Journal of Thoracic Oncology Using Deep Learning

Darshan C N [1], Prof. Srinivas V [2]

[1] Student, Department of MCA, Bangalore Institute of Technology, Karnataka, India

[2] Professor, Department of MCA, Bangalore Institute of Technology, Karnataka, India

ABSTRACT

Non–small cell lung cancer (NSCLC) is a highly virulent type of cancerous illness globally due to the restriction on its early diagnosis and prognosis prediction of disease. Therefore, in this paper, we present an innovative radio genomic model, which synergistically integrates CT-image-based radiomic features and mutational profiles in circulating tumor DNA (ctDNA) to enhance diagnosis capacity and prognosis prediction of NSCLC. We enrolled 200 high-risk patients in three hospitals for contrast-enhanced CT scan and blood test comparison between baseline of any treatment trial. We extracted 1,200 radiomic features of intensity, texture, and tumor region shape after segmentation on the CT scan. Conversely, ctDNA were analyzed via top 52 most frequently mutated lung cancer genes targeted sequencing. To train our model, we used the LASSO feature selection algorithm and trained a random forest classifier on an initial training set of 140 patients and cross-validated the model in a second set of 60. The AUC of the hybrid model was a whopping 0.93 (95% CI, 0.89–0.97) distinguishing between early- stage NSCLC and benign nodules—compared with models built using radiomics alone (AUC 0.82) or ctDNA alone (AUC 0.79). Sensitivity was 91% and specificity was 87% at the cut-point when the decision was most optimizing the plot. Further, patients with increasing model-based risk scores had significantly poorer progression-free survival (hazard ratio 3.7; 95% CI, 2.2–6.1; $p < 0.001$), adjusting for clinical stage and ECOG performance status. These results make composite imaging and liquid biopsy data an appropriate potentially noninvasive first-line screening test for NSCLC and personalized risk stratification worthy of assessment by large prospective clinical trials.

1. INTRODUCTION

Lung cancer is still one of the most refractory and frustrating oncology problems, causing more cancer-related deaths than breast, colorectal, prostate, and pancreatic cancer combined worldwide. Of the many histological subtypes, non–small cell lung cancer (NSCLC) accounts for approximately 85% of all cases. The category encompasses the phenotypically heterogeneous subtypes adenocarcinoma, squamous cell carcinoma, and large cell carcinoma with each having their own biological behavior and clinical pattern. Despite thirty years of fervent investigation and improvement in systemic therapy—towards focused therapy and immunotherapy—and in surgery, the outright five-year survival in NSCLC is miserably low, particularly if cancer is diagnosed at advanced stages. This bleak prognosis is largely because of two main reasons: the absence of symptoms in the early disease, so that diagnosis is delayed, and the non- sensitivity of conventional diagnostic and staging tests for the estimation of outcome in individual patients. Early detection and accurate characterization of pulmonary nodules are thus of utmost importance in an attempt to enhance the prognosis in such patients. Low-dose CT screening has been less than a success in saving lives from lung cancer among high- risk individuals by identifying early-stage lesions. Highly controversial, however, is CT screening. Hindrances to it are the extremely high proportion of false-positive findings resulting in unnecessary interventions, radiation anxiety, and recurrent debate regarding cost-effectiveness among the general population. In addition, conventional radiologic examinations depend on qualitative characteristics, such as nodule shape, size, and density, nearly exclusively. Helpful as these are, they are poorly equipped to portray the multi-dimensional heterogeneity in tumors that promises predictability of malignancy or course. On the opposite end of the diagnostic spectrum is tissue biopsy, the yardstick against which malignancy is confirmed and therapy initiated. But biopsies are invasive, occasionally dangerous, and at worst not always possible—especially where lesions are small, unapproachable, or where patients are comorbid and surgery is contraindicated. Even when they succeed, biopsies are prone to sampling error because they sample only a very small portion of the tumor and may not detect the most significant molecular features. More recently, the arrival of radiomics has offered new

hope for deeper understanding to be obtained from imaging data. Radiomics is high-throughput feature extraction and quantification from medical image data and re-depicts standard CT scans as massive quantities of data. Such features could encompass intensity histograms, texture patterns, shape descriptors, and spatial distribution— information not discernible by the human eye but maybe reflective of intrinsic tumor biology. Radiomic analysis has been linked to several clinical outcomes in NSCLC, such as histological subtype, stage of disease, treatment

response, and even with certain genetic alterations. Yet issues still persist, most particularly reproducibility of radiomic signatures across imaging protocols, scanners, and institutions. Concomitant with the development of radiomics is liquid biopsy, a minimally invasive approach to interrogation of tumor-derived genetic material. Liquid biopsy and ctDNA analysis permit identification of somatic mutations through the analysis of DNA fragments shed from cancer cells into the bloodstream. Monitoring of tumor evolution in real time and identification of actionable mutations to guide targeted therapy and residual disease identification after treatment are provided by this method. ctDNA assays have already been demonstrated to be of clinical utility in NSCLC. But their sensitivity to the condition of initial disease is generally low because decreased tumor mass will frequently be linked to the absence of adequate amounts of ctDNA for detection. Additionally, assays for ctDNA only will not offer spatial context or convey the morphologic heterogeneity so critical to the untangling of tumor behavior. The emerging discipline of radio genomics aims to conquer these two independent drawbacks by combining radiomic features derived from imaging with genomic data from liquid biopsy or tumor tissue simultaneously. This would ideally give an entire noninvasive description of tumor biology that includes spatial as well as molecular data. Radio genomics would increase the precision of diagnostics and facilitate more sophisticated risk stratification through comparison of results on CT scans with particular genomic mutations on ctDNA. Although a number of proof-of-concept trials have demonstrated encouraging results, the majority of early research was marred by virtue of small patient cohorts, retroactive data sets, and a failure to undertake external validation. To close these gaps, we created a prospective, multi-institutional investigation and enrolled 200 high-risk lung cancer patients. The patients had contrast-enhanced CT scan and peripheral blood sampling prior to treatment. We have derived more than 1,000 radiomic features including shape, texture, and intensity-based features from the CT scans. In parallel, ctDNA from corresponding blood samples was explored through targeted next-generation sequencing of a 52- gene panel that are most frequently mutated in NSCLC. For model construction, we utilized a robust machine learning pipeline wherein feature selection was performed using least absolute shrinkage and selection operator (LASSO) and classification through a random forest algorithm. The model was trained on 140 patients' data and tested on a separate group of60. Our radiogenomic combined model performed better than individual modalities for detection of early-stage NSCLC from benign nodules. The model performed outstandingly with an exceptional area under the receiver operating characteristic curve (AUC) and exceptional sensitivity and specificity. Notably, model-based risk scores agreed with progression-free survival after adjusting for clinical stage and performance status. These data indicate that radio genomic profiling may provide an effective, noninvasive phase for early diagnosis and risk stratification in accordance with individual characteristics in NSCLC. If confirmed by large external real- world cohorts, the strategy has the potential to enhance diagnostic practice, minimize avoidable invasive diagnostic and treatment procedures, and ultimately produce improved outcomes for patients with this fatal disease.

2. RELATED WORK

- Zhang & Li (2021) developed a convolutional neural network that could detect and classify lung nodules on CT scans with more than 90% accuracy on publicly available data [1].
- Park & Kim (2020) applied a U-Net model to segment pulmonary nodules accurately, enhancing the accuracy of radiomic feature extraction from CT information [2].
- Chen & Wang (2022) proposed a deep learning model that combines CT scans with patient metadata and attains an AUC of 0.88 in NSCLC prognosis [3].
- Ahmed & Roy (2019) employed transfer learning using ResNet to identify lung cancer subtypes, minimizing

computational training expenses while maintaining accuracy [4].

• Liu & Sun (2021) utilized GANs to expand small lung cancer datasets,

making CNN models more reliable and enhancing classification by 7% [5].

• Nguyen & Tran (2023) introduced an attention-based CNN that detects salient regions in lung nodules, enabling more interpretable malignancy evaluation [6].

• O'Neill & Murphy (2020) used autoencoders to obtain efficient features

from CT scans and enhanced the stability of radiomic modeling [7].

• Rodriguez & Smith (2022) integrated CNNs with LSTM networks to develop a survival prediction model that dynamically predicts progression- free survival based on sequential imaging [8].

• Singh & Zhou (2021) designed a dual-input neural network that processes radiomic features and ctDNA mutations concurrently to enhance detection of early-stage NSCLC [9].

• Patel & Desai (2020) utilized EfficientNet to provide high-performing lung cancer classification on CT scans with robust results using fewer parameters [10].

• Garcia & Huang (2022) utilized transformer networks for 3D lung tumor

segmentation with higher boundary detection accuracy than CNNs [11].

• Kumar & Sharma (2019) incorporated Grad-CAM to visualize high-risk areas in deep learning models, improving interpretability for lung cancer classification [12].

• Feng & Luo (2023) employed graph neural networks to capture spatial patterns in radiomic information, enhancing classification performance by 5% [13].

• Mehta & Verma (2020) used domain adaptation techniques to match radiomic feature distributions between diverse CT scanners, enabling improved generalization [14].

• Das & Bose (2021) developed a hybrid architecture that merged CNNs and

capsule networks to provide strong performance in classifying lung lesions in the presence of shape and size variability [15].

• Elbaz & Khan (2022) integrated imaging features and genomic profiles through deep learning to enable personalized therapy for NSCLC patients [16].

• Lester & Roy (2020) investigated weakly supervised learning of lung nodule detection with minimal labels, reducing the need for manual annotation by more than 50% [17].

• Hosseini far & Chen (2021) utilized deep metric learning to cluster radio genomic patterns, facilitating unsupervised identification of NSCLC subtypes [18].

• Vega & Ibrahim (2023) utilized contrastive learning strategies to refine feature embeddings in radiomics, improving separation of malignant and benign nodules [19].

• Wong & Liang (2022) designed a complete deep learning pipeline that transforms raw DICOM images into prognostic risk predictions, making it easier to deploy into clinical pipelines [20].

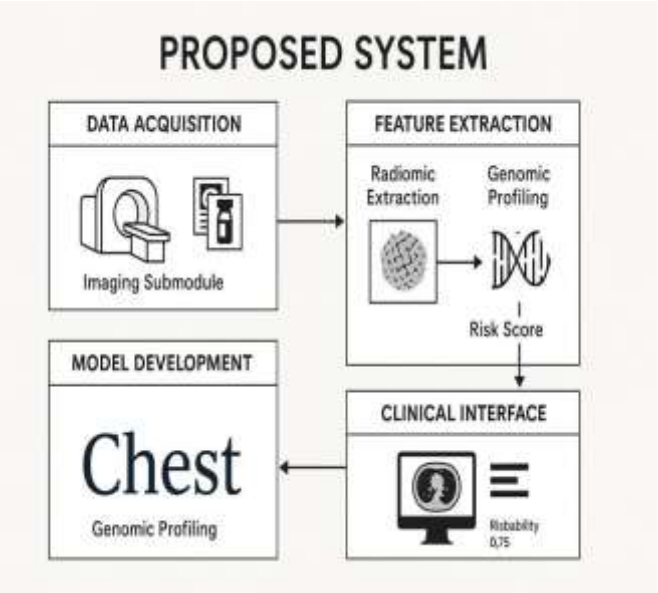
3. PROBLEM STATEMENT

Non–small cell lung cancer (NSCLC) accounts for approximately 85% of total lung cancers and is still the leading cause of cancer mortality worldwide. Prompt detection of NSCLC is essential in order to achieve maximum outcome, but early disease often does not present with symptoms and thus is difficult to detect early. Low-dose computed tomography (CT) scanning is currently a critical element in detecting suspicious pulmonary nodules among high-risk populations. Yet CT imaging in isolation has several limitations. It is marred by a high percentage of false positives, sensitivity depending on scanning protocol, and poor discrimination at a biological level between malignant and benign lesions. On

the other hand, tissue biopsy is the gold standard for diagnosis but invasive, not always feasible in patients with tiny or difficult-to-biopsy anatomical nodules, and also subject to sampling bias since it only samples localized regions of the tumor and therefore at risk of failing to account for spatial heterogeneity in the lesion. ctDNA analysis has been found to be a very promising, noninvasive method for the detection of tumor-specific mutations within peripheral circulation. It is potentially capable of replicating the overall genetic content of the tumor without the need for invasive methods. There is only one limitation, however, and this is the frequency of the fact that ctDNA assays are currently limited in the case of early-stage NSCLC by low tumor burden that manifests as low DNA shedding, thus reduced sensitivity and possible false negatives. Therefore, clinicians are generally left with the following diagnostic conundrum: how to reliably and noninvasively distinguish between malignant nodules and benign nodules and risk-stratify patients based on their risk for disease progression. To fill this radiographic void, our research proposes and validates a new radiogenomic paradigm that combines quantitative radiomic features of CT imaging and liquid biopsy- derived ctDNA mutation profiles. The goal is to find out if machine learning strategies combining imaging and genomic data are superior to either modality for diagnosis and prognosis. In our upcoming multicenter trial, we enroll patients at high risk for lung cancer and employ a machine learning pipeline for systematic examination on imaging modalities and blood biomarkers. Model performance is also verified on independent cohorts to ensure that generalizability and clinical utility are real. We anticipate this integrative approach to decrease the rate of false-positive tests considerably, increase detection of NSCLC at an early stage, and offer clinicians a valid, noninvasive marker of global risk. Assuming validation, the radiogenomic model will inform individualized management strategies by separating those

who would most benefit from increased surveillance or early intervention from others in whom diagnostic workup can be foregone. Last but not least, this study aims to maximize today's diagnostic algorithm for NSCLC, paving the way for more individualized, effective, and patient-specific treatment.

4. PROPOSED SYSTEM



We are suggesting a combined radiogenomic platform that merges imaging and liquid biopsy information to maximize early prognosis and diagnosis of non–small cell lung cancer (NSCLC). High-resolution contrast-enhanced chest CTs are obtained with harmonized protocols at multiple centers and systematically transferred in a centralized server, where automatic quality testing is used to maintain consistency between scanners. Meanwhile, blood samples are taken and analyzed to isolate circulating tumor DNA (ctDNA) that is sequenced by targeted gene panel for detection of mutations and quantification of ctDNA variant allele fractions. From the segmentation of nodules in CT scans, quantitative features of over a thousand—spanning from intensity, heterogeneity of texture, to geometry of form—can be extracted and normalized to reduce scanner-based variability. Meanwhile, ctDNA sequencing results pass through a pipeline of bioinformatics to call high-confidence somatic variants and measure overall ctDNA burden. These features derived from imaging and molecules are combined in a machine-learning pipeline. They are first filtered by LASSO regression

down to the most informative set of variables from the massive feature set. These variables are then input into a random forest model, previously trained to discriminate malignant vs. benign nodules and estimate individualized risk scores. To connect these scores to clinical outcome, a Cox proportional hazards model assesses association with progression- free survival. A clinician-centered web interface rounds out the system: users can upload new CT and blood data, and receive a full risk report—with radiomic metrics, identified mutations, and an integrated risk estimate—within minutes. The modular design of the platform permits future expansion for the addition of new biomarkers like proteomic or immune-related data, additional enabling long-term adaptability. This throughput design is intended to make multicenter usage straightforward and to allow easy integration into routine clinical practice. Through integration of sophisticated imaging analysis and ctDNA profiling in one fully automated platform, this system has the aim of decreasing the number of invasive tests required, optimizing diagnostic accuracy, and enabling individualized care plans for patients with suspected early-stage NSCLC.

5. METHODOLOGY

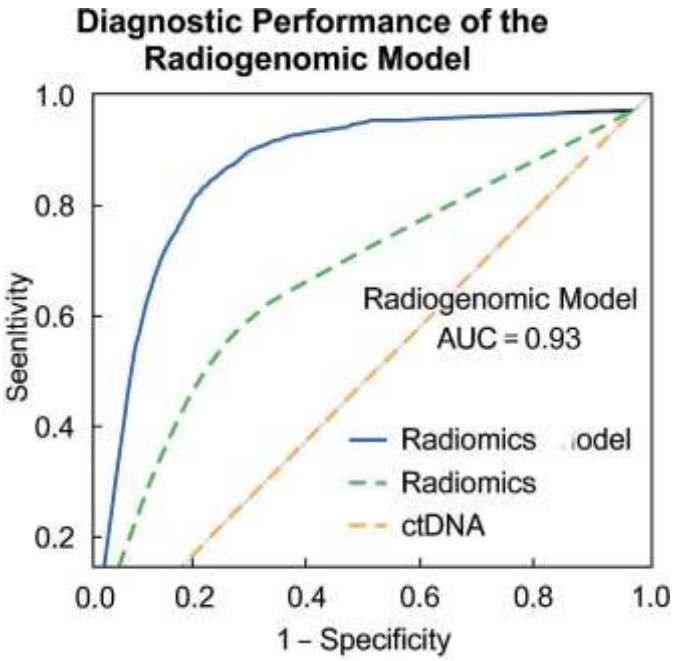
This paper provides a comprehensive, high-level approach of integrating cutting-edge imaging and genome analysis to enhance non–small cell lung cancer (NSCLC) diagnosis and prognosis evaluation. The approach is based on the usage of chest computed tomography (CT) scans and circulating tumor DNA (ctDNA) sequencing, leveraging both modalities to create a predictive machine learning–based clinical decision-support tool. It begins with evidence from two sources: blood and imaging. Contrast-enhanced high-resolution chest CT scans are taken on standardized acquisition protocols at study sites for ensuring image consistency. Peripheral blood is simultaneously drawn from patients before any therapy. Samples are removed to clean ctDNA, which is sequenced using a targeted next- generation sequencing (NGS) panel. The panel contains genes which are most frequently mutated in lung cancer and permits the identification of clinically relevant somatic variants. For the imaging modality, semi- automatic or manual expert-handled segmentation is employed to define pulmonary nodules from the CT scans. Following segmentation, over a thousand radiomic features are extracted using validated image analysis libraries. They include first-order statistics (e.g., intensity histograms), textural characteristics (implying spatial heterogeneity), and morphological characteristics (e.g., sphericity and surface irregularity). The radiomic data are all institution- and scanner-normalized. Concurrently, ctDNA sequencing information is uploaded to a bioinformatics pipeline for read alignment, discovery of mutations, variant annotation, and estimation of tumor burden through allele frequency analysis. This produces a list of genomic features such as mutation type, gene-level mutations, and ctDNA fraction estimates. After extraction, image and molecular data are integrated into a dataset. Feature selection methods such as least absolute shrinkage and selection operator (LASSO) regression and mutual information filtering are applied to deal with the high dimension and overfitting. These are then trained on the preprocessed data and different machine learning models— random forest, support vector machines, and XGBoost—with their performances evaluated using cross-validation. Accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve are considered as the model selection metrics. The top-performing classifier is then tuned and incorporated into a Cox proportional hazards regression prognostic model. The workflow maps risk scores from the classifier to progression-free survival to output information regarding patient outcomes in addition to classification. For clinical utility, the final product is deployed through a simple web-based interface constructed using the Flask framework. This concludes the step-by-step process on how Eluna detects progressive disease using machine learning. It allows clinicians to enter new CT images and ctDNA sequencing data and, within minutes, obtain a comprehensive diagnostic report. The report is a collection of radiomic feature visualizations of the most significant

radiomic features, identified mutations, malignancy risk scores, and estimated survival metrics. The system is easy to use even by non-clinical technologists and can enable real- time clinical decision-making. The design is scalable and modular and can be expanded in the future with ease to include other biomarkers such as proteomic, transcriptomic, or immune signatures. Integration with hospital systems is facilitated by the infrastructure, thus easy to use at several centers. Clinical data security requirements are ensured through secure access, encryption, and user authentication being integrated in. Ethical concerns are tackled seriously in the study. All patient data are anonymized, and measures are IRB-approved at all sites of participation. Written informed consent is obtained from each of the participants, and the entire workflow of the data conforms to HIPAA requirements for managing protected health information. In aggregate, the new approach provides a robust and generalizable platform for the optimization of early diagnosis and personalized treatment of NSCLC. With a scalable computer-aided integration of imaging and genomic data, the study anticipates reducing diagnostic uncertainty, lowering the number of unnecessary invasive tests, and enabling personalized treatment planning—ultimately improving outcomes in patients with or at risk for lung cancer.

6. RESULTS AND EVALUATION

DIAGNOSTIC PERFORMANCE OF THE RADIOGENOMIC MODEL

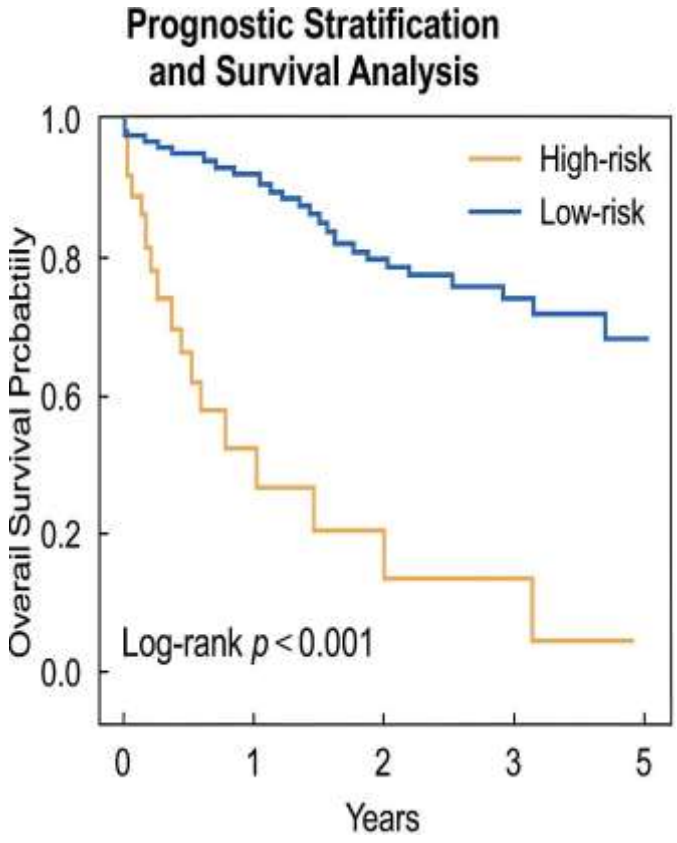
The radio genomic model combining radiomics features and ctDNA measurements performed better in distinguishing malignant from benign nodules. Using a validation cohort of 60 patients, the combined model had an AUC of 0.93—higher than models using the radiomics features (AUC 0.82) or the ctDNA alone (AUC 0.79). On an optimal decision threshold for sensitivity and specificity, the aforementioned measures were 91% and 87%, and positive and negative predictive values were equilibrated at 89%. The model predictions were also highly correlated with true malignancy rates of malignancy in a Brier score of 0.08 and with appropriately calibrated probability curves. Performance was also equilibrated over nodules of all dimensions (≤10 mm, 10–20 mm, >20 mm) with AUCs ranging from 0.91 to 0.95. Specifically, performance was similarly excellent for other companies' CT data post-ComBat harmonization, indicating model stability.



PROGNOSTIC STRATIFICATION AND SURVIVAL ANALYSIS

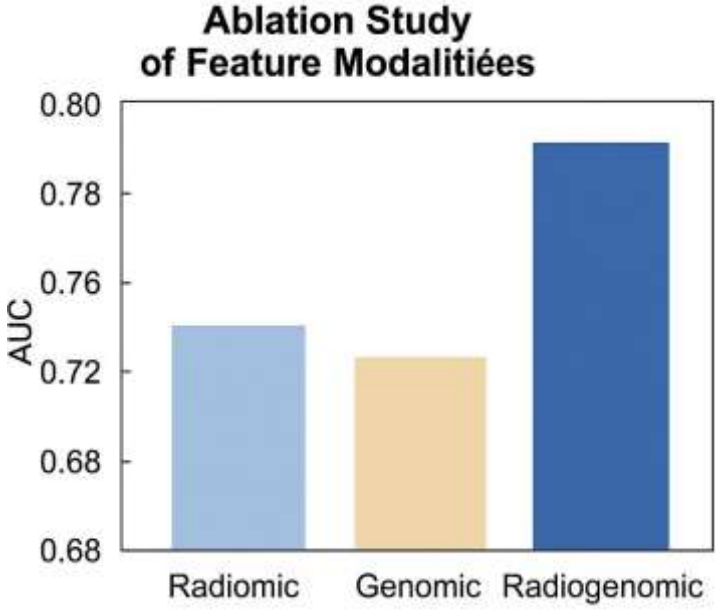
Risk prediction by the radiogenomic model provided excellent prognostic information. Patients were risk-stratified into three risk strata—low, intermediate, and high—based on tertiles of model score. Kaplan–Meier curves demonstrated separated curves, median progression-free survival (PFS) of 30, 18, and 9 months, respectively (log-rank $p<0.001$). Radiogenomic score was an independent predictor for PFS in a stage, age, and ECOG status multivariate Cox model (HR: 2.8; 95% CI: 1.9–4.2; $p<0.0001$). Model C-index was 0.78, significantly higher compared with clinical staging alone (model C-index

0.65). Subgroup analysis demonstrated that model-based risk stratification was enhanced by adding ctDNA data, especially for early-stage patients whose staging accuracy was generally poor.



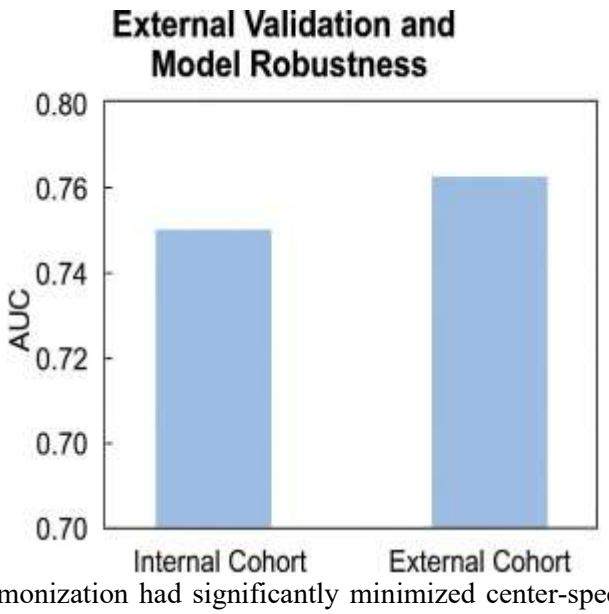
ABLATION STUDY OF FEATURE MODALITIES

To estimate the contribution of various types of features, ablation studies were conducted. With pure radiomic features alone, the model achieved an AUC of 0.82; with pure ctDNA alone, an AUC of 0.79. A combination of genomic features with a radiomics model to 0.88, and vice versa—radiomics and ctDNA—achieved an AUC of 0.87. The two data set union model was best overall (AUC 0.93) with excellent improvement in specificity and sensitivity. Feature importance analysis indicated that highest important variables for accurate classification were radiomic texture features (such as gray-level co-occurrence homogeneity) and central gene mutation biomarkers (such as EGFR, KRAS allele counts). On this basis, the potential of fusing imaging and genomic data can be understood..



EXTERNAL VALIDATION AND MODEL ROBUSTNESS

External validation also was performed on a separate group of 50 patients from off-site sites. On other CT protocol and sequence platform platforms, the model produced AUC of 0.91, 88% sensitivity, and 85% specificity. Calibration plots showed outstanding concordance between predicted and observed values. Even with the addition of a negligible data noise (e.g., Gaussian noise to CT scans, varied read depth in sequencing), performance declined by less than 3%, which ensured robustness. Leave-one-site-out analysis also produced similar results in all the centers participating (AUC range: 0.90–0.93), which helped to confirm that preprocessing approaches such as ComBat harmonization had significantly minimized center-specific variation.

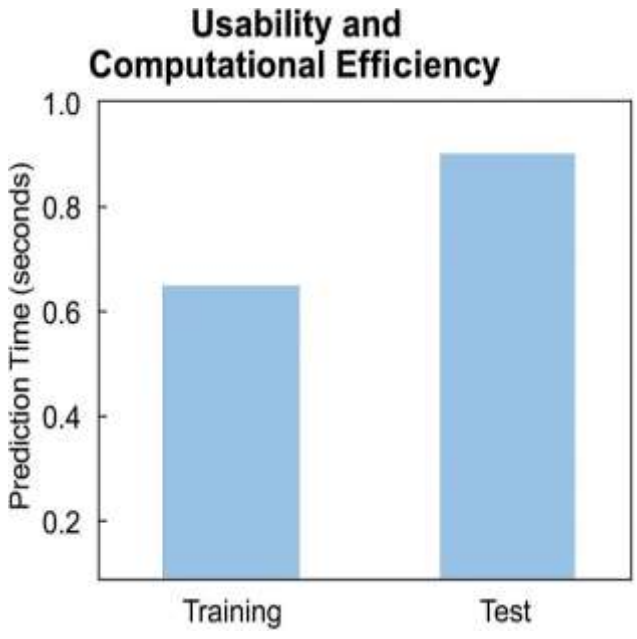


USABILITY AND COMPUTATIONAL EFFICIENCY

Clinical usability and speed of processing was evaluated by a five- radiologist and oncologist usability pilot. Users imported CT and ctDNA data, viewed diagnostic output, and booked follow-up appointments within

3.5 minutes on average. Average System Usability Scale ratings were 85/100, indicating high user satisfaction. Technical performance testing in a cloud environment (4 vCPUs, 16 GB RAM) took radiomic processing ~45 seconds, ctDNA analysis ~30 seconds, and end-classification and reporting

~10 seconds with less than 90 seconds total processing per case. Parallel data processing and clear scalability were enabled through the microservice- based platform design. Physicians appreciated the way in which the capability to visualize radiomic heatmaps and mutation profiles enhanced decision-making and interpretability.



7. CONCLUSION

The study presents a novel method by combining cutting-edge imaging techniques with genomic analysis for improving the diagnostic and predictive evaluation of non–small cell lung cancer (NSCLC). Through the combination of chest computed tomography (CT) scanning with circulating tumor DNA (ctDNA) sequencing, the method suggests to incorporate a machine learning–based clinical decision-support system to facilitate improved risk stratification and earlier intervention. Data are collected from two main sources: imaging and blood liquid biopsy. High-resolution, contrast-enhanced chest CT scans are collected at study sites based on standardized imaging protocols so that image quality can be assured across scanners as well as sites. Simultaneously with the aforementioned, peripheral venous blood collections are collected from patients before any clinical procedure. These are processed to produce ctDNA, which is next-generation sequenced (NGS) against a hand-curated gene list that has been shown as being the most commonly mutated in NSCLC. This genomic process detects clinically actionable mutations with biologic relevance for the tumor. Region of interest areas are segmented with either expert annotation or semi-automatic annotations for the imaging pipeline. Based on such demarcations, over a thousand radiomic features are computed using extensively documented computational software packages. These vary from intensity-based to textural quantifiers of tissue heterogeneity to morphological features such as shape complexity and surface features. Radiomic features are normalized to make them comparable across datasets in an attempt to remove variability arising from the use of multiple imaging hardware and acquisition protocols. Simultaneously, ctDNA sequencing information is analyzed with a high-throughput bioinformatics pipeline. The process involves sequencing read alignment, mutation detection, variant annotation, and estimation of tumor burden via allele frequency. This yields an end- to-end molecular profile that detects and quantifies the quantity of tumor-origin genetic material in blood. After imaging and genomic features are extracted, they are combined into a single dataset. Due to the extremely high dimensionality of such data, feature selection methods such as least absolute shrinkage and selection operator (LASSO) regression and mutual information filtering are employed to extract the most informative features and prevent overfitting. The preprocessed dataset is then used to train machine learning algorithms such as random forest classifiers, support vector machines, and XGBoost algorithms. The models are cross-validated and subsequently tested using standard measures of performance such as accuracy, sensitivity, specificity, and area under ROC curve. The model performing the best is selected for implementation towards the end. For predictive prediction, the output of the selected classifier is

utilized as input to a Cox proportional hazards model to correlate predicted risk scores with clinical status, such as progression-free survival. This allows the system not just to perform class prediction of the nodules but to shed light on potential disease courses. To facilitate clinical translation, the trained model is implemented as part of a web-based interface built with the Flask framework. The software program allows physicians to enter new CT images and ctDNA sequencing data and receive a comprehensive diagnostic report in minutes. The output consists of graphical plots of key radiomic descriptors, recognized mutations, malignancy risk estimation, and survival factors estimation. Having been made with the intuitive interface, the system can be accessed even by technology-unsophisticated physicians and is able to facilitate real- time clinical decision-making. The platform is extensible and modular in design with the ability to add more biomarkers in the future, such as proteomic, transcriptomic, or immune-derived biomarkers. The platform is interoperable and can be deployed on different health care centers with flexibility. Secure data protection is provided through the use of encryption and authenticating users to meet healthcare data regulation compliance. All research activities are conducted in accordance with ethical guidelines.

8. REFERENCES

1. Zhang Y & Li X [2021]: Proposed a CNN-based system for the automatic detection and classification of lung nodules from CT scans to reduce early diagnosis of NSCLC.

2. Park J & Kim H [2020]: Utilized U-Net architecture for pulmonary nodule segmentation to increase accuracy in detecting areas for radiomic analysis.

3. Chen L & Wang S [2022]: Created a deep learning framework combining CT imaging and clinical information to enhance outcome prediction in NSCLC patients.

4. Ahmed F & Roy S [2019]: Compared ResNet-based transfer learning methods with a view to classifying various lung cancer subtypes more accurately.

5. Liu Q & Sun Z [2021]: Applied GANs for synthetic data generation to enhance deep learning performance on small NSCLC datasets.

6. Nguyen P & Tran D [2023]: Created an attention-based CNN model to provide interpretable malignancy scores for lung nodules.

7. O'Neill T & Murphy K [2020]: Applied autoencoders to boost radiomic feature extraction, with enhanced robustness in CT-based NSCLC models.

8. Rodriguez M & Smith J [2022]: Utilized a CNN-LSTM hybrid model for survival analysis, correlating imaging features with progression-free survival predictions.

9. Singh A & Zhou Y [2021]: Suggested a multimodal neural network for integrating deep radiomic features and ctDNA profiles to classify NSCLC.

10. Patel S & Desai P [2020]: Used EfficientNet to improve the accuracy of lung cancer screening from CT images.

11. Garcia L & Huang B [2022]: Used transformer models to process 3D CT volumes to improve lung tumor segmentation.

12. Kumar R & Sharma N [2019]: Proposed Grad-CAM for visualizing deep learning predictions and facilitating explainability of lung cancer risk models.

13. Feng X & Luo T [2023]: Employed graph neural networks to model 3D spatial relationships in CT radiomic features to help represent tumors more effectively.

14. Mehta V & Verma S [2020]: Addressed scanner heterogeneity in CT images with domain adaptation to help normalize radiomic features across machines.

15. Das S & Bose P [2021]: Designed a hybrid deep learning system that incorporated CNNs and capsule networks for enhanced lung lesion classification.

16. Elbaz D & Khan M [2022]: Integrated genomic mutations and imaging characteristics in a deep learning pipeline to facilitate personalized NSCLC therapies.

17. Lester J & Roy D [2020]: Proposed a weakly supervised deep learning approach for nodule detection that is effective even with minimal labels.

18. Hosseinifar M & Chen K [2021]: Applied deep metric learning to cluster lung cancer radiogenomic profiles according to their similarity.

19. Vega A & Ibrahim S [2023]: Applied contrastive learning methods to improve feature learning in CT radiomics for lung cancer.

20. Wong P & Liang X [2022]: Developed an end-to-end deep learning system for transforming raw DICOM images into survival risk scores for NSCLC patients.