A Hybrid Deep Learning Framework for Combining Quantitative Radiomics Features with Serum Numerical Biomarkers in Non-Small Cell Lung Cancer Staging

Authors

Hassan Bilal, Robert Ardagh, Edafe Fego, Eli Sanchez, Sunday Oladele

Date: June 25, 2026

Abstract

Non-small cell lung cancer (NSCLC) remains the leading cause of cancer-related mortality worldwide, with accurate staging being paramount for determining optimal treatment strategies and predicting patient outcomes. Traditional staging methods relying on invasive procedures and qualitative assessments are time-consuming, subject to inter-observer variability, and may not capture the full extent of tumor heterogeneity. Recent advances in radiomics and machine learning offer promising non-invasive alternatives, yet existing approaches typically employ unimodal data sources, limiting their discriminative power. This study addresses this gap by proposing a hybrid deep learning framework that synergistically integrates quantitative radiomic features extracted from computed tomography (CT) images with serum numerical biomarkers, including C-reactive protein (CRP), tumor mutation burden (TMB), and lactate dehydrogenase (LDH). The framework employs a multimodal fusion architecture combining a deep neural

network for radiomic feature processing with gradient boosting for biomarker integration. Evaluation on a retrospective dataset of 422 NSCLC patients demonstrated superior performance over unimodal approaches, achieving a classification accuracy of 89.4% (95% CI: 86.2-92.1%) and an AUC-ROC of 0.93 (95% CI: 0.90-0.95). The hybrid model significantly outperformed radiomics-only (78.0% accuracy) and biomarker-only (85.2% accuracy) approaches ($p < 0.01$). Feature importance analysis identified wavelet-transformed texture features and TMB as the most influential predictors. This framework provides a replicable, non-invasive tool for enhancing NSCLC staging accuracy, with implications for personalized treatment planning and improved clinical decision-making.

Keywords: Non-Small Cell Lung Cancer, Radiomics, Numerical Biomarkers, Deep Learning, Multimodal Fusion, Cancer Staging, Machine Learning

1. Introduction

1.1 Background

Lung cancer constitutes the foremost cause of cancer-related mortality globally, with non-small cell lung cancer (NSCLC) accounting for approximately 80-85% of all primary lung malignancies [1]. The accurate staging of NSCLC is fundamental to clinical management, as it directly informs treatment selection, prognostic assessment, and therapeutic stratification. The Tumor-Node-Metastasis (TNM) classification system, endorsed by the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC), serves as the international standard for cancer staging [2]. However, pathological staging, while considered the gold standard, is constrained by several limitations including invasiveness, sampling errors, and unfeasibility in advanced disease or patients with significant comorbidities [3].

In recent years, medical imaging has emerged as a pivotal non-invasive modality for cancer characterization. Computed tomography (CT) remains the primary imaging tool for lung cancer evaluation, providing anatomical information essential for diagnosis and staging. However, conventional visual interpretation of CT images is inherently subjective and may not fully exploit the quantitative information embedded within medical images [4]. This limitation has catalyzed the development of radiomics—a high-throughput approach that extracts hundreds of quantitative features from medical images to capture tumor heterogeneity and phenotype non-invasively [5].

Concurrently, advances in molecular diagnostics have identified numerous serum biomarkers with prognostic and predictive significance in NSCLC. Among these, C-reactive protein (CRP), tumor mutation burden (TMB), and lactate dehydrogenase (LDH) have demonstrated

correlations with tumor stage, aggressiveness, and patient outcomes [6]. The Neutrophil-to-Lymphocyte Ratio (NLR), Platelet-to-Lymphocyte Ratio (PLR), and other inflammatory indices have similarly been associated with NSCLC prognosis, reflecting the interplay between systemic inflammation and tumor progression [7].

The integration of artificial intelligence (AI) and machine learning techniques has transformed the analysis of both imaging and biomarker data in oncology. Deep learning methods, particularly convolutional neural networks and multilayer perceptrons, have shown remarkable capabilities in extracting discriminative patterns from high-dimensional data [8]. These approaches offer opportunities for comprehensive analysis of multi-dimensional data, potentially improving prognostic and predictive accuracy [3].

1.2 Problem Statement

Despite significant advances in both radiomics and serum biomarker analysis for NSCLC characterization, existing approaches predominantly employ unimodal data sources, thereby failing to leverage the complementary information offered by different data modalities [9]. Studies utilizing radiomic features alone, while demonstrating potential for histological subtype prediction and survival modeling, have yielded variable performance with accuracy ranging from 78% to 83% [10][11]. Similarly, machine learning models based on numerical biomarkers have achieved classification accuracies of approximately 85-89% for cancer stage prediction [6]. However, the integration of these complementary data sources into a unified predictive framework remains underexplored.

The research conducted by Sunny et al. (2024) demonstrated the efficacy of machine learning models, including Support Vector Machines (SVM), Random Forest (RF), and Multi-Layer Perceptron (MLP), in classifying cancer stages using numerical biomarker data, achieving a maximum accuracy of 91.4% with MLP [6]. While these results are promising, the study focused exclusively on biomarker data, leaving the potential synergy with imaging-based radiomic features untapped. Furthermore, existing radiomic studies often focus on specific clinical endpoints such as histological subtyping or survival prediction rather than comprehensive stage classification [10][11].

A critical gap exists in the development of validated, multimodal frameworks that can effectively integrate quantitative radiomic features with serum numerical biomarkers for NSCLC staging. The challenge lies not only in data integration but also in addressing the inherent heterogeneity of NSCLC, which manifests in both imaging phenotypes and molecular profiles. The combination of these data sources through artificial intelligence techniques holds promise for providing more accurate staging and outcome prediction, as suggested by emerging clinical evidence [3][9].

1.3 Objectives of the Study

General Objective:
To develop and validate a hybrid deep learning framework that integrates quantitative radiomic features from CT imaging with serum numerical biomarkers for accurate non-invasive staging of non-small cell lung cancer.

Specific Objectives:

1. To identify the most predictive radiomic features and serum biomarkers for NSCLC stage classification through comprehensive feature selection and importance analysis.
2. To design and implement a hybrid deep learning architecture incorporating a deep neural network for radiomic feature processing and gradient boosting for serum biomarker integration, with multimodal fusion at the decision level.
3. To validate the proposed framework against unimodal approaches (radiomics-only and biomarkers-only) using rigorous cross-validation and statistical testing, quantifying performance improvements in terms of accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC).

1.4 Research Questions

1. What combination of quantitative radiomic features and serum numerical biomarkers most accurately predicts NSCLC stage classification?
2. How does the proposed hybrid deep learning framework compare to unimodal approaches (radiomics-only and biomarkers-only) in terms of classification accuracy and discriminative power?
3. What are the most influential features contributing to the hybrid model's predictive performance, and how do imaging-based and biomarker-based features complement each other in NSCLC staging?

1.5 Significance of the Study

For Clinicians and Healthcare Administrators:
This study provides a non-invasive, replicable framework for enhancing NSCLC staging accuracy, potentially reducing the need for invasive diagnostic procedures and enabling more timely treatment decisions. The quantitative nature of the approach minimizes inter-observer variability and provides objective staging information that can complement clinical judgment.

For Patients:
Accurate staging through non-invasive means reduces patient exposure to invasive procedures and associated risks. Improved staging accuracy directly translates to more appropriate treatment selection, potentially improving outcomes and reducing unnecessary therapeutic interventions.

For Academic Literature:
This research extends the existing body of knowledge by demonstrating the complementary value of integrating imaging-based radiomics with serum biomarkers through deep learning. It contributes a methodological framework for multimodal data fusion in oncological applications and provides benchmark performance metrics for future comparative studies.

For Future Researchers:
The proposed framework and the identified feature importance rankings offer a foundation for further investigation into multimodal cancer characterization. The methodology can be extended to other cancer types, integrated with additional data modalities (e.g., genomic, proteomic), or refined with larger prospective cohorts.

1.6 Scope and Limitations

Scope:
This study focuses on NSCLC stage classification using CT-based radiomic features and serum numerical biomarkers (CRP, TMB, LDH). The dataset comprises 422 NSCLC patients with confirmed pathological diagnoses and available staging information. The analysis is retrospective in nature, utilizing publicly available and institutional data sources. The proposed framework is evaluated using rigorous cross-validation and compared against unimodal baselines.

Limitations:
The retrospective nature of the study may introduce selection bias, and the results require prospective validation. The use of a single imaging modality (CT) may limit the generalizability of radiomic features, whereas multimodal imaging (e.g., PET/CT) could provide additional discriminatory information. Furthermore, the framework's performance on specific NSCLC subtypes (adenocarcinoma vs. squamous cell carcinoma) remains to be investigated. The sample size, while sufficient for initial validation, may limit the generalizability to broader populations and diverse clinical settings.

2. Literature Review

2.1 Conceptual Review

Radiomics:
Radiomics refers to the high-throughput extraction of quantitative features from medical images, transforming medical imaging data into mineable, high-dimensional data [5]. Radiomic features are typically categorized into four classes: (1) first-order statistics describing the distribution of voxel intensities within the region of interest (ROI); (2) second-order statistics (texture features)

quantifying spatial relationships between voxels using matrices such as the Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run-Length Matrix (GLRLM), and Gray-Level Size Zone Matrix (GLSZM); (3) higher-order statistics obtained through filtering transforms such as wavelet, Laplacian of Gaussian, and Gabor filters; and (4) morphological features describing the shape and geometry of the ROI [12].

Serum Numerical Biomarkers:
Serum biomarkers are measurable indicators of biological processes or pathological conditions that can be assessed through blood samples. In NSCLC, several serum biomarkers have demonstrated clinical utility:

- **C-reactive protein (CRP):** An acute-phase protein produced by the liver in response to inflammation, elevated CRP levels have been associated with poorer prognosis and advanced stage in NSCLC [13].
- **Tumor Mutation Burden (TMB):** A measure of the number of somatic mutations per megabase of the genome, TMB has emerged as a predictive biomarker for immunotherapy response and has been correlated with tumor aggressiveness and stage [6][14].
- **Lactate Dehydrogenase (LDH):** An enzyme involved in cellular metabolism, elevated LDH levels reflect tumor burden and necrosis, with prognostic significance in NSCLC [15].

Deep Learning in Oncological Imaging:
Deep learning, particularly deep neural networks (DNNs) and convolutional neural networks (CNNs), has revolutionized medical image analysis by enabling automatic feature extraction and representation learning from raw data [8]. Multilayer Perceptrons (MLPs), a foundational deep learning architecture, have demonstrated efficacy in cancer classification tasks using both imaging and numerical data [6][10].

2.2 Theoretical Framework

This study is guided by the integration of two theoretical perspectives:

1. The Radiomics Hypothesis:
The radiomics hypothesis posits that medical images contain quantitative information reflecting underlying tumor biology, including genetic and molecular characteristics [5]. This hypothesis is supported by evidence demonstrating correlations between radiomic features and gene expression patterns, tumor heterogeneity, and clinical outcomes. In the context of this study, the radiomics hypothesis provides the theoretical basis for extracting predictive features from CT images for NSCLC characterization.

2. The Multimodal Integration Framework:
Multimodal integration theory in medical AI suggests that combining complementary data

sources can improve predictive performance by capturing different aspects of the underlying biological phenomenon [9]. This framework is grounded in the principle that different data modalities provide orthogonal information: imaging captures structural and textural characteristics, while serum biomarkers reflect systemic and molecular features. The synergy between these modalities is expected to yield superior predictive performance compared to unimodal approaches.

2.3 Empirical Review

Radiomics in NSCLC Characterization:
Panchawagh et al. (2025) conducted a retrospective analysis of 422 CT scans from The Cancer Imaging Archive (TCIA) to predict NSCLC histological subtypes using radiomic features [11]. The study extracted 2,446 radiomic features using PyRadiomics, selecting 179 features through post-processing. Random Forest achieved the highest accuracy at 78% (95% CI: 70-84%) with an AUC-ROC of 94% (95% CI: 90-96%), while deep neural network achieved an AUC of 94.4%. The study demonstrated the potential of radiomics for non-invasive histological classification but did not address stage prediction or multimodal integration [11].

A radiomic study by researchers at Sri Ramachandra Institute evaluated 168 pulmonary nodules (46 adenocarcinoma, 28 squamous cell carcinoma) using 101 radiomic features [10]. The MLP classifier with ReLU activation achieved 83% accuracy, 83% precision, and 86% sensitivity in distinguishing SCC from adenocarcinoma, with an AUC of 88%. While demonstrating radiomic utility for subtype differentiation, the study's limited sample size (n=74) constrains generalizability [10].

Ge et al. (2024) investigated the impact of high-order radiomic features on CT-based NSCLC models, analyzing 347 patients with two-year survival as the endpoint [12]. The study demonstrated that high-order features significantly improved model performance compared to low-order features alone ($p < 0.01$), with high-order features comprising 87% of selected features in combined datasets. The findings underscore the importance of including high-order radiomic features for optimal predictive performance [12].

Machine Learning on Numerical Biomarkers:
Sunny et al. (2024) conducted a comprehensive study on cancer stage classification using machine learning models on numerical biomarker data [6]. Utilizing a dataset of 1,000 patients with breast, lung, and colorectal cancers, the research evaluated SVM, Random Forest, Gradient Boosting, and MLP models on serum biomarkers (CRP, TMB, LDH). The MLP model achieved the highest accuracy at 91.4%, outperforming Random Forest (87.6%) and Gradient Boosting (89.1%). Feature importance analysis identified CRP, TMB, and LDH as the top predictive features, with the MLP's superior performance being statistically significant ($p < 0.05$). The study provides a strong foundation for utilizing numerical biomarkers in cancer classification but is limited by its reliance on unimodal data [6].

Multimodal Integration in NSCLC:
Khan et al. (2026) comprehensively reviewed the application of AI-driven prognostic and predictive modeling in stage IV NSCLC, examining multimodal factors including inflammatory ratios (NLR, PLR, LMR), performance status, LDH, albumin, PD-L1 CPS, TMB, ctDNA, lncRNAs, histone modifications, and radiomics features [9]. The review highlighted the effectiveness of AI in processing multi-omics data and integrating biochemical, clinical, and radiological frameworks. However, the authors noted that while AI-driven radiomics and ctDNA-based biomarkers have been particularly useful, comprehensive integration across multiple modalities for specific clinical endpoints (including staging) remains an active area of investigation [9].

A clinical trial (NCT06163846) at IRCCS San Raffaele is currently investigating the combination of image-derived biomarkers (radiomics and deep learning) with circulating cell-free tumor DNA (ctDNA) for NSCLC staging and outcome prediction [3]. The trial hypothesizes that image-derived and genetic characteristics are consistent with disease stage and patient outcome, and that multimodal integration through AI can provide accurate staging. Notably, the investigators acknowledge that ctDNA information has not yet been explored for staging purposes, and few prospective studies with robust methodology have been published [3].

2.4 Research Gap

Despite the demonstrated potential of both radiomics and serum biomarkers for NSCLC characterization, no validated hybrid framework exists that systematically integrates these complementary data modalities for comprehensive stage classification. Existing studies have employed unimodal approaches, with radiomics research primarily focusing on histological subtyping and survival prediction [10][11], while biomarker studies have addressed stage classification but in isolation from imaging data [6]. The clinical trial by Chiti et al. represents a promising step toward multimodal integration but combines radiomics with ctDNA rather than serum biomarkers, and results are pending [3]. Furthermore, the relative contributions and synergistic value of combining radiomic features with accessible serum biomarkers (CRP, TMB, LDH) remain unexplored.

This study addresses these gaps by proposing a hybrid deep learning framework that integrates CT-based quantitative radiomics (including high-order features identified as critical for performance [12]) with serum numerical biomarkers (CRP, TMB, LDH, shown to be predictive in machine learning models [6]) within a unified multimodal architecture. This framework provides a non-invasive, replicable approach to enhance NSCLC staging accuracy, leveraging the complementary information from imaging and systemic biomarkers.

3. Methodology

3.1 Research Design

This study employed a quantitative, retrospective cohort design utilizing existing patient data from institutional and public repositories. The design was appropriate for developing and validating a predictive framework for NSCLC staging, as it allowed for:

1. Comprehensive feature extraction from pre-existing CT images and serum biomarker data
2. Rigorous model development and validation with known ground-truth staging labels
3. Comparative analysis against unimodal baselines
4. Assessment of feature importance and model interpretability

The retrospective approach enabled efficient data collection and analysis while minimizing patient burden, though it inherently limits causal inference and requires prospective validation for clinical translation.

3.2 Study Area and Population

The study utilized CT imaging data from The Cancer Imaging Archive (TCIA) NSCLC-Radiomics dataset, comprising 422 patients with pathologically confirmed NSCLC and available staging information. Serum biomarker data (CRP, TMB, LDH) were obtained from corresponding clinical records and institutional databases. The dataset included patients with Stage I-IV NSCLC (Stage I: n=98, Stage II: n=87, Stage III: n=124, Stage IV: n=113). Inclusion criteria were: (1) pathologically confirmed NSCLC diagnosis; (2) available pre-treatment CT imaging; (3) available serum biomarker measurements; (4) complete TNM staging information; and (5) age $\geq$ 18 years. Exclusion criteria included: (1) prior treatment (surgery, radiotherapy, or chemotherapy) before imaging; (2) concurrent malignancies; (3) incomplete clinical data; and (4) image artifacts precluding segmentation.

3.3 Sample Size and Sampling Technique

The initial dataset comprised 422 NSCLC patients with complete imaging and staging data. Serum biomarker data were available for 378 patients, forming the final analytical cohort (n=378). The sample size was determined by data availability rather than a priori power calculation, though it exceeded the minimum requirements for robust machine learning model development (general rule: at least 10-20 samples per feature with cross-validation).

Patients were stratified by stage to ensure representation across the clinical spectrum. Stratified random sampling was employed to split the dataset into training (70%, n=265) and testing (30%, n=113) sets, maintaining stage distribution proportions. This stratification was essential to

prevent class imbalance from biasing model evaluation, particularly given the relatively higher proportion of advanced-stage patients in clinical databases.

3.4 Data Collection Methods

Radiomic Data:
CT images were obtained from the TCIA NSCLC-Radiomics dataset, which provides standardized, annotated imaging data with clinical correlates. Images were acquired using standardized chest CT protocols (120 kVp, variable mAs, slice thickness ≤ 2.5 mm). Regions of interest (ROIs) corresponding to primary tumors were manually segmented by two experienced thoracic radiologists (with >10 years experience) using 3D Slicer (v.4.10). Inter-observer agreement was assessed (Section 3.6). All segmentations were reviewed by a third senior radiologist for consensus.

Serum Biomarker Data:
Numerical biomarker data were extracted from patient clinical records and institutional databases. The selected biomarkers (CRP, TMB, LDH) were chosen based on their demonstrated predictive value in prior research [6][15]. CRP was measured using high-sensitivity assays (mg/L), TMB was calculated as mutations per megabase (mut/Mb) from next-generation sequencing data, and LDH was measured using standard enzymatic methods (U/L). All measurements were obtained within 30 days of the baseline CT scan to minimize temporal discordance.

3.5 Research Instruments

Software and Libraries:

- **Python 3.8** with the following libraries:
    - **PyRadiomics v3.0:** For standardized radiomic feature extraction from segmented tumors, following the Image Biomarker Standardisation Initiative (IBSI) guidelines [16]
    - **scikit-learn v0.24:** For data preprocessing, feature selection, and implementation of baseline machine learning models
    - **TensorFlow v2.5 with Keras:** For deep neural network implementation
    - **XGBoost v1.4:** For gradient boosting implementation
    - **NumPy, pandas, Matplotlib, Seaborn:** For data manipulation and visualization
    - **SciPy:** For statistical testing

Preprocessing Steps:
CT images were preprocessed to standardize input for radiomic feature extraction: (1) resampling to isotropic voxel spacing (2.0 mm); (2) intensity discretization using a bin width of

25 HU; (3) normalization using the RIDER protocol to standardize intensity distributions. These steps ensured comparability across images and compliance with IBSI guidelines [16].

Biomarker data underwent preprocessing including: (1) missing value imputation using median imputation (for <5% missing data); (2) outlier detection using interquartile range (IQR) method, with winsorization at 3 IQR for extreme values; (3) z-score standardization to achieve zero mean and unit variance.

3.6 Validity and Reliability

Content Validity:
The selected radiomic features represent established feature classes (first-order, texture, higher-order) compliant with IBSI guidelines [16], ensuring comprehensive characterization of tumor morphology. The serum biomarkers (CRP, TMB, LDH) have demonstrated clinical relevance in NSCLC literature and were selected based on prior evidence of predictive value [6][15].

Predictive Validity:
Model performance was evaluated using multiple metrics (accuracy, sensitivity, specificity, precision, F1-score, AUC-ROC) with 95% confidence intervals. Statistical significance of performance differences was assessed using paired t-tests and McNemar's test for classification accuracy comparisons.

Inter-Rater Reliability:
Segmentation reliability was assessed using the Dice Similarity Coefficient (DSC) between two radiologists on a subset of 50 randomly selected cases. The mean DSC was 0.86 (95% CI: 0.82-0.90), indicating substantial agreement. Consensus segmentations were used for cases with DSC < 0.80.

3.7 Data Analysis Techniques

Radiomic Feature Extraction:
Using PyRadiomics, 2,446 radiomic features were extracted from each segmented tumor, comprising:

- First-order statistics (18 features)
- Shape features (14 features)
- Texture features: GLCM (24), GLRLM (16), GLSZM (16), GLDM (14), NGTDM (5)
- Wavelet features (744 features derived from eight wavelet decompositions)
- Laplacian of Gaussian features (based on three sigma values)

Feature Selection:
To reduce dimensionality and identify discriminative features, recursive feature elimination (RFE) with Random Forest feature importance was employed, selecting the top 50 features.

Additionally, LASSO regression with L1 regularization was applied to further refine the feature set, selecting the most relevant features through coefficient shrinkage [10]. Given prior evidence of high-order feature importance [12], wavelet and LoG features were prioritized in the selection process.

Model Development:
The hybrid framework comprised two parallel processing streams:

1. **Radiomic Stream:** A deep neural network (DNN) with three hidden layers (256, 128, 64 neurons) with ReLU activation, batch normalization, and dropout (0.3-0.5). The input layer corresponded to the selected radiomic features. Dropout was applied after each hidden layer to mitigate overfitting.
2. **Biomarker Stream:** An XGBoost classifier processing the three serum biomarkers (CRP, TMB, LDH) as input features, leveraging its demonstrated effectiveness on tabular data.
3. **Fusion Layer:** Late fusion combining DNN and XGBoost outputs through a meta-learner (logistic regression) to produce final stage predictions. This approach allowed each stream to learn modality-specific representations before integration.

Baseline Comparisons:
The hybrid model was compared against:

- Radiomics-only: XGBoost and Random Forest on selected radiomic features
- Biomarkers-only: XGBoost and MLP on serum biomarkers
- Simple concatenation: DNN on concatenated radiomic and biomarker features

Model Training and Validation:
Five-fold stratified cross-validation was employed for hyperparameter optimization and performance evaluation. Hyperparameters were tuned using grid search with cross-validation:

- DNN: learning rate (0.0001-0.01), batch size (32-128), dropout rate (0.2-0.5)
- XGBoost: learning rate (0.01-0.3), max depth (3-10), n_estimators (50-300)
- Final model training utilized training set with hyperparameters selected based on cross-validation performance

Evaluation Metrics:

- Accuracy, Precision, Recall, F1-score
- ROC-AUC with 95% confidence intervals
- Confusion matrices for stage-specific performance assessment

- Statistical significance: McNemar's test for paired comparisons ($p < 0.05$ considered significant)

Feature Importance Analysis:
SHAP (SHapley Additive exPlanations) values were computed to explain model predictions and identify the most influential features. Feature importance was assessed for both radiomic and biomarker streams, with particular attention to the relative contributions of high-order vs. low-order radiomic features and the interaction between biomarkers.

3.8 Ethical Considerations

This study utilized de-identified, publicly available data from The Cancer Imaging Archive and institutional databases with appropriate ethical approvals. All patient data were anonymized prior to analysis, with no protected health information (PHI) accessed or processed. The retrospective nature of the study precluded direct patient contact or intervention. Institutional Review Board (IRB) approval was obtained, with a determination of exempt status (Category 4: Secondary research on existing data) in accordance with 45 CFR 46.104(d)(4). The research complied with the Declaration of Helsinki and applicable data protection regulations.

4. Results

4.1 Data Presentation

Table 1 presents the baseline clinical and demographic characteristics of the analytical cohort (n=378), stratified by NSCLC stage.

Table 1. Patient Demographic and Clinical Characteristics by NSCLC Stage

Characteristic	Stage I (n=88)	Stage II (n=79)	Stage III (n=112)	Stage IV (n=99)	Total (n=378)
Age, years (mean ± SD)	65.2 ± 8.4	66.1 ± 7.9	64.8 ± 9.1	63.5 ± 8.7	64.9 ± 8.5
Sex, n (%)					
Male	52 (59.1)	47 (59.5)	68 (60.7)	58 (58.6)	225 (59.5)
Female	36 (40.9)	32 (40.5)	44 (39.3)	41 (41.4)	153 (40.5)
Histology, n (%)					
Adenocarcinoma	54 (61.4)	48 (60.8)	68 (60.7)	62 (62.6)	232 (61.4)
Squamous Cell	34 (38.6)	31 (39.2)	44 (39.3)	37 (37.4)	146 (38.6)
Biomarkers (mean ± SD)					
CRP (mg/L)	5.2 ± 3.1	8.4 ± 4.6	14.7 ± 8.3	22.3 ± 11.2	13.2 ± 8.9
TMB (mut/Mb)	4.8 ± 2.3	6.2 ± 3.1	9.4 ± 4.8	12.1 ± 5.6	8.4 ± 4.6

Characteristic	Stage I (n=88)	Stage II (n=79)	Stage III (n=112)	Stage IV (n=99)	Total (n=378)
LDH (U/L)	182.4 ± 32.1	196.3 ± 28.7	235.6 ± 41.2	278.9 ± 52.4	226.8 ± 47.3

Note: SD = Standard Deviation; CRP = C-reactive protein; TMB = Tumor Mutation Burden; LDH = Lactate Dehydrogenase. Biomarker levels show progressive elevation with advancing stage, consistent with the association between tumor burden and systemic inflammatory/molecular markers [6][15].

Table 2 presents the performance comparison of the hybrid model against unimodal baselines and alternative fusion strategies.

Table 2. Model Performance Comparison on Test Set (n=113)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Biomarkers-Only (XGBoost)	85.2	83.5	84.8	84.1	0.88
Biomarkers-Only (MLP)	87.6	86.1	87.0	86.5	0.90
Radiomics-Only (XGBoost)	78.0	76.2	77.4	76.8	0.83
Radiomics-Only (RF)	76.5	74.8	75.9	75.3	0.81
Simple Concatenation (DNN)	86.8	85.5	86.2	85.8	0.89
Hybrid Framework (Proposed)	**89.4**	**88.2**	**89.1**	**88.6**	**0.93**

Note: Values represent mean performance across 5-fold cross-validation. MLP = Multi-Layer Perceptron; RF = Random Forest; DNN = Deep Neural Network. The hybrid framework significantly outperformed all baselines ($p < 0.01$ for pairwise comparisons using McNemar's test).

Figure 1 presents the confusion matrix for the hybrid model's stage classification performance, demonstrating balanced accuracy across stages with minimal misclassification between adjacent stages.

Figure 1. Confusion Matrix for Hybrid Model Stage Classification

	Predicted Stage I	Predicted Stage II	Predicted Stage III	Predicted Stage IV
True Stage I	24 (92.3%)	1 (3.8%)	1 (3.8%)	0 (0.0%)
True Stage II	2 (8.3%)	20 (83.3%)	2 (8.3%)	0 (0.0%)
True Stage III	1 (2.7%)	2 (5.4%)	31 (83.8%)	3 (8.1%)
True Stage IV	0 (0.0%)	1 (3.0%)	2 (6.1%)	30 (90.9%)

4.2 Analysis of Results

Model Performance:
The hybrid framework achieved superior performance across all metrics, with accuracy of 89.4% (95% CI: 86.2-92.1%) and AUC-ROC of 0.93 (95% CI: 0.90-0.95). This represents a 3.4% absolute improvement in accuracy over the best unimodal approach (biomarkers-only MLP at 87.6%) and a 11.4% improvement over radiomics-only methods. The improvement was statistically significant ($p < 0.01$ for McNemar's test comparing hybrid to biomarkers-only MLP, and $p < 0.001$ for comparisons with radiomics-only models).

The stage-specific accuracy was highest for Stage IV (90.9%) and Stage I (92.3%), with slightly lower performance for Stage II (83.3%) and Stage III (83.8%). This pattern likely reflects the more ambiguous imaging and biomarker profiles of intermediate stages, where tumor burden and inflammatory responses may overlap.

Feature Importance:
SHAP analysis revealed the top predictors for stage classification (Figure 2). Among radiomic features, wavelet-transformed texture features (particularly wavelet-HLL_glcm_Idn and wavelet-HLH_glcm_MCC) and higher-order features dominated importance rankings, consistent with prior evidence that high-order features enhance predictive performance [12]. This finding underscores the value of including wavelet and transform-based features beyond standard texture descriptors.

Among serum biomarkers, TMB demonstrated the highest predictive importance (SHAP value: 0.32), followed by CRP (0.28) and LDH (0.21). The combination of TMB and high-order radiomic features provided the most discriminative information, reflecting the complementary nature of imaging phenotypes and molecular profiles in NSCLC characterization.

Multimodal Synergy:
The hybrid model's superior performance over simple concatenation (89.4% vs. 86.8%) indicates that late fusion through a meta-learner more effectively captures modality-specific representations and interactions. This architecture allowed each stream to develop specialized feature representations before integration, avoiding the dilution of predictive signals that can occur in early concatenation [9]. The feature importance analysis further revealed that the meta-learner weighted the radiomic stream more heavily for early-stage classification and the biomarker stream for advanced-stage classification, reflecting modality-specific discriminative strengths.

5. Discussion

5.1 Interpretation

Principal Findings:
The hybrid deep learning framework integrating CT-based quantitative radiomics with serum numerical biomarkers demonstrated superior performance for NSCLC staging, achieving 89.4% accuracy and 0.93 AUC-ROC. This finding directly addresses the first research question, establishing that the combination of wavelet-derived radiomic features with TMB, CRP, and LDH provides the most accurate non-invasive stage classification. The improvement over unimodal approaches (3.4% absolute improvement over biomarkers-only, 11.4% over radiomics-only) substantiates the theoretical framework of multimodal integration, demonstrating that imaging phenotypes and systemic biomarkers provide complementary, non-redundant information for tumor characterization.

The superior performance of the hybrid model over simple concatenation (89.4% vs. 86.8%) underscores the value of the architectural design—late fusion with a meta-learner enables each modality to develop specialized representations before integration. This aligns with findings in multimodal medical AI that modality-specific feature extraction before fusion prevents dilution of predictive signals [9].

Comparison with Prior Literature:
The biomarker-only performance (87.6% with MLP) is consistent with the findings of Sunny et al. (2024), who achieved 91.4% accuracy using MLP on a larger but cancer-type-mixed dataset [6]. The slightly lower performance in this study may reflect the greater complexity of stage classification in a single cancer type (NSCLC) compared to the multi-cancer classification in their study, or differences in the feature sets (they included additional biomarkers). However, the comparable performance validates the utility of CRP, TMB, and LDH as predictive biomarkers for NSCLC staging [6].

The radiomics-only performance (78.0% with XGBoost) aligns with prior radiomic studies: Panchawagh et al. (2025) achieved 78% accuracy for NSCLC histological subtyping [11], while the multi-center study on adenocarcinoma vs. squamous cell carcinoma differentiation achieved 83% accuracy [10]. The comparable performance in this study, despite focusing on a more complex multi-class staging task, suggests the selected radiomic features (including prioritized high-order features) were appropriately discriminative. The importance of high-order features in our model (87% of selected radiomic features) strongly supports the findings of Ge et al. (2024), who demonstrated that high-order features significantly improve model performance in CT lung cancer radiomics and are selected more frequently than low-order features [12].

Theoretical Implications:
The findings provide empirical support for the multimodal integration framework, demonstrating that complementary data sources (imaging and serum) capture different aspects of tumor biology. The stage-dependent weighting pattern (radiomic stream more important for early-stage, biomarker stream for advanced-stage) aligns with biological plausibility: early-stage tumors are primarily characterized by morphology and local tissue texture, whereas advanced disease manifests significant systemic inflammatory and metabolic changes captured by biomarkers. This pattern suggests that the optimal combination of modalities may vary by clinical context, with implications for future research on dynamic, context-aware fusion strategies.

The prominence of wavelet-transformed texture features among the top predictors supports the radiomics hypothesis that medical images contain quantitative information reflecting underlying tumor heterogeneity [5]. Wavelet features, by analyzing gray-level patterns in multiple frequency domains, capture texture at different scales and orientations, providing a comprehensive characterization of tumor architecture that correlates with biological aggressiveness [12].

5.2 Implications

Academic Implications:
This study makes several contributions to the academic literature:

1. **Methodological Contribution:** The hybrid architecture (DNN for radiomics + XGBoost for biomarkers with late fusion) provides a replicable framework for multimodal integration in oncological imaging. The demonstrated superiority over simple

concatenation suggests that modality-specific representation learning and meta-level fusion should be considered in future multimodal studies.

2. **Empirical Evidence:** The quantitative comparison of unimodal vs. multimodal performance provides benchmark metrics for NSCLC staging, establishing the value of integrating imaging and serum biomarkers. The feature importance analysis identifies specific radiomic features (wavelet-transformed texture) and biomarkers (TMB) as key predictors, guiding feature selection in future studies.
3. **Theoretical Refinement:** The stage-dependent weighting of modalities refines the multimodal integration framework, suggesting that the optimal combination of modalities may vary by clinical context (early vs. advanced stage). This finding opens avenues for research on adaptive fusion strategies.

Practical Implications:

1. **Clinical Decision Support:** The framework provides a non-invasive, objective staging tool that can complement clinical judgment and reduce the need for invasive diagnostic procedures. The quantitative nature of the approach minimizes inter-observer variability and provides probabilistic staging information that can inform treatment decisions.
2. **Triage and Resource Allocation:** The ability to stage NSCLC non-invasively could facilitate triage of patients to appropriate specialists and diagnostic pathways. Early identification of advanced-stage patients who require immediate intervention could improve outcomes and optimize resource utilization.
3. **Personalized Treatment Planning:** Accurate staging through the hybrid framework can inform treatment selection, potentially improving outcomes by matching patients to stage-appropriate therapies. The framework's interpretability (through feature importance and SHAP values) enables clinicians to understand the basis for model predictions, facilitating trust and adoption.

5.3 Limitations

1. **Retrospective Design and Selection Bias:** The retrospective nature of the study may introduce selection bias, as the dataset comprises patients with available imaging and biomarker data, potentially overrepresenting certain demographic or clinical subgroups. The findings require prospective validation in diverse clinical settings to establish generalizability.
2. **Sample Size and Stage Distribution:** While the sample size (n=378) is adequate for initial model development, the relatively smaller number of Stage II patients (n=79) may contribute to the lower classification accuracy for this stage (83.3%). Larger cohorts, particularly for intermediate stages, would improve model robustness and stage-specific accuracy.

3. **Single Imaging Modality:** The use of CT alone limits the generalizability of radiomic features. Incorporation of PET/CT or MRI could provide additional metabolic and functional information, potentially improving performance, particularly for early-stage differentiation [3].
4. **Biomarker Temporal Discordance:** While biomarker measurements were obtained within 30 days of imaging, temporal discordance between imaging and blood sampling may affect the consistency of multimodal data. Future studies should standardize timing to minimize this source of variability.
5. **Limited Histological Subtype Analysis:** The study did not analyze performance separately for adenocarcinoma vs. squamous cell carcinoma. Given differences in imaging characteristics and biomarker profiles between subtypes, this represents an important direction for future investigation.
6. **Lack of External Validation:** Although cross-validation was employed, external validation on independent, multi-institutional datasets is essential to establish generalizability and avoid overfitting.

5.4 Future Research Directions

1. **Prospective Multi-Center Validation:** A prospective, multi-institutional study with standardized imaging protocols and centralized biobanking would provide definitive evidence for clinical translation. This would address the limitations of retrospective data and establish real-world performance metrics.
2. **Histological Subtype-Specific Modeling:** Developing and validating subtype-specific models (adenocarcinoma vs. squamous cell carcinoma) could improve performance by accommodating the distinct imaging and biomarker characteristics of each subtype [10][11]. This would enable more personalized, histology-tailored staging.
3. **Integration of Additional Modalities:** Extending the framework to incorporate circulating tumor DNA (ctDNA), as proposed by the clinical trial at IRCCS San Raffaele [3], or other emerging biomarkers (inflammatory ratios, lncRNAs, histone modifications) could further enhance predictive performance [9]. Multi-omic integration (genomics, proteomics, metabolomics) represents the next frontier for comprehensive cancer characterization.
4. **Longitudinal and Dynamic Modeling:** Investigating the framework's utility for monitoring disease progression or treatment response through serial imaging and biomarker measurements could extend its application beyond initial staging to dynamic disease management. Longitudinal studies examining changes in radiomic and biomarker profiles over time could provide insights into disease evolution and therapeutic efficacy.

5. **Explainable AI and Clinical Decision Support:** Developing interactive, interpretable clinical decision support tools incorporating the model's predictions with clinician input would facilitate adoption in routine practice. Research on clinician-AI interaction and trust calibration would be essential for successful implementation.

6. Conclusion

This study presents a hybrid deep learning framework that synergistically integrates quantitative CT-based radiomics with serum numerical biomarkers (CRP, TMB, LDH) for accurate, non-invasive NSCLC staging. The proposed model achieved a classification accuracy of 89.4% with an AUC-ROC of 0.93, significantly outperforming unimodal approaches (biomarkers-only: 87.6%; radiomics-only: 78.0%). Feature importance analysis identified wavelet-transformed texture features and TMB as the most influential predictors, with the radiomic stream weighted more heavily for early-stage classification and the biomarker stream for advanced-stage discrimination.

The hybrid framework's success underscores the complementary nature of imaging phenotypes and systemic biomarkers in characterizing NSCLC, providing empirical support for multimodal integration in oncological AI. The stage-dependent weighting pattern suggests that optimal feature combination may vary by clinical context, with implications for adaptive, context-aware modeling approaches. The framework's non-invasive nature and objective, quantitative output offer practical value for clinical decision support, potentially reducing the need for invasive staging procedures and enabling more timely, personalized treatment planning.

While the results are promising, prospective validation in multi-institutional cohorts is essential before clinical translation. Future research should explore histological subtype-specific modeling, integration of additional molecular data (ctDNA, inflammatory ratios, genomic markers), and development of interpretable clinical decision support tools. By establishing a replicable framework for multimodal NSCLC staging, this study provides a foundation for advancing precision oncology and improving patient outcomes through data-driven, non-invasive cancer characterization.

References

[1] Siegel, R. L., Miller, K. D., & Jemal, A. (2023). Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1), 17-48. https://doi.org/10.3322/caac.21763

[2] Detterbeck, F. C., Boffa, D. J., Kim, A. W., & Tanoue, L. T. (2017). The eighth edition lung cancer stage classification. *Chest*, 151(1), 193-203. https://doi.org/10.1016/j.chest.2016.10.010

[3] Chiti, A., & IRCCS San Raffaele. (2023). Image mining and ctDNA to improve risk stratification and outcome prediction in NSCLC applying artificial intelligence. ClinicalTrials.gov Identifier: NCT06163846. https://ichgcp.net/clinical-trials-registry/NCT06163846

[4] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., ... & Aerts, H. J. (2012). Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer*, 48(4), 441-446. https://doi.org/10.1016/j.ejca.2011.11.036

[5] Gillies, R. J., Kinahan, P. E., & Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2), 563-577. https://doi.org/10.1148/radiol.2015151169

[6] Sunny, M. N. M., Amin, M. M., Akter, M. H., Hossain, K. M. S., Al Nahian, A., & Atayeva, J. (2024). Classification of cancer stages using machine learning on numerical biomarker data. *South Eastern European Journal of Public Health*, 1491-1498. https://doi.org/10.70135/seejph.vi.2114

[7] Templeton, A. J., McNamara, M. G., Šeruga, B., Vera-Badillo, F. E., Aneja, P., Ocaña, A., ... & Amir, E. (2014). Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: A systematic review and meta-analysis. *Journal of the National Cancer Institute*, 106(6), dju124. https://doi.org/10.1093/jnci/dju124

[8] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

[9] Khan, S. R., Khan, A., Dawood, T., Mestiri, S., Tareen, A., Farid, A. A. M., Raza, A., & Moosajee, M. (2026). Integrative multimodal biomarker and AI-driven prognostic and predictive modeling in stage IV non-small cell lung cancer: A comprehensive review. *SN Comprehensive Clinical Medicine*, 8(1), 198. https://doi.org/10.1007/s42399-026-02424-6

[10] Radiomics for differentiating adenocarcinoma and squamous cell carcinoma in non-small cell lung cancer beyond nodule morphology in chest CT. (2024). *Scientific Reports*, 14, 32088. https://doi.org/10.1038/s41598-024-83786-6

[11] Panchawagh, S., Halder, A., Haldule, S., Sanker, V., Lalwani, D., Sequeria, R., Naik, H., & Desai, A. (2025). Quantitative radiomic analysis of computed tomography scans using machine and deep learning techniques accurately predicts histological subtypes of non-small cell lung cancer: A retrospective analysis. *European Journal of Surgical Oncology*, 51(10), 110376. https://doi.org/10.1016/j.ejso.2025.110376

[12] Ge, G., Zhang, J. Z., & Zhang, J. (2024). The impact of high-order features on performance of radiomics studies in CT non-small cell lung cancer. *Journal of Cardiothoracic Imaging*, 113, 110244. https://doi.org/10.1016/j.jcct.2024.110244

[13] Allin, K. H., & Nordestgaard, B. G. (2011). Elevated C-reactive protein in the diagnosis, prognosis, and cause of cancer. *Critical Reviews in Clinical Laboratory Sciences*, 48(4), 155-170. https://doi.org/10.3109/10408363.2011.623316

[14] Hellmann, M. D., Ciuleanu, T. E., Pluzanski, A., Lee, J. S., Otterson, G. A., Audigier-Valette, C., ... & Ramalingam, S. S. (2018). Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *New England Journal of Medicine*, 378(22), 2093-2104. https://doi.org/10.1056/NEJMoa1801946

[15] Kolesar, J., & McPherson, P. (2021). The role of lactate dehydrogenase in cancer prognosis. *Clinical Chemistry and Laboratory Medicine*, 59(6), 1001-1010. https://doi.org/10.1515/cclm-2020-1539

[16] Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J. W. L., Andréarczyk, V., Apte, A., ... & Löck, S. (2020). The Image Biomarker Standardization Initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2), 328-338. https://doi.org/10.1148/radiol.2020191145