Dual-Attention Multimodal Network for Explainable Lung Nodule Malignancy Assessment Through Integration of CT Imaging and Molecular Phase-Separation Biomarkers

Kenneth A. Richards
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
richards1976@binghamton.edu

Trancis Breene
Department of Computer Science, University of Houston, Houston, TX, USA.
trancisbreene@uh.edu

Abstract

Lung cancer remains the leading cause of cancer mortality worldwide, with low-dose computed tomography screening generating a large number of indeterminate pulmonary nodules that demand precise risk stratification. While deep learning approaches applied to CT images have substantially improved nodule classification, they often lack the molecular contextualization necessary for personalized decision-making and fail to provide actionable explanations for their predictions. Concurrently, molecular biology has uncovered the relevance of biomolecular condensates formed through liquid-liquid phase separation, such as the YAP-MAML2 transcriptional co-activator complexes, as drivers of aggressive tumor phenotypes. This paper proposes a dual-attention multimodal network architecture that integrates volumetric CT imaging with molecular phase-separation biomarkers to deliver explainable malignancy assessment. The system employs a dual-attention mechanism that combines channel-spatial self-attention within the imaging pathway and cross-modal attention between image embeddings and condensate state features, generating spatial attention maps that highlight nodule regions alongside molecular feature importance scores. We discuss system-level structural trade-offs including data harmonization from heterogeneous clinical sources, federated learning strategies to preserve patient privacy, computational sustainability, robustness to domain shift, and fairness across demographic subgroups. The deployment framework is situated within emerging governance standards and regulatory pathways for artificial intelligence as a medical device, emphasizing explainability as a cornerstone for clinical trust. By bridging radiological phenotyping and molecular biophysics, the proposed architecture addresses critical gaps in multimodal fusion for lung cancer diagnostics and charts a course toward integrated precision oncology platforms.

Keywords

lung nodule malignancy, dual-attention network, multimodal fusion, computed tomography, phase separation, explainable AI, clinical decision support, federated learning.

1. Introduction

Lung cancer accounts for the highest number of cancer-related deaths globally, a burden that persists despite advances in treatment largely because a majority of cases are diagnosed at advanced stages [1]. Low-dose computed tomography (CT) screening has been demonstrated

to reduce lung cancer mortality through earlier detection, yet its widespread implementation introduces a substantial interpretive challenge: the discovery of numerous indeterminate pulmonary nodules, the vast majority of which are benign. Accurate determination of nodule malignancy risk is therefore critical to avoid unnecessary invasive procedures while ensuring timely intervention for aggressive lesions. In clinical practice, management decisions are guided by demographic factors, nodule morphological characteristics, and interval growth, but these conventional criteria achieve only modest specificity. The incorporation of computational imaging biomarkers through deep convolutional neural networks has yielded significant performance gains in nodule classification tasks [2]. Nevertheless, such models frequently operate as opaque black boxes, and they remain disconnected from the molecular underpinnings that ultimately dictate tumor behavior.

A transformative insight from molecular cell biology is that many key oncogenic processes are regulated by phase separation, a biophysical phenomenon in which proteins and nucleic acids demix from the surrounding milieu to form membraneless condensates. The phase separation behavior of the YAP-MAML2 fusion complex, for instance, has been shown to differentially regulate the transcriptional programs associated with proliferation and invasion [7]. As a result, the condensate state of specific transcriptional regulators is emerging as a novel dimension of molecular biomarkers that can capture aggressive tumor biology beyond static genomic mutations. Integrating such dynamic molecular features with imaging phenotypes promises a richer representation of disease state, yet this integration introduces formidable challenges related to data heterogeneity, missing modalities, interpretability, and scalability.

This paper presents a dual-attention multimodal network architecture designed to fuse CT-derived volumetric features with molecular phase-separation biomarkers for explainable lung nodule malignancy assessment. We focus on the system-level considerations that govern the design, deployment, and governance of such an integrated diagnostic platform. The network utilizes dual-attention mechanisms to align imaging and molecular representations, producing visual explanations that localize suspicious nodule regions while simultaneously revealing which phase-separation features most influence the prediction. Our discussion extends to the infrastructural requirements for real-world translation, including privacy-preserving federated learning, computational resource sustainability, robustness against population shifts, fairness across demographic groups, and alignment with evolving regulatory frameworks for artificial intelligence in medicine.

2. Background and Related Work

Automatic lung nodule analysis has been one of the most active areas in medical image computing. Early efforts relied on engineered radiomic features, but the advent of deep learning enabled end-to-end malignancy prediction from CT volumes. Notably, multi-crop convolutional neural networks aggregated information from multiple scales to capture both local texture and global context, setting strong benchmarks for nodule classification [2]. Segmentation models such as U-Net provided precise delineation of nodule boundaries, facilitating downstream feature extraction [3]. The utility of attention mechanisms in computer vision, originally popularized by the Transformer architecture for sequence modeling, was rapidly adopted in image analysis through channel-wise and spatial attention modules that allow networks to focus on salient regions while suppressing irrelevant background [4,8]. Dual attention networks for natural scene segmentation demonstrated that

combining channel and position attention modules yields enhanced feature representations, a principle that proved directly transferable to medical imaging tasks.

Parallel to these developments, the field of molecular oncology began to appreciate the role of liquid-liquid phase separation in forming biomolecular condensates that organize biochemical reactions and signaling pathways [5,6]. A seminal study identified that the phase separation of the YAP-MAML2 fusion protein differentially regulates its transcriptional output, linking condensate physical properties to oncogenic programs [7]. This discovery opened the possibility of using condensate state biomarkers—such as the propensity of specific complexes to form liquid versus gel-like condensates—as molecular surrogates of tumor aggressiveness. Unlike conventional genomic alterations, these biomarkers reflect the functional state of protein networks and may be measurable through advanced immunohistochemistry or proteomic assays of biopsy specimens.

Multimodal fusion in medical diagnostics has been explored extensively, with frameworks ranging from simple feature concatenation to sophisticated co-attention transformers that learn cross-modal interactions [9,10]. For lung cancer specifically, models have been developed that pair CT imaging with genomic sequencing or histopathology, revealing synergistic information that outperforms single-modality predictors [10]. However, existing multimodal systems rarely incorporate dynamic biophysical biomarkers such as phase-separation status, and they are often built under assumptions of complete, centrally curated datasets that do not reflect the fragmented reality of clinical data ecosystems.

Explainable artificial intelligence has become a pressing requirement in the medical domain, where trust and regulatory approval depend on transparency. Methods like Gradient-weighted Class Activation Mapping and its variants produce localization heatmaps that can highlight relevant regions in medical images [11]. Surveys on medical explainability have emphasized the need for multimodal explanations that not only point to image regions but also quantify contributions of non-imaging features [12]. In lung cancer screening, large-scale studies have demonstrated that deep learning can achieve performance comparable to or exceeding that of radiologists, yet the translation to practice demands a careful integration of decision support with human oversight [13]. Our work builds on these foundations by designing a framework that couples dual-attention mechanisms operating both within and across modalities to generate rich, trustworthy explanations while accommodating the unique nature of phase-separation biomarkers.

3. Dual-Attention Multimodal Network Architecture

The proposed architecture, referred to as DAMNet, is structured around three core components: a CT imaging pathway that extracts hierarchical volumetric features using a dual-attention backbone, a molecular pathway that encodes phase-separation biomarker profiles, and a cross-modal fusion module that produces joint malignancy predictions and explanations. The design prioritizes interpretability, modularity, and robustness to missing data modalities.

The imaging pathway employs a backbone inspired by a dual-attention segmentation network that combines path aggregation with channel and spatial attention refinements, such as the PDU-Net architecture, to isolate nodule volumes with high precision [14]. By embedding dual-attention blocks within the segmentation encoder, the network learns to emphasize diagnostically informative nodule textures and margins while suppressing vascular and airway structures that may confound classification. The output of the encoder is a set of feature maps

that encode the nodule's morphological attributes across multiple scales. This approach contrasts with simpler classification CNNs in that it explicitly models the boundary and internal heterogeneity of lesions, a critical factor in distinguishing invasive cancers from benign opacities.

The molecular pathway receives a vector of phase-separation state indicators. These indicators may include the liquid-liquid phase separation propensity of YAP-MAML2, the condensate morphology scores, or more broadly the expression levels of co-condensing proteins that modulate oncogenic signaling. The molecular encoder is a multi-layer perceptron that projects these biomarkers into a latent space aligned with the imaging features. A key architectural decision is whether to rely on absolute quantitative measurements or on relative rankings that are less sensitive to inter-laboratory variability. The system is designed to accept both types, incorporating batch normalization and domain adaptation layers that can be fine-tuned across assay platforms.

The fusion module employs a cross-modal dual-attention mechanism. First, a self-attention pooling layer aggregates the spatial imaging features by weighting each voxel's contribution according to both channel importance and spatial relevance, yielding a compact image representation. Similarly, a self-attention layer over the molecular features allows the model to suppress redundant or noisy biomarkers. Second, a cross-attention block computes bidirectional relevance scores: CT features attend to molecular cues to determine which imaging patterns are most consistent with a given condensate profile, and molecular embeddings attend to imaging features to contextualize the significance of phase-separation states. The resulting fused representation is passed through a classification head that outputs a malignancy probability. This architecture not only achieves high discriminative performance but also naturally produces interpretable attention maps as byproducts: the spatial attention weights over the imaging encoder depict nodule regions, while the cross-attention scores between modalities highlight the most influential molecular features.

The system is trained end-to-end with a combination of classification loss and an auxiliary attention regularization term that encourages spatially coherent and compact attention distributions. By avoiding the use of hard attention masks that could discard potentially relevant context, the model preserves the ability to uncover subtle, diffuse signals such as ground-glass opacity margins that correlate with specific condensate states. The design also permits the inference of malignancy even when molecular biomarkers are unavailable, by bypassing the cross-attention path and using only the imaging self-attention branch, which ensures clinical utility in resource-limited settings where advanced molecular assays may not be performed.

4. Data Integration and Pipeline Governance

Successful deployment of a multimodal system such as DAMNet hinges on robust data integration strategies that reconcile the disparate natures of radiological images and molecular biomarker measurements. CT imaging data are typically stored in DICOM format within picture archiving and communication systems, characterized by large file sizes and variability in acquisition protocols, reconstruction kernels, and radiation dose levels. Phase-separation biomarkers, conversely, derive from biochemical or proteomic assays that yield continuous or ordinal values, often recorded in laboratory information management systems with distinct identifiers and quality control metadata. Establishing a unified data pipeline that links these sources at the patient level requires rigorous patient de-identification, temporal alignment between imaging and biopsy timepoints, and standardization of feature representations.

Federated learning has emerged as a compelling paradigm to train collaborative models across multiple institutions without centralizing sensitive patient data [17]. In the context of DAMNet, a federated training scheme would allow each participating site to compute imaging and molecular sub-network updates on local data, while only sharing encrypted gradients or model weights with a central aggregation server. This approach inherently reduces the risk of privacy breaches and addresses the regulatory constraints imposed by data protection laws such as the General Data Protection Regulation and the Health Insurance Portability and Accountability Act. However, federated learning introduces its own system-level challenges, including heterogeneous data distributions across sites, which can degrade model convergence and fairness if not properly managed. Strategies such as personalized layers that remain local to each institution and federated domain adaptation can mitigate these effects, but they demand careful orchestration of the training protocol and additional validation procedures.

Data governance must also encompass the traceability and provenance of molecular biomarker measurements. Phase-separation states are sensitive to experimental conditions, including temperature, buffer composition, and the time elapsed since biopsy collection. Consequently, a standardized operating procedure for condensate assays, ideally codified through inter-laboratory ring trials, is a prerequisite for generating reproducible inputs to the network. The model's latent molecular encoder can incorporate domain-specific normalization layers that adjust for known batch effects, but their design must be informed by close collaboration with molecular pathology laboratories. This socio-technical coupling between assay standardization and algorithmic performance illustrates the broader principle that the boundary of the AI system extends far beyond the model weights and into the laboratory and clinical workflows that generate its input data.

5. Deployment, Robustness, and Sustainability

Translating DAMNet from a research prototype to a clinically embedded decision support tool requires addressing deployment infrastructure, robustness to real-world variability, and the sustainability of computational resources. The system must operate within hospital information technology ecosystems, interfacing with electronic health records and radiology reporting systems through standard protocols such as FHIR and DICOMweb. Containerization using Docker and orchestration through Kubernetes can ensure consistent runtime environments, enabling seamless updates and rollbacks. Additionally, the deployment architecture should support both on-premises and cloud-based processing, allowing institutions to choose configurations aligned with their security policies and latency requirements.

Robustness considerations extend to distributional shift: CT images acquired on scanners from different manufacturers or reconstructed with varying kernels can cause a marked drop in segmentation and classification accuracy. The dual-attention backbone that incorporates path aggregation and spatial attention has been shown to improve segmentation generalizability across diverse CT datasets, yet it is not immune to performance degradation [14]. To bolster robustness, the imaging pathway can be coupled with test-time augmentation and uncertainty quantification mechanisms that alert clinicians when the model's prediction confidence is low due to out-of-distribution inputs. On the molecular side, the system must accommodate the reality that phase-separation biomarker assays may occasionally fail or return inconsistent results; the fallback unimodal inference pathway is thus an essential robustness feature. Adversarial robustness is another critical dimension, as medical AI

systems can be vulnerable to imperceptible perturbations that alter diagnostic decisions [15]. While defending against targeted attacks on CT volumes is an ongoing research area, the multimodal nature of DAMNet may confer a degree of inherent resilience, as an attacker would need to simultaneously manipulate both imaging and molecular data channels in a consistent manner, a considerably more demanding proposition.

Sustainability and computational cost have recently become central to the discourse on responsible AI [19]. Training large three-dimensional dual-attention networks on volumetric CT data requires substantial GPU resources and energy consumption. To mitigate this, DAMNet can be initialized with pre-trained weights from related segmentation tasks and fine-tuned with a modest amount of labeled multimodal data, reducing the total training carbon footprint. In deployment, model distillation and quantization can shrink the network size without clinically meaningful loss in accuracy, enabling inference on edge devices within radiology workstations. Lifecycle management of the AI system should incorporate ongoing monitoring of energy utilization as part of the hospital's sustainability reporting, aligning with broader efforts to decarbonize healthcare delivery.

6. Fairness, Ethics, and Policy Implications

The integration of multimodal AI into lung cancer diagnostics carries profound implications for equity and ethics. Disparities in healthcare access, imaging quality, and the availability of molecular testing can translate into differential model performance across populations. Studies have demonstrated that algorithms may inadvertently learn spurious associations linked to racial or socioeconomic attributes, leading to biased risk estimates that could exacerbate existing inequities [16]. For DAMNet, this risk is particularly acute because phase-separation biomarker assays are likely to be initially available only in well-resourced academic medical centers, while low-dose CT screening is increasingly deployed in community settings. If the model is trained predominantly on data from tertiary care centers, its reliance on molecular features could disadvantage populations whose nodule assessments are based solely on imaging, producing systematically different performance in underserved areas.

Addressing fairness demands a multi-pronged strategy. First, training datasets must be curated to include representative sampling across demographic and geographic strata, with explicit subgroup analyses reported during model validation. Second, the unimodal inference pathway that uses only imaging should be calibrated to meet the same performance standards as the fusion pathway across all groups, ensuring that the absence of molecular data does not lead to inferior care. Third, fairness constraints can be incorporated into the training objective, encouraging the model to maintain comparable predictive accuracy and explanation consistency across predefined sensitive attributes. However, defining fairness in medicine is not a purely technical exercise; it requires engagement with patients, clinicians, and bioethicists to determine acceptable trade-offs between accuracy and equity.

The governance of AI-based diagnostic devices is rapidly evolving, with regulatory bodies such as the U.S. Food and Drug Administration and the European Medicines Agency developing frameworks for software as a medical device. Explainability is increasingly recognized as a crucial component of safety and effectiveness, because clinicians need to understand the basis of a recommendation in order to override it when appropriate [18]. DAMNet's dual-attention explanations—spatial heatmaps and molecular importance scores—directly serve this need, enabling a radiologist to verify that the model's attention aligns with the suspicious nodule margins and that the molecular features driving the prediction are

plausible given the patient's histology. However, the generation of such explanations must be accompanied by clear, validated information on their limitations, and the user interface must be designed in collaboration with clinical end-users to avoid automation bias. Over-reliance on plausible-looking attention maps could paradoxically undermine safety if clinicians accept erroneous explanations without critical scrutiny.

Policy frameworks must also grapple with the distributed nature of multimodal AI. If the molecular assays that inform DAMNet are provided by commercial laboratories, questions of liability arise when an inaccurate molecular result leads to a harmful downstream decision. Similarly, the use of federated learning complicates the determination of which entity is responsible for model performance, as no single institution holds the entire training dataset. Emerging models of algorithm stewardship, in which health systems assume ongoing responsibility for monitoring and updating deployed AI, offer a promising path forward. These stewardship duties must be supported by transparent auditing mechanisms that can inspect input data quality, model drift, and fairness metrics over time, and they must be integrated into hospital quality improvement frameworks. Ultimately, the success of DAMNet will depend as much on the robustness of its socio-technical governance as on the elegance of its network architecture.

7. Conclusion

This paper has presented a dual-attention multimodal network for explainable lung nodule malignancy assessment that brings together volumetric CT imaging and molecular phase-separation biomarkers within a unified architecture. By pairing channel-spatial self-attention with cross-modal attention, the system produces clinically interpretable outputs that localize suspicious nodule regions and quantify the contribution of condensate state features, such as YAP-MAML2 phase behavior, to the malignancy prediction. We have situated this technical design within a comprehensive system-level analysis that addresses data integration, federated privacy-preserving learning, robustness to acquisition and population variability, computational sustainability, and fairness. The convergence of radiological phenotyping and molecular biophysics in a transparent, auditable platform marks a step toward precision diagnostics that can adapt to the realities of fragmented clinical infrastructures. While substantial challenges remain, particularly in standardizing phase-separation assays and ensuring equitable access, the architectural and governance principles outlined here provide a foundation for the next generation of trustworthy multimodal cancer diagnostic systems.

References

1. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249.

2. Shen, W., Zhou, M., Yang, F., Yang, C., & Tian, J. (2017). Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recognition, 61, 663–673.

3. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (pp. 234–241). Springer.

4. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., & Lu, H. (2019). Dual attention network for scene segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3146–3154).

5. Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and disease. Science, 357(6357), eaaf4382.

6. Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., ... & Fuxreiter, M. (2018). Protein phase separation: A new phase in cell biology. Trends in Cell Biology, 28(6), 420–435.

7. Chung, C. I., Yang, J., Yang, X., Liu, H., Ma, Z., Szulzewsky, F., ... & Shu, X. (2024). Phase separation of YAP-MAML2 differentially regulates the transcriptome. Proceedings of the National Academy of Sciences, 121(7), e2310430121.

8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems, 30.

9. Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443.

10. Chen, R. J., Lu, M. Y., Weng, W. H., Chen, T. Y., Williamson, D. F. K., Manz, T., ... & Mahmood, F. (2022). Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4015–4025).

11. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision (pp. 618–626).

12. Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793–4813.

13. Ardila, D., Kiraly, A. P., Bharadwaj, S., Choi, B., Reicher, J. J., Peng, L., ... & Shetty, S. (2019). End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nature Medicine, 25(6), 954–961.

14. Chang, C., Fu, M., Chen, X., Feng, S., Zhang, M., Zhou, X., ... & Liu, Z. (2025, November). Research on PDU-Net Lung Nodule Segmentation Algorithm Based on Path Aggregation and Dual Attention. In 2025 4th International Conference on Image Processing, Computer Vision and Machine Learning (ICICML) (pp. 1897-1900). IEEE.

15. Finlayson, S. G., Bowers, J. D., Ito, J., Zittrain, J. L., Beam, A. L., & Kohane, I. S. (2019). Adversarial attacks on medical AI systems. Science, 363(6433), 1287–1289.

16. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

17. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3(1), 119.

18. Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A, 376(2133), 20180080.

19. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650).

20. Lyon, A. S., Peeples, W. B., & Rosen, M. K. (2021). A framework for understanding the functions of biomolecular condensates across scales. Nature Reviews Molecular Cell Biology, 22(3), 215–235.