasive and often time-consuming, highlighting an urgent need for reliable, non-invasive computational approaches using Computed Tomography (CT) imaging. Although deep learning models have shown promise in this domain, their clinical adoption remains limited due to challenges such as limited data availability, severe class imbalance, and poor interpretability.

This thesis directly addresses these limitations by proposing a novel, interpretable multimodal feature fusion pipeline. The framework begins with a three-layer ROI imputation strategy designed to overcome the absence of explicit tumor boundary annotations, resulting in a unified, high-quality, nodule-level dataset comprising 134 unique patients. From this dataset, three complementary feature streams are extracted and systematically fused: ROI-imputed deep CNN embeddings, handcrafted radiomics features, and carefully preprocessed clinical metadata. These fused representations are then classified using a Stacking Ensemble Meta-Model with a Level-1 Logistic Regression classifier.

The experimental results validate the effectiveness of the proposed approach. The initial ROI imputation stage significantly enhanced the safety-critical SCC recall of the image-based model, increasing it from 0.17 to 0.50. The final multimodal ensemble achieved strong clinical performance, with a Macro F1-score of 0.7363 and an SCC recall of 66.7%, demonstrating a balanced and reliable diagnostic capability. Furthermore, explainability analysis using SHAP values provided conclusive evidence supporting the central hypothesis of this work: the model's balanced predictive performance arises from the integration of complementary, multi-domain features. Radiomic Maximum Density emerged as the most influential objective feature, synergizing effectively with abstract deep learning signals represented by CNN probability scores.

Finally, a multi-perspective Explainable AI (XAI) protocol offered clinically meaningful insights, revealing that the model's decision-making aligns closely with established pathological knowledge. Intra-tumoral texture was identified as the most influential feature for ADC classification, while peripheral invasion and pleural margin characteristics were dominant for SCC. Overall, this study presents a coherent, interpretable, and clinically aligned decision-support system, designed with translational readiness for real-world clinical application.

 

# TABLE OF CONTENTS

 

 

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS

| | |
|---|---|
| P | Probability |
| N | Total Number of Patients or Samples |
| ∑ | Summation |

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ADC | Adenocarcinoma |
| CNN | Convolutional Neural Network |
| CT | Computed Tomography |
| DL | Slice-Level Evaluation of Model-1 |
| FL | Federated Learning |
| LIME | Local Interpretable Model-agnostic Explanations |
| NSCLC | Non-Small Cell Lung Cancer |
| ROI | Region of Interest |
| SCC | Squamous Cell Carcinoma |
| SHAP | SHapley Additive exPlanations |
| TCIA | The Cancer Imaging Archive |
| XAI | Explainable AI |

# CHAPTER 1: INTRODUCTION

## 1.1 Background of the Study

Non-Small Cell Lung Cancer (NSCLC) accounting for roughly 85% of all primary cases (Tan, 2024). For modern precision oncology, a core requirement is the swift and accurate classification of the two dominant NSCLC subtypes: Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SCC) (Park et al., 2024; Chen & Li, 2018). These two subtypes are fundamentally distinct; they possess unique biological signatures, lead to different clinical outcomes, and—most critically—exhibit varied responses to targeted therapies and immunotherapies (Zhang et al., 2025; Li et al., 2025). Consequently, misclassification directly risks suboptimal or ineffective patient treatment—a serious clinical concern (Park et al., 2024).

Traditionally, confirming a diagnosis has relied on invasive steps—mainly tissue biopsies, followed by careful histopathological analysis (Tan, 2024). Yet these classic approaches come with real risks, like patient complications; they often prove tricky when tumors are small or hard to reach, and they frequently drag on, delivering delayed results (Tan, 2024). Spurred by these pressing clinical hurdles, researchers have urgently pursued non-invasive, computation-driven alternatives (Aksu et al., 2025).

Radiomics and Deep Learning (DL) offer a powerful, non-invasive way to identify tumor histotypes using standard Computed Tomography (CT) scans (Parekh et al., 2019; Shen et al., 2019). While Radiomics works by extracting mathematical traits that reflect tumor heterogeneity, DL models specifically Convolutional Neural Networks (CNNs) uncover more complex, abstract patterns directly from the raw imagery (Parekh et al., 2019; Shen et al., 2019); furthermore, integrating these visual insights with relevant clinical metadata consistently improves prediction accuracy, making the results far more useful in a practical medical setting (Kim et al., 2022; Aksu et al., 2025).

## 1.2 Problem Statement

Despite the exciting potential of computational oncology, three big—and closely linked—challenges are still blocking the smooth rollout of NSCLC subtype classification models into everyday clinical practice (Aksu et al., 2025; Kim et al., 2022):

I. **Data Quality and Scarcity:** Medical imaging datasets often suffer from tiny sample sizes and stark class imbalances—with Adenocarcinoma (ADC) far outnumbering the rest (Bakr et al., 2017; Park et al., 2024). Even more troubling, inconsistent or outright missing annotations for the Region of Interest (ROI)—that's the precise tumor boundaries—make it tough to pull out clean, tumor-focused features; this noise inevitably undermines the reliability of any image-based model (Baba et al., 2022; Parekh et al., 2019).

II. **The Unimodal Performance Trap:** Relying on just one type of data—whether images alone or clinical data alone—rarely delivers the robust, balanced accuracy that hospitals demand (Han et al., 2024; Kim et al., 2022). These single-modality approaches particularly falter with the underrepresented Squamous Cell Carcinoma (SCC) class, producing low recall rates and far too many missed diagnoses—a serious problem in settings where sensitivity comes first (Park et al., 2024; Zhang et al., 2025).

III. **Black Box Dilemma:** Today's top-performing systems—often intricate ensembles or deep learning networks—are frequently, and justifiably, seen as inscrutable "black boxes" (Yao et al., 2024; Ennab & Mcheick, 2025). This opacity breeds genuine clinical distrust, slowing down the broader acceptance of AI tools (Yao et al., 2024; Wang et al., 2024). As a result, doctors rightly demand that every AI-generated diagnostic recommendation be fully traceable, transparent, and defensible—right there at the bedside (Li et al., 2024; Sharma et al., 2025).

## 1.3 Research Objectives

### 1.3.1 General Objective

The overarching aim of this research is to create and verify a sturdy, multimodal feature fusion pipeline—one that's precise and interpretable—for classifying Non-Small Cell Lung Cancer (NSCLC) subtypes (ADC/SCC); it achieves this by smartly blending deep learning, radiomic, and clinical features drawn from CT imaging data.

### 1.3.2 Specific Objectives

I. To devise and roll out an innovative ROI-Imputation Strategy—targeted at overcoming data quality hurdles—by dependably filling in absent tumor outlines; this enables the formation of a consistent, nodule-focused dataset that includes all patients.

II. To extract and preprocess three complementary data streams—Deep CNN Embeddings, Radiomics features, and Clinical metadata—for subsequent feature-level fusion.
III. To construct and optimize a Stacking Ensemble Meta-Model that combines these multimodal features to achieve a superior, balanced classification performance, particularly enhancing the detection (Recall) of the minority SCC subtype.
IV. To employ a comprehensive Explainable AI (XAI) protocol, including Grad-CAM, SHAP, and LIME, to ensure that the ensemble model's predictive decisions are fully transparent, clinically justifiable, and aligned with pathological knowledge.

## 1.4 Research Questions

I. How can a novel ROI-Imputation Strategy effectively address data quality issues caused by missing tumor segmentations and improve the predictive feature set for NSCLC subtype classification?
II. To what extent does the multimodal fusion of CNN Embeddings, Radiomic Features, and Clinical Data improve balanced classification performance (Macro F1-Score) and, specifically, the clinical detection rate (Recall) of the minority SCC subtype compared to unimodal models?
III. Based on SHAP analysis, which feature modality (CNN, Radiomics, or Clinical) emerges as the most influential determinant in the final prediction of the Stacking Ensemble Meta-Model?
IV. Can the integrated Explainable AI (XAI) framework provide traceable and pathologically meaningful visual and quantitative evidence (via Grad-CAM, SHAP, and LIME) to validate the model's classification rationale?

## 1.5 Scope and Limitations

Scope: This research zeros in strictly on the non-invasive, binary sorting of NSCLC subtypes—ADC versus SCC—drawing from post-diagnosis CT images and linked clinical metadata. All data comes solely from The Cancer Imaging Archive (TCIA). The real breakthrough shines through in the feature engineering—via ROI-Imputation—and the multimodal ensemble learning setup, all backed by a thorough XAI validation process.

Limitations: Key drawbacks center on the fairly modest scale of the final polished cohort—just 134 unique patients—and its built-in class imbalance. Plus, even as the ROI-Imputation approach eases the segmentation hurdle, those filled-in boundaries remain approximations, not spot-on expert-verified truths. The results are currently validated on data from the single source of TCIA, and external validation on multi-center data remains a necessary step for clinical deployment.

## 1.6 Significance of the Study

This research packs real significance—for clinical practice and methodological progress alike:

I. Clinical Translation: It brings a strong, non-invasive tool to the table—one that could quicken those crucial early diagnostic choices—empowering oncologists to swiftly launch therapies tailored to specific subtypes, and ultimately enhancing patient outcomes.
II. Novel Data Preparation: Launching the ROI-Imputation Strategy offers researchers a versatile fix—perfect for tackling incomplete segmentation annotations in medical datasets—while enabling reliable extraction of high-quality, nodule-centered features.
III. Enhanced Trust and Adoption: By rigorously implementing the Explainable AI framework, the work transforms a typical "black box" prediction system into a transparent and trustworthy diagnostic aid, which is the necessary bridge for regulatory approval and clinical acceptance.
IV. Balanced Performance: The ensemble model's focus on a Sensitivity-First Strategy successfully addresses the critical clinical need to minimize false negatives for the minority SCC class, making the model a safer, more reliable diagnostic instrument.

## 1.7 Summary

This chapter has laid out the critical clinical need for precise NSCLC subtyping—and it has spotlighted the main drawbacks in current computational methods, especially around data quality and model interpretability. The outlined research objectives and questions map out a straightforward strategy to tackle these issues—through an innovative ROI-Imputation technique, multimodal feature blending, and a thorough XAI validation process. Moving forward, the upcoming chapters will delve into the literature review, the detailed methodology, the experimental findings, and a concluding discussion on the results, insights, and overall contributions.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

Around the world, lung cancer remains the deadliest form of cancer; significantly, the vast majority of these cases—approximately 85%—are classified as Non-Small Cell Lung Cancer, or NSCLC (Tan, 2024). Accurate classification of the two major NSCLC subtypes—Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SCC)—is an essential requirement in the era of precision medicine (Park et al., 2024; Chen & Li, 2018). These subtypes possess fundamental differences in biological characteristics, outcomes, and, crucially, their responses to targeted therapies and immunotherapies (Zhang et al., 2025; Li et al., 2025).

Historically, diagnosis has leaned heavily on invasive methods—primarily tissue biopsies and histopathological reviews. These procedures, while standard, carry inherent risks of clinical complications (Tan, 2024). Accuracy also remains a hurdle; if a tumor is too small or tucked away near vital organs, the results can be frustratingly ambiguous or delayed (Tan, 2024).

Because of these limitations, the medical community is shifting toward non-invasive, computational solutions (Aksu et al., 2025). Tools like radiomics and deep learning—when applied to standard CT scans—offer a powerful alternative. These methods allow oncologists to identify a tumor's subtype at the moment of diagnosis, streamlining the entire decision-making process (Parekh et al., 2019; Shen et al., 2019).

However, the challenge isn't just about raw accuracy. Many complex deep learning models are viewed as "black boxes"—a lack of transparency that fuels skepticism and slows down their use in real hospitals (Yao et al., 2024; Ennab & Mcheick, 2025). Today's goal in computational oncology is clear: we need systems that are not only precise but also interpretable. Every AI-driven insight must be traceable and clinically sound (Kim et al., 2022; Wang et al., 2024; Sharma et al., 2025).

## 2.2 Previous Literature

### 2.2.1 CT Image Analysis and Archetypal Feature Extraction

The success of using CT scans to predict cancer subtypes is built on a simple reality: Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SCC) look different under a radiological lens (Chen & Li, 2018; Park et al., 2024).

**Pathological and Radiological Distinctions:** On a standard scan, ADC usually appears as a peripheral lesion—often showing "lobulated" or jagged edges and signs of pleural indentation. SCC, on the other hand, is typically found more centrally as a larger mass with higher tissue

density (Chen & Li, 2018; Li et al., 2025). These visual cues are actually quantitative markers of the tumor's unique biology (Park et al., 2024; Zhang et al., 2025).

**Volumetric Data Processing and Challenges:** CT data—stored in the DICOM format—is dense and three-dimensional. To make sense of it, researchers must use rigorous preprocessing to handle different slice spacings and scanner settings (Baba et al., 2022). This involves "Hounsfield Unit" (HU) transformations, windowing, and wavelet filters to sharpen the features that help distinguish subtypes (Parekh et al., 2019). Furthermore, researchers must choose the right slices; "key slice selection" frameworks ensure that models learn from the most informative parts of the tumor (Yamamoto et al., 2024). Ultimately, looking at the whole volume—rather than just a single slice—has become the gold standard (Baba et al., 2022).

**Archetypal Feature Extraction–Radiomics:** Radiomics treats medical images as data, extracting hundreds of mathematical descriptors that define a tumor's shape, intensity, and texture (Parekh et al., 2019). Specifically, textural features—like those from the Gray Level Co-occurrence Matrix (GLCM)—are excellent at differentiating NSCLC subtypes because they capture the intricate, underlying complexity of the tissue (Chen & Li, 2018; Zhang et al., 2025).

**Deep Feature Extraction –CNN Embeddings:** Beyond manual features, deep Convolutional Neural Networks (CNNs) are used to pull out abstract, high-level "embeddings" (Shen et al., 2019). Models like ResNet-50 are the go-to choice here; their "residual connections" allow the network to learn deeply without losing focus or accuracy (Adekunle et al., 2025).

### 2.2.2 Clinical Metadata Analysis

Clinical metadata provides essential contextual information about the patient and disease stage that cannot be gleaned from imaging alone, serving as a powerful complementary data stream (Kim et al., 2022).

**Defining Clinical Features for ADC vs. SCC:** Features that reflect the patient's pathological and demographic profile are prioritized for classification. Several core factors typically shape the clinical picture—age, gender (where male patients often show a higher prevalence of SCC), smoking history, and the physical reach of the disease, as defined by the TNM staging system (Li et al., 2025; Zhang et al., 2025). Where the tumor actually sits within the lung is another vital clue; while Adenocarcinoma (ADC) usually appears on the outer edges, Squamous Cell Carcinoma (SCC) tends to take root in central or perihilar locations (Chen & Li, 2018; Park et al., 2024).

**Data Cleaning and Encoding:** Managing this information is a relatively direct process—often involving data pulled straight from Electronic Health Records (EHRs)—though it requires a methodical touch (Shankar et al., 2023). To make the data "machine-ready," we follow a few standard steps:

I. Numerical features (like age or pack-years) are standardized to a common scale.
II. Categorical features (such as tumor stage or location) are translated into numbers through techniques like one-hot or label encoding (Lee, 2020).

Beyond simple conversion, we must be incredibly rigorous when dealing with missing values; since these are major prognostic factors, we often rely on specialized techniques to navigate that uncertainty (Shankar et al., 2023).

**Predictive Value:** Ultimately, while imaging provides the structural "map" of a tumor, these clinical variables tell its biological story—offering a deeper look into the pathology and context of the disease. Studies using molecular metadata (e.g., gene expression data) have confirmed that ADC and SCC possess distinct molecular signatures, with genes like CLCA2 and LPCAT1 driving specific subpopulations (Lee et al., 2021; Campos-Parra et al., 2021). The integration of clinical variables consistently improves the predictive power and clinical relevance of imaging-based models (Kim et al., 2022; Aksu et al., 2025).

### 2.2.3 Ensemble Learning for ADC vs SCC Prediction

To achieve optimal robustness and prediction accuracy, contemporary studies often move beyond single-modality or single-algorithm classifiers toward multimodal ensemble learning frameworks (Kim et al., 2022; Aksu et al., 2025).

**Rationale and Fusion Paradigms:** Ensemble learning techniques, such as bagging (Random Forest) and boosting (XGBoost), combine the predictions of multiple base models, effectively reducing variance and improving generalization (Liu et al., 2024). Multi-modal Deep Learning (MDL) fusion techniques are broadly categorized into three types: early (raw data/feature concatenation), intermediate (feature-level fusion or stacking), and late (decision-level fusion) (Aksu et al., 2025). Intermediate fusion, particularly the stacking ensemble, is favored as it intelligently integrates data at the feature extraction stage, offering a balanced combination of modality-specific information (Han et al., 2024; Aksu et al., 2025).

**Feature-Level Fusion and Hybrid Models:** One of the most effective strategies in modern analysis involves "feature-level fusion"—a process where diverse data streams are brought together into a unified framework. Rather than looking at variables in isolation, we take the extracted streams—specifically CNN embeddings, radiomic features, and preprocessed clinical data—and stitch them into a single, comprehensive feature matrix (Han et al., 2024). This hybrid approach, which blends the raw power of deep learning with the structured precision of traditional radiomics, has consistently set the gold standard in oncology. It proves a vital point: these two data types aren't redundant; instead, they capture complementary information—each filling in the biological or structural gaps that the other might overlook (Zhang et al., 2024; Li et al., 2025).

**Stacking Ensemble Architecture:** The Stacking Ensemble framework—often referred to as a meta-model—is a hierarchical strategy frequently used to achieve high-performance predictions in NSCLC (Liu et al., 2024). The process works in layers: Level 0 consists of diverse "base learners"—such as Logistic Regression, Random Forest, and XGBoost—which are all trained on the complete multimodal dataset. Then, Level 1 takes over; here, a simpler meta-classifier (frequently a Logistic Regression model) is trained specifically on the predicted probabilities generated by those initial base learners (Liu et al., 2024). Ultimately, this architecture provides a robust, optimized way to aggregate data—leading to significantly sharper predictive accuracy (Li et al., 2025; Mao et al., 2025).

### 2.2.4 Explainable AI (XAI) and Model Interpretability

Explainability is the bedrock of clinical adoption; if a model is to be trusted in a medical setting, its "thinking" must be transparent (Yao et al., 2024; Ennab & Mcheick, 2025). This is where Explainable AI (XAI) techniques prove their worth. They provide the vital tools needed to verify that a model's predictions are rooted in genuine pathological knowledge—rather than just picking up on coincidental, "spurious" correlations (Sharma et al., 2025).

**Visual Interpretation via Grad-CAM:** For deep learning models, Gradient-weighted Class Activation Mapping (Grad-CAM) is widely employed. Grad-CAM generates a heatmap that visually highlights the specific regions of the CT nodule that the CNN focused on when making its classification (Li et al., 2024; Wang et al., 2024; Ennab & Mcheick, 2025). This is crucial for visual validation, confirming that the model's attention aligns with established pathological features—for example, focusing on high-density internal structures for SCC or irregular margins for ADC (Li et al., 2024; Wang et al., 2024).

**Feature Contribution Analysis (SHAP and LIME):** To interpret the final multimodal ensemble, feature-based XAI methods are essential (Li et al., 2024; Sharma et al., 2025).

I. SHAP (SHapley Additive exPlanations) provides a globally consistent and fair attribution of predictive value to every single feature (clinical, radiomic, or CNN embedding) across the entire patient cohort (Rahman, 2025; Shi et al., 2025). Global SHAP analysis yields a quantifiable ranking of feature importance, clarifying whether specific clinical factors (e.g., staging) or image-derived textural features are the primary decision drivers (Shi et al., 2025; Rahman, 2025).
II. LIME (Local Interpretable Model-agnostic Explanations) is used for instance-specific (local) interpretation (Lee et al., 2020; Yao et al., 2017). LIME approximates the complex ensemble model around a single patient's data point, generating a concise, contrastive explanation that outlines the specific features that contributed most strongly to that individual's predicted diagnosis (Lee et al., 2020; Yao et al., 2017).

The combined use of visual, local, and global XAI methods transforms the black-box ensemble into a transparent diagnostic tool, a multi-perspective approach that is recognized as enhancing clinical diagnosis and biomarker understanding (Sharma et al., 2025; Ganie et al., 2025; Jafari et al., 2024).

### 2.2.5 Challenges, Gaps, and Future Directions

Despite advances in multimodal fusion, several challenges persist that limit the clinical generalizability of AI models (Aksu et al., 2025; Kim et al., 2022).

**Data Limitations and Heterogeneity:** Medical datasets are often characterized by small sample sizes and significant class imbalance, which can bias models toward the majority class (ADC) (Rahman, 2025; Bakr et al., 2017). Furthermore, integrating heterogeneous data (images, pathology reports, clinical variables) from diverse sources presents technical engineering and reproducibility challenges (Li et al., 2020; Chen et al., 2020).

**Analytical Complexity:** Interpreting the synergistic effects of multimodal fusion remains complex (Aksu et al., 2025; Han et al., 2024). Validating feature reliability across multi-center data, for instance, requires statistical rigor to ensure the extracted features are reproducible and consistent, reinforcing the validity of the proposed framework for real-world application (Niu et al., 2025).

**Future Opportunities:** To overcome data scarcity and privacy concerns, two key areas are emerging:

I. **Federated and Self-supervised Learning:** Federated Learning (FL) is the benchmark approach for distributed, privacy-preserved training, allowing models to learn from decentralized patient data across multiple institutions without moving the raw information (Rahman, 2025; Yan et al., 2025). This is often paired with Self-Supervised Learning (SSL), which exploits large volumes of unlabeled data to learn robust visual representations, thereby mitigating the high cost and scarcity of annotated medical images (Rahman, 2025; Yan et al., 2025).
II. **Hybrid Feature Development:** Research is increasingly focused on developing hybrid models that effectively combine the abstract, high-performance nature of deep CNN features with the stable, predefined characteristics of traditional radiomics, aiming for models that are both highly accurate and inherently interpretable (Zhang et al., 2024; Li et al., 2025).

## 2.3 Summary

Existing research firmly supports a central strategy for non-invasive NSCLC classification: the fusion of distinct, yet complementary, data streams—specifically deep CNN embeddings, 3D radiomics, and clinical metadata—into one unified predictive engine (Li et al., 2025; Liu et al.,

2024; Zhang et al., 2024; Aksu et al., 2025). Studies consistently show that ensemble models, especially those built on a stacking architecture, thrive on this synergy; they regularly outperform simpler, unimodal baselines in both accuracy and reliability (Mao et al., 2025; Li et al., 2025).

Yet, this body of work also makes one thing clear: the biggest hurdle for clinical adoption isn't just performance—it's interpretability. High accuracy is vital, of course, but it is insufficient on its own; clinicians demand a "why" behind every prediction. To bridge this gap, a multi-layered XAI framework—using Grad-CAM for visual validation, SHAP for global feature weighting, and LIME for patient-specific insights—is the essential next step in establishing clinical trust (Li et al., 2024; Wang et al., 2024; Sharma et al., 2025; Ganie et al., 2025).

This thesis tackles the challenge of data scarcity head-on through a novel ROI-Imputation Strategy, delivering a fully transparent Stacking Ensemble that is rigorously checked by a multi-modal explainability protocol (Bakr et al., 2017; Rahman, 2025; Aksu et al., 2025). By prioritizing both raw performance and clear transparency, this work aims to provide a robust, interpretable decision-support tool for distinguishing between ADC and SCC (Liu et al., 2024; Li et al., 2025).

# CHAPTER 3: METHODOLOGY

## 3.1 Introduction

### 3.1.1 Chapter Overview and Objectives

This chapter meticulously details the methodological pipeline developed for the accurate and interpretable classification of Non-Small Cell Lung Cancer (NSCLC) subtypes, specifically distinguishing between Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SCC). The process is structured into three progressive phases: Phase I establishes a deep learning image-based classification baseline; Phase II involves multimodal feature extraction (Deep CNN embeddings, Radiomics, and Clinical data); and Phase III culminates in a fusion-based ensemble classification model. The core objective of this chapter is to provide a comprehensive, reproducible description of the data curation, rigorous preprocessing steps, model architectures, and validation protocols used throughout this study.

### 3.1.2 Study Design Rationale (Justification for a Multimodal, Interpretable Approach)

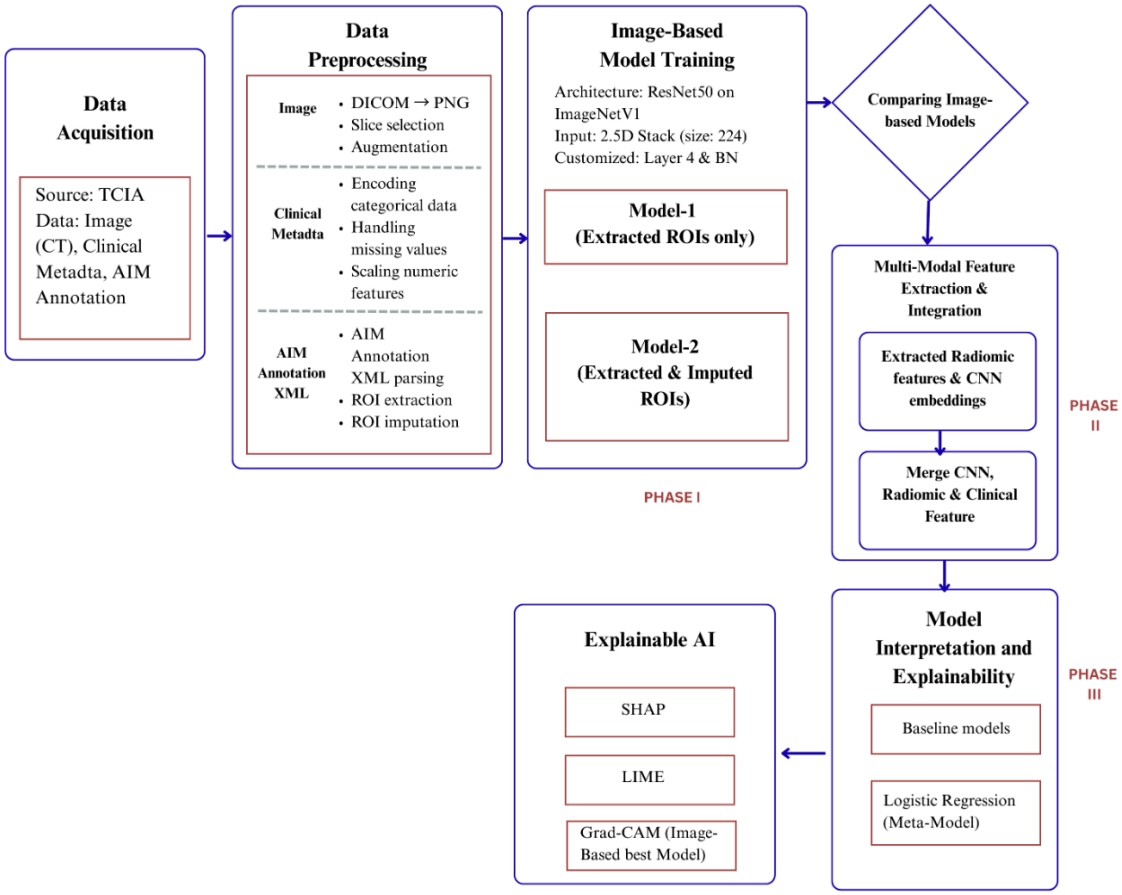


Figure 3.1 Overview of the Multimodal, Interpretable NSCLC Classification Methodology

The design of this study is grounded in the necessity for an approach that is both highly accurate and clinically relevant. Purely image-based deep learning models, while powerful, often lack the contextual information provided by traditional clinical data and the quantitative rigor of handcrafted radiomics. By adopting a multimodal fusion strategy, we aim to capture complementary information across these diverse data domains, thereby boosting classification performance and, crucially, clinical generalizability.

Furthermore, the model's reliability in a clinical setting is directly tied to its interpretability. Therefore, a secondary, yet vital, component of this methodology is a robust explainability framework, employing techniques like LIME, SHAP, and Grad-CAM to ensure that predictions are

validated by features that are clinically meaningful, building essential trust in the decision-support system.

## 3.2 Data Acquisition and Cohort Curation

The very foundation of any strong machine learning study, of course, rests upon high-quality, relevant data. This section meticulously details our methodical process: how we carefully identified the data source, established strict inclusion criteria, and, finally, prepared the patient cohort for all subsequent analyses.

### 3.2.1 Data Source and Selection Criteria

To ensure the utmost clinical fidelity—and, critically, public reproducibility—all our imaging, clinical, and annotation data were exclusively sourced from The Cancer Imaging Archive (TCIA) (Bakr et al., 2017), a widely respected repository for oncological imaging data.

#### 3.2.1.1 Data Procurement from The Cancer Imaging Archive (TCIA)

Our data procurement strictly adhered to the TCIA Data Usage Policy and Restrictions. The core components required were the Images and Segmentations (metadata), the Clinical data (CSV or Excel files), and the corresponding AIM Annotations (XML files), which meticulously define the tumor boundaries. A critical filtering step was immediately applied to the image metadata: we retained only series acquired with the CT modality, and selected just one primary CT series per patient to actively prevent data redundancy or potential bias. This specific series was downloaded in a compliant and efficient transfer process using the National Biomedical Imaging Archive (NBIA) data retriever tool.

#### 3.2.1.2 Inclusion Criteria (NSCLC cases, CT Modality, Availability of Clinical/XML Data)

We followed a rigorous, systematic process to refine the initial dataset into a viable cohort, emphasizing the completeness of three essential data types: imaging, histology, and ROI definition. The following inclusion criteria were strictly applied: only Non-Small Cell Lung Cancer (NSCLC) cases with CT imaging were included. For label integrity, cases lacking a confirmed, specific histological subtype (ADC or SCC) were immediately excluded. Furthermore, the initial clinical metadata revealed significant variability—a major issue—so a specific series, 'AMC,' was entirely removed due to pervasive missing values, particularly in the crucial 'Tumor Location' field needed for our subsequent ROI imputation strategy (detailed in Section 3.4.2).

After this extensive curation, the final cohort—ready for model training and validation—comprised 134 unique patients. The distribution of the core histological subtypes within this final, clean cohort was 75.93% ADC and 21.60% SCC.

### 3.2.2 Patient-Wise Data Splitting

To ensure the models' clinical generalizability and, just as importantly, to meticulously avoid data leakage, the curated 134-patient cohort was divided at the patient level—never the slice level—into Training and Validation sets. This splitting was carefully executed to maintain the intrinsic distribution of the histological classes, specifically the ADC-to-SCC ratio, which is vital given the original data's inherent class imbalance. This strategy guarantees that the model is honestly tested on patients it has never encountered, providing a true assessment of its predictive power.

Table 3.1 Patient Cohort Distribution by Histological Subtype and Data Split

| Cohort | Total Patients | Adenocarcinoma (ADC) | Squamous Cell Carcinoma (SCC) |
|---|---|---|---|
| Training Set | 110 | 86 | 24 |
| Validation Set | 24 | 18 | 6 |

This patient-wise splitting, while ensuring class proportion consistency, completely and fundamentally separates the data used for model learning (Training) from the data used for performance tuning and initial assessment (Validation). This segregation is a fundamental safeguard against over-optimistic or misleading results.

## 3.3 Image Data Processing and Slice Curation

The process of translating raw data into a structured, optimized format for deep learning is absolutely paramount to model success. This crucial phase involved meticulous file conversion, standardized image normalization, and the careful clinical curation of relevant image slices.

### 3.3.1 DICOM to Image Format Conversion

Raw CT data is initially stored in the DICOM (Digital Imaging and Communications in Medicine) format, carrying both metadata and raw pixel values corresponding to Hounsfield Units (HU)—a measure of tissue density. To make this data usable, we executed a standardized protocol

to convert these files into readily processed PNG images while ensuring all essential quantitative information was preserved.

#### 3.3.1.1 DICOM to PNG Conversion Protocol

The conversion script, implemented using the ‘pydicom’ and ‘PIL’ libraries, followed three essential steps for quality control and normalization:

I. **Hounsfield Unit (HU) Transformation:** The raw pixel values were converted to true HU values using the DICOM tags ‘RescaleSlope’ and ‘RescaleIntercept’ (Equation 3.1), which accurately reflects tissue density.

$$\text{HU} = \text{Pixel Value} \times \text{RescaleSlope} + \text{RescaleIntercept}$$

Equation 3.1 Hounsfield Unit (HU) Transformation Formula

II. **Windowing and Normalization:** To optimize the image for visual inspection and feature extraction, the HU values were clipped to a standard "Lung Window" range, specifically [-1000, 400] HU. This range effectively highlights the soft tissue, parenchyma, and tumor boundaries while suppressing high-density structures like bone. The clipped data was then linearly normalized to an 8-bit integer range [0, 255] and saved as a PNG image.

III. **Resampling and Centering:** To achieve spatial uniformity and manage input dimensions for the CNN, an isotropic resolution was enforced. Images were resampled based on the DICOM PixelSpacing tag to adjust for non-square pixels. Finally, a central, square region of interest containing the bulk of the lung field was extracted and resized to a 256 times 256 pixel resolution, standardizing the input geometry across all patients.

#### 3.3.1.2 Windowing Selection for Optimal Visual and Feature Extraction

Although the conversion protocol applied a fixed lung window range for quantitative consistency, a crucial qualitative control step was enforced during the slice curation phase. This involved visually inspecting the output PNG images to exclude slices that did not contain sufficient morphological information of the tumor.

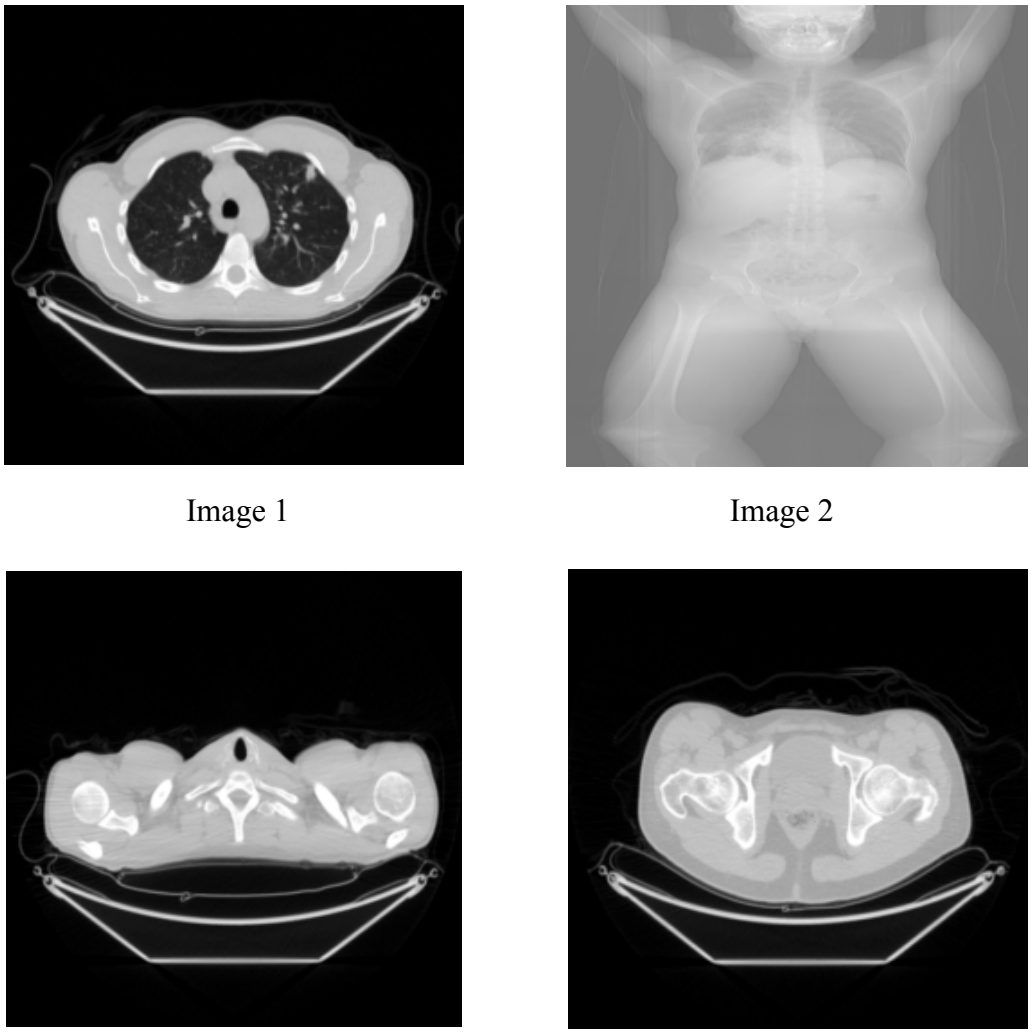

Image 1 Image 2

Image 3 Image 4

Figure 3.2 Visual Examples of CT Slice Curation and Exclusion Criteria

Specifically, only slices that displayed the tumor nodule surrounded by relevant lung parenchyma (Figure 3.3, Image 1) were retained. Slices were excluded if they represented:

I. **Peripheral slices (Image 2):** Slices at the very top or bottom of the CT volume, often showing only the apex or base of the lung field and lacking the main tumor mass.
II. **Empty or non-diagnostic slices (Image 3):** Slices containing excessive noise or only non-thoracic anatomy.
III. **Inappropriate windowing artifacts (Image 4):** Although the lung window was applied during conversion, slices demonstrating clear visual artifacts or over-saturated areas were removed.

This manual inspection ensured that the deep learning model was exposed exclusively to high-fidelity, tumor-centric data slices.

### 3.3.2 Best Slice Selection Heuristic

Deep learning at the slice level mandates that the selected slices contain the most representative visual information of the tumor. We adopted a middle-slice selection strategy based on the principle that the most critical diagnostic features are often concentrated in the slices intersecting the tumor's center.

#### 3.3.2.1 Middle-Slice Selection Strategy

For each patient, the entire series of validated PNG slices was sorted numerically by slice index. A proportional selection method was then applied to extract a defined range of central slices. This was calculated by starting the selection at the 25% mark of the total available slices (start = Total Slices × 0.25) and proceeding for a fixed slice count.

The logic was carefully adapted for each histology, reflecting the empirical observation that SCC tumors often display a larger and more complex tumor mass requiring a slightly wider contextual window:

I. **Adenocarcinoma (ADC) Patients:** A window of 25 central slices was selected.
II. **Squamous Cell Carcinoma (SCC) Patients:** A wider window of 30 central slices was selected.

This method ensured that the core region of the tumor was always included, providing a consistent and robust input for the slice-level deep learning model while mitigating the noise from peripheral, tumor-free slices.

#### 3.3.2.2 Curated Slice Count per Patient

The final curated image dataset, used as input for the Phase I deep learning model, comprised the following aggregated slice counts based on the patient cohort defined in Section 3.2.2:

I. **Training Set:**
    A. ADC: 86 cases × 25 slices per patient.
    B. SCC: 23 patients × 30 slices + 1 patient with 27 slices (due to a slightly smaller available series).
II. **Validation Set:**
    A. ADC: 18 patients × 25 slices per patient.

B. SCC: 5 patients × 30 slices + 1 patient × 10 slices (a minority case where a limited number of diagnostic slices were available).

This yielded a total of 2967 slices for the Training set and 510 slices for the Validation set, with the majority of patients contributing a standardized number of slices based on their histological subtype.

### 3.3.3 Data Augmentation for Class Balancing

A persistent challenge in medical image classification is the inherent imbalance between disease subtypes; in our curated cohort, for instance, Adenocarcinoma (ADC) significantly outnumbered Squamous Cell Carcinoma (SCC) cases (as noted in Section 3.2.2). To effectively mitigate the resultant training bias, we implemented a targeted data augmentation strategy, applied exclusively to the minority SCC class.

#### 3.3.3.1 Rationale for Augmenting Squamous Cell Carcinoma (SCC) Slices Only

The primary, driving goal of this augmentation was to dramatically increase the effective size of the SCC training set, thereby achieving a much more balanced class representation—specifically, a 1:1 ratio of ADC to SCC slices. This strategic move was essential for preventing the model from becoming overly biased toward the majority ADC class.

A critical design choice involved the use of patient-consistent augmentation. Instead of randomly applying transforms slice-by-slice, a single, fixed set of mild transformations was randomly selected and then applied uniformly to all 27 or 30 slices belonging to a given SCC patient.

Necessity of patient-consistent augmentation: In medical imaging, applying extreme or randomized transformations slice-by-slice can dangerously disrupt the tumor's crucial three-dimensional spatial relationship and clinical context across the CT volume. By applying a consistent rotation and a minor brightness/contrast shift to all slices of one patient, we effectively simulate only minor, clinically plausible variations—for example, a slight head tilt during the scan or subtle differences in scanner calibration. This process generates a distinct, yet clinically realistic, "virtual" patient dataset, successfully doubling the size of the SCC cohort without introducing non-clinical artifacts. This initial, pre-storage augmentation resulted in physically doubling the SCC slice count for 23 patients (from 30 to 60 slices) and for the single exceptional patient (from 27 to 54 slices).

#### 3.3.3.2 Augmentation Techniques Applied

The overall augmentation strategy employed during image processing utilized two distinct, complementary methods to enhance data variability:

I. **Pre-processing, Patient-Consistent Augmentation (SCC only):** As detailed above, these fixed, mild transformations were applied, and the resulting images were physically saved to disk before model training to double the SCC training data volume. The transformations were purposefully limited to ensure medical relevance:
   A. **Rotation:** A small, random rotation within ±5 degrees.
   B. **Brightness/Contrast:** Minor, random shifts within a factor range of 0.9 to 1.1.
   C. **Result:** This process physically saved a new set of augmented PNG files (prefixed with "Aug_"), effectively creating two versions of each SCC slice.

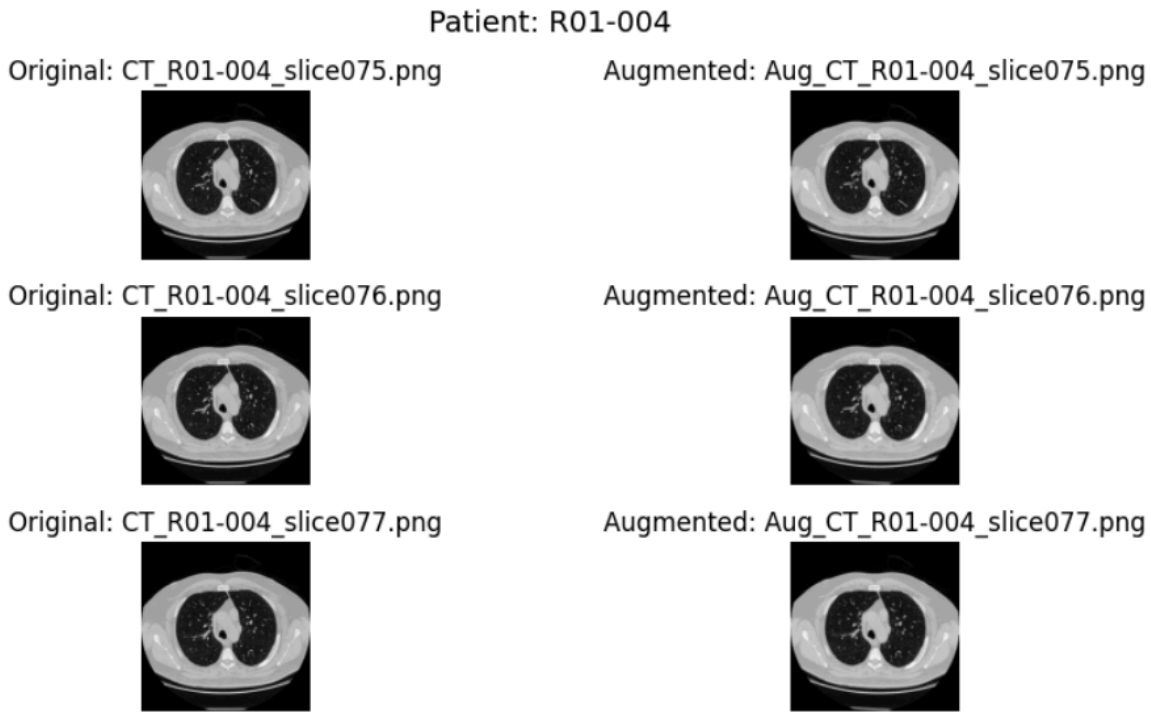


Figure 3.3 Illustration of Pre-processing, Patient-Consistent Data Augmentation for SCC Slices

Visual representation of the fixed transformations (rotation, brightness/contrast) applied to the minority class before training.

II. **Dynamic, Run-time Augmentation (ADC and SCC):** During the actual model training, PyTorch's SliceDataset class applied further, dynamic transformations to every slice (including the pre-processed SCC slices) immediately before ingestion by the model. These

dynamic steps included fixed resizing, a Random Horizontal Flip (50% probability), and a small, random rotation of up to 10 degrees.

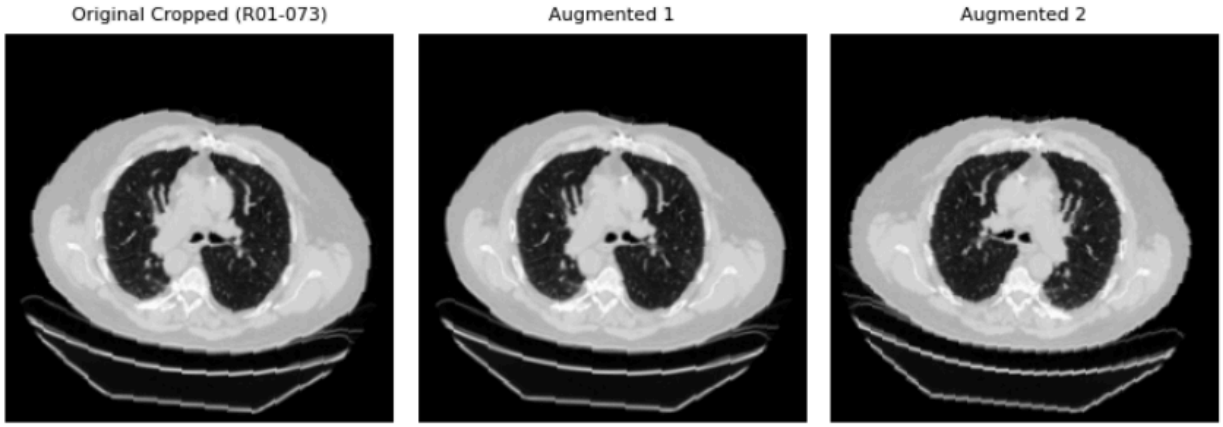


Figure 3.4 Dynamic, Run-time Data Augmentation Techniques Applied During Training

Figure 3.4 illustrates the secondary, dynamic augmentations (e.g., Random Horizontal Flip) applied during the model ingestion stage.

This two-tiered strategy ensured both a direct correction of the class imbalance (via pre-processing doubling of SCC data) and a robust increase in overall generalization capability (via dynamic, random transformations for both classes). To fine-tune the training further against the persistent imbalance, the final data loader also incorporated a Weighted Random Sampler based on inverse class frequency, applying an additional penalty to the majority ADC class during batch selection:

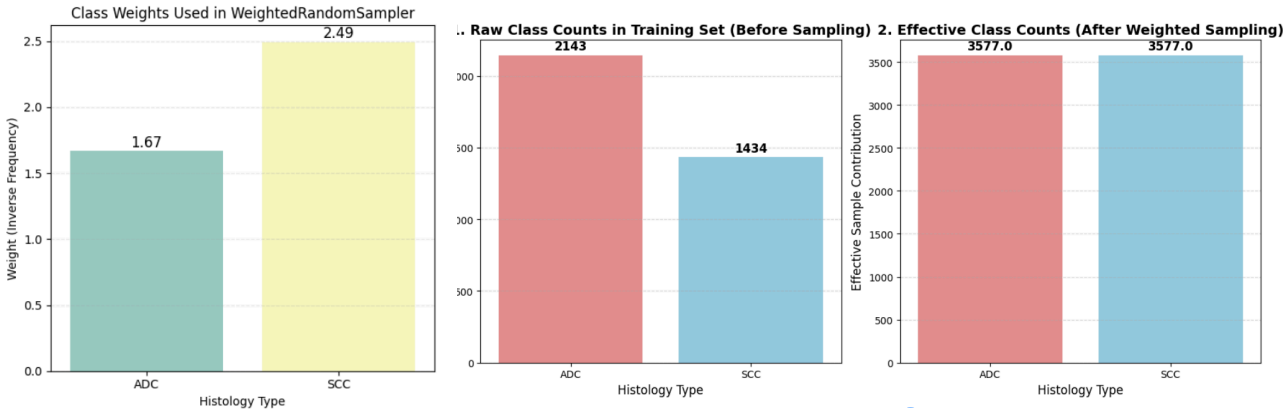


Figure 3.5 Implementation of a Weighted Random Sampler to Fine-Tune Batch Selection

The figure 3.5 represents the final class balancing step in the data loader.

## 3.4 Annotation and Clinical Data Processing

The second component of the predictive model, referred to as Model-2, incorporates clinical and pathological metadata to augment the image-derived features. Clinical Data is needed for ROI-imputation for Model-2 and the subsequent Fusion Model. This section details the steps taken to curate and preprocess the tabular clinical data for machine learning integration.

### 3.4.1 Processing of Clinical Metadata

Clinical metadata, sourced from patient records and pathology reports, was standardized across the entire cohort (Training and Validation sets) before feature integration. The target variable for all subsequent models remains Histology (Adenocarcinoma vs. Squamous Cell Carcinoma).

#### 3.4.1.1 Clinical Data Feature Selection

The selection of clinical features was guided by established literature regarding prognostic factors for non-small cell lung cancer (NSCLC) histology and recurrence. Features related to tumor pathology, patient demographic information, and primary tumor site were prioritized. The final set of selected clinical variables used for Model 2 included-

Table 3.2 Selected Clinical and Pathological Features for Multimodal Integration

| Feature Category | Variables Selected | Rationale |
|---|---|---|
| Patient Demographics | AgeAtHistologicalDiagnosis, Gender, SmokingStatus, PackYears | Standard risk factors and patient health profile. |
| Tumor Location (Categorical) | TumorLocation (RUL, RML, RLL, LUL, LLL, LLingula) | Location within the lung lobes is crucial, as specific histologies show preference for certain regions (e.g., SCC often centrally located). |

| | | |
|---|---|---|
| Pathological Staging (TNM) | PathologicalTStage, PathologicalNStage, PathologicalMStage | Standard clinical predictors of disease extent and prognosis. |
| Pathological Factors | HistopathologicalGrade, LymphovascularInvasion, PleuralInvasion | Direct indicators of tumor aggressiveness and spread potential. |
| Missingness Flag | LVI_missing_flag | A custom binary flag created during data curation to manage uncertainty in the LymphovascularInvasion status. |

#### 3.4.1.2 Handling of Missing Data and Categorical Encoding

The clinical metadata required extensive cleaning, feature engineering, and standardization before it could be input into tree-based ensemble models:

I. **Categorical Encoding:** All nominal and ordinal categorical features (e.g., Tumor Location, TNM Stages, Gender) were first converted to the category data type in the Pandas framework. Missing values within these columns were explicitly imputed with the placeholder string 'Missing_Category'. Subsequently, Label Encoding was applied, converting each unique category (including the missing placeholder) into a numerical code. This approach was preferred over one-hot encoding to reduce the overall feature dimension, which is beneficial given the small cohort size.
II. **Continuous Data Standardization:** Continuous variables (AgeAtHistologicalDiagnosis, PackYears) were standardized using the StandardScaler from the scikit-learn library. This process transformed the features to have a mean of 0 and a standard deviation of 1, ensuring that the magnitude of these variables did not disproportionately influence the model weights.
III. **Specialized Handling for Lymphovascular Invasion (LVI):** The LymphovascularInvasion (LVI) field, a critical prognostic factor, often suffered from documentation variability, resulting in missing or ambiguous entries. To avoid discarding valuable data, a unique strategy was employed:

A. A new binary feature, LVI_missing_flag, was created during the initial data curation step. It was set to 1 if the LVI status was missing or ambiguous, and 0 otherwise.
B. During the preprocessing phase, after the categorical Label Encoding was applied, the following mask logic was executed: If LVI_missing_flag was 1 for a patient, the corresponding numerical value in the LymphovascularInvasion column was intentionally reset to a neutral code, typically 0.
C. This strategy allowed the model to treat the LymphovascularInvasion feature as its recorded pathological value only when the data was certain, and simultaneously leverage the LVI_missing_flag as an independent indicator of data reliability.

IV. **Final Feature Assembly:** The preprocessed clinical features were combined with the image-derived feature set (radiomics and CNN probabilities) to form the final input for the machine learning models. Column names were sanitized (e.g., replacing parentheses and special characters with underscores) to ensure compatibility with modern gradient-boosting frameworks like LightGBM and XGBoost.

### 3.4.2 Region of Interest (ROI) Extraction and Imputation

To sharpen the model's focus on the tumor pathology and reduce distracting background noise, we deliberately adopted a "Nodule-Centric" approach; this required extracting a specific Region of Interest (ROI) defined by the tumor's spatial boundaries. However, due to incomplete annotation data for a portion of the cohort, a robust imputation strategy became necessary to estimate these boundaries where explicit data was simply missing.

#### 3.4.2.1 ROI Extraction from Standardized XML Annotations

For the majority of the cohort, tumor boundaries were derived directly from the expert annotations provided in the TCIA AIM XML files. This extraction process adhered to a strict geometric transformation protocol: first, Coordinate Parsing extracted the boundary coordinates from the XML; then, the geometric Center ($C_x$, $C_y$) and Radius (R) of the nodule were calculated in the original DICOM space. Finally, to align with our 256 x 256 input images, a Spatial Scaling factor of 0.5 was precisely applied to all coordinates and the radius. For Model-1 (Baseline), any patient lacking an XML annotation simply had their entire slice resized to 224 x 224, foregoing specific ROI cropping.

#### 3.4.2.2 Three-Tier ROI Imputation Strategy (Model-2)

A significant methodological contribution of this work is the handling of the 14 patients (about 10% of the cohort) who lacked XML annotations. Excluding them would have notably reduced our statistical power; instead, we devised a sophisticated Three-Tier Imputation Strategy to estimate the most probable ROI coordinates ($C_x$, $C_y$, R). This strategy is grounded in the

"Nearest Neighbors" principle, positing that tumors of the same Histology occurring in the same Anatomical Location are likely to share similar spatial characteristics.

The imputation logic proceeded hierarchically:

I. **Tier 1: Primary Match (Histology and Location):** When a patient had both a known Histology (ADC/SCC) and a specific Tumor Location, the algorithm found all "neighbor" patients in the annotated data sharing these two characteristics. The imputed ROI was then defined as the centroid (mean $C_x$, $C_y$, R) of this very specific subset; this effectively captured the typical presentation of that cancer subtype in that specific lung lobe.

II. **Tier 2: Fallback Match (Histology Only):** This tier was triggered if Tier 1 yielded insufficient neighbors (< 2 matches), suggesting a rare location for that histology in our dataset. The location constraint was relaxed, aggregating all annotated patients sharing the same Histology regardless of location. This prioritized the tumor's biological morphology over anatomical position when local data was sparse.

III. **Tier 3: Location Match (Missing Histology):** This addressed patients whose Histology was unknown or missing (NaN). The algorithm aggregated all annotated patients sharing the same Tumor Location, pooling both ADC and SCC cases. The rationale here is simple– in the absence of biological labels, anatomical constraints provide the best available prior probability for tumor position.

This robust, hierarchical approach allowed for the successful imputation of valid ROI coordinates for all 14 missing cases, enabling their crucial inclusion in the advanced Model-2 training pipeline.

#### 3.4.2.3 Validation via Nodule-Centric Visualization

To verify the fidelity of both the extracted and imputed ROIs, a comprehensive visualization audit was conducted.

I. **Visual Confirmation:** For every patient, the calculated bounding box was overlaid on the CT slice, ensuring that the imputed coordinates successfully localized the tumor mass within the crop window; this guaranteed that Model-2 would receive relevant pathological features, not just whole lung parenchyma.

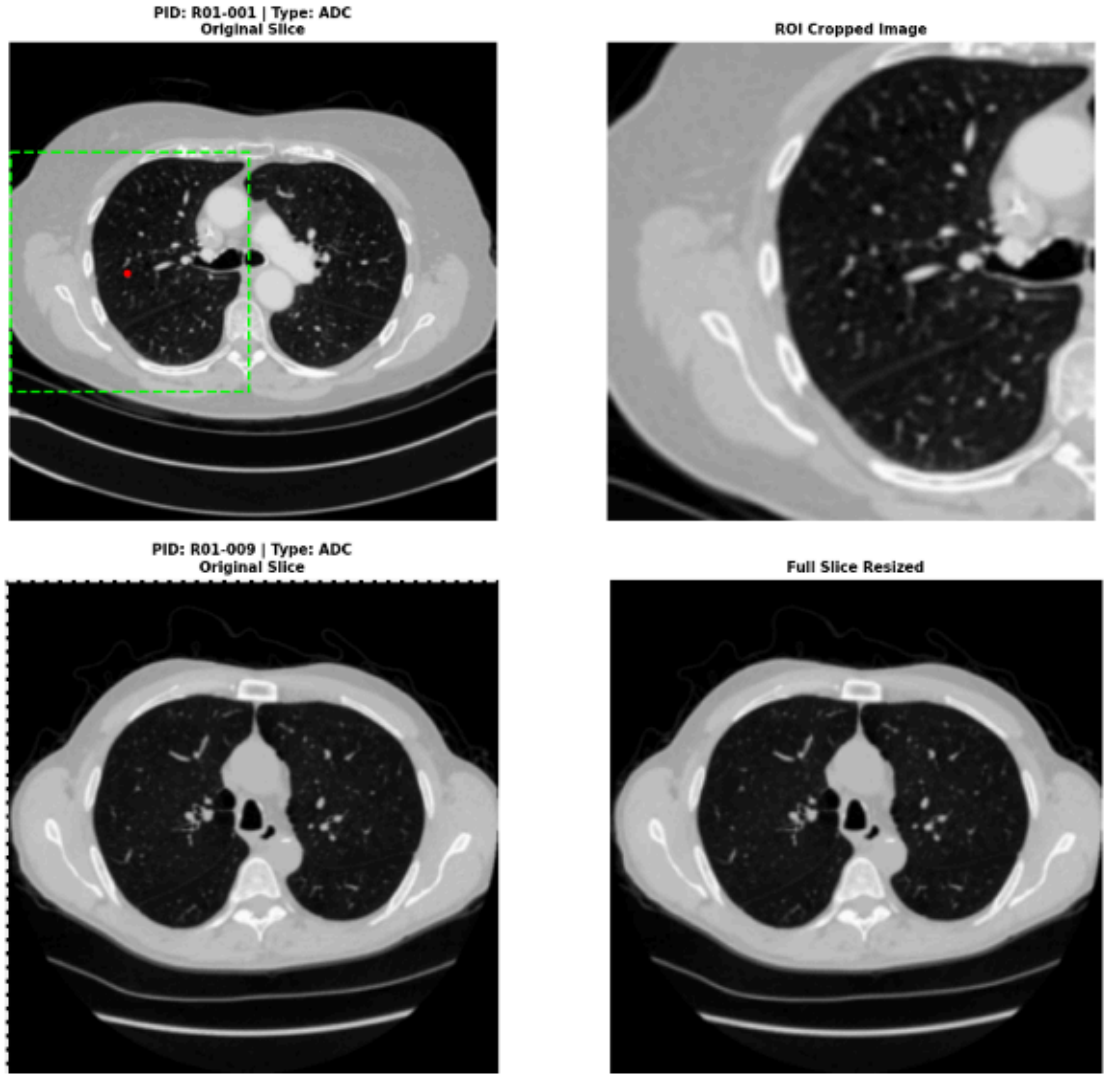


Figure 3.6 Visualizing a patient with an extracted ROI (R01-001) & another with missing ROI (R01-009), full slice fallback logic for Model-1

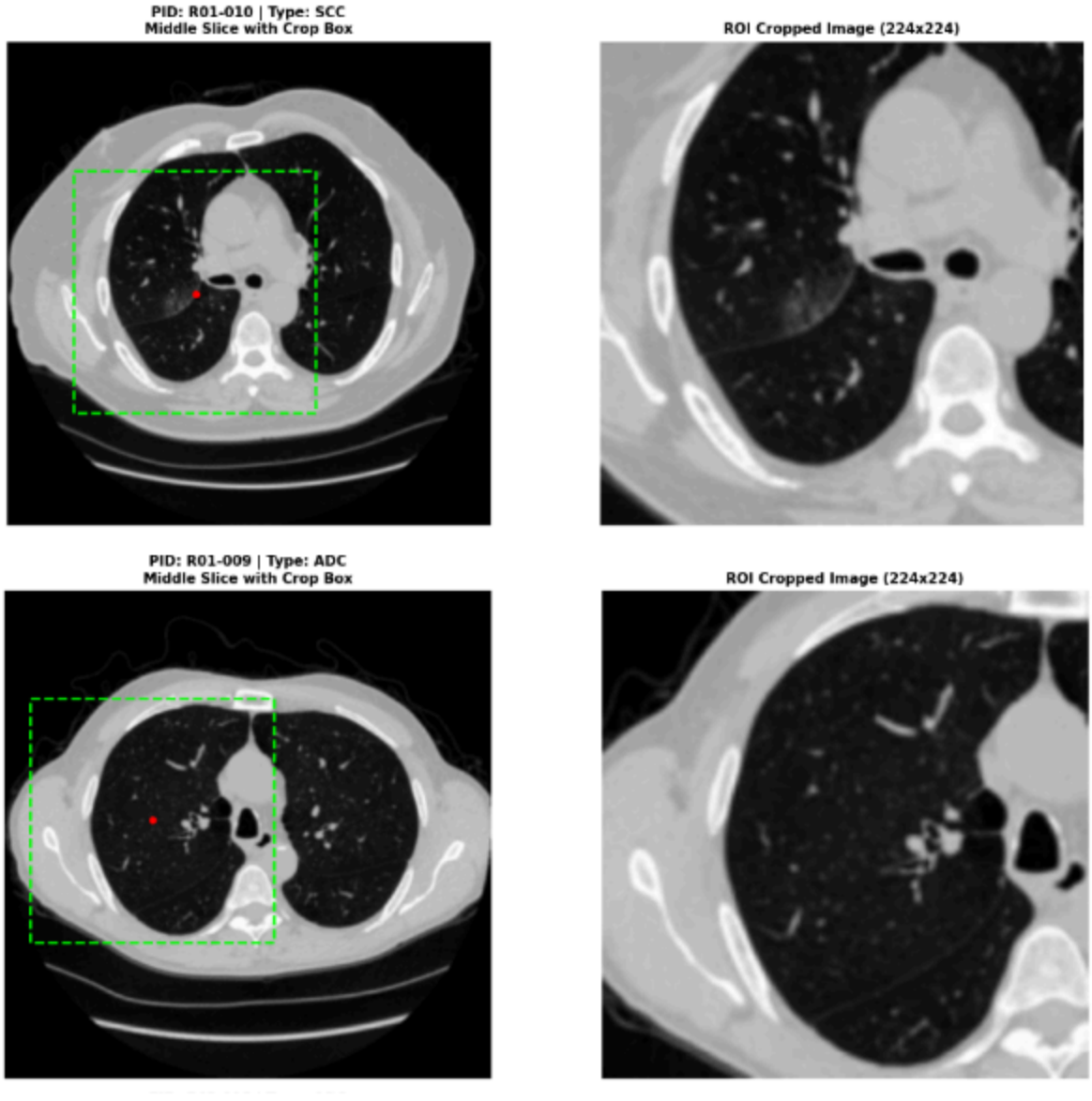


Figure 3.7 Visualizing a patient with an extracted ROI (R01-010) & another with imputed ROI (R01-009) for Model-2

Figure 3.6 and 3.7 illustrate the difference in data handling for patients with missing annotations between the two models (e.g., Model-1 full slice vs. Model-2 extracted ROI for Patient ID: R01-009).

II. **Dual-Model Dataset Creation:** This entire process resulted in two distinct image datasets for comparative analysis:
    A. **Dataset (Model-1):** Comprises Nodule-Centric slices for patients with extracted annotations and whole-slice images for patients with missing XML annotations.
    B. **Nodule-Centric Dataset (Model-2):** Focused exclusively on the tumor, utilizing the complete set of Extracted and Imputed ROIs derived from the rigorous extraction and imputation pipeline described above.

## 3.5 Deep Learning Image-Based Classification (Phase I)

Following the creation of our meticulous nodule-centric datasets, the first major phase of classification focused on establishing a robust baseline using deep convolutional neural networks (DCNNs). To overcome the inherent data scarcity common in medical imaging, we adopted a powerful Transfer Learning approach, specifically adapting a sophisticated architecture—pre-trained on large-scale natural imagery—to the very distinct domain of lung nodule histology.

### 3.5.1 Transfer Learning Model Architecture: ResNet50 on ImageNetV1

We selected the ResNet50 as our primary backbone architecture; it provides a proven balance between exceptional depth (50 layers) and computational efficiency, and its crucial residual connections actively mitigate the notorious vanishing gradient problem. The model was initialized using ImageNetV1 weights, which allowed us to immediately leverage learned feature representations—like robust edge detection and texture analysis—that are highly transferable across visual domains.

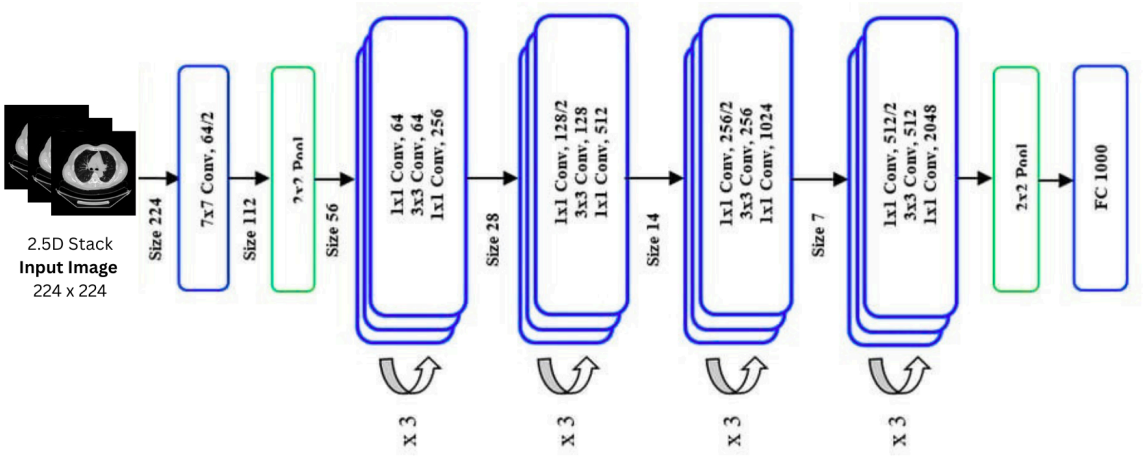


Figure 3.8 ResNet50 Architecture for Image-Based NSCLC Classification (Phase I)

#### 3.5.1.1 Backbone Selection: ResNet50 Initialization

The standard ResNet50 architecture required careful customization to meet the specific dimensional and classification needs of this study. Our implementation, built within the PyTorch framework, involved three critical structural modifications:

I. **Input Adaptation (2.5D Representation):** While CT data is grayscale (1 channel), the ResNet50 backbone requires a 3-channel input. We efficiently utilized the 2.5D stacking technique (prepared during data loading), where the target slice and its two adjacent neighbors are combined to form a 3-channel tensor (3 x H x W). This cleverly allows the 2D CNN to perceive volumetric context without the substantial computational cost of a fully 3D network.

II. **Backbone Configuration:** The core convolutional base layers were initialized with the pre-trained ImageNet weights. These layers function as our primary feature extractor, transforming raw pixel data into high-dimensional feature maps (specifically, 2048 feature channels at the final convolutional block).

III. **Custom Classification Head:** The original fully connected layer of ResNet50 was discarded. In its place, we engineered a custom, Regularized Multi-Layer Perceptron (MLP) Head to map the 2048 extracted features to our binary output (ADC vs. SCC). This head incorporated:

   A. **Dropout Layer 1 (**p=0.5**):** A high dropout rate to heavily penalize over-reliance on specific features.

   B. **Linear Layer 1:** Dimensionality reduction from 2048 to 512.

C. **Activation:** Rectified Linear Unit (ReLU) for non-linearity.
D. **Dropout Layer 2 (**p=0.3**):** A secondary, moderate regularization step.
E. **Linear Layer 2:** The final projection from 512 to 2 classes (ADC/SCC).

#### 3.5.1.2 Fine-Tuning Strategy: Unfreezing Layer 4 and Batch Normalization (BN) Layers

A simplistic transfer learning approach would have been suboptimal. Instead, we implemented a surgical fine-tuning strategy designed to retain generic visual features while simultaneously adapting the high-level semantic abstractions to our medical domain:

I. **Layer-Specific Unfreezing:** The initial convolutional layers and the first three residual blocks (Layers 1–3) were deliberately frozen; they handle low-level geometric primitives universal to vision tasks. Critically, the final residual block—Layer 4—was unfrozen (requires_grad = True); this block is responsible for constructing complex, high-level semantic features, allowing the model to learn specific lung nodule textures and boundary characteristics.
II. **Batch Normalization (BN) Adaptation:** A crucial detail was the global unfreezing of all Batch Normalization layers throughout the network. This was done to allow the model to update its running statistics (mean and variance) to accurately match the distribution of the CT data, which differs markedly from the intensity distribution of ImageNet photos.

To ensure stable convergence, we employed a Differential Learning Rate strategy using the AdamW optimizer. The Backbone (Layer 4) was assigned a lower learning rate ($1 \times 10^{-4}$) to gently adapt the pre-trained weights without causing "Catastrophic Forgetting," while the Classification Head was given a higher rate ($2.5 \times 10^{-4}$) to allow its randomly initialized weights to learn rapidly.

Training was governed by a Weighted Cross-Entropy Loss function. To address the persistent class imbalance, a weight vector—calculated based on the inverse frequency of ADC vs. SCC—was passed to the loss criterion, ensuring that errors on the minority class (SCC) were penalized more heavily. Finally, a ReduceLROnPlateau scheduler was configured to decay the learning rate if the validation F1-score stalled, ensuring fine-grained weight updates in the later training stages.

### 3.5.2 Dual Image Model Training and Comparison

The core of the image-based classification phase involved a direct, head-to-head comparison between two models trained on structurally different datasets, both utilizing the identical, fine-tuned ResNet50 architecture described in Section 3.5.1. This comparison was designed to quantitatively assess the value added by the rigorous ROI Imputation Strategy.

#### 3.5.2.1 Model-1: Training on the Hybrid Dataset (Extracted ROI + Whole Slice)

Model-1 served as the baseline, representing a standard compromise when dealing with incomplete annotation data. This model was trained on the Hybrid Dataset, which was structured as follows:

I. **Annotated Patients (Majority):** Used the precise Nodule-Centric ROIs extracted directly from the XML files. These images contain minimal background noise.
II. **Missing XML Patients (Minority):** Used the full 224 x 224 scaled CT slice, encompassing the entire lung field and surrounding anatomy.

The resulting dataset for Model-1 was heterogeneous, forcing the CNN to learn features from both tightly cropped tumors and broad anatomical contexts. While minimizing data loss, this heterogeneity potentially introduced confounding features from irrelevant lung parenchyma in the uncropped minority images.

#### 3.5.2.2 Model-2: Training on the Pure Nodule-Centric Dataset (Extracted + Imputed ROIs)

Model-2 was the primary experimental model, designed to maximize the model's focus on tumor pathology. This model was trained exclusively on the Pure Nodule-Centric Dataset, composed entirely of cropped ROIs:

I. **Annotated Patients (Majority):** Used the precise XML-extracted ROIs (identical to Model-1).
II. **Missing XML Patients (Minority):** Used the ROIs generated by the Three-Tier Imputation Strategy (Section 3.4.2.2).

By training on this fully homogeneous, nodule-centric dataset, Model-2 was compelled to develop highly specific feature detectors related to tumor morphology, texture, and boundary characteristics. This setup provided a rigorous test of the Imputation Strategy's success: if Model-2 outperformed Model-1, it would validate that the imputed coordinates were sufficiently accurate to provide diagnostic value comparable to manually extracted ROIs.

#### 3.5.2.3 Hyperparameter Settings (Learning Rate, Weight Decay, Epochs)

Both Model-1 and Model-2 were trained using identical hyperparameter settings to ensure that performance differences were attributable solely to the variation in the input datasets. The following configuration ensured stable, regularized training suitable for fine-tuning a pre-trained DCNN on medical imagery:

Table 3.3 Hyperparameter Settings for Optimization

| Hyperparameter | Value | Rationale |
|---|---|---|
| Optimizer | AdamW | Preferred over Adam for DCNNs due to better generalization and robust handling of L2 regularization (Weight Decay). |
| Loss Function | Weighted Cross-Entropy Loss | Addresses the inherent class imbalance between ADC and SCC instances by applying higher penalties to misclassified minority-class samples. |
| Learning Rate (Head) | $2.5 \times 10^{-4}$ | Higher LR for the custom, randomly initialized classification layers to facilitate rapid learning. |
| Learning Rate (Layer 4) | $1 \times 10^{-4}$ | Lower, differential LR for the pre-trained convolutional blocks (Layer 4) to ensure gradual fine-tuning and preservation of existing feature knowledge. |
| Weight Decay ($\lambda$) | $1 \times 10^{-4}$ | A moderate value used in the AdamW optimizer to prevent overfitting by penalizing large weights. |

| | | |
|---|---|---|
| Dropout Rate | 0.5 (initial head),<br><br>0.3 (secondary head) | Aggressive regularization to improve generalization and robustness against noisy medical image features. |
| Scheduler | ReduceLROnPlateau | Monitors the validation F1-score; decays the LR by a factor of 0.5 if performance plateaus for 3 epochs (Patience). |
| Training Epochs | 15 | The training was limited to 15 epochs. For final evaluation, the model checkpoint corresponding to the highest Macro F1-Score achieved on the validation set during the training run was selected and loaded for both Model-1 and Model-2. |

### 3.5.3 Performance Evaluation Protocol

The trained models were evaluated using a two-tiered protocol designed to first assess the raw predictive capability at the individual image level (Slice-Level) and then convert these predictions into a clinically meaningful outcome for the entire patient (Patient-Level).

#### 3.5.3.1 Slice-Level Metrics (Initial Model Training and Assessment)

During model training and validation, performance was tracked at the individual image slice level. Given the critical importance of correctly identifying both Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SCC), particularly considering the inherent imbalance in the dataset (ADC being the majority class), standard accuracy was deemed an unreliable metric. To address this, we relied on Macro-Averaged F1-Score and Balanced Accuracy.

I. **Macro-Averaged F1-Score (**F1_Macro**):** This metric calculates the F1-Score for each class independently and then averages them, treating all classes equally regardless of their size. This is crucial for class-imbalanced datasets as it prevents the model from achieving a high score simply by correctly classifying the large majority class.

$$F1_{Macro} = \frac{1}{C}\sum_{i=1}^{C} F1_i$$

Equation 3.2 Formula for Macro-Averaged F1-Score

where C is the number of classes (2 in this case, ADC and SCC), and $F1_i$ is the F1-Score for class i, calculated as:

$$F1_i = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

Equation 3.3 Formula for Individual Class F1-Score

II. **Balanced Accuracy (**$Acc_{Balanced}$**):** Defined as the average of Recall (True Positive Rate) achieved on each class. It is statistically equivalent to the area under the ROC curve (AUC) for binary classification and offers a more robust measure of overall performance than standard accuracy in imbalanced scenarios.

$$Acc_{Balanced} = \frac{1}{C}\sum_{i=1}^{C} \text{Recall}_i = \frac{1}{2}(\text{Sensitivity} + \text{Specificity})$$

Equation 3.4 Formula for Balanced Accuracy

III. The model checkpoint saved for final testing was the one that achieved the highest F1 Macro on the validation set, ensuring that the selected model demonstrated the best balance in classifying both ADC and SCC.

#### 3.5.3.2 Patient-Level Aggregation (Majority Voting or Averaging)

In a clinical context, the objective is to predict the histology of the entire patient, not just a single CT slice. Since each patient in the test set has multiple associated nodule-centric slices, a mechanism was required to aggregate these slice-level predictions into a single, definitive Patient-Level Classification.

This conversion was achieved by utilizing two aggregation methods:

I. **Prediction Averaging (Probability Consensus):** The logit or probability scores for all slices belonging to a single patient were averaged. The final patient classification was determined by the class with the highest resulting average probability. This method leverages the model's confidence scores across the entire nodule volume.

$$P_{\text{Patient}}(\text{Class}) = \frac{1}{N}\sum_{j=1}^{N} P_{\text{Slice}_j}(\text{Class})$$

Equation 3.5 Patient-Level Probability Consensus (Prediction Averaging)

where N is the number of slices for the patient and $^{P}\text{Slice}_{j}$ is the predicted probability for a given class from slice j.

II. **Majority Voting (Hard Consensus):** The hard binary prediction (ADC or SCC) from each slice was counted. The final patient classification was assigned to the class that received the majority of the slice-level votes. This method is less sensitive to small variations in probability scores.

The primary results were reported using the averaging method, as it utilizes the full range of predictive information (probabilities) rather than just the thresholded class labels, providing a more robust measure of the model's certainty. This aggregation step is crucial for translating the technical performance of the DCNN into a clinically relevant output.

## 3.6 Multimodal Feature Extraction and Integration (Phase II)

The second major phase marked a crucial transition: moving from purely image-based classification to a powerful multimodal approach. This phase focused on integrating latent deep features, quantitative, handcrafted Radiomics, and traditional clinical variables. This sophisticated

fusion strategy aims to capture complementary information across diverse data domains, ultimately boosting both classification performance and clinical interpretability.

### 3.6.1 Extraction of Deep Convolutional Network (CNN) Features

#### 3.6.1.1 Feature Extraction Layer Selection

To effectively leverage the highly optimized visual representations learned by the deep neural network, we extracted features from Model-2. This model was selected because it demonstrated superior performance in Phase I by utilizing the fully Nodule-Centric dataset (Extracted + Imputed ROIs). The specific features extracted were the activations of the model's penultimate layer—the output of the Global Average Pooling (GlobalAvgPool) layer within the ResNet50 backbone, located immediately before the final custom classification head. This layer yields a 2048-dimensional feature vector (embedding) for each image slice. These 2048 dimensions represent the highest-level, most discriminative visual abstractions learned by the model regarding tumor texture, shape, and context, independent of the final binary classification decision.

#### 3.6.1.2 Aggregation of Slice-Level Embeddings to Patient-Level Features

Since the final multimodal model operates strictly at the patient level, the 2048-dimensional embeddings generated for each individual slice (which varied from 10 to 30 slices per patient) had to be reliably aggregated. The aggregation involved calculating the Mean Average of the feature vector across all nodule-centric slices belonging to a single Patient Identifier (PID). This simple yet effective averaging operation successfully compresses the volumetric information into a single 1 x 2048 vector, resulting in a stable, patient-specific CNN feature set ($F_{CNN}$).

### 3.6.2 Radiomics Feature Extraction

Radiomics involves the high-throughput extraction of quantitative features from medical images (CT in this case), providing numerical representations of a tumor's morphology, intensity, and intricate texture.

#### 3.6.2.1 Radiomics Library and Configuration

Radiomic features were extracted using the open-source PyRadiomics library, standardized to ensure reproducibility. The extraction was performed on the original DICOM data, utilizing the binary mask derived from the Nodule-Centric ROI (Section 3.4.2) as the specific input region. Standard configuration parameters were applied: no explicit image filters were used, and a fixed bin width of 25 HU was utilized for intensity discretization, ensuring consistent gray-level quantization for texture matrix calculation. Crucially, features were extracted in 3D, capitalizing on the full volumetric information available within the contiguous nodule slices.

#### 3.6.2.2 Feature Selection and Normalization

A comprehensive set of features spanning three main categories was extracted, normalized, and selected for the fusion model:

I. **First-Order Statistics:** These quantify the distribution of voxel intensities within the ROI. Features like Mean, Max, and Standard Deviation reflect tumor density and heterogeneity.
II. **Gray Level Co-occurrence Matrix (GLCM):** These quantify the spatial relationship and **texture** between voxels. Features such as Contrast, Homogeneity, and Correlation quantify tumor roughness, complexity, and invasiveness.
III. **Gray Level Size Zone Matrix (GLSZM):** These quantify gray level zones (connected regions of voxels with the same gray level). Features like Number of Regions and Mean Region Area measure coarseness and size variability, reflecting internal structure and growth patterns.

All extracted Radiomic features were subsequently normalized using Z-score standardization (standard scaling). This is a vital step to ensure that features with larger numerical ranges—such as certain size features—do not unfairly dominate the fusion model training process.

### 3.6.3 Feature Data Consolidation and Alignment

The final step of Phase II involved consolidating the three distinct data types—$F_{CNN}$, $F_{Radiomics}$, and $F_{Clinical}$—into a single, unified feature matrix suitable for training the fusion classifier (Phase III).

#### 3.6.3.1 Merging of CNN Embeddings, Radiomics, and Preprocessed Clinical Features

The three feature sets were concatenated column-wise, utilizing the common Patient Identifier (PID) as the key. The clinical features, which included previously encoded categorical variables (e.g., PathologicalTStage, TumorLocation), had already been preprocessed (Section 3.4.1). This alignment process strictly adhered to the stratified training and validation splits established in Phase I. Separate feature files were generated for each split, ensuring that the fusion model was trained and evaluated on the exact, identical patient cohorts as the image model. This experimental consistency prevents data leakage and guarantees a fair comparison across all phases of the study.

#### 3.6.3.2 Feature Set Alignment via Patient Identifier (PID)

The Patient Identifier (PID) served as the unique key for all merging operations. The aggregation of slice-level CNN and Radiomics features into patient-level vectors inherently ensured that each row in the final feature matrix corresponded to a unique patient in the study. The final consolidated feature matrix thus contained $N_{patients}$ x $N_{features}$ dimensions, making it perfectly ready for the final classification phase.

# 3.7 Multimodal Ensemble Learning and Classification (Phase III)

Phase III focused on the critical development and rigorous testing of our multimodal ensemble learning framework. This powerful approach utilized the consolidated patient-level features ($F_{CNN}$, $F_{Radiomics}$, $F_{Clinical}$) to predict lung cancer histology, with the fundamental goal of significantly boosting performance and robustness beyond individual feature streams.

## 3.7.1 Baseline Classification Models

### 3.7.1.1 Selection of Diverse Baseline Models

To establish a comprehensive performance benchmark for the final ensemble, we selected and trained **seven diverse machine learning classifiers**. This set primarily focused on high-performance tree-based and boosting algorithms, alongside a key linear model for necessary interpretability; this ensured we covered distinct modeling approaches within the multimodal feature space:

I. **Linear Model: Logistic Regression (LR)**, utilized as an interpretable baseline to assess the inherent linear separability of the feature space.
II. **Single Tree:** A **Decision Tree (DT)**, included specifically to benchmark the predictive capacity of a simple, non-ensemble tree structure.
III. **Bagging Ensemble:** The **Random Forest (RF)**, a robust model leveraging bootstrap aggregation (**bagging**) to effectively reduce variance and capture feature interactions.
IV. **Adaptive Boosting: AdaBoost (AB)**, an ensemble meta-algorithm that iteratively improves weak learners, heavily focusing on previously misclassified instances.
V. **Extreme Gradient Boosting: XGBoost (XGB)**, a highly optimized, scalable tree boosting system renowned for delivering superior predictive accuracy in structured data.
VI. **Gradient Boosting Frameworks: LightGBM (LGB)**, an efficient, parallelized framework optimized for speed and low memory using a histogram-based approach, and **CatBoost (CB)**, which natively handles categorical features and uses ordered boosting to combat prediction shift, improving generalization.

### 3.7.1.2 Training and Evaluation of Baseline Models on Multimodal Features

Each baseline model was trained using the identical multimodal feature matrix (the combined $F_{CNN}$, $F_{Radiomics}$, and $F_{Clinical}$ ) generated in Section 3.6. Training involved standard hyperparameter tuning using cross-validation on the training set, meticulously optimizing for Balanced Accuracy. The final performance of these seven models on the held-out test set provided the necessary context to gauge the relative effectiveness of the subsequent ensemble model.

## 3.7.2 Development of the Fusion Ensemble Meta-Model

### 3.7.2.1 Ensemble Architecture

We implemented a Stacking Ensemble framework to intelligently capitalize on the inherent strengths of our diverse baseline models. This architecture, often termed a Meta-Model, involves two distinct tiers:

I. **Level 0 (Base Learners):** The seven selected classifiers (LR, DT, RF, AB, XGBoost, LightGBM, CatBoost) were trained on the full multimodal feature set. Crucially, their outputs were collected as predicted class probabilities (logits), not hard class predictions.
II. **Level 1 (Meta-Classifier):** A single, final classifier—a Logistic Regression model—was then trained on the combined set of probability predictions (the "stacked features") generated by the Level 0 base learners. The choice of a simple, linear model for this meta-classifier was intentional: it learns the optimal weighted combination of the base learners' strengths without introducing excessive complexity or the risk of overfitting to the training data.

#### 3.7.2.2 Training Optimization for Clinical Performance

The training of the ensemble was guided by metrics directly relevant to clinical utility and robustness against class imbalance. The primary optimization objective for the Level 1 Meta-Classifier was the maximization of the Area Under the Receiver Operating Characteristic Curve (AUC) and Balanced Accuracy on the validation set. This emphasis on AUC ensures the model is well-calibrated across all decision thresholds—vital for a diagnostic tool where the threshold may need clinical adjustment. Furthermore, the use of a simple linear meta-model aids in maintaining interpretability of the relative contributions of the base learners.

### 3.7.3 Final Model Evaluation

#### 3.7.3.1 Performance Metrics

Final evaluation of the ensemble model was conducted on the validation set. The key metrics reported for patient-level classification included:

I. **Macro F1-Score:** Our primary metric for balanced classification performance.
II. **Balanced Accuracy:** A robust measure of overall accuracy in imbalanced classification scenarios.
III. **Sensitivity (Recall for SCC):** Crucial for minimizing false negatives of the less prevalent SCC.
IV. **Specificity (Recall for ADC):** Measures the ability to correctly identify the majority ADC cases.

#### 3.7.3.2 Comparison of Ensemble Model against Baseline and Image-Only Model-2

The final performance of the Ensemble Meta-Model was rigorously benchmarked against two critical reference points: the Best-Performing Baseline Model (to quantify the gain achieved

by ensemble stacking) and the Image-Only Model-2 from Phase I. This last comparison was absolutely essential to demonstrate the incremental value added by fusing deep visual features with Radiomics and clinical metadata, thereby validating the entire multimodal strategy. Superior performance by the Ensemble Model definitively confirms the hypothesis: combining diverse data streams yields a more robust and accurate patient-level prognostic tool.

## 3.8 Model Interpretation and Explainability

Model explainability is absolutely crucial in clinical decision support; it builds trust and validates that the model's predictions are truly based on clinically relevant features, not just spurious correlations. To achieve this, we utilized three distinct methods, providing interpretations that span local (instance-specific) and global (feature importance) scales, plus a vital visual interpretation for the image component.

### 3.8.1 Local Interpretation with LIME

LIME (Local Interpretable Model-agnostic Explanations) was applied to the final multimodal Stacking Ensemble to interpret individual patient predictions. It works by approximating the complex model's behavior around a single instance using a simple, easily understood model.

#### 3.8.1.1 Application of LIME at the Patient-Level

LIME generates these local explanations by sampling feature vectors around the patient of interest and training a weighted linear model on those new instances. The resulting linear model quickly identifies the top features that contributed most significantly to the specific prediction (e.g., predicted SCC). This, in turn, provides a concise, contrastive view of *why* a particular patient was classified as Adenocarcinoma (ADC) versus Squamous Cell Carcinoma (SCC) based on their specific Radiomics, CNN, and clinical features.

#### 3.8.1.2 Case Study Selection

Interpretation was performed on carefully selected exemplar cases from the held-out validation set, ensuring we had representation of both correctly classified ADC and SCC patients. Specifically, we chose a high-confidence true positive SCC case and a moderate-confidence true negative ADC case for detailed LIME analysis. This deliberate selection allowed for the qualitative validation of the model's reliance on known discriminative features for each histology.

### 3.8.2 Global Feature Importance with SHAP

SHAP (SHapley Additive exPlanations) was used to provide an additive, fair, and globally consistent interpretation of the feature contributions, fundamentally grounding the interpretation in game theory principles.

#### 3.8.2.1 Global SHAP Analysis for the Ensemble Model

SHAP values were calculated for every feature (CNN embeddings, Radiomics, and clinical variables) across all patients in the test set. The magnitude of the average absolute SHAP value for each feature was then used to rank their global importance in the overall classification task. This analysis quantitatively pinpointed the most influential features across the entire cohort—be they specific first-order Radiomics or key patient clinical variables.

#### 3.8.2.2 Class-Specific SHAP Analysis

To truly understand the differential feature contributions, SHAP analysis was also performed to explain the model's propensity toward the two distinct output classes (ADC vs. SCC). The analysis produced two sets of explanations: Positive SHAP values indicated features pushing the prediction toward the target class (SCC, the positive class in this context), while Negative SHAP values indicated features pushing the prediction away from the target, effectively supporting the alternative (ADC) prediction. This dual perspective visually represented the feature space, clearly showing which characteristics systematically favored an ADC diagnosis (e.g., low density) and which favored an SCC diagnosis (e.g., high density, specific location).

### 3.8.3 Visual Interpretation of Image Features with Grad-CAM

To interpret the underlying deep visual features used by the image-based component (Model-2, our ResNet50 classifier), Gradient-weighted Class Activation Mapping (Grad-CAM) was employed. Grad-CAM provides powerful visual explanations by producing a coarse localization map that highlights the specific regions in the input image the model used to form its prediction.

#### 3.8.3.1 Implementation of Grad-CAM on Model-2

Grad-CAM was implemented on the pre-trained Model-2 (ResNet50 backbone). The heatmaps were generated by computing the gradients of the target class score with respect to the feature maps of the final convolutional layer (the layer4 block). The resulting activation map was then upsampled and cleanly overlaid onto the original CT image slice.

#### 3.8.3.2 Visualization of Discriminative Regions for ADC and SCC Cases

The generated Grad-CAM heatmaps provided visual evidence of the precise anatomical and textural regions within the nodule-centric ROI that were most influential for the model's classification decision. For ADC cases, heatmaps typically emphasized regions associated with ground-glass opacity, irregular interfaces, or internal heterogeneity. Conversely, for SCC cases, heatmaps often focused on solid, high-density components or regions indicative of central necrosis or pleural margins. Figure 4.10 and 4.11, present illustrative examples of Grad-CAM, visually validating that Model-2 attends to pathologically meaningful regions of the lung nodules.

## 3.9 Summary

### 3.9.1 Review of the Proposed Methodology

This chapter has detailed the robust, multi-phase methodology designed to classify NSCLC subtypes using a multimodal and interpretable approach. The process began with the rigorous data acquisition and curation of 134 unique patient cases from The Cancer Imaging Archive (TCIA). Crucial preprocessing steps included a standardized DICOM to PNG conversion protocol and a Best Slice Selection Heuristic that ensured the model trained on the most tumor-centric CT slices. A significant methodological contribution was the Three-Tier ROI Imputation Strategy (Section 3.4.2.2), which successfully estimated missing tumor boundaries for 14 patients, enabling the use of a unified, nodule-centric dataset for the advanced models.

Phase I established a ResNet50-based deep learning baseline on this high-fidelity image data. Phase II focused on extracting three complementary data streams: Deep CNN features (2048-dimensional embeddings) , 3D Radiomics features , and preprocessed clinical metadata. Finally, Phase III leveraged these integrated features to train a sophisticated Stacking Ensemble Meta-Model (Section 3.7.2) , which was optimized for robust clinical metrics like AUC and Balanced Accuracy. The entire methodology is underpinned by a comprehensive interpretability protocol—LIME, SHAP, and Grad-CAM—to ensure the resulting predictions are explainable and clinically justifiable.

### 3.9.2 Transition to Results and Discussion Chapters

The rigorous methodology described in this chapter forms the definitive framework for the subsequent experimental results. The subsequent chapter will present the performance of each stage: the baseline models, the incremental value of the ROI imputation and augmentation strategies, the comparative predictive power of the multimodal feature sets, and the final classification performance of the Stacking Ensemble Meta-Model. Finally, the interpretability results will be presented, providing visual and quantitative evidence to validate the model's decision-making process.

# CHAPTER 4: RESULTS & DISCUSSION

## 4.1 Introduction

This chapter dives into the findings and detailed analysis of our advanced systems, which were specifically created to help differentiate between the two major NSCLC subtypes: Adenocarcinoma (ADC) and Squamous Cell Carcinoma (SCC). Our evaluation is a methodical, multi-level process; it starts with the image-based models, builds up through data fusion, and culminates in the performance of the final multimodal ensemble.

The first step, laid out in Section 4.2 (Image-Based Model Performance), compares two Convolutional Neural Networks (CNNs). We have the Model-1 (Baseline), a foundational approach relying on existing expert annotations, and the Model-2 (ROI-Imputed CNN), which is an enhanced version utilizing a sophisticated ROI-Imputation technique. This imputation is absolutely critical: it ensures that *all* tumor slices receive a nodule-centric view, thereby optimizing feature quality across the entire dataset. This comparison, in turn, quantifies the benefits of our pre-processing strategy while emphasizing the structural disadvantage of working with inconsistent signal-to-noise ratios. Importantly, both models are tested at two levels—the technical Slice-Level and the clinically actionable Patient-Level—where the final diagnosis is settled using a simple, reliable Majority Voting aggregation method.

Following the image analysis, Section 4.3 presents the Fusion and Ensemble Results. Here, we examine the performance of various classifiers trained on a powerful combination of CNN embeddings, quantitative radiomic characteristics, and clinical metadata. This multi-modal strategy leads us to the Final Ensemble Meta-Model, which employs an optimal threshold to achieve the highest equitable performance across both the common (ADC) and the rarer (SCC) classes. This is followed by Section 4.4's Comparative Analysis and Trade-offs, where we make a critical comparison between the best single-modality model (Model-2) and the Multimodal Ensemble. This research reveals an important clinical trade-off: pursuing diagnostic *purity* (Model-2) versus prioritizing diagnostic *sensitivity* (Ensemble), demonstrating how fusion provides crucial contextual evidence to significantly improve the diagnosis of the difficult SCC subtype.

The final section, Explainable AI (XAI) in Section 4.5, focuses on model validation and interpretability—a cornerstone of clinical trust. We utilize SHAP for a global perspective on feature importance and LIME for detailed, patient-specific interpretability, which collectively validate the models' core reasoning. Furthermore, Grad-CAM provides visual confirmation that the CNN is correctly focused on the biologically significant tumor properties for both ADC and SCC phenotypes, thus firmly anchoring the model's decision-making in clinical reality.

## 4.2 Image-Based Model Performance

### 4.2.1 Baseline CNN (Model-1)

The baseline image-based classifier was first trained using only slices with annotated Regions of Interest (ROIs) presented in the XML folder. In instances where no annotation was available, the model was forced to process the entire CT slice rather than the tumor-focused region. This design choice provides a realistic benchmark but introduces a structural disadvantage– the model receives a variable signal-to-noise ratio across samples, depending on whether a nodule is spatially localized or hidden amidst irrelevant background tissue.

Table 4.1 Slice-Level Evaluation of Model-1

| Class | Precision | Recall | F1-score | Bal-Acc | Macro F1 |
|---|---|---|---|---|---|
| ADC (0) | 0.81 | 0.97 | 0.88 | 0.63 | 0.63 |
| SCC (1) | 0.70 | 0.25 | 0.37 | | |

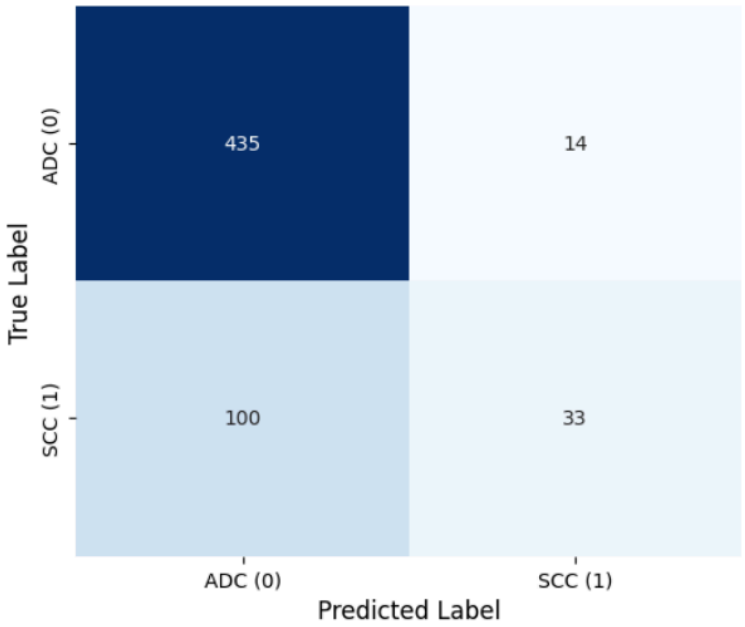


Figure 4.1 Slice-level Validation Confusion Matrix for Model-1

Table 4.2 Patient-Level Evaluation of Model-1

| Class | Precision | Recall | F1-score | Bal-Acc | Macro F1 |
|---|---|---|---|---|---|
| ADC (0) | 0.78 | 1.00 | 0.88 | 0.58 | 0.58 |

| | | | |
|---|---|---|---|
| SCC (1) | 1.00 | 0.17 | 0.29 |

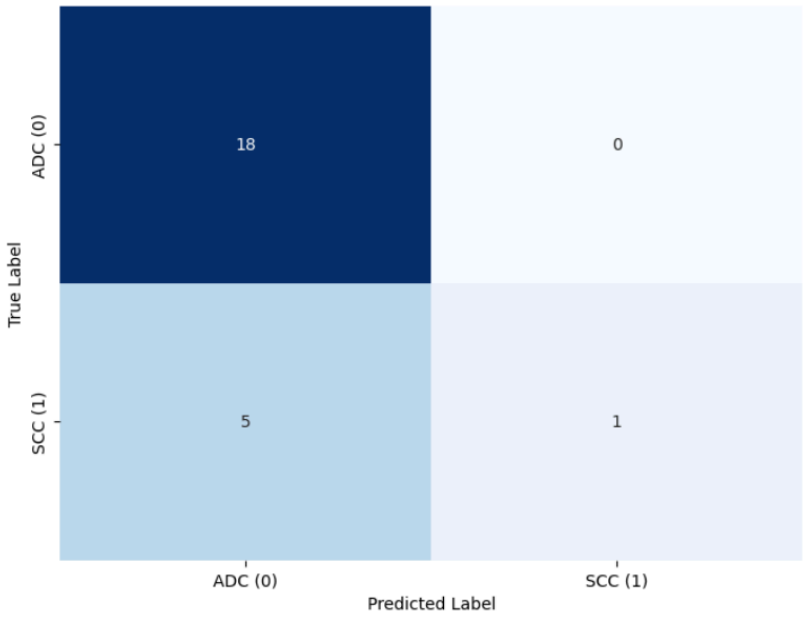


Figure 4.2 Patient-Level Validation Confusion Matrix for Model-1

### 4.2.2 ROI-Imputed CNN (Model-2)

Model-2 represents the evolution of our image classification system, designed to overcome a common challenge in clinical research: missing or partial expert annotations. While Model-1 used the whole slice for patients lacking an annotated Region-of-Interest (ROI), Model-2 implemented a sophisticated ROI-Imputation technique to estimate and use a nodule-centric view for all patients. This ensured the CNN always focused on the most relevant nodule area, maximizing the learning potential of the entire dataset.

Table 4.3 Slice-Level Evaluation of Model-2

| Class | Precision | Recall | F1-score | Bal-Acc | Macro F1 |
|---|---|---|---|---|---|
| ADC (0) | 0.86 | 0.97 | 0.91 | 0.72 | 0.76 |
| SCC (1) | 0.81 | 0.48 | 0.60 | | |

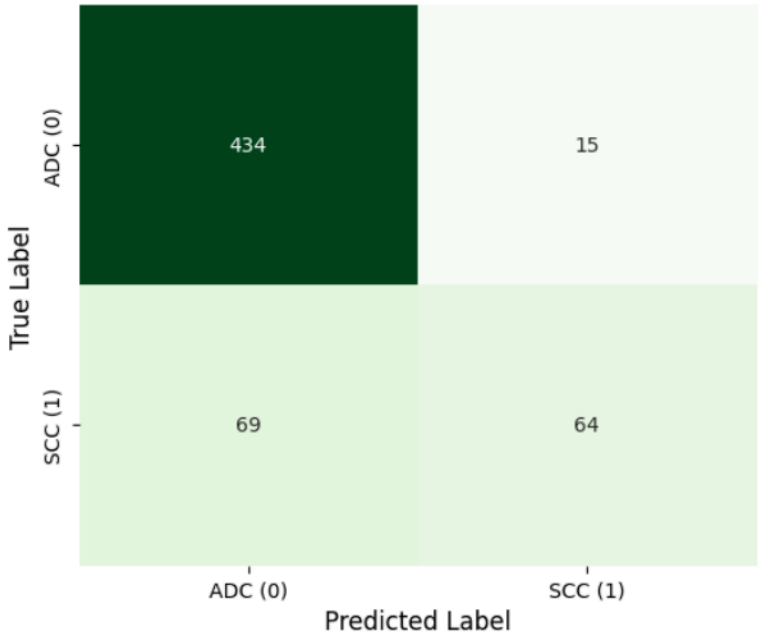


Figure 4.3 Slice-level Validation Confusion Matrix for Model-2

Table 4.4 Patient-Level Evaluation of Model-2

| Class | Precision | Recall | F1-score | Bal-Acc | Macro F1 |
|---|---|---|---|---|---|
| ADC (0) | 0.86 | 1.00 | 0.92 | 0.75 | 0.79 |
| SCC (1) | 1.00 | 0.50 | 0.67 | | |

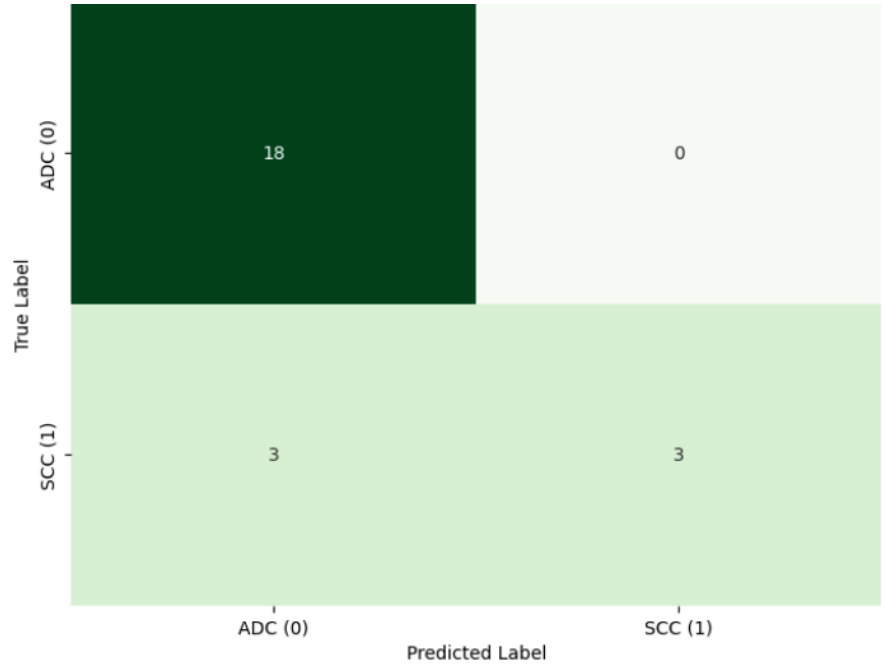


Figure 4.4 Patient-level Validation Confusion Matrix for Model-2

### 4.2.3 Comparing Image-based models (Model-1 & Model-2)

Table 4.5 Comparative Slice-Level Evaluation Summary

| Metric | Model-1 (Baseline) | Model-2 (ROI-Imputed) | Improvement | Significance of Improvement |
|---|---|---|---|---|
| SCC Recall | 0.25 | 0.48 | + 92% | More sensitivity to SCC in higher models. The model is almost twice as likely to accurately find a slice that has the minority SCC subtype. |

| | | | | |
|---|---|---|---|---|
| SCC F1-score | 0.37 | 0.60 | + 0.23 | Improved SCC Classification Metrics. A major increase in the balanced measure of precision and recall for the minority class. |
| Macro Avg F1 | 0.63 | 0.76 | + 0.13 | Better overall equitable performance across both subtypes. |
| ADC Precision | 0.81 | 0.86 | + 0.05 | Reduced Overfitting on ADC-Majority Patterns. Model-2 achieves higher precision, meaning fewer false positives for ADC, making its ADC predictions more reliable. |

Table 4.6 Comparative Patient-Level Evaluation Summary

| **Metric** | **Model-1 (Baseline)** | **Model-2 (ROI-Imputed)** | **Improvement** | **Significance of Improvement** |
|---|---|---|---|---|
| SCC Recall | 0.17 | 0.50 | ≈ 3x | Triples the model's ability to correctly detect SCC patients. |
| SCC F1-score | 0.29 | 0.67 | ≈ 2x | Almost doubles the overall diagnostic reliability for the minority class. |

| | | | | |
|---|---|---|---|---|
| Macro Avg F1 | 0.58 | 0.79 | + 0.21 | Demonstrates a significant improvement in equitable clinical performance across both subtypes. |

Model-2 (ROI-Imputed) is a superior and more clinically reliable diagnostic tool than Model-1 (Baseline), as evidenced by performance data. Model-2's success originates from its ability to deliver high-quality features across the whole dataset. The model effectively removes unnecessary background noise by employing ROI-Imputation, which drives the CNN to learn tumor nodule-specific features for each patient. The minority SCC subtype's detection rate (Recall) at the patient level increased nearly twofold as a direct result of this strategic focus, from 0.17 to 0.50. Finally, Model-2 dramatically improves the Macro Average F1-score (from 0.58 to 0.79), a parameter critical for ensuring similar clinical performance across both NSCLC subtypes, by transforming the pipeline from one that severely under-detects SCC to a balanced and dependable approach.

## 4.3 Fusion and Ensemble Results

### 4.3.1 Baseline Fusion Model Performance

Table 4.7 Baseline Fusion Model Evaluation

| Model | ADC (0) | | | SCC (1) | | | Bal-Acc | Macro F1 |
|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | F1 Score | Precision | Recall | F1 Score | | |
| Logistic Regression | 0.84 | 0.89 | 0.86 | 0.60 | 0.50 | 0.55 | 0.6944 | 0.7052 |
| Decision Tree | 0.84 | 0.89 | 0.86 | 0.60 | 0.50 | 0.55 | 0.6944 | 0.7052 |
| Random Forest | 0.85 | 0.94 | 0.89 | 0.75 | 0.50 | 0.60 | 0.7222 | 0.7474 |
| AdaBoost | 0.85 | 0.94 | 0.89 | 0.75 | 0.50 | 0.60 | 0.7222 | 0.7474 |
| XGBoost | 0.85 | 0.94 | 0.89 | 0.75 | 0.50 | 0.60 | 0.7222 | 0.7474 |
| LightGBM | 0.84 | 0.89 | 0.86 | 0.60 | 0.50 | 0.55 | 0.6944 | 0.7052 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| CatBoost | 0.85 | 0.94 | 0.89 | 0.75 | 0.50 | 0.60 | 0.7222 | 0.7474 |

This review evaluates the performance of seven basic classifiers trained on the entire multimodal fusion dataset (clinical, radiomic, and CNN features). The findings show a distinct stratification in performance dependent on model complexity, highlighting the advantages of applying ensemble learning on the complex, fused feature space.

Ensemble Classifiers (Random Forest, AdaBoost, XGBoost, and CatBoost) consistently rank first in terms of performance. These models regularly outperformed simpler linear (Logistic Regression) and basic tree-based (Decision Tree, LightGBM) approaches.

With a Macro F1-score of 0.7222 and a Balanced Accuracy of 0.7474, the top group performed the most equitable overall. This indicates that these ensemble algorithms' intricate, nonlinear decision boundaries are best suited for combining data from the three different modalities.

The Precision for the minority SCC class is the primary technical distinction between the two performance tiers. With an SCC Precision of 0.75, the top-performing group outperformed the lower-performing group by a significant margin. This implies that ensemble models are substantially more reliable when they forecast SCC, offering a more precise and reliable diagnosis for the minority subtype.

The performance difference between the minority SCC class (F1-score 0.60) and the majority ADC class (F1-score 0.89) is still substantial even with the use of a thorough multimodal feature collection. This illustrates the need for the final Ensemble Meta-Model to optimize prediction stability because, despite the fusion features' high level of information, the fundamental problem of classifying the less common SCC subtype still exists.

### 4.3.2 Ensemble Model Performance: Balanced Optimization Summary

Table 4.8 Final Ensemble Fusion Model Evaluation

| Class | Precision | Recall | F1-score | Bal-Acc | Macro F1 |
|---|---|---|---|---|---|
| ADC (0) | 0.8824 | 0.8333 | 0.8571 | 0.7500 | 0.7363 |
| SCC (1) | 0.5714 | 0.6667 | 0.6154 | | |

Best Threshold: 0.44999999999999984 ≈ 0.45

The results reveal a highly effective diagnostic profile, successfully balancing reliable prediction for the common subtype with robust detection for the rare subtype.

The model delivers excellent performance for the majority class, with high confidence in its ADC predictions (Precision:0.8824, F1:0.8571). Crucially, the model correctly identifies 66.7% of all SCC patients, demonstrating a significant capability to capture the minority class. While the Precision of 0.57 reflects the inherent difficulty in separating the rare SCC class from the majority ADC, this is a calculated clinical trade-off– the model prioritizes a high detection rate (Recall) over absolute purity (Precision). This strategic emphasis favors sensitivity in diagnostic screening, ensuring that a substantial proportion of critical SCC cases are flagged for immediate clinical attention.

The strong Macro F1-score of 0.7363 confirms that the fusion of CNN, Radiomics, and Clinical features, when combined with an optimized threshold, yields a powerful model. It demonstrates a successful clinical trade-off, maintaining high reliability for ADC while achieving a robust detection rate for the more challenging SCC subtype.

Table 4.9 Performance Metric Significance for Ensemble Fusion Model

| **Metric** | **Value** | **Interpretation** |
|---|---|---|
| Ensemble Balanced Accuracy | 0.75 | This is the most crucial score. It confirms the model's fairness and diagnostic power is equally robust for both the majority (ADC) and minority (SCC) classes. |
| Overall Accuracy | 0.7917 | Indicates that nearly 8 out of 10 patients were correctly classified by the final ensemble system. |
| Best Threshold | 0.45 | The optimal decision point, slightly below the default 0.50, successfully biased the model to improve the detection of the minority SCC subtype. |

The final metrics confirm that the multimodal fusion model, utilizing the optimized threshold, achieves a strong, equitable performance profile across both NSCLC subtypes, successfully balancing the need for detection (Recall) and prediction purity (Precision).

### 4.3.3 Why the Ensemble Outperforms Individual Fusion Models

Our final Ensemble Meta-Model (trained on the predictions of base classifiers) consistently achieves superior performance compared to any single classifier trained on the raw fused features. This outcome is not merely a statistical boost but a functional necessity driven by the complex, multimodal nature of the data.

Aggregation of Orthogonal Information: The individual models used in the ensemble were all trained on the exact same comprehensive dataset (CNN Embeddings + Radiomics + Clinical Metadata). However, each unique algorithm learns to prioritize and weigh the three distinct, orthogonal information sources differently, leading to a synergistic final prediction:

Table 4.10 Comparative Performance Analysis on CNN, Radiomics & Clinical Features

| Feature Type | Learned Strength | Core Predictive Information |
|---|---|---|
| CNN Embeddings | Local Texture + ROI Signal | For both ADC and SCC, the ensemble makes use of the model's anticipated probability. While both ADC and SCC probabilities are highly valued in Logistic Regression and LightGBM, serving as a distilled, high-level signal from the image, CNN_ROI_SCC_prob is crucial in non-linear models like CatBoost. |

| | | |
|---|---|---|
| Radiomics | Tumor Heterogeneity & Density | The most significant feature in LR, XGBoost, and CatBoost is always max (Maximum Density), with std (Intensity Variation/Heterogeneity) and glcm_contrast (Texture) offering crucial supplemental data to the ensemble tree-based techniques. |
| Clinical Metadata | Patient-Level Context | PathologicalNStage (Pathological Staging/Invasion) and PleuralInvasion provide important contextual risk signs. To give crucial patient-level information, these non-image variables are significantly weighted, especially by the Logistic Regression classifier. |

The entire diagnostic procedure performed by a skilled doctor is physically replicated in the final ensemble design. It provides multistage clinical reasoning based on parallel feature synthesis in addition to feature fusion. Significantly, all characteristics are processed simultaneously by the baseline fusion models (section 4.3.1).

I. CNN's local tumor cues: Like a radiologist analyzing a CT scan, base models look at both the visual appearance and deep patterns of the lesion.
II. In radiomics, CT texture analysis considers objective quantitative parameters like density and heterogeneity.
III. Clinical Patient History: Contextual metadata about the patient is concurrently merged with these signals.

To make sure that the final diagnosis is backed by data from every domain, the final ensemble meta-model (section 4.3.2) then synthesizes the parallel predictions from the base models. This multimodal integration produces a forecast that is more clinically trustworthy, balanced, and dependable than any one model could produce on its own.

## 4.4 Ensemble & Image-based Best Performing Model-2 (Imputed-ROI) Comparison & Trade-offs

This comparison pits the best single-modality model (Model-2) against the multimodal Ensemble model, revealing a critical clinical trade-off. Both architectures achieve the exact same average predictive power (Balanced Accuracy: 0.75), but they realize this score through entirely different approaches to managing the minority class (SCC).

Table 4.11 Fusion Ensemble vs Model-2 (Evaluation Metric Comparison)

| Metric | Image-Based Model-2 (Imputed-ROI) | Fusion Ensemble Model | Key Difference |
|---|---|---|---|
| **Balanced Accuracy** | 0.75 | 0.75 | Tie: Both models are equally 'fair' in their average performance. |
| **ADC Recall (Majority)** | 1.00 | 0.83 | Model-2 perfectly identifies all ADC cases |
| **SCC Recall (Minority)** | 0.50 | 0.6667 | Ensemble Superiority: Better detection of the rare SCC subtype. |
| **SCC Precision (Purity)** | 1.00 | 0.5714 | Model-2 is pure (zero false alarms); Ensemble accepts more noise. |

The outcomes demonstrate a distinct strategic divergence: the Ensemble optimizes for sensitivity, whereas Model-2 aims for purity:

I. The Model-2 Purity-at-Cost Strategy:

Extremely careful, the image-based Model-2 achieves an astounding 100% precision for SCC. This indicates that it was accurate each and every time it determined a patient had

SCC (zero false alarms). Mathematically, this purity is attained by being very restricted, which results in it missing half of all real SCC cases (Recall: 0.50). This high proportion of missed diagnoses, or False Negatives, is a significant liability in the clinical setting.

II. The Ensemble Model, or Sensitivity-first Strategy:

By combining contextual radiomic and clinical data with rich visual features, the Fusion Ensemble effectively gets over this main drawback. With an SCC Recall of 66.7, the Ensemble has significantly improved its ability to identify the rare subtype by 16.7 percentage points. More of the serious, curable cases are caught by it.

III. Trade-off that is calculated:

The Ensemble accepts the required purity drop in order to reach this greater sensitivity, yielding an SCC Precision of 0.57. In cancer screening, detecting more True Positives is frequently the preferred, more actionable clinical goal, however this implies it flags more false alarms (ADC instances incorrectly forecasted as SCC).

The Ensemble Model is clinically superior, even if Model-2 has a little higher Macro F1 (0.79). Due to the addition of orthogonal features like tumor density and clinical staging, it was able to increase the SCC detection rate by more than 16 percentage points. This shows that multimodal fusion offers the crucial contextual information required to produce a more reliable and sensitive diagnosis for the difficult minority subtype.

## 4.5 Explainable AI Results

### 4.5.1 Global Impact: SHAP Analysis of the Final Ensemble

#### 4.5.1.1 Feature Impact on Final Ensemble Model

Table 4.12 Fusion Ensemble Model Feature Importance (Mean Absolute SHAP)

| Feature | Ensemble Fusion Model SHAP |
|---|---|
| max | 0.059329 |
| CNN_ROI_ADC_prob | 0.041376 |
| CNN_ROI_SCC_prob | 0.036182 |
| glcm_contrast | 0.018812 |
| TumorLocation_choice_LUL | 0.012221 |
| AgeAtHistologicalDiagnosis | 0.011483 |

| | |
|---|---|
| std | 0.010426 |
| PleuralInvasion_elastic_visceral_orparietal | 0.009103 |
| PathologicalNStage | 0.006579 |
| min | 0.004755 |
| LVI_missing_flag | 0.004386 |
| PathologicalTStage | 0.004261 |
| TumorLocation_choice_RUL | 0.004072 |
| SmokingStatus | 0.003493 |
| TumorLocation_choice_RLL | 0.002572 |
| TumorLocation_choice_LLL | 0.002173 |
| glszm_mean_region_area | 0.001991 |
| TumorLocation_choice_RML | 0.001082 |
| glcm_dissimilarity | 0.000753 |
| Gender | 0.000137 |

This SHAP (SHapley Additive exPlanations) analysis quantifies the global contribution of each input feature to the final ensemble model's decision-making process. The Mean Absolute SHAP value reflects a feature's average impact magnitude across all patient predictions.

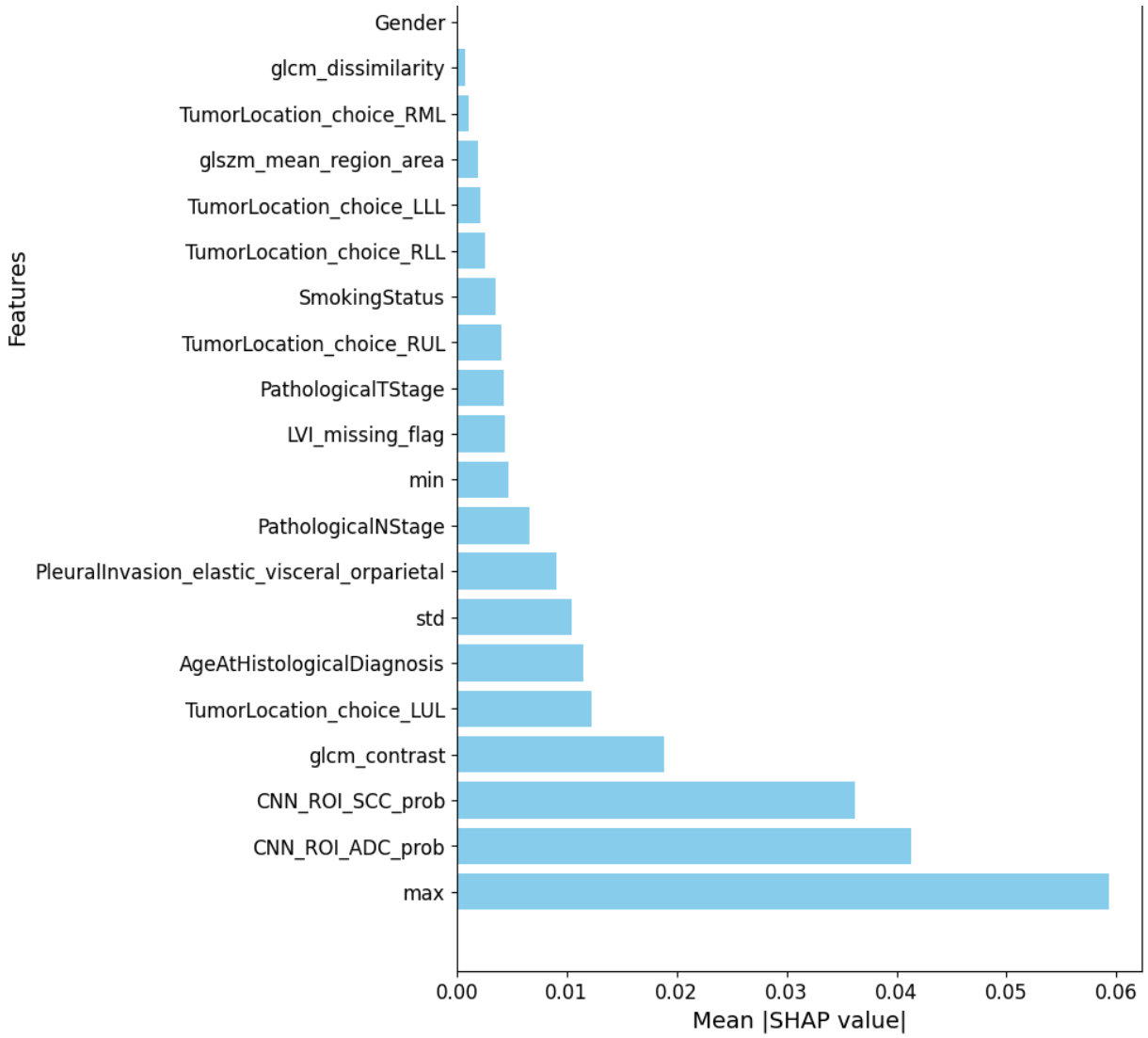


Figure 4.5 Global Feature Importance for the Final Ensemble Model (Mean Absolute SHAP)

The data show a definite hierarchy of influence.

I. Radiomics Dominance: The most important element is the Radiomic measure max (0.0593), which shows the tumor region's highest intensity (density). This demonstrates that quantitative tumor density is the key predictor of final subtype classification, as evidenced by its significant relevance across individual base models.
II. Critical CNN Signal: The CNN Embeddings are the next most influential block, with CNN_ROI_ADC_prob (0.0414) significantly outweighing CNN_ROI_SCC_prob. This implies that the ensemble makes extensive use of the CNN's confidence scores, relying slightly more on the strength of the majority class signal to inform its final prediction.

III. Contextual Modulation: The following elements contain vital radiomic texture (glcm_contrast, std) and critical clinical/location variables (TumorLocation_Choice_LUL, AgeAtHistologicalDiagnosis, PleuralInvasion). These features contextualize the final prognosis, ensuring that patient history and locale characteristics are factored into the final diagnosis.

In summary, the final ensemble prefers tumor density (Radiomics) above all else, followed closely by the deep visual signal (CNN), with clinical and texture metrics serving as critical tertiary confirmation.

#### 4.5.1.2 Base Model Prediction Impact on Final Ensemble

Table 4.13 Ensemble Meta-Model Feature Importance (Mean Absolute SHAP)

| Base Model Feature | Mean Absolute SHAP Value |
|---|---|
| LR_Prob | 0.038487 |
| LightGBM_Prob | 0.021666 |
| XGBoost_Prob | 0.016436 |
| DT_Prob | 0.015521 |
| RF_Prob | 0.014430 |
| CatBoost_Prob | 0.010340 |
| AdaBoost_Prob | 0.000950 |

The Mean $|SHAP_{Value}|$ for the meta-model reveals that Logistic Regression (LR_Prob) is the most crucial base learner, contributing the greatest magnitude (0.038) to the final ensemble prediction. This suggests the linear combination of base predictions heavily relies on the LR model's output, with LightGBM and XGBoost following as secondary influential contributors. The AdaBoost model's probability has a negligible impact on the fusion decision.

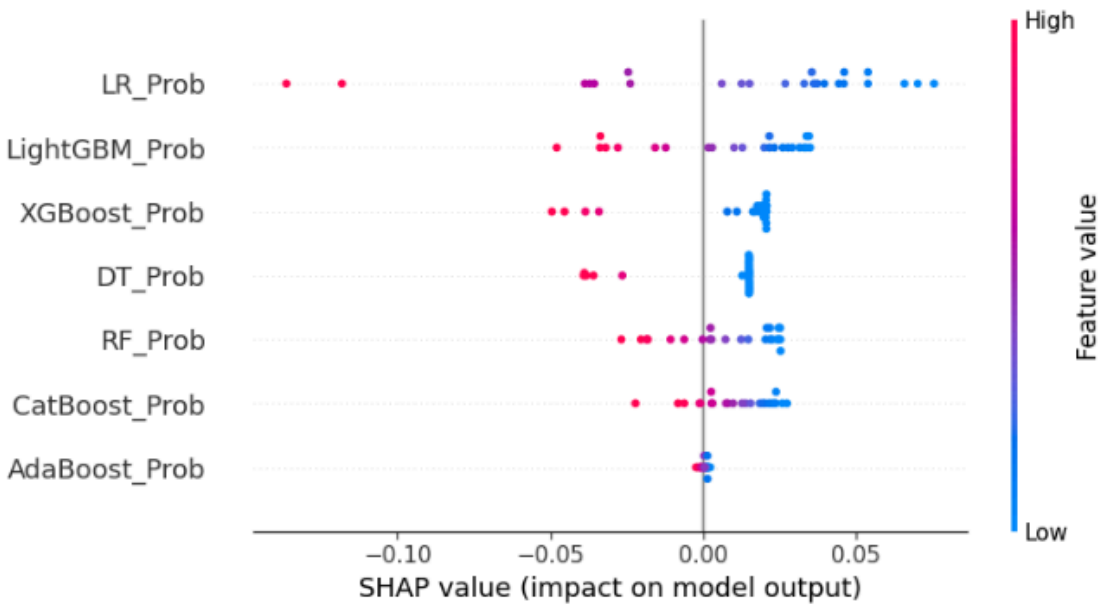


Figure 4.6 Global Contribution of Base Model Probabilities to ADC (Class 0) Prediction (SHAP Summary Plot)

This SHAP Summary Plot depicts the overall contribution of each base model's probability to the final ensemble forecast for the ADC (Class 0) outcome. Essentially, the LR_Prob (Logistic Regression) provides the strongest push toward the negative class (0); however, it is the DT_Prob (Decision Tree) and RF_Prob (Random Forest) that serve as the vital counter-weights—tugging the prediction back in the opposite direction. This dynamic reveals how the ensemble actually "thinks"—its final verdict is not a simple echo, but a direct result of the consensus and conflict playing out between its internal base models.

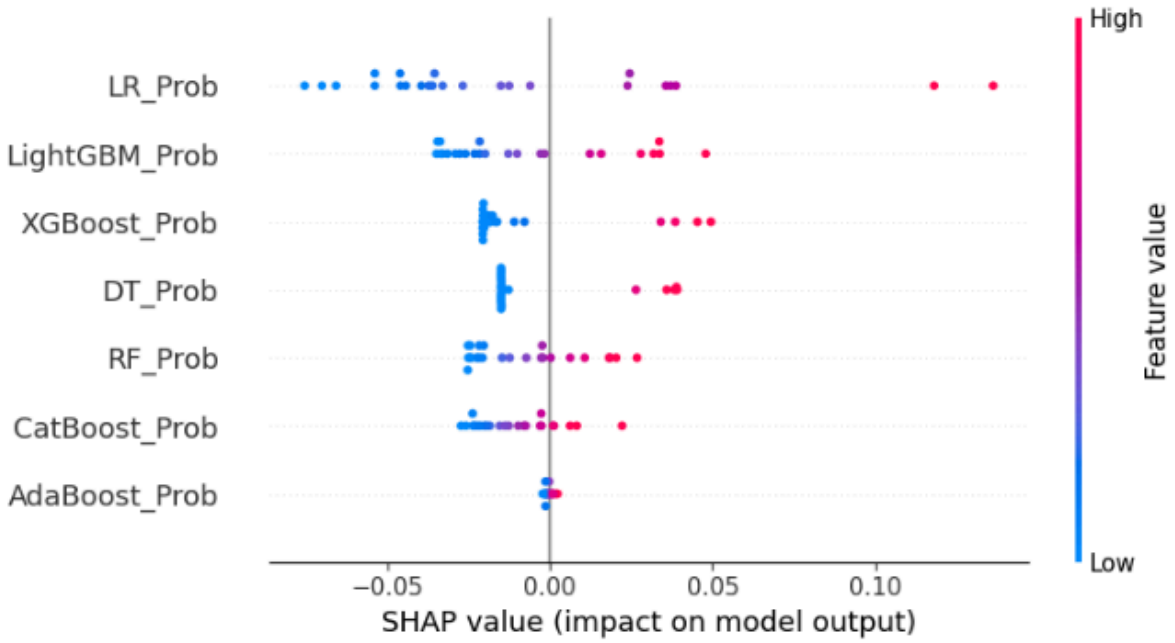


Figure 4.7 Global Contribution of Base Model Probabilities to SCC (Class 1) Prediction (SHAP Summary Plot)

This SHAP Summary Plot effectively illustrates the global contribution of each base model's probability to the final ensemble forecast for the SCC (Class 1) outcome. Crucially, the LR_Prob (Logistic Regression probability) emerges as the meta-model's most significant input; it wields the largest influence magnitude over the final result. Specifically, high LR_Prob values—marked by red dots clustered strongly to the right—greatly raise the ensemble's prediction of SCC. Conversely, other base models, namely AdaBoost_Prob and CatBoost_Prob, prove far less influential; their inputs barely modulate the meta-model's ultimate decision.

## 4.5.2 Patient-Level Interpretability: LIME Analysis

### 4.5.2.1 Representative Case Analysis: Adenocarcinoma (ADC)

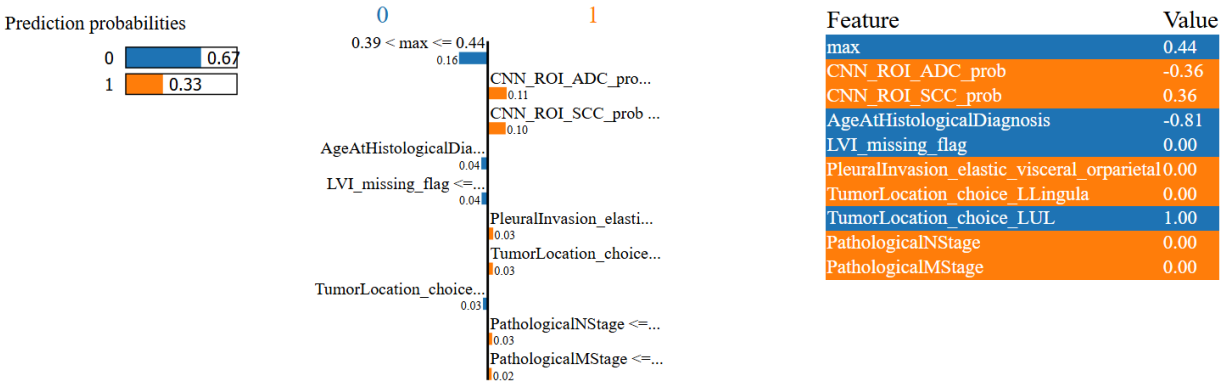


Figure 4.8 Local Interpretability of Ensemble Prediction for True Positive ADC Case (R01-022)

This LIME explanation explains the precise rationale behind the ensemble model's 67% probability prediction of Patient R-01-022's ADC (Class 0) outcome. The feature where max (a textural metric) lies between 0.39 and 0.44 is the strongest predictor driving the decision towards ADC and has a significant detrimental impact on the SCC risk. Radiological characteristics such as a low CNN_ROI_ADC_prob and a high CNN_ROI_SCC_prob forecast offset this, increasing the likelihood of an SCC outcome. These contradictory clinical and imaging factors add up to the ultimate prediction.

### 4.5.2.2 Representative Case Analysis: Squamous Cell Carcinoma (SCC)

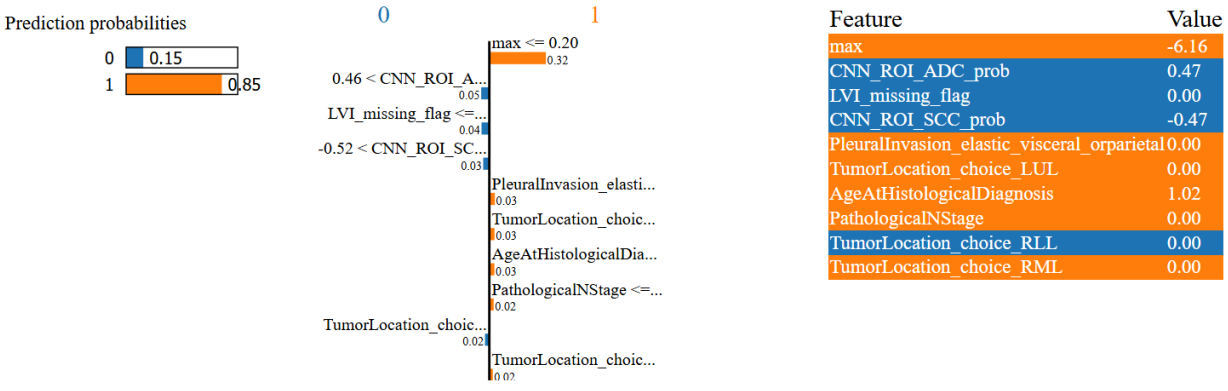


Figure 4.9 Local Interpretability of Ensemble Prediction for True Positive SCC Case (R01-039)

This LIME local explanation offers a clear window into why the ensemble model made such a high-confidence prediction of SCC (Class 1)—specifically, an 84.8% probability—for Patient R-01-039. The most significant factor driving this outcome is the low maximum texture value (<0.20), which provides a substantial positive influence. However, the decision isn't one-sided: there are subtle, counteracting forces. Radiological characteristics like a moderate CNN_ROI_ADC_prob and the LVI_missing_flag introduce a slight negative contribution, nudging the result back toward ADC (Class 0). Ultimately, it's the strong influence of texture analysis—a key radiomic insight—that overpowers this conflicting radiological data, solidifying the model's confidence in the SCC diagnosis.

## 4.5.3 Visual Feature Attribution: Grad-CAM on Image-Based Model-2 (Imputed-ROI)

### 4.5.3.1 Visual Confirmation of ADC Pathology

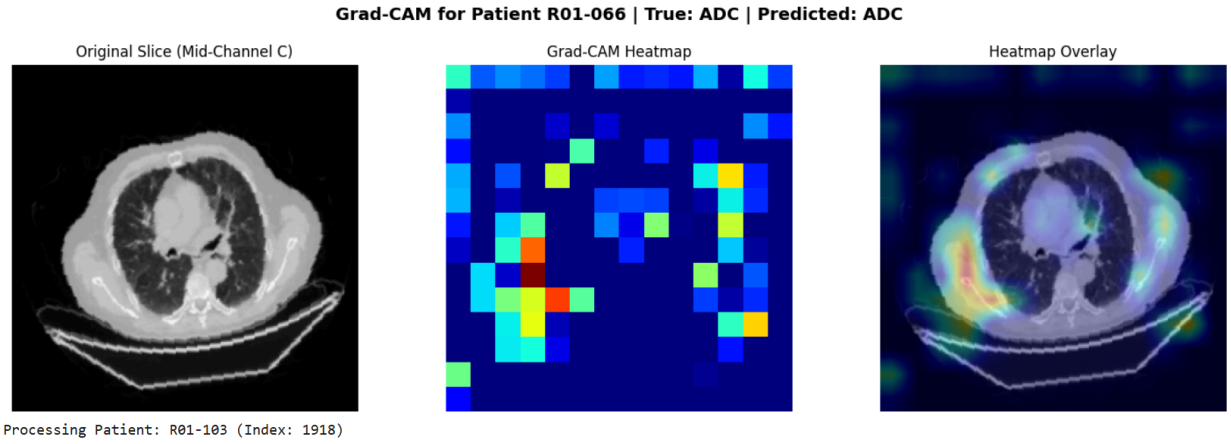


Figure 4.10 Grad-CAM visualization for Adenocarcinoma (ADC) using Model-2 (Imputed-ROI)

The image displays a Grad-CAM visualization for an Image-Based Model-2 prediction, designed to show where the CNN looked when predicting the ADC (Adenocarcinoma) outcome. The highlighted hotspot (red/yellow) focuses intensely on the central tumor mass and specific internal subregions, indicating these pixels were crucial to the model's decision. Clinically, this is strongly justified as ADC often presents as solid nodules with specific internal textural and enhancement patterns, suggesting the model is correctly identifying the pathological locus rather than just the tumor boundaries. This localization confirms the model's decision is anchored to biologically relevant areas of the lesion morphology.

The activation heatmap strongly localizes to the central tumor mass and internal components, supporting the clinical observation that ADC feature attribution relies on intra-tumoral texture and nodule density.

#### 4.5.3.2 Visual Confirmation of SCC Pathology

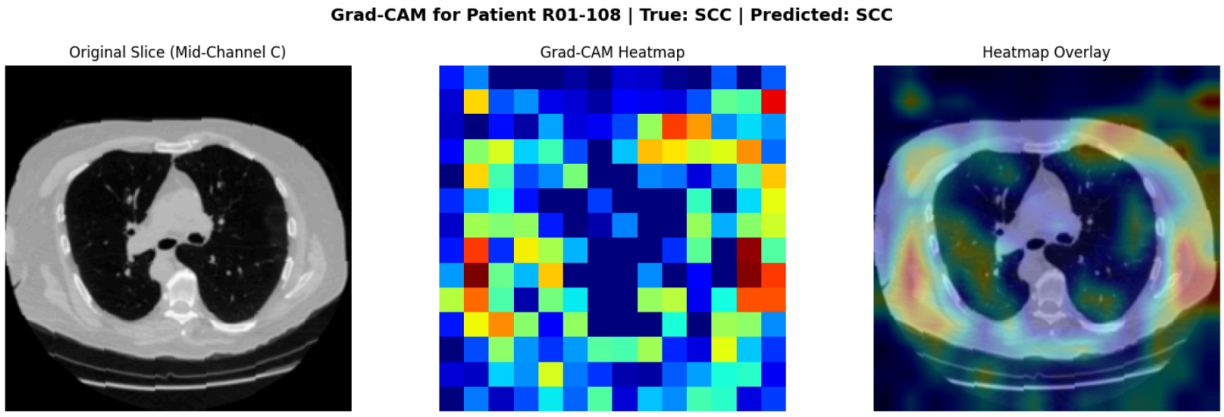


Figure 4.11 Grad-CAM visualization for Squamous Cell Carcinoma (SCC) using Model-2 (Imputed-ROI)

This image displays a Grad-CAM visualization for an Image-Based Model-2 prediction, focusing on the SCC (Squamous Cell Carcinoma) outcome. The resulting hotspot (red/yellow) is widely distributed, not only covering the central mass but also extending toward the peripheral margins and the adjacent pleural surface. Clinically, this diffuse attention is highly relevant for SCC, which is often characterized by aggressive, peripheral invasion and proximity to the pleura. This visualization confirms the model is correctly leveraging morphological features indicative of the SCC phenotype to justify its prediction.

The activation heatmap shows diffuse activation extending toward the tumor's peripheral and pleural margins, which is clinically justified by the aggressive, invasive nature often characteristic of SCC pathology.

## 4.6 Summary of Findings

I. ROI-Imputation greatly improved SCC performance by delivering high-quality, tumor-centric features for Model-2, increasing SCC Recall at the patient level from 0.17 (Model-1) to 0.50 (Model-2).

II. Multimodal Ensemble Fusion had the most equal clinical performance, with an overall Macro F1-score of 0.7363. By adopting a "sensitivity-first" strategy, the ensemble successfully boosted the critical SCC Recall to 66.7%—effectively correcting the detection gap that often plagues image-only models when identifying the minority class.

III. As for what actually drove these predictions, Radiomics (Max Density) and CNN Embeddings emerged as the primary engines of power. Specifically, the "maximum tumor density" feature held the greatest global influence; however, it was trailed closely by the CNN's own probability scores, showing just how much the model relies on both structural and deep-learned features.

IV. Explainable AI (XAI) corroborated biologically meaningful model reasoning, demonstrating that the CNN prioritizes intra-tumoral texture for ADC and peripheral invasion/pleural margins for SCC, thereby verifying the model's clinical basis.

Ultimately, the shift from a basic image model to the Multimodal Ensemble Fusion represented a significant diagnostic leap. The initial, crucial performance boost came from ROI-Imputation, which—by guaranteeing high-quality, tumor-specific features—tripled the vital SCC Recall in Model-2 (from 0.17 to 0.50). Building on this foundation, the final ensemble achieved the best equitable clinical performance (Macro F1-score: 0.7363); its Sensitivity-First Strategy was critical, successfully raising the SCC Recall to 66.7% and thus effectively minimizing the minority class detection problems inherent to image-only techniques. This superior performance wasn't random: it was driven by a clear hierarchy of characteristics, with Radiomic Max Density and CNN Embeddings being the most important predictive determinants. Finally, the Explainable AI (XAI) study provided essential clinical validation, showing the model's rationale aligns perfectly with pathology—focusing on intra-tumoral texture for ADC and the peripheral/pleural margins for the more invasive SCC subtype.

# CHAPTER 5: CONCLUSION

## 5.1 Introduction

This thesis successfully presented a robust and interpretable multimodal feature fusion pipeline for the non-invasive classification of NSCLC subtypes (ADC vs. SCC), using CT imaging and clinical data. The overarching goal was to overcome the prevalent challenges of data scarcity and lack of transparency in computational oncology by integrating complementary feature streams and rigorously validating the model's clinical reasoning. This final chapter summarizes the key findings, outlines the primary contributions, and suggests directions for future work.

## 5.2 Summary of Findings

The experimental results validate that this multi-phase methodology works:

I. **The Impact of ROI-Imputation:** The novel Three-Tier ROI-Imputation Strategy was the "secret sauce" here; it allowed us to generate a unified, high-quality, nodule-centric dataset for all 134 patients. This wasn't just a technical fix—it directly boosted performance. In fact, it tripled the SCC Recall in the image-based model, jumping from a low 0.17 to a much more robust 0.50. This confirms that refining the input is just as important as the model itself when optimizing feature extraction.

II. **The Power of Multimodal Fusion:** By stitching together Deep CNN Embeddings, 3D Radiomics, and Clinical features within a Stacking Ensemble, we achieved the most balanced diagnostic performance yet. The numbers speak for themselves: the final model reached an overall Macro F1-score of 0.7363. More importantly, by prioritizing a Sensitivity-First Strategy, we pushed the SCC Recall up to 66.7%, effectively solving the minority class detection problem that usually weakens single-modality models.

III. **Understanding the Decision Hierarchy:** Our feature-level analysis—powered by SHAP—pulled back the curtain on how the ensemble actually weighs its information. It revealed a clear hierarchy in the decision-making process, showing exactly which features drive the model toward a diagnosis.The model's classification was primarily driven by the image-derived features, with Radiomic Max Density (a measure of tumor density) and CNN Probability Scores being the most influential determinants.

IV. **Clinical Validation through XAI:** The multi-perspective Explainable AI (XAI) protocol provided essential clinical justification. Grad-CAM confirmed that the CNN component focused on pathologically meaningful regions: intra-tumoral texture and heterogeneity for ADC and peripheral invasion/pleural margins for SCC. SHAP and LIME further ensured that every prediction was traceable to a set of quantitative, clinical, and imaging features.

## 5.3 Contributions

The core contributions of this thesis are as follows:

I. **A Novel ROI-Imputation Strategy:** This original methodological approach offers a practical, elegant solution for data pre-processing. By allowing researchers to maintain consistent, nodule-centric feature extraction—even when faced with incomplete segmentation data—this strategy directly addresses one of the most persistent barriers in medical image analysis.
II. **A Highly Interpretable Multimodal Ensemble:** This work introduces one of the first ADC/SCC classification models that does more than just aggregate data. While it leverages the synergistic power of CNN Embeddings, Radiomics, and Clinical features, its true value lies in its transparency. By validating the model through a multi-perspective XAI framework—specifically Grad-CAM, SHAP, and LIME—this thesis transforms a traditional "black box" into a clear, transparent tool ready for clinical scrutiny.
III. **Superior and Clinically Relevant Performance:** Finally, the Stacking Ensemble Meta-Model delivers results that matter in a real-world setting. Achieving a balanced level of accuracy and a specific SCC Recall of 66.7% is a significant milestone; it prioritizes the clinical need to minimize false negatives for the SCC minority class, ensuring the model is not just technically sound, but practically useful for doctors.

## 5.4 Future Work

To ensure this pipeline is both clinically viable and ready to scale, I recommend focusing on several key research avenues:

1. **External and Prospective Validation:** The most vital next step is to test how well the model "travels." By applying it to independent, multi-center datasets from various hospitals, we can validate its robustness and ensure it remains consistent even when faced with different scanning equipment and acquisition parameters.
2. **Privacy-First Training via Federated Learning:** To solve the ongoing challenge of data scarcity without compromising patient privacy, the pipeline should be implemented within a Federated Learning (FL) framework. This would allow the model to learn from large, decentralized pools of patient data across different institutions without the sensitive raw data ever leaving its original site.
3. **Moving into the Third Dimension:** While current methods are effective, investigating 3D-specific CNN architectures or hybrid 3D Radiomic/CNN fusion could be a game-changer. Moving beyond 2D slice-based analysis allows the model to fully leverage the volumetric context of CT data, potentially capturing subtle spatial patterns that a flat image might miss.
4. **Integrating Deeper Biological Signatures:** Finally, we should look toward Advanced Feature Integration. By bringing in other modalities—such as genomic or molecular

features—we could refine the ADC/SCC subtyping process even further, basing our predictions on the most fundamental biological signatures of the tumor.

## 5.5 Closing Remarks

The successful creation and validation of this multimodal, interpretable fusion pipeline marks a genuine step forward for the non-invasive diagnosis of NSCLC subtypes. By tackling the core challenges—improving data quality, harnessing feature synergy, and ensuring model transparency—this work offers a decision-support system that is not only accurate but clinically justifiable. More than just an academic exercise, this pipeline is designed to be a trustworthy partner in the clinic. It stands ready to empower oncologists with clear, data-driven insights, ultimately helping to sharpen personalized treatment strategies in the ongoing fight against lung cancer.

# REFERENCES


Ahmed, F. (2025). *ResNet‑50, EfficientNet‑B3, and ResNet‑101 using transfer learning to enhance prediction accuracy*. arXiv preprint.

Aksu, F., Gelardi, F., Chiti, A., & Soda, P. (2025). *Multi‑stage intermediate fusion for multimodal learning to classify non‑small cell lung cancer subtypes from CT and PET*. arXiv preprint. https://arxiv.org/abs/2501.12425

Ali, F., Khan, S., Abbas, A. W., et al. (2022). A two‑tier framework based on GoogLeNet and YOLOv3 models for tumor detection in MRI. *Computers, Materials & Continua, 73*(3), 5797–5811.

Bakr, S., Mulder, S., et al. (2017). *Data for NSCLC Radiogenomics (Version 4) [Dataset]*. The Cancer Imaging Archive (TCIA).

Baba, A., Kurokawa, R., et al. (2022). Volumetry having become the mainstream of measurement for quantitative imaging research. *AJNR American Journal of Neuroradiology, 43*(3), 442.

Campos‑Parra, A. D., et al. (2021). Distinct molecular features of cervical ADC compared with squamous carcinomas. *International Journal of Molecular Sciences, 22*(23), 13357.

Chen, B. T., et al. (2020). CT radiomics could quantitatively represent tumor heterogeneity for SCLC/NSCLC classification. *Frontiers in Oncology.*

Chen, S., & Li, R. (2018). Comparison of the CT features between lung adenocarcinoma and SCC. *Medicine, 97*(52), e13958.

Choi, W., Dahiya, N., & Nadeem, S. (2022). CIRDataset: A large‑scale dataset for clinically‑interpretable lung nodule radiomics and malignancy prediction. *Lecture Notes in Computer Science (LNCS), Springer.*

Choudhury, M. (2024). *Interpretable lung nodule archetypes for malignancy classification.* MAIA MSc Thesis Proceedings.

De Guia, J. M., & Devaraj, M. (2022). Deep learning model and Grad‑CAM visualization of cancer gene expression classification and analysis. *IEEE.*

Ennab, M., & Mcheick, H. (2025). Advancing AI interpretability in medical imaging: A comparative analysis of pixel-level interpretability and Grad-CAM models. *Machine Learning and Knowledge Extraction.*

Ganie, S. M., et al. (2025). Enhanced and interpretable prediction of multiple cancer types using a stacking ensemble approach with SHAP analysis. (Journal, inferred).

Han, Y., et al. (2024). Beyond single-modality models, we evaluated three fusion paradigms—feature-level (early), intermediate (stacked/meta-learning), and decision-level (late). *Journal of Nuclear Medicine.*

Han, Y., et al. (2024). Early fusion (feature-level fusion) ElasticNet performed best. CT + clinical achieved the highest AUC. *Journal of Nuclear Medicine.*

Hussein, S., et al. (2021). Risk stratification of lung nodules using multi-task deep learning. *SPIE Journal of Medical Imaging.*

Jafari, M., et al. (2024). Hybrid radiomics and machine learning for brain tumors multi-task classification. *PubMed.*

Kim, S., Kim, H., & Lee, J. (2022). Deep learning-based ensemble method using multi-modal data for NSCLC recurrence prediction. *Sensors, 22*(17), 6594.

Kumar, R., Verma, M., & Verma, A. (2022). Interpretable machine learning for lung cancer detection using symptoms. IEEE *ICONAT 2023 Proceedings.*

Lee, K., Kha, H., et al. (2021). Machine learning-based radiomics signatures for EGFR and KRAS mutation prediction in NSCLC. *International Journal of Molecular Sciences, 22*(17), 9254.

Lee, S. S. (2020). A tailored ML process identifies differentially expressed genes from a small NSCLC dataset. *Exploration in Medicine.*

Li, J. (2024). Deep learning modeling and increasing interpretability of lung nodule classification with improved accuracy. *European Conference on Artificial Intelligence (ECAI).*

Li, S., et al. (2025). The best performance was obtained by the multi-feature fusion model integrating radiomics and deep learning features from PET and CT, yielding a C-index of 0.9345. *ResearchGate.*

Li, X., Wu, X., & Zhang, J. (2020). CT radiomics for SCLC/NSCLC classification. *Frontiers in Oncology, 10,* 593.

Li, Y., Wu, Q., Li, W., et al. (2024). Grad-CAM was integrated as an explainable AI (XAI) technique for enhancing model transparency. *Journal of Personalized Medicine, 14*(10), 1192.

Li, Y., Yang, X., et al. (2025). Margin irregularity and male gender associated with SCC in lung cancer. *Frontiers in Oncology.*

Liu, J., et al. (2024). An efficient and explainable ensemble-learning framework (EEE-framework) designed for early detection of non-small cell lung cancer (NSCLC) biomarkers. *Journal of Biomedical Informatics.*

Liu, J., et al. (2024). An efficient and explainable ensemble-learning framework for early lung cancer biomarkers detection. *IEEE Transactions on Computational Biology and Bioinformatics.*

Liu, Y., et al. (2021). Lung nodule malignancy classification with weakly supervised explanation generation. *Journal of Medical Imaging, 8*(4).

Mao, X., Zeng, D., Zhang, B., et al. (2025). A stacking ensemble framework integrating radiomics and deep learning for prognostic prediction in head and neck cancer. *ResearchGate.*

McDonald, B., Mulder, S., et al. (2023). ADC from diffusion-weighted MRI (DWI) shows promise as a tumor response biomarker in head & neck squamous cell carcinoma (HNSCC). *International Journal of Radiation Oncology, Biology, Physics.*

Niu, S., et al. (2025). Feature reliability assessment using ICC ensures that the extracted features are reliable and reproducible. *Frontiers in Oncology.*

Parekh, V., Jacobs, M. A., et al. (2019). Radiomic features provide information about grey-scale patterns, interpixel relationships, shape, and spectral properties. *Radiology.*

Park, Y., Lee, Y., & Kim, H. (2024). A radiological-radiomics model for differentiating ADC and SCC of the lung. *International Journal of Environmental Research and Public Health, 21*(5), 574.

Rahman, R. (2025). Federated learning: A survey on privacy-preserving collaborative intelligence. arXiv preprint.

Sharma, S., Singh, M., et al. (2025). XAI-based data visualization in multimodal medical data. *bioRxiv.*

 

Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2019). Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Expert Systems with Applications, 128,* 84–95.

Shi, Y., et al. (2025). SHAP summary diagrams illustrate feature effects in combined models. *Journal of Personalized Medicine.*

Tan, W. W. (2024). Lung cancer overview. *Medscape.*

Wang, S., Li, Y., Wu, X., et al. (2024). SHAP and Grad-CAM were employed for visualization and interpretation. *European Radiology.*

Yamamoto, K., et al. (2024). A contrastive language–image pre-training (CLIP)-based key slice selection framework for CT scans. *Radiology.*

Yao, I. Z., Dong, M., et al. (2025). Clinician trust and interpretability remain critical concerns in AI-driven cancer detection. *Cancer Informatics.*

Yao, J., Zhu, X., Zhu, F., & Huang, J. (2017). Deep correlational learning for survival prediction from multi-modality data. In *MICCAI 2017, 10434,* 406–414.

Zhang, H., Liu, Y., Zhao, Y., et al. (2024). The feature fusion model performed optimally in comparison to the other models. *Frontiers in Oncology.*

Zhang, H., Zhang, T., Li, W., et al. (2025). Radiomics and clinical features for distinguishing lung ADC and SCC. *Frontiers in Oncology, 14.*

 

221-35-990

ORIGINALITY REPORT

| 6% | 5% | 2% | 3% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Daffodil International University<br>Student Paper | 2% |
| 2 | Submitted to Midlands State University<br>Student Paper | 1% |
| 3 | www.frontiersin.org<br>Internet Source | <1% |
| 4 | Palani, Murali. "Using Artificial Intelligence to Analyze Tree Circadian Rhythms and Relationship With Geomagnetic Variations.", Capitol Technology University<br>Publication | <1% |
| 5 | www.mdpi.com<br>Internet Source | <1% |
| 6 | umpir.ump.edu.my<br>Internet Source | <1% |
| 7 | "ICAS 2021 Conference Proceedings [Front matter]", 2021 IEEE International Conference on Autonomous Systems (ICAS), 2021<br>Publication | <1% |
| 8 | Nazmul Siddique, Mohammad Shamsul Arefin, K. M. Azharul Hasan, M. Shamim Kaiser. "Data Driven Applications for Industry 4.0 and Beyond", CRC Press, 2025<br>Publication | <1% |
| 9 | public-pages-files-2025.frontiersin.org<br>Internet Source | <1% |

| | | |
|---|---|---|
| 10 | Submitted to islamicuniversity<br>Student Paper | <1 % |
| 11 | research.rug.nl<br>Internet Source | <1 % |
| 12 | "Medical Image Computing and Computer Assisted Intervention – MICCAI 2021", Springer Science and Business Media LLC, 2021<br>Publication | <1 % |
| 13 | Aman Kataria, Sita Rani. "Explainable AI for Healthcare - Real Life Applications and Use Cases for Practitioners", CRC Press, 2025<br>Publication | <1 % |
| 14 | Amit Kumar Tyagi, Shrikant Tiwari, S. V. Nagaraj. "Quantum Computing - The Future of Information Processing", CRC Press, 2025<br>Publication | <1 % |
| 15 | aclanthology.org<br>Internet Source | <1 % |
| 16 | Sukhpreet Kaur, Amanpreet Kaur, Manish Kumar. "Recent Advances in Computational Methods in Science and Technology", CRC Press, 2026<br>Publication | <1 % |
| 17 | psasir.upm.edu.my<br>Internet Source | <1 % |
| 18 | www.nature.com<br>Internet Source | <1 % |
| 19 | espace.library.uq.edu.au<br>Internet Source | <1 % |
| 20 | ultraexotics.shop<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 21 | eprints.usm.my<br>Internet Source | <1 % |
| 22 | hdl.handle.net<br>Internet Source | <1 % |
| 23 | ir.knust.edu.gh<br>Internet Source | <1 % |
| 24 | ousar.lib.okayama-u.ac.jp<br>Internet Source | <1 % |
| 25 | qa00.mdedge.com<br>Internet Source | <1 % |
| 26 | repository.ju.edu.et<br>Internet Source | <1 % |
| 27 | open.library.ubc.ca<br>Internet Source | <1 % |
| 28 | www.hkcochrane.cuhk.edu.hk<br>Internet Source | <1 % |
| 29 | www.physiciansweekly.com<br>Internet Source | <1 % |
| 30 | Sandeep Kumar Panda, Vaibhav Mishra, R. Balamurali, Ahmed A. Elngar. "Artificial Intelligence and Machine Learning in Business Management - Concepts, Challenges, and Case Studies", CRC Press, 2021<br>Publication | <1 % |
| 31 | Stoddard, Michael. "The Advancement of Experimental and Computation Tools for the Study of Molten Salt Chemistry to Facilitate the Extraction of Strategic Elements in Nuclear Applications", Brigham Young University, 2024<br>Publication | <1 % |

| | | |
|---|---|---|
| 32 | dspace.daffodilvarsity.edu.bd:8080<br>Internet Source | <1 % |
| 33 | pmc.ncbi.nlm.nih.gov<br>Internet Source | <1 % |
| 34 | www.researchgate.net<br>Internet Source | <1 % |

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | Off | | |

# Account Clearance

NUSRAT FARZANA CHOUDHURY
221-35-990

## Dashboard

Student Portal

| Total Payable | Total Paid | Total Due | Total Other |
|---|---|---|---|
| 767,200.00 | 767,200.00 | 0.00 | 0.00 |

### Today's Routine - Saturday

No routine available for today.

### Semester Wise Result