Multimodal Data Fusion for Lung Cancer Diagnosis: A Review Study

Gunjan Mittal Roy[1], Ajay Kumar[2]
[1]Lincoln University College Malaysia, [2]IILM University, Greater Noida, India
[1]Email ID: gunjanmittal2013@gmail.com; [2]Email ID: ajay.phdcse@gmail.com

Abstract: Lung Cancer is the world's leading cause of cancer related deaths so the early-stage diagnosis is crucial to improve the survival rate. Multimodality fusion of diverse data modalities such as organized clinical records, molecular profiles (genomics, proteomics), and radiological imaging (CT/PET), is applied for lung cancer diagnosis. It may enhance prognostication, therapeutic decision-making, and diagnostic accuracy. Recent developments (2020–2025) in multimodal data fusion techniques for lung cancer diagnosis are summarized in this review, along with important multimodal public datasets, representative models and architectures, a thorough comparison table of significant studies, research gaps, and technical difficulties that should direct future investigations, are taken care of. This study suggests specific research avenues with a focus on clinical translation, interpretability, and robustness.

Keywords: Lung cancer; Multimodal Data Fusion; Deep Learning; Artificial Intelligence in Healthcare

Introduction

One of the main causes of cancer death globally is lung cancer. To direct treatment and enhance results, it is crucial to promptly identify and characterize the tumor subtype (such as adenocarcinoma or squamous cell carcinoma) and actionable genetic changes (such as EGFR mutations). Imaging techniques like computed tomography (CT), positron emission tomography (PET), and molecular tests like proteomics and genomes have historically been used separately. But while genomes and proteomics identify genetic causes and potential treatment targets, imaging records macroscopic, phenotypic features like tumor form and heterogeneity. For prognostic modeling, clinical metadata—such as demographics, comorbidities, and treatment records—provides crucial information. Compared to single-modality techniques, multimodal fusion—the combination of imaging, molecular, and clinical data—promises to generate diagnostic and prognostic models that are more precise, reliable, and clinically meaningful. With an emphasis on datasets, fusion strategies, representative studies, evaluation practices, obstacles, and future research pathways, this review explores the state-of-the-art in multimodal approaches for lung cancer diagnosis which focuses on datasets, fusion strategies, representative studies, evaluation practices, challenges, and avenues for future research.

Literature Survey

In the recent years, scientists and researchers have worked hard to use a variety of prediction tools to diagnose lung cancer. Bakr et al. [1] gave the crucial multimodal collection connecting CT to TCGA molecular profiles in the field of radiogenomic biomarker prediction and its subtype classification. This allowed for later investigations showing a linkage between CT symptoms and genomic alterations (e.g., EGFR). This dataset remains a benchmark for methods predicting molecular status from imaging or fused

representations. Chen et al. [2] and related radiogenomics studies applied radiomics-derived features combined with gene expression profiles to predict outcomes and identify imaging biomarkers correlated with molecular pathways illustrating the potential of statistical fusion pipelines when coupled with robust feature selection and validation. In the Attention & Cross-Attention Fusion Architectures category, a number of research works are proposed. Verma et al. [3] and Deng et al. [4] demonstrated cross-attention modules that align image-derived representations with gene-expression vectors. Their intermediate fusion architectures showed improved survival prediction and subtype classification compared to unimodal baselines, highlighting attention's ability to focus on relevant image regions and gene sets. In the domain of Self-Supervised Pretraining for Imaging Encoders Wolf et al. [5] reviewed on self-supervised learning (SSL) and showed that pretraining imaging encoders on large unlabeled CT collections (e.g., LIDC/LUNA) using contrastive learning or masked autoencoders significantly boosts downstream performance on smaller radiogenomic cohorts. SSL mitigates label scarcity, a frequent bottleneck in multimodal studies. For Proteogenomics & Multiomic Integration category of research for predicting lung cancer Gillette et al. [6], Soltis et al. [7] illustrated the value of proteogenomic integration; proteomics provides complementary information to genomics and can be fused with imaging metadata for better biomarker discovery and prognostication. Multiomic studies emphasize the need for sophisticated normalization and batch-effect correction. In the area of Clinical Translation and Prognostication, several studies such as Vanguri et al. [8], Bourbonne et al. [9] used multimodal fusion (CT + pathology + genomics or PET/CT + genomics) to predict treatment response and mutation status, reporting clinically meaningful AUC/C-index improvements when compared to unimodal baselines, but many of these works remain single-center or lack broad external validation.

Table 1. Compares this work with the related work or previous research by other researchers

Sl. No.	Study (short)	Modalities	Dataset(s)	Fusion Method	Task / Metric	Key Limitation
1	Bakr et al. / NSCLC Radiogenomics (2018).[1]	CT + Genomics + Clinical	NSCLC Radiogenomics (TCIA↔TCGA)	Feature-level fusion / radiogenomics	Biomarker association, classification	Moderate cohort size; missing modalities
2	Gillette / CPTAC-LUAD (2020). [6]	Imaging + Proteomics + Genomics	CPTAC-LUAD	Statistical multiomic integration	Biomarker discovery	Small proteomics cohort; cost
3	Verma et al., 2024.[3]	CT + Gene expression + Clinical	TCIA/TCGA cohorts	Cross-attention intermediate fusion	Survival prediction (C-index)	Limited paired samples; interpretability
4	Wolf et al., 2023 (SSL CT). [5]	CT (pretraining)	LIDC/LUNA	SSL pretrain → fine-tune	Improved downstream accuracy	Pretraining compute & domain transfer
5	Bourbonne et al., 2024 (PET/CT). [9]	PET + CT + Genomics	Institutional / TCIA	Radiomic + ML fusion	Mutation prediction (AUC)	Cost & multi-modality availability
6	Soltis et al.,	Genomics +	CPTAC cohorts	Multiomic	Prognostic	Sample size;

	2022 (proteogenomics).[7]	Proteomics + Clinical		pipelines	biomarker discovery	heterogeneity
7	Vanguri et al., 2022. [8]	CT + Pathology + Genomics	Institutional cohort	Multi-stage intermediate fusion	Immunotherapy response (AUC)	Single-center; generalizability
8	Zheng et al., 2024 (graph-attention). [10]	Pathology + Omics	TCGA / institutional	GNN + attention fusion	Survival prediction	Need co-registration & complex pipeline
9	Jiang et al., 2024 (systematic review). [11]	Review	Multiple studies	N/A	Meta-analysis findings	Notes heterogeneity & small cohorts
10	Pei et al., 2023 (multimodal DL review).[12]	Review	—	N/A	Fusion taxonomy & best practice	Calls for benchmarks
11	He et al., 2024 (radiogenomics review). [13]	Review	—	N/A	Radiogenomics methods overview	Calls standardization
12	Simon et al., 2024/25 (narrative). [14]	Review	—	N/A	Future of multimodal models	Emphasizes benchmarks

Materials & Methods

To execute the lung cancer prediction using multi modal fusion approach, a number of phases are required to be accomplished. The very first stage is to acquire datasets and other resources. Therefore, a major enabler for multimodal research is the availability of datasets that link imaging, molecular, and clinical data. Some of the most widely used datasets are worth notify. The NSCLC Radiogenomics (TCIA↔TCGA) dataset is a widely used multimodal resource that maps CT scans to genomic profiles and clinical metadata for non-small cell lung cancer (NSCLC) patients. It remains the primary benchmark for many radiogenomic fusion studies. In a similar way, another dataset which is quite popularly used is LIDC-IDRI & LUNA16 dataset. This is a large public CT dataset focused on lung nodules (detection/segmentation). While imaging-only, these datasets are invaluable for pretraining imaging encoders or for imaging tasks in fusion pipelines. Another popular dataset; TCGA (TCGA-LUAD/ TCGA-LUSC), are a rich genomic, transcriptomic, and clinical datasets for lung adenocarcinoma and squamous carcinoma. When aligned with TCIA imaging, these data enable deeper multimodal studies. Next in this category are CPTAC-LUAD datasets. These are proteogenomic datasets that complement genomic data with high-quality proteomics and associated imaging metadata. Useful for multiomic fusion research. Apart from these datasets some more collection of datasets includes NLST, QIN Lung CT, and institutional cohorts sometimes provide imaging plus clinical endpoints useful for prognostic modeling and survival analysis. Each dataset has strengths (public availability, multimodal linkage) and limitations (moderate sample sizes, missing modalities, heterogeneity), which inform experimental design choices in the literature.

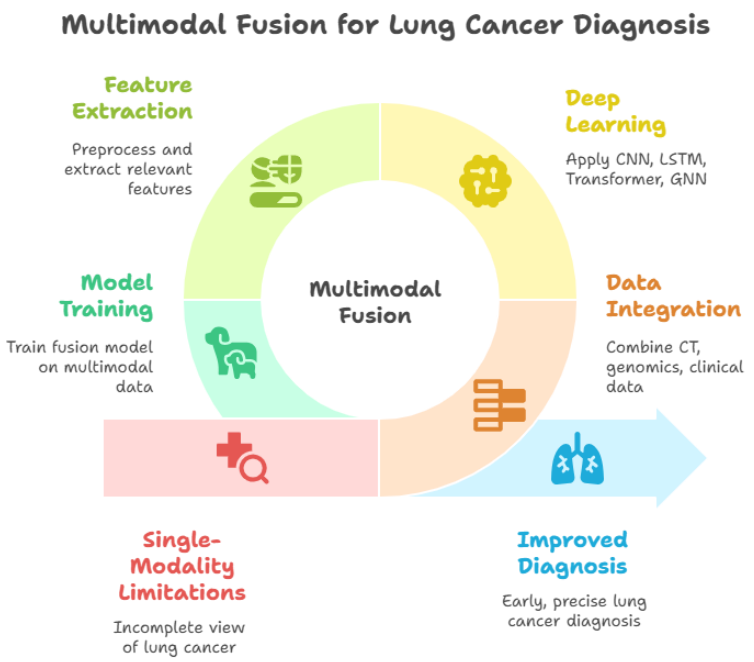

Fig. 1 Multimodal fusion for lung cancer diagnosis

The next phase is the fusion for which various strategies and model architectures are used. In general, there are three classes of multimodal fusion methods namely early (feature-level) fusion, late (decision-level) fusion and intermediate/hybrid fusion. Feature-level or early fusion concatenates features from multiple modalities before feeding them into a joint model. This is straightforward when modalities have compatible representations (e.g., tabular clinical features and vectorized genomic signatures), but may be challenged by scale differences and modality-specific noise. Late fusion combines predictions or scores from modality-specific models. Advantages include modularity and interpretability; however, it may fail to capture synergistic cross-modal interactions. Lastly, intermediate fusion (sometimes called joint or representation-level fusion) learns modality-specific encoders and merges their latent representations within a joint network (e.g., via concatenation, attention, or gating). This approach offers a balance, it models cross-modal interactions while allowing modality-specific pretraining. Recent work emphasizes attention-based cross-modal layers, cross-attention transformers, and graph-based fusion for richer inter-modal alignment. Apart from fusion models there are various architectural components used for fusion process. In this review work, three of them are considered. The first one is the imaging encoders, such as CNNs (2D/3D), U-Net/V-Net variants (for segmentation), and Vision Transformers (ViT) or masked autoencoders for representation learning. Next type are Genomic/proteomic encoders that includes MLPs, autoencoders, and attention-based modules for gene expression or multiomic vectors. Graph neural networks (GNNs) also model gene-gene interactions. Lastly, Cross-modal interaction component is cross-attention, gated multimodal units, and fusion layers that adaptively weight modalities.

Figure 1 depicts the multimodal fusion pathway for lung cancer diagnosis, that combines many data sources such as genetics, CT imaging, and clinical information to enable more precise and early detection. In order to capture intricate correlations, deep learning algorithms such as CNNs, LSTMs, Transformers, or GNNs are used after feature extraction, which involves extracting pertinent characteristics from each

modality. The system may then learn thorough diagnostic patterns thanks to the integration of these features using fusion techniques, which produce a single representation for model training. Multimodal fusion delivers a comprehensive perspective, which improves the accuracy and dependability of lung cancer diagnosis and prognosis in contrast to single-modality techniques that only offer a partial picture of the illness.

Research gaps and Challenges in Multimodal Lung Cancer Fusion

There are multiple research gaps in the multimodal lung cancer fusion including scarcity of large, truly paired multimodal cohorts. Most public multimodal collections (e.g., NSCLC Radiogenomics) contain only hundreds of patients with fully paired CT + genomics + clinical records, limiting statistical power for deep models and robust external validation. This scarcity forces heavy reliance on small cohorts or on synthetic/partial matching [15]. So, by following various steps like taking pooled multi-center initiatives, by standard data-sharing agreements, or by using federated benchmarks sample sizes can be scaled down. Another challenge is the heterogeneity & batch effects across sites, scanners, and omics platforms. Site/scanner differences and omics batch effects produce domain shifts that degrade cross-site performance. Techniques adapted from genomics (ComBat) and image harmonization reduce site effects but are still an active area of work for CT/radiomics and multiomic fusion [16]. This challenge can be mitigated by adopting and benchmarking the harmonization pipelines (ComBat variants, deep harmonization) and reporting model sensitivity to harmonization choices. In the different datasets a number of modalities remain missing and therefore the records are incomplete. Real clinical datasets frequently lack one or more modalities (e.g., imaging without RNAseq). Few studies systematically evaluate robustness to missing modalities or compare imputation, modality-dropout, and gated fusion strategies under realistic missingness patterns [17]. By benchmarking missing-modality scenarios (MCAR/MAR/MNAR) across fusion paradigms and publish standardized protocols, this research gap can be taken care of. Another research gap is the lack of standardized benchmarks, protocols & metrics for fusion evaluation. In general, there is no universally adopted benchmark that tests early/intermediate/late fusion, missing-modality scenarios, harmonization stress tests, and external validation for NSCLC. Reviews call for standardized datasets and reproducible evaluation pipelines [18]. So, the solution to this challenge could be to create a reproducible benchmark (train/val/test splits, withheld external centers) and require reporting of calibration (Brier), discrimination (AUC/C-index), and robustness metrics.

Result and Conclusion

Multimodal data fusion that is integrating CT imaging, genomics/proteomics, and clinical data, offers significant potential to improve lung cancer diagnosis and prognosis. Public multimodal resources (NSCLC Radiogenomics, CPTAC) have catalyzed research, while methodological advances (self-supervised pretraining, cross-attention fusion, graph-based architectures) offer powerful modeling tools. Critical challenges remain in data harmonization, limited paired sample sizes, missing modalities, interpretability, and external validation. Addressing these through standardized benchmarks, modality-aware architectures, explainability, and privacy-preserving collaborative learning will be crucial to translate multimodal fusion models into clinical practice.

Future Scope

Although significant progress has been made in multimodal fusion for lung cancer diagnosis, several open challenges continue to inspire future research. In order to close the gap between experimental research and practical application, the next stage of development should concentrate on creating interpretable fusion models, standardized multimodal datasets, and clinically verified frameworks. Thus, a variety of methods could be applied in the same way. For example, provide a common baseline for NSCLC Radiogenomics, CPTAC-LUAD, and LIDC/LUNA to systematically assess early, intermediate, and late fusion, taking into account heterogeneity stress testing and missing-modality scenarios. To close the clinical-ML gap, another strategy might be to create models that provide gene-level attributions and lesion-level saliency maps that are verified by pathology and molecular tests. Implementing federated training pipelines across several centers is another step in this direction. These pipelines allow training on imaging + omics without sharing raw data, protecting privacy while expanding cohort sizes. Additionally, to better align representations from the image and molecular domains, simultaneously pretrain cross-modal encoders and extend SSL paradigms to omics (masked feature reconstruction). In order to evaluate the usefulness of multimodal prediction models in therapy selection and result enhancement, incorporate them as decision-support instruments in prospective clinical trials.

References

1. Bakr, S., Gevaert, O., Echegaray, S., et al. "A radiogenomic dataset of non-small cell lung cancer" *Scientific data,* 5(1), 1-9, 2018.
2. Chen W, Xiong M, Zhou R, et al. Integrating Radiomics with Genomics for Non-Small Cell Lung Cancer Survival Analysis. Journal of Oncology. 2022 https://doi:10.1155/2022/5131170
3. Verma, S., Magazzu, G., Eftekhari, N., et al. "Cross-attention enables deep learning on limited omics-imaging-clinical data of 130 lung cancer patients", *Cell reports methods*, *4*(7), 2024.
4. Deng, R., Shaikh, N., Shannon, G., & Nie, Y. "Cross-modality attention-based multimodal fusion for non-small cell lung cancer (NSCLC) patient survival prediction", In *Medical Imaging, Digital and Computational Pathology* (Vol. 12933, pp. 46-50). SPIE, 2024
5. Wolf D., Payer T., Lisson C.S., et al. "Self-supervised Pre-Training with Contrastive and Masked Autoencoder Methods for Convolutional Models on CT Images", 2023
6. Gillette, M. A., Satpathy, S., Cao, S., Dhanasekaran, S. M., et al. "Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma", *Cell*, *182*(1), 200-225, 2020
7. Soltis, A. R., Bateman, N. W., Liu, J., et al. "Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways" *Cell Reports Medicine*, *3*(11), 2022
8. Vanguri, R. S., Luo, J., Aukerman, A. Tet al. "Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L) 1 blockade in patients with non-small cell lung cancer", *Nature cancer*, *3*(10), 1151-1164, 2022
9. Bourbonne, V., Morjani, M., Pradier, O., et al., "PET/CT-Based Radiogenomics Supports KEAP1/NFE2L2 Pathway Targeting for Non–Small Cell Lung Cancer Treated with Curative Radiotherapy", *Journal of Nuclear Medicine*, *65*(4), 630-634, 2024

10. Y. Zheng, R. D. Conrad, E. J. Green, E. J. Burks, M. Betke, J. E. Beane, and V. B. Kolachalama, "Graph attention-based fusion of pathology images and gene expression for prediction of cancer survival," *IEEE Transactions on Medical Imaging*, vol. 43, no. 9, pp. 3085–3097, Sep. 2024 https://doi: 10.1109/TMI.2024.3386108.
11. Luo, X., Zheng, R., Zhang, J., et al., "CT-based radiomics for predicting Ki-67 expression in lung cancer: a systematic review and meta-analysis", *Frontiers in Oncology*, *14*, 1329801, 2024
12. Pei, X., Zuo, K., Li, Y., & Pang, Z. "A Review of the Application of Multi-modal Deep Learning in Medicine: Bibliometrics and Future Directions*"* *International Journal of Computational Intelligence Systems*, 16, Article 44, 2024 https://doi.org/10.1007/s44196-023-00225-6
13. He, W., Huang, W., Zhang, L., et al. "Radiogenomics: bridging the gap between imaging and genomics for precision oncology", *MedComm*, *5*(9), e722, 2024
14. Simon, B. D., Ozyoruk, K. B., Gelikman, D. G., et al. "The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review", *Diagnostic and Interventional Radiology*, *31*(4), 303, 2025
15. Bakr, S., Gevaert, O., Echegaray, S., et al. "A radiogenomic dataset of non-small cell lung cancer", *Scientific data*, *5*(1), 1-9, 2018
16. Fortin, J. P., Cullen, N., Sheline, Y. I., et al. "Harmonization of cortical thickness measurements across scanners and sites" *Neuroimage*, *167*, 104-120, 2018
17. Vanguri, R. S., Luo, J., Aukerman, A. T.et al. "Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L) 1 blockade in patients with non-small cell lung cancer", *Nature cancer*, *3*(10), 1151-1164, 2022
18. Pei, X., Zuo, K., Li, Y. et al. "A review of the application of multi-modal deep learning in medicine: bibliometrics and future directions", *International Journal of Computational Intelligence Systems*, *16*(1), 44, 2023.