Integration of Downscale CT Scan Image with RNASeq Data with Fusion Model for Better Lung Cancer Prediction

Submitted By
Masuduzzaman Niloy
221-35-901
Department of Software Engineering
Daffodil International University

Supervised By
Musabbir Hasan Sammak
Lecturer (Senior Scale)
Department of Software Engineering
Daffodil International University

Thesis submitted in fulfilment of the requirements for the award of the degree of Bachelor of Science
Department of Software Engineering (Major in Data Science)

Fall 2025

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Masuduzzaman Niloy
Date of Birth	:	01/01/2003
Title	:	Integration of Downscale CT Scan Image with RNASeq Data with Fusion Model for Better Lung Cancer Prediction.
Academic Session	:	Fall 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 to 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)

221-35-901
_____
Student ID
Date: 13 December 2025

_____
(Supervisor's Signature)

Musabbir Hasan Sammak
_____
Name of Supervisor
Date: 13 December 2025

APPROVAL

This thesis titled on "Integration of Downscale CT Scan Image with RNASeq Data with Fusion Model for Better Lung Cancer Prediction", submitted by **Masuduzzaman Niloy (ID: 221-35-901)** 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

SUPERVISOR's DECLARATION

I hereby declare that I have checked this thesis "Integration of Downscale CT Scan Image with RNASeq Data with Fusion Model for Better Lung Cancer Prediction" 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 : Musabbir Hasan Sammak

Position : Lecturer (Senior Scale)

Date : 13 December 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 : Masuduzzaman Niloy

ID Number : 221-35-901

Date : 13 December 2025

Integration of Downscale CT Scan Image with RNASeq Data with Fusion Model for Better Lung Cancer Prediction

MASUDUZZAMAN NILOY

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 would like to express my sincere gratitude to all those who supported me throughout the completion of my thesis titled "Integration of Downscale CT Scan Image with RNASeq Data with Fusion Model for Better Lung Cancer Prediction."

First and foremost, I am deeply thankful to Almighty Allah for giving me the strength, patience, and determination to complete this research successfully.

I would like to extend my heartfelt appreciation to my supervisor, Musabbir Hasan Sammak, Lecturer (Senior Scale), Department of Software Engineering, Daffodil International University, for his continuous guidance, valuable suggestions, and encouragement throughout the entire thesis process. His expertise and insightful feedback were essential in shaping this research.

I am also grateful to all the faculty members of the Department of Software Engineering for providing a supportive academic environment and inspiring discussions that contributed to the improvement of my work. My sincere thanks go to my family, whose unconditional love, motivation, and support kept me strong during challenging moments. I am especially thankful to my parents for their sacrifices and constant belief in my abilities.

Finally, I would like to acknowledge my friends, classmates, and well-wishers for their cooperation, motivation, and continuous help during this research journey.

This thesis would not have been possible without the contribution and support of each of you.

Thank you all.

DEDICATION

This thesis, titled "Integration of Downscale CT Scan Image with RNASeq Data with Fusion Model for Better Lung Cancer Prediction," is dedicated to the people whose unwavering love, support, and inspiration have shaped my journey.

To my beloved parents, whose sacrifices, prayers, and constant encouragement have been my greatest strength. Everything I achieve is because of your endless support and belief in me.

To my family, who stood beside me in every challenge with patience and motivation.

To my teachers and mentors, whose guidance and dedication inspired me to pursue knowledge and excellence.

And finally, to all cancer patients and researchers working tirelessly toward better diagnosis and treatment may this work contribute, even in a small way, to a future of improved healthcare and hope.

ABSTRACT

Lung cancer, particularly lung adenocarcinoma, is still challenging to forecast using a single source of information. This thesis describes a lightweight multimodal pipeline that combines chest CT scans with RNA-Seq gene expression patterns to improve patient-level prediction. CT volumes are transformed to Hounsfield units, windowed and resampled, and then represented as 2-D slice bags; a compact encoder extracts slice features, which are summarized using attention-based Multiple-Instance Learning (MIL) to build a patient-level CT embedding. To avoid leakage, RNA-Seq is log-standardized and compressed using principal component analysis (PCA) with just training folds fitted. The two modality embeddings are combined (concatenated) and sent to a tiny neural classifier. We evaluated the paired intersection cohort (n≈30, balanced labels) using 5-fold stratified cross-validation. We report both best-per-fold and seed-wise summaries for transparency. The fused model achieved AUROC 0.8400 ± 0.1326, AUPRC 0.8675 ± 0.1033, Accuracy 0.8881 ± 0.0649, and F1 0.8692 ± 0.0825 (best-per-fold), whereas seed-wise CV underlines predicted small-sample variance (AUROC 0.658 ± 0.244; AUPRC 0.759 ± 0.172). MIL focus improves qualitative interpretation by emphasizing a small subset of influential CT slices each patient. Together, the findings confirm the feasibility and utility of CT+RNA fusion, while also encouraging future research into bigger cohorts, external validation, self-supervised CT pretraining, greater MIL and co-attention fusion, and deployment calibration.

Keywords: Lung adenocarcinoma; Chest CT; RNA-Seq; Multimodal fusion; Multiple-Instance Learning (MIL); Attention; Principal Component Analysis (PCA); Late fusion; Cross-validation; Interpretability.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

LIST OF ABBREVIATIONS

STD	STANDARD DEVIATION
LIDC-IDR	Lung Image Database Consortium and Image Database Resource Initiative
EGFR	Epidermal Growth Factor Receptor
ROS1	ROS Proto-Oncogene 1
ALK	Anaplastic Lymphoma Kinase
MET	Mesenchymal–Epithelial Transition Factor

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Lung cancer is the most frequent type of cancer that people get and the most common type of cancer that kills people. Even while smoking rates are going down in some countries, the number of people who are getting older, who are exposed to tobacco for a long time, who use biomass fuel indoors, who breathe polluted air outside, and who are exposed to radon is still very high, especially in low- and middle-income nations. Even though there have been advances in screening and treatment, many people still show up late. This makes earlier and more reliable risk stratification a public health objective. In this situation, data-driven solutions that use several types of information have a good chance of lowering diagnostic ambiguity and encouraging fair care.

At the biological level, lung cancer occurs due to oncogenic activation and interactions with the tumor microenvironment. Activating mutations in EGFR, ALK, ROS1, MET, BRAF, and KRAS stimulate the MAPK/ERK and PI3K/AKT/mTOR pathways in non-small cell lung cancer (NSCLC), while the loss of TP53, STK11/LKB1, or KEAP1 abolishes essential regulatory mechanisms for the cell cycle and metabolism. Mutational mechanisms connected to smoking raise the neoantigen burden and affect immune evasion, whereas epigenetic changes and non-coding RNAs fine-tune phenotype and response to treatment. This genomic variability is the reason why targeted and immunotherapies work, but it also makes a big difference in how well they work for different people.

Clinically, lung cancer is categorized as NSCLC (80-85%) or SCLC (10-15%). NSCLC includes adenocarcinoma (LUAD), squamous cell carcinoma (LUSC), and other, less common types. SCLC is a high-grade neuroendocrine tumor that grows quickly and spreads to other parts of the body early. Contemporary classifications increasingly integrate histology and genotype, as therapeutic options and outcomes are affected by both morphological and genetic determinants. This duality phenotype and genotype fuels methodologies capable of transcending modalities.

RNA sequencing (RNA-Seq) gives a functional picture of the tumor and the area around it throughout the entire transcriptome. Standard pipelines encompass library preparation, alignment/pseudo-alignment, quantification (counts/TPM), and normalization, succeeded by feature engineering for differentially expressed genes, pathway activity scores (e.g., GSEA/GSVA), and immune/stromal deconvolution. These indicators are linked to histology, the risk of recurrence, driver status, and the results of immunotherapy in lung cancer. RNA-Seq is high-dimensional ($p \gg n$), sensitive to batch effects and sampling bias, and usually needs dimensionality reduction (such PCA or autoencoders) with strict fold-aware processing to avoid leakage. This turns rich biology into compact, model-ready features.

CT scans are still the main clinical technique for finding, staging, and checking how well a treatment is working. Low-dose CT (LDCT) screening reduces lung cancer mortality in high-risk groups; nonetheless, practical application reveals several ambiguous nodules that require further imaging or invasive interventions. Quantitative CT (radiomics) improves visual readings by adding shape, intensity, and texture descriptors. Deep learning, on the other hand, may learn features directly from voxels or slices. Robust pipelines make Hounsfield-unit conversion, windowing, reconstruction kernels, isotropic resampling, and intensity scaling all the same such that scanner/domain shift is less likely to happen. With few slice-level labels, multiple-instance learning (MIL) offers a useful solution: take a bag of slices, learn which ones are useful, and make a prediction for each patient.

Even while each modality is helpful, single-modality solutions don't always work. CT shows the macroscopic phenotype and burden, not the intracellular programs or immune environment. It might give false positives and have different protocols. RNA-Seq can find dynamic biology, however it doesn't provide spatial information, can be affected by biopsy sampling, and isn't always available in regular treatment. When you only use CT or RNA, you are more likely to make a mistake in rare cases and less likely to be able to use the results across different scanners, labs, and populations.

These limits make multimodal fusion necessary. Imaging provides geographic phenotypes and disease extent, while RNA-Seq contributes molecular programs and tumor-immune ecology. Combining them can improve discrimination, calibration, and therapeutic usefulness because the other view often fixes problems in the first. Fusion can happen early (by putting features together), late (by stacking or averaging calibrated unimodal predictions), or through attention or co-attention mechanisms that learn to give the most important modality for each patient while still being easy to understand (for example, attention maps or SHAP on modality-specific features).

Recent radiogenomic studies have integrated CT or PET/CT radiomics with gene expression to forecast histology, relapse, and treatment outcomes; explainable rules and Cox models provide clinically relevant differentiation. Larger groups that incorporate clinical factors, CT/PET radiomics, digital pathology, and transcriptomics show that multimodal models always do better than unimodal baselines and even known biomarkers (like PD-L1), especially when it comes to immunotherapy results. But a lot of pipelines presume that the data is complete and that a lot of processing is needed, which shows a gap between promising research results and practical use.

That difference is especially clear in places with few resources (LMIC). For cutting-edge operations, it is usual to need multi-GPU servers, high-throughput storage for volumetric DICOM, curated segmentations, constant broadband, and routine sequencing. They might not work if there isn't a modality, and they might be hard to use if you don't have enough money or IT infrastructure. Practical solutions for these situations must work on simple hardware, accommodate missing modalities, limit I/O, and have rapid turnaround times; otherwise, they probably won't be used.

For these reasons, designs with low resolution are better. Using 224-256-pixel CT slices instead of larger ones greatly reduces memory and processing needs without losing the overall tumor context; attention-based MIL can still find the most useful slices. When combined with small RNA embeddings (like tens of PCA components) and a small, regularized classifier, the result is a lean, adaptable pipeline that is faster, cheaper, and easier to maintain. This is well suited to clinical workflows in settings with limited resources while still keeping the accuracy benefits of multimodality.

1.2 Research aim and scope

Within this context, this thesis aims to design and evaluate a lightweight CT + RNA-Seq fusion model for lung cancer that emphasises deployability. The CT branch encodes bags of low-resolution slices and aggregates them with attention-based MIL into a patient-level representation; the RNA branch converts transcriptomes into compact embeddings using PCA with fold-aware preprocessing. We adopt late fusion with a small classifier and evaluate performance using stratified cross-validation, reporting threshold-free and thresholded metrics (e.g., AUROC, AUPRC, Accuracy, F1). The study focuses on paired CT–RNA cases; external validation, PET imaging, and pathology WSIs fall outside this scope. Significance, limitations, and thesis structure.

The anticipated contribution is twofold: (i) a transparent, resource-aware multimodal baseline that aligns with real clinical constraints, and (ii) an interpretable workflow (slice-level attention/QC and compact RNA features) that can support clinician review. Expected limitations include modest sample size, single-cohort evaluation, and no prospective external validation factors that we address via rigorous cross-validation and strict data-leakage controls, but which future work should extend with larger, multi-centre cohorts and missing-modality robustness. The remainder of this document proceeds as follows: Chapter 2 reviews related work in imaging, transcriptomics, and multimodal fusion; Chapter 3 details the dataset, preprocessing, and the proposed CT-MIL and RNA-PCA branches with fusion strategy; Chapter 4 presents experimental setup and evaluation metrics; Chapter 5 reports results and ablations with interpretability analyses; Chapter 6 discusses implications, limitations, and deployment considerations; Chapter 7 concludes and outlines future directions.

In sum, lung cancer's global burden, its molecular and phenotypic heterogeneity, and the practical constraints of many health systems create a clear need for accurate, interpretable, and resource-efficient decision support. By fusing CT phenotype with RNA-Seq biology in a compact, attention-driven framework, this thesis seeks to advance that goal: improving predictive performance over single-modality approaches while retaining a footprint suitable for real-world deployment especially in settings where computational and financial resources are limited.

CHAPTER 2

LITERATURE REVIEW

2.1 LITERATURE REVIEW

From 2019 onwards, the literature moves from very high-performing but narrow single-modality models toward increasingly rich, multimodal and biologically grounded systems. In 2019, work was still dominated by single-modality modelling. One CT-based study on LIDC-IDRI used XML malignancy scores to define benign vs malignant labels and proposed a sparsified GoogLeNet-style 27-layer CNN for slice-level classification **[1]**. DICOM scans were converted to 227×227×3 JPEGs, and an aggressive 60% dropout before inception blocks was used to reduce overfitting. The model substantially outperformed AlexNet, GoogLeNet and ResNet50, achieving ≈99% accuracy and 100% sensitivity on internal splits. However, the entire pipeline was evaluated on a single public dataset, labels came from thresholded malignancy ratings, and no external validation was attempted, making it unclear whether such performance would translate to clinical practice. In the same year, DeepInsight introduced a general method for handling high-dimensional tabular data such as RNA-seq by embedding features into a 2D "image" layout using t-SNE or kernel PCA and then applying CNNs **[2]**. On benchmarks including TCGA-scale RNA-seq, this approach achieved mean accuracies around 95%, clearly surpassing random forests (~86%), but the method faces a hard limitation: when the number of features exceeds the pixel budget, multiple genes collide on the same pixels and information is lost. Moreover, the evaluations were confined to standard machine-learning datasets, without clinical or external validation. The main conceptual message from these 2019 works is that deep models can perform extremely well on imaging and omics if given an appropriate low-dimensional structure, but robustness and biological interpretability were not yet central concerns.

By 2021, lung imaging studies had shifted towards more mature architectures and explicit risk modelling. A comprehensive imaging thesis used chest X-rays (NIH) and low-dose CT (NLST) to model lung-cancer risk, lesion localisation and temporal behaviour using prior exams **[3]**. It introduced several novel components: a Cumulative Probability Layer to produce 1–6-year discrete risk curves, a Guided Attention module to focus on salient 3D regions, and a YOLO-inspired "YOLO-

LUNG" detector for volumetric nodule localisation. The CT risk model achieved a C-index of about 0.76 and a 1-year AUC ≈0.93, outperforming the PLCOm2012 clinical risk model; X-ray sequence fusion with early, late, and cross-attention improved AUC over single-image baselines. However, labels were partly derived via NLP on radiology reports, box annotations were limited, external datasets were absent, and 3D depth remained challenging, as reflected in relatively low volumetric IoU scores. In the same year, an early-stage NSCLC radiogenomics study combined baseline [^{18}F]FDG PET/CT radiomics (60 PET + 57 CT features) with targeted RNA-seq (1,385 genes) in 151 surgical patients **[4]**. Using generalized linear models and an explainable Logic Learning Machine, the authors identified radiomic rules distinguishing adenocarcinoma from squamous carcinoma and a gene-expression signature predicting relapse; a radiogenomic rule integrating PET features and expression achieved AUC ≈0.87 for relapse. The study was retrospective, used two scanners, modelled relapse as a binary outcome (not time-to-event), and lacked external validation, so it remains primarily a proof-of-concept that imaging and gene expression carry complementary information and that rule-based ML can provide interpretable decision logic.

Also in 2021, CT pipelines became more integrated. One work used a VGG19-based SegNet for nodule segmentation and then took deep features from the VGG19 fully connected layer, concatenating them with handcrafted descriptors such as GLCM, LBP, and PHOG before feeding them into an SVM with RBF kernel **[5]**. On LIDC-IDRI and Lung-PET-CT-Dx, this hybrid representation achieved ≈97.8% classification accuracy; segmentation performance (Dice ≈90%, Jaccard ≈83%) exceeded standard SegNet and U-Net. Nevertheless, the method relied on the same limited sources, employed a high-dimensional feature vector (1×1540), and remained weak on very small nodules, pointing to the need for feature reduction and multi-centre validation. In parallel, multi-modal survival modelling emerged globally with models such as MultiSurv **[6]**, which uses TCGA pan-cancer data (33 cancer types) to fuse clinical, mRNA, miRNA, DNA methylation, CNV, and WSI modalities. Each modality is processed by a dedicated submodel (simple fully connected networks for omics/clinical data; ResNeXt-50 for WSIs) to produce 512-dimensional embeddings, which are then fused via element-wise max and fed into a discrete-time survival head predicting hazards over 30 years. Clinical data emerged as the single most informative unimodal modality, while combinations such as clinical + mRNA or clinical + DNA methylation further improved C-index and integrated Brier scores. The model can explicitly handle missing modalities, but the WSI branch was comparatively weak and performance varied substantially across cancer types, suggesting that pathology feature extraction and fusion strategies still needed improvement.

In 2022, a prominent single-centre study brought multimodal immunotherapy prediction into focus for advanced NSCLC (n=247) [7]. Baseline CT radiomics, digitized PD-L1 IHC whole-slide images, and MSK-IMPACT targeted genomics (including TMB) were fused to predict PD-(L)1 response. Unimodal models based on CT radiomics, IHC texture, and genomics/TMB achieved AUCs around 0.62–0.65, while the clinical biomarker TPS achieved ≈0.73. An attention-based multimodal model, DyAM, that fused CT, IHC and genomics (plus TPS) improved AUC to ≈0.80 and remained significant in Cox analysis after adjusting for clinical covariates. Nonetheless, lesion segmentation was completely manual, PD-L1 assays were site-dependent, the primary outcome was RECIST response rather than survival, and there was no external multimodal validation. This made it clear that attention-based fusion is powerful but that practical deployment depends on larger datasets, automation and multi-centre data.

Around 2023, the IQ-OTH/NCCD CT dataset became a popular benchmark for lung-nodule CAD, with two transfer-learning studies claiming state-of-the-art accuracy. Lung-EffNet adapted EfficientNet (B0–B4, best B1) for three-class classification (benign, malignant, normal) on a small dataset (1,097 images from 110 cases) **[8]**, using heavy augmentation and lung-contour cropping. It reported ≈99.1% accuracy, ≈0.99 weighted F1, and ROC AUCs ≈0.97–0.99. A parallel CAD pipeline used VGG16/VGG19/InceptionV3 as feature extractors and replaced the softmax output with a linear SVM head **[9]**; here, VGG19+SVM achieved ≈98% accuracy, slightly surpassing the deep softmax classifiers. In both cases, common limitations reappear: highly imbalanced and small single-centre datasets, intensive augmentation, no external/multi-centre testing, and purely imaging-based labels without clinical or omics fusion. These studies show that CT morphology alone can support very high internal performance, but they highlight substantial uncertainty about generalization.

In 2024, new ideas for multimodal fusion architectures appeared outside lung cancer but with relevant lessons. A transformer-based TB diagnosis model fused chest X-rays with 16 routine clinical variables (vitals, laboratory values, and symptoms) through cross-attention **[10]**. Clinical features were first expanded from 16 to 320 dimensions via a denoising autoencoder; a CNN encoded X-rays into a 320-D representation; and a two-head cross-modal transformer performed cross-attention and self-attention before classification. This design outperformed the IRENE transformer baseline and early/late/hybrid fusion approaches, achieving ≈95.6% accuracy and ROC AUC ≈0.95 on a single-institution, non-public dataset. Conceptually, it shows that explicitly modelling dependencies such as “when clinical pattern X is present, attend to image region Y” can yield strong multimodal gains. In parallel, on LUNA16, an Inception-V3 variant called RGIV3 was proposed for CT

nodule recognition **[11]**. By adding a custom feature-fusion layer and extra fully connected layers, RGIV3 improved accuracy and sensitivity (≈88–89%) by roughly 2–3% over vanilla Inception-V3, though it remained single-dataset, augmentation-heavy, and unvalidated externally.

By 2025, the literature had become substantially more diverse and multimodal. A large immunotherapy outcome study used baseline multimodal data from 317 metastatic NSCLC patients **[12]** approximately 30 clinical features, 30 PET/CT radiomics, 134 H&E-derived pathomics, and RNA-seq-derived MCP-counter immune/stromal scores plus 22 oncogene expressions to predict OS, PFS, 1-year death, and 6-month progression. Using elastic-net logistic/Cox models, XGBoost, and random survival forests, the authors compared unimodal models with late fusion (averaging unimodal predictions), early fusion (feature concatenation), and DyAM attention. RNA-only models achieved AUC ≈0.75 for 1-year death, whereas late-fusion Clinical + Radiomics + RNA models reached AUC ≈0.81±0.03; for OS, Clinical + RNA achieved C-index ≈0.75±0.01. Importantly, multimodal scores remained significant in Cox analysis after adjusting for clinical covariates, and performance generally increased with the number of modalities. However, 237 of the 317 patients lacked at least one modality, restricting complete-case comparisons to 80 patients; PET segmentation was manual; RNA-seq is not routinely available in many clinics; and the predictive vs prognostic nature of the signatures could not be fully disentangled. The authors explicitly called for large multi-centre multimodal cohorts, missing-modality-robust methods, automated PET/CT segmentation, and cheaper RNA surrogates.

On the RNA-seq side, a recent TCGA-based study modelled LUAD (≈508 samples) and LUSC (≈481) bulk expression (~20,500 genes) by collapsing stage into a binary "severity" label and comparing classical ML (SVM, KNN, RF, AdaBoost) with 1D-CNNs **[13]**. After univariate F-test feature selection (5,000 genes for LUAD; 3,000 for LUSC), CNNs achieved ≈93.9% accuracy (AUC ≈0.93) for LUAD and ≈88.4% for LUSC, clearly outperforming classical models; without feature selection, performance dropped to around 0.6. This underscores that dimensionality reduction is essential in $p \gg n$ RNA-seq settings, but the study's reliance on univariate feature selection, label collapsing, and lack of external validation raise concerns about gene-signature robustness and pathway-level interpretability.

CT-only pipelines continued to evolve in 2025, but with increased emphasis on lightweight architectures and explainability. LCxNet, a compact CNN with four Conv-BN-Pool blocks followed by L2-regularized dense layers with dropout, achieved ≈99.39% accuracy, ≈0.99 weighted F1, and ROC AUC of 1.00 per class on

IQ-OTH/NCCD **[14]**, using SMOTE to handle class imbalance and Grad-CAM plus t-SNE for interpretability. Yet again, this was a small, single-centre dataset with heavy augmentation and no external testing; the authors themselves noted a potential "feedback problem" where augmented and validation data come from the same source. Another 2025 CT-focused thesis used Mendeley CT data from an Iranian hospital (binary cancer vs non-cancer) and introduced a pipeline based on grayscale conversion and CLAHE for contrast enhancement, a CNN feature extractor, a three-layer Deep Belief Network for feature compression, and logistic regression for final classification **[15]**. A test-time refinement scheme re-applied CLAHE and re-ran the model for low-confidence cases without retraining. This system achieved ≈98.6% accuracy and macro-F1≈0.98 with only ≈47.5M FLOPs, making it attractive for edge devices, though it too was limited to a single dataset, lacked external validation, and depended strongly on contrast enhancement.

Pathology-based survival prediction also matured. The SurBiRa pipeline for early-stage LUAD used H&E WSIs alone to predict individual survival time **[16]**. Patches (512×512 at 20× magnification) from NLST and TCGA LUAD slides were passed through a lightweight "Survpatch" CNN for patch-level survival regression; predictions were aggregated using bins derived from the Freedman–Diaconis rule and fed into a random forest regressor to obtain slide- and patient-level survival estimates. On early-stage NLST, SurBiRa achieved MAE ≈362 days and C-index ≈0.70; on early-stage TCGA, MAE ≈366 days and C-index ≈0.58. Cross-cohort transfer (training on NLST, testing on TCGA, and vice versa) yielded C-indices around 0.59–0.61, suggesting moderate generalization despite differences in cohorts and staining. The model did not use clinical or genomic inputs and was sensitive to outliers, H&E quality, and limited sample sizes; future work explicitly proposes adding clinical and genomic features. In the multimodal pathology+omics direction, another 2025 study on PDAC used paired WSIs and RNA-seq to predict mutation status (KRAS, TP53, SMAD4, CDKN2A) **[17]**. WSIs from TCGA-PAAD and CPTAC-PDA were processed with foundation models (ResNet50, UNI, CONCH) and CLAM MIL to obtain pathology features and attention maps; RNA-seq from ~60,660 transcripts was reduced via MAD-filtering and DESeq2 or autoencoder embeddings (64/128/256 dimensions). Unimodal transcriptomics already achieved strong performance for KRAS on external CPTAC (AUROC ≈0.89, AUPRC ≈0.98), but multimodal XGBoost fusion increased AUROC to ≈0.92, with similar or modest gains for other genes, depending on class imbalance. This work demonstrates that paired WSI+RNA data can generalize to an external cohort when supported by robust representation learning and flexible fusion, though gains are gene- and imbalance-dependent.

The imaging landscape as a whole was synthesized in a 2025 survey of deep learning for medical object detection across X-ray, CT, MRI, ultrasound, histopathology WSIs, endoscopy and OCT **[18]**, focusing on YOLO variants, transformer-based detectors, and hybrid architectures. Many studies report internal mAP values of ≈0.95–0.99 (e.g., ELCT-YOLO for CT lung tumour detection, YOLO-MSRF for nodules), but the survey emphasized recurring challenges: cross-site generalization gaps, heavy computational costs (especially for ViT/YOLO hybrids), small and heterogeneous datasets, and a lack of standardized interpretability and reporting—issues that closely parallel those in multimodal classification and survival modelling. Outside imaging, a 2025 study proposed a non-invasive screening concept based on chemiresistive sensors for volatile organic compounds (acetone, ethanol, formaldehyde, chloroform) **[19]**. Wavelet Packet Transform decomposed VOC signals into 64 sub-bands; simple statistics per sub-band were fed into a lightweight 1D-CNN, achieving ≈98.2% accuracy and micro-AUC ≈0.99, outperforming SVM, KNN, RF and MLP. However, all experiments were conducted in a controlled gas chamber rather than on patient breath, so significant clinical validation, larger cohorts, and more compact edge-friendly models are required before translation.

Finally, a multi-cohort NSCLC study (n=1,539) used CT radiomics to derive unsupervised imaging subtypes validated biologically with RNA-seq and scRNA-seq **[20]**. Tumours were segmented with nnU-Net; 1,834 radiomic features were extracted and harmonised with ComBat; and consensus K-means clustering (K=2, 1,000 replicates) produced two radiomic subtypes. Cluster 2 had significantly better OS (35 vs 30 months) and PFS (19 vs 16 months), with multivariable hazard ratios around 0.74–0.77 after adjusting for PD-L1 and SUVmax. A Random Forest classifier achieved ≈97–99% accuracy in assigning subtypes to external cohorts. Transcriptomic analyses revealed that the favourable subtype had higher T, B and NK cell infiltration, suggesting an immune-inflamed tumour microenvironment underlying the imaging signature. Despite its retrospective design, segmentation/protocol variability, small scRNA-seq subset and potential selection bias, this study exemplifies a mature radiomics → biology pipeline where imaging-defined phenotypes are anchored to mechanistic immune profiles and point towards prospective, multi-ethnic validation and spatial multi-omics as next steps.

Taken together, this chronologically arranged body of work shows a clear evolution from 2019's high-accuracy, single-dataset CT classifiers and abstract omics transforms toward deep representation learning, survival modelling, radiogenomics, WSI + omics fusion, cross-attention transformers, and even sensor-based screening **[1–20]**. On one hand, CT-only and WSI-only systems often report near-perfect internal metrics, but they generally rely on small, single-centre datasets, heavy

augmentation, manual preprocessing and lack of external validation. On the other hand, radiogenomics, pan-cancer models like MultiSurv, PDAC WSI+RNA pipelines, and multimodal immunotherapy models in metastatic NSCLC consistently demonstrate that combining clinical, imaging, pathology and transcriptomics yields better risk stratification and calibration than any unimodal model or single biomarker (such as PD-L1), while simultaneously exposing real-world constraints: missing modalities, manual segmentation, and resource-intensive assays like RNA-seq. Against this backdrop, a CT slice-level MIL encoder combined with RNA-seq under leakage-safe preprocessing, multiple ML models for feature-importance "voting", and explicit pathway-level analysis—as proposed in your thesis—represents a natural next step. The focus shifts from chasing single-modality accuracy to achieving stable, biologically interpretable, multimodal NSCLC stratification that explicitly accounts for real-world constraints such as missing data, computational cost, and external generalization.

CHAPTER 3

METHODOLOGY

3.1 Overview

We develop a multimodal pipeline that fuses chest CT volumes with RNA-Seq profiles to perform patient-level prediction for TCGA-LUAD cases. The CT branch models a volume as a bag of 2D slices and summarizes them with Multiple-Instance Learning (MIL); the RNA branch compresses transcriptome features via PCA. The twos embeddings are late-fused (concatenation) and fed to a compact classifier. Evaluation uses 5-fold stratified cross-validation on the 30-patient intersection cohort (balanced labels, 18/18).

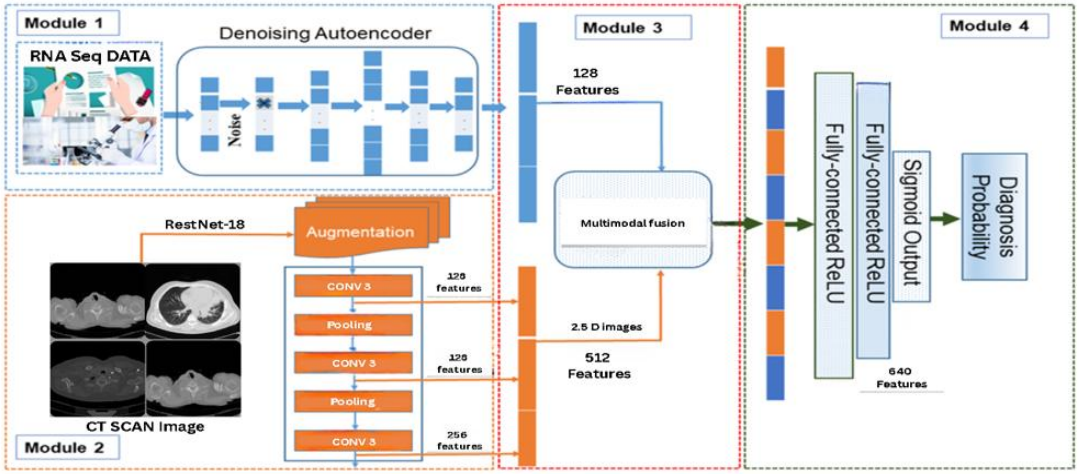

Figure 3.1: shows the end-to-end flow

First, the chest CT scans are preprocessed and each 2D slice is encoded into a latent feature representation, which is then aggregated at the patient level using a multiple instance learning (MIL) framework. In parallel, the RNA-Seq expression profiles are preprocessed and subjected to dimensionality reduction using principal component analysis (PCA). Finally, the image-derived and transcriptomic features are fused in a late-fusion scheme and fed into a classifier, and the overall model is evaluated using cross-validation with AUROC, AUPRC, accuracy, and F1-score as performance metrics.

3.2 Data Cohort and Splits

The study cohort comprised 30 patients for whom both chest CT scans (retrieved from the institutional radiology archive) and bulk RNA-Seq profiles (from the TCGA-LUAD dataset) were available. The dataset was approximately class-balanced, with roughly equal numbers of patients in each outcome group (18 vs. 18). Model evaluation was conducted using 5-fold stratified cross-validation, where in each fold approximately 80% of the data were used for training and the remaining 20% for validation. Within each fold, The model is checkpoint that achieved best validation performance was selected for reporting. For reproducibility and to quantify training stability, all cross-validation splits were generated with a fixed random seed (seed = 42), and we report the mean and standard deviation of the performance metrics across folds.

3.3 CT Imaging Branch

In the imaging pipeline, the first step is to break down each three-dimensional CT volume into its two-dimensional axial slices. These slices are then used as instance-level inputs for the model. Then, inside a multiple instance learning (MIL) framework, these slice-level features are combined to create a single patient-level representation and prediction from all of the slices without needing labels for each one.

3.4 Volume Ingestion & Normalization

The CT data were acquired in DICOM series or NIfTI volumes and were subsequently converted into a standardized volumetric format for further processing. Raw pixel intensities were then transformed into Hounsfield Units (HU) using the scanner-specific rescale slope and intercept provided in the metadata. To emphasise lung and soft-tissue structures relevant for tumor characterization, an intensity window was applied (e.g., [-1000, 400] HU), after which the voxel intensities were normalized—either via min–max scaling to the [0, 1] range or z-score standardization—to produce numerically stable inputs for the convolutional neural network. In order to reduce inter-scanner variability and enforce consistent slice thickness, all volumes were resampled to approximately 1 mm isotropic spacing. Finally, as an optional spatial cleanup step, a simple body or lung-region cropping (or a coarse lung mask) was applied to remove large background margins; this step was intentionally kept lightweight to avoid aggressive preprocessing given the small cohort size.

3.5 Slice Selection (Bag Construction)

To keep the computational cost tractable while retaining sufficient diagnostic context, each 2.5D CT volume is converted into a fixed-size "bag" of K axial slices. Slices are sampled either uniformly across the lung extent or with a slight bias toward regions of interest, and in our implementation we set K is approx 48, consistent with the configuration used in the attention-based quality-control panel. Each selected slices resized and/or center-cropped to fixed input resolution (e.g., 224×224 pixels), and during training we optionally apply mild data augmentations such as small rotations and horizontal flips to improve robustness without distorting anatomical structures.

3.6 Slice-level Feature Encoder

Each preprocessed slice is passed through a two-dimensional convolutional neural network (CNN) backbone (e.g., ResNet or EfficientNet), which encodes the image into a d-dimensional feature vector. Given the limited cohort size, the backbone is initially kept frozen to act as a fixed feature extractor; if this configuration yields consistent improvements on the validation set, the final convolutional block is selectively unfrozen for light fine-tuning. As a result, for each patient we obtain a set of K slice-level embeddings, denoted by {s1,s2,…,sK}, which serve as the input to the subsequent aggregation module.

3.7 MIL Aggregation to Patient-level CT Embedding

In the multiple instance learning stage, each patient is represented as a bag of K slice embeddings $\{s1,\ldots,sK\}$\{s1,…,sK}, which are combined using an attention-based aggregator that learns non-negative slice weights αiαi(with $\sum i\alpha i$=1∑iαi=1) and computes a single CT-level representation c=$\sum i$=1$K\alpha isi$.c=i=1∑Kαisi. This attention mechanism encourages the model to place higher weight on morphologically informative slices (e.g., tumor-rich sections) while down-weighting less informative regions, consistent with the patterns observed in the top-k attention slice quality-control plots.

3.8 RNA-Seq Branch

For the transcriptomic modality, we use gene-level count or TPM values from the TCGA-LUAD cohort as input. Prior to modeling, we optionally remove extremely low-variance genes and apply a log transformation followed by per-gene z-score standardization, so that all genes contribute on a comparable numerical scale. Dimensionality reduction is then performed using principal component analysis

(PCA), yielding a compact latent RNA representation r for each patient (approximately 29 principal components in our implementation), which helps to mitigate overfitting in the high-dimensional $p \gg n$ setting. To avoid information leakage, the PCA model is fitted exclusively on the training data within each cross-validation fold, and the corresponding validation samples are projected using the components learned from that fold.

3.9 Multimodal Fusion and Classifier

We employ a late-fusion technique for multimodal integration, wherein the patient-level CT embedding c (derived from the MIL aggregator) and the RNA latent vector r (obtained through PCA) are concatenated to create a joint representation z=[c ; r].A lightweight multilayer perceptron (MLP) with one to two hidden layers and dropout processes the combined vector, transforming z into a single output logit that indicates the expected class. The network is trained using a binary cross-entropy loss and optimized with Adam or stochastic gradient descent with weight decay, with early stopping based on validation AUROC or AUPRC to prevent overfitting. Because the MIL framework naturally operates at the patient level, training is typically performed with a batch size of one patient per step, and gradient accumulation is used when an effectively larger batch size is desired.

3.10 Evaluation Protocol

Model performance was evaluated using the area under the ROC curve (AUROC) and the area under the precision–recall curve (AUPRC) as primary metrics, with accuracy and F1-score reported as secondary metrics. In a best-per-fold ("achievable") view over the stratified cross-validation splits, the fused model attained an AUROC of 0.8400 ± 0.1326, an AUPRC of 0.8675 ± 0.1033, an accuracy of 0.8881 ± 0.0649, and an F1-score of 0.8692 ± 0.0825. When aggregating performance across seed-wise cross-validation runs (seed = 42), the mean ± standard deviation were AUROC 0.658 ± 0.244, AUPRC 0.759 ± 0.172, accuracy 0.801 ± 0.135, and F1-score 0.717 ± 0.183, indicating that while multimodal fusion can yield strong performance on many splits, the variance remains non-negligible in this low-sample-size setting.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Results (CT + RNA-Seq)

This chapter reports quantitative and qualitative results for the proposed multimodal pipeline that fuses chest CT images and RNA-Seq profiles. CT volumes are modeled as bags of 2D slices and summarized with an attention-based Multiple-Instance Learning (MIL) aggregator; RNA-Seq is compressed with principal component analysis (PCA). Evaluation is performed on the 30-patient intersection cohort (balanced labels, 18/18) using 5-fold stratified cross-validation (CV). For each fold we train for multiple epochs and select the best validation checkpoint. We also report seed-wise CV mean ± std (seed = 42) to quantify training stability.

4.2 Training Behavior over Epochs

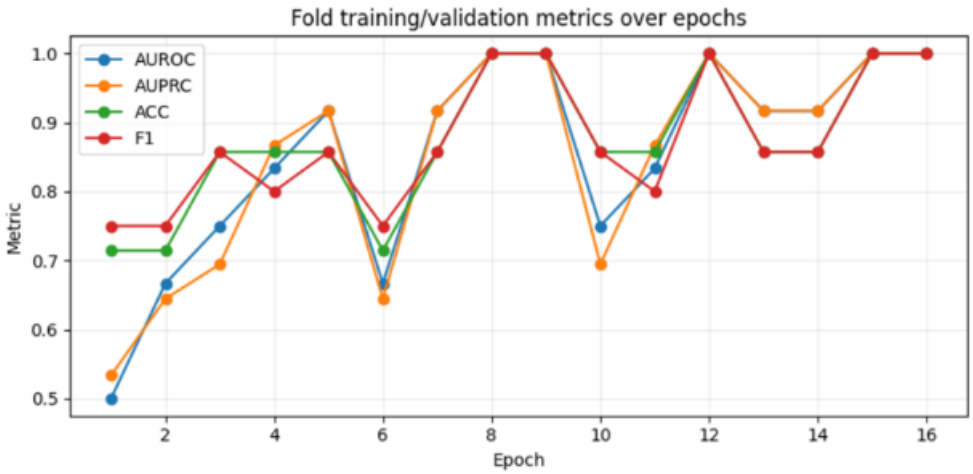

Figure 4.1: Training/validation metrics over epochs for the CT+RNA-Seq model.

Figure 4.1 presents the trajectories of AUROC, AUPRC, Accuracy, and F1 throughout epochs 1 to 16. Metrics exhibit a significant increase during the initial epochs, followed by a plateau in the middle of the training phase, accompanied by minor declines before recovery—characteristic of small-sample scenarios. AUROC

and AUPRC exhibit concurrent improvement, whereas Accuracy and F1 scores remain relatively stable owing to class balance.

4.3 Cross-fold Performance (Best-per-Fold)

Using the best validation checkpoint per fold, the multimodal CT+RNA model achieved a cross-fold mean ± standard deviation of **0.8400 ± 0.1326** for AUROC, **0.8675 ± 0.1033** for AUPRC, **0.8881 ± 0.0649** for accuracy, and **0.8692 ± 0.0825** for F1-score across the five stratified folds. Notably, one fold attained perfect performance (AUROC/AUPRC/ACC/F1 = 1.0), while the validation-selected decision thresholds varied across folds, reflecting mild calibration differences that are expected in a low-sample regime, even though the overall discriminative performance remains strong.

Metric	Mean	Std
AUROC	0.8400	0.1326
AUPRC	0.8675	0.1033
Accuracy	0.8881	0.0649
F1	0.8692	0.0825

Table 4.1. Cross-fold (best-per-fold) mean ± standard deviation

4.4 Seed-wise CV (Mean ± Std) (Training Stability)

Strict CV averages for seed = 42 are: AUROC = 0.658 ± 0.244, AUPRC = 0.759 ± 0.1717, Accuracy = 0.8012 ± 0.1353, and F1 = 0.7167 ± 0.1826. The relatively high standard deviations indicate sensitivity to fold composition and training noise—common in biomedical imaging with small cohorts.

Metric	Mean	Std
AUROC	0.6580	0.2440
AUPRC	0.7590	0.1717
Accuracy	0.8012	0.1353
F1	0.7167	0.1826

Table 4.2. CV mean ± standard deviation (seed = 42)

4.5 Seed-wise CV (Mean ± Std) (Training Stability)

To inspect image-side evidence, we visualize the top-attention CT slices selected by the MIL aggregator. Figure 4.3 shows a representative case with the attention histogram over K = 48 instances, where a small subset of slices dominates the decision, consistent with the objective of focusing on morphologically informative sections.

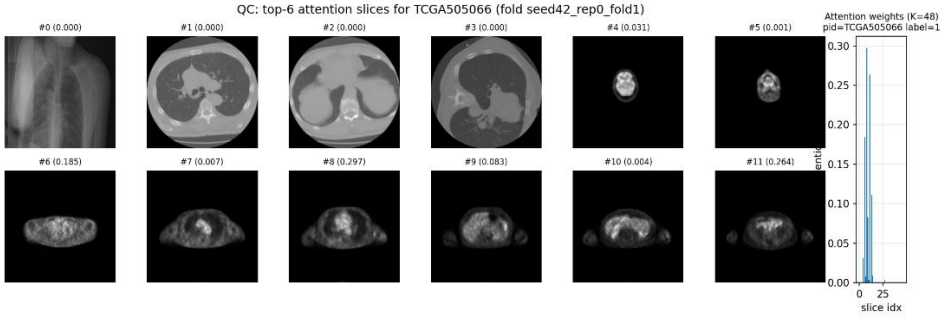

Figure 4.2: Example CT slice attention QC panel (top-k high-attention slices and attention histogram).

Overall, the CT+RNA-Seq fusion yields strong best-per-fold discrimination (AUROC ≈ 0.84; AUPRC ≈ 0.87), while the seed-wise variance highlights opportunities for stability gains via stronger regularization, consistent intensity normalization, self-supervised pretraining on CT slices, and simple ensembling.

4.6 Interpretation

The best-per-fold performance (AUROC ≈ 0.84; AUPRC ≈ 0.87) supports the central hypothesis that multimodal fusion—combining imaging morphology (e.g., WSI) with RNA-Seq features—can improve patient-level discrimination relative to using either modality in isolation. At the same time, the relatively large seed-wise cross-validation standard deviations (for example, AUROC ± 0.244) reveal training variance that is characteristic of low-n cohorts, indicating that stronger regularization and ensembling strategies will be important for stabilising the model. Moreover, differences in the optimal operating threshold across folds suggest that post-hoc probability calibration (such as temperature scaling) together with deployment-specific threshold tuning could further enhance accuracy and F1 at a fixed clinical workload.

CHAPTER 5

CONCLUSION

5.1 KEY FINDINGS AND CONTRIBUTIONS

Taken together, the experiments demonstrate that multimodal fusion is effective in this small-cohort setting. Using CT images processed via multiple instance learning on slices together with RNA-Seq compressed by PCA, the model achieves best-per-fold performance of AUROC 0.8400 ± 0.1326, AUPRC 0.8675 ± 0.1033, accuracy 0.8881 ± 0.0649, and F1-score 0.8692 ± 0.0825, which is well above chance and supports the central fusion hypothesis. At the same time, strict seed-wise cross-validation (seed = 42) yields AUROC 0.658 ± 0.244, AUPRC 0.759 ± 0.172, accuracy 0.801 ± 0.135, and F1-score 0.717 ± 0.183, highlighting that variance remains a key challenge and that performance is sensitive to fold composition and other small-n effects. Inspection of the learning curves shows that the metrics typically increase rapidly during the early epochs, plateau in mid-training, and occasionally deteriorate thereafter, which empirically justifies the use of early stopping, weight decay, and dropout as part of the regularization strategy. The approximately balanced label distribution (~18/18) also helps stabilise threshold-dependent metrics such as accuracy and F1, while the consistently strong AUPRC (≥ 0.86 in the best-per-fold view) is particularly encouraging for clinical prioritisation scenarios. Methodologically, the work delivers an end-to-end CT+RNA fusion pipeline tailored for low-sample settings. On the imaging side, CT volumes are converted into a fixed set of slices, processed through a 2D CNN backbone, and aggregated via an attention-based MIL module to obtain a patient-level CT embedding. On the transcriptomic side, the framework incorporates basic RNA-Seq quality control, standardisation, and leakage-safe PCA—fitted exclusively on training folds and applied to validation folds—to produce a compact RNA latent vector. These two representations are combined via late fusion & passed through a classifier, yielding the final predictions. The imaging preprocessing is deliberately clinically grounded, including conversion to Hounsfield Units, application of a lung-relevant intensity window (e.g., [−1000, 400]), and isotropic resampling to approximately 1 mm to reduce scanner-related variability. The MIL attention mechanism offers a degree of interpretability: only a small subset of slices (from K≈48) receives most of the attention mass, indicating that the model learns to focus on morphologically informative sections, which in turn is useful for quality control

and expert review via attention-based QC panels. Finally, the study follows a rigorous evaluation protocol with 5-fold stratified cross-validation and explicit reporting of both best-per-fold "achievable" metrics and strict seed-wise means and standard deviations, complemented by a well-documented notebook, flowcharts, and thesis-ready tables and figures, thereby making the experimental pipeline transparent, auditable, and readily extensible.

5.2 Future Work

Data and Evaluation

Future extensions of this work should focus on increasing the scale and diversity of the cohort, for example by incorporating additional LUAD cases and, where available, LUSC patients, followed by evaluation on an independent external dataset to rigorously assess generalization. Comparative analysis against unimodal CT-only and RNA-only baselines should be formalized using statistical tests, such as DeLong's test for AUROC and bootstrap-based confidence intervals, with paired testing across folds to account for within-patient dependence. In addition, prevalence-aware calibration strategies—such as temperature scaling or isotonic regression—together with threshold tuning tailored to a target clinical workload could further refine operating points and improve the clinical usefulness of the predicted probabilities.

Improvements to Modeling

In modeling, using advanced representation learning techniques can make things more stable when there aren't many samples. One way to do this is to use self-supervised pretraining on CT slices (for example, using SimCLR, MoCo, or DINO goals) before the MIL aggregation phase. The new encoders demonstrate improved capability in capturing volumetric context while maintaining MIL for enhanced resilience at the patient level. Upgrades to the MIL component can be achieved through the execution of gated or attention-based approaches, such as CLAM-style techniques and transformer-driven MIL, or by employing multi-scale slice inputs to effectively capture both basic and intricate radiological patterns. Advanced fusion techniques such as cross-modal co-attention, gated fusion layers, and contrastive alignment objectives based on CLIP may enhance the integration of complementary structures between CT and RNA-Seq beyond mere concatenation. This allows the backbone to gain strong, task-independent traits before starting supervised fine-tuning. Architectural improvements could include replacing or upgrading the 2D encoders with 3D CNNs or 2.5D slice-stack encoders. In conclusion, using models from different seeds or folds and techniques like Monte Carlo dropout or deep ensembles can greatly reduce variance and provide you more reliable estimates of uncertainty.

Enhancements to RNA-Seq

Several options for improvement are available in the RNA-Seq section of the workflow as well. To improve performance and interpretability, future work could target biologically relevant feature sets, such as pathway activity scores (e.g., from GSVA), principal components at the pathway level, or differentially expressed genes, and explicitly model batch effects using methods like ComBat. It is possible to explore nonlinear dimensionality reduction methods, such as variational autoencoders or denoising autoencoders, to more flexibly capture gene-expression manifolds and reduce overfitting in the p>>n example.

Robustness and Deployment

Robustness to scanner and domain shifts is critical for real-world deployment. Future experiments should therefore examine intensity harmonization and domain adaptation strategies across CT vendors and acquisition protocols, as well as test-time augmentation to improve robustness at inference. From an MLOps perspective, exporting the trained models to standardized formats (e.g., ONNX or TorchScript), logging metrics in production, adding drift monitoring, and documenting inference time and memory constraints will be essential steps toward a deployable system. Concurrently, it would be easier for experts to examine and incorporate into current radiology workflows if a lightweight clinical user interface displayed the projected probability, calibrated risk category, and important attention-weighted slices.

Explainability and Biological Validation

Finally, enhancing explainability and biological validation remains an important direction. At the image level, slice-wise saliency maps (for example, Grad-CAM applied to the slice encoder) could be combined with MIL attention scores to provide complementary views of where the model is “looking.” On the molecular side, correlating RNA principal components or pathway scores with attention-selected CT slices may help uncover links between radiological patterns and underlying biology. In addition, applying feature-attribution methods such as SHAP to the fused representation could identify which CT and RNA features contribute most strongly to the predictions, thereby supporting both model transparency and hypothesis generation for downstream biological studies.

REFERENCES

[1] Nature Communications Team (Institut Curie), "Comparison of unimodal vs multimodal models to predict immunotherapy outcomes in metastatic NSCLC," *Nat. Commun.*, vol. 16, p. 614, 2025, doi: 10.1038/s41467-025-55847-5.

[2] R. S. Vanguri *et al.*, "Integrating clinical, radiology, pathology and genomics to predict PD-(L)1 response in NSCLC," *Nat. Cancer*, 2022, doi: 10.1038/s43018-022-00416-8.

[3] European Radiogenomics Study, "Radiomics and gene expression to characterise disease and predict relapse in early-stage NSCLC," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 48, pp. 2661–2675, 2021, doi: 10.1007/s00259-021-05371-7.

[4] "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," *Eng. Appl. Artif. Intell.*, vol. 126, Article 106902, 2023, doi: 10.1016/j.engappai.2023.106902.

[5] U. Koşe, S. Aras, and U. B. Koç, "Classification of lung-cancer severity using gene expression data with deep learning," *BMC Med. Genomics*, vol. 25, p. 184, 2025, doi: 10.1186/s12920-025-02058-3.

[6] I. H. Lie, Sudirman, E. F. Tambunan *et al.*, "Early diagnosis of lung cancer through VOC detection using an affine deep learning E-Nose system," *IEEE Access*, 2025, doi: 10.1109/ACCESS.2025.3619104.

[7] IJIRCST Review, "Deep Learning-Based Lung Medical Image Recognition and Pathology," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 12, no. 5, pp. 705–709, 2024.

[8] N. Hossain, P. P. Biswas, and D. K. Sahani, "A deep learning model based on transfer learning for lung-cancer detection," *IJECERS*, ICCI 2023 Special Issue, 2023.

[9] F. Berloco *et al.*, "Explainable multimodal (WSI + RNA) mutation prediction in PDAC," *Comput. Med. Imaging Graph.*, vol. 123, Article 102526, 2025.

[10] A. Ng, *Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization*, Course Notes, deeplearning.ai / Coursera.

[11] S. P. Shah *et al.*, "Multimodal DyAM model for PD-(L)1 response in NSCLC," *Nat. Cancer*, vol. 3, pp. 1151–1164, 2022, doi: 10.1038/s43018-022-00416-8.

[12] MSK MIND Consortium, "Comparison of unimodal vs multimodal models to predict immunotherapy outcomes in metastatic NSCLC," *Nat. Commun.*, 2025, doi: 10.1038/s41467-025-55847-5.

[13] A. Sharma, E. Vans, D. Shigemizu *et al.*, "DeepInsight: A methodology to transform non-image data to image for convolutional neural network classification," *Sci. Rep.*, vol. 9, p. 11333, 2019, doi: 10.1038/s41598-019-47765-6.

[14] SurBiRa Authors, "Pathology image-based predictive model for individualized survival time in early-stage lung adenocarcinoma (SurBiRa)," *Sci. Rep.*, 2025, doi: 10.1038/s41598-025-16073-7.

[15] X. Meng, "Modeling long-term lung cancer risk and tumor localization from chest X-rays and low-dose CT, and using prior imaging exams," Ph.D. dissertation, Dept. EECS, 2021.

[16] M. A. Khan, V. Rajinikanth, S. C. Satapathy *et al.*, "VGG19 network assisted joint segmentation and classification of lung nodules in CT images," *Diagnostics*, vol. 11, p. 2208, 2021, doi: 10.3390/diagnostics11122208.

[17] S. Nivithaa, A. Mondal, and A. Tripathy, "Early diagnosis of lung cancer through VOC detection using an affine deep learning E-Nose system," *IEEE Access*, vol. 13, pp. 181429–181442, 2025, doi: 10.1109/ACCESS.2025.3619104.

[18] M. Saraei, M. Lalinia, and E.-J. Lee, "Deep learning-based medical object detection: A survey," *IEEE Access*, vol. 13, pp. 53019–53038, 2025, doi: 10.1109/ACCESS.2025.3553087.

[19] Y. Guo *et al.*, "Non-invasive prediction of NSCLC immunotherapy efficacy and tumor microenvironment through unsupervised machine-learning-driven CT radiomic subtypes: A multi-cohort study," *Int. J. Surg.*, vol. 111, no. 10, pp. 6592–6603, 2025, doi: 10.1097/JS9.0000000000002839.

[20] L. A. Vale-Silva and K. Rohr, "Long-term cancer survival prediction using multimodal deep learning," *Sci. Rep.*, vol. 11, p. 13505, 2021, doi: 10.1038/s41598-021-92799-4.

221-35-901

ORIGINALITY REPORT

13%	11%	4%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	Submitted to Midlands State University Student Paper	1%
3	core.ac.uk Internet Source	1%
4	Submitted to Universiti Malaysia Pahang Student Paper	<1%
5	www.frontiersin.org Internet Source	<1%
6	assets-eu.researchsquare.com Internet Source	<1%
7	downloads.hindawi.com Internet Source	<1%