ORIGINAL RESEARCH **Open access**

Multimodal Fusion Network Combining Whole-Slide Histopathology Images and Genomic Expression Data for Predicting Immunotherapy Response in Non-Small Cell Lung Cancer Patients

Maria Silva[1*], Joao Pereira[1]

Abstract

Immunotherapy with immune checkpoint inhibitors is a standard treatment for advanced non-small cell lung cancer (NSCLC), with durable responses in selected patients. Whole-slide histopathology images provide morphological and immune microenvironment information, while genomic expression data capture pathway activity and resistance mechanisms. Single-modality approaches based on either histopathology or genomics fail to capture complementary tumor information, limiting accurate stratification of responders and non-responders and leading to suboptimal treatment selection. We propose a multimodal fusion network that integrates whole-slide histopathology images and genomic expression data to predict immunotherapy response in NSCLC. Separate encoders process each modality, followed by cross-attention for joint representation learning in an end-to-end framework. The system includes a multiple instance learning-based WSI module, a gene expression encoder with attention over gene sets, and a cross-attention fusion module. The model outputs a binary or probabilistic prediction of treatment response using paired slide and genomic data. The model captures complementary morphological and molecular signals, linking immune infiltration patterns with transcriptomic activity. Attention mechanisms enhance interpretability by highlighting key tissue regions and gene pathways, while also improving robustness to partial modality missingness. This multimodal framework improves NSCLC immunotherapy response prediction by integrating histopathology and genomic data, offering a step toward more precise patient stratification in precision oncology.

Keywords Multimodal fusion, Attention mechanisms, Whole-slide images, Genomic expression, Immunotherapy response, NSCLC

*Correspondence:
Maria Silva
maria.silva@gmail.com
[1] Department of Artificial Intelligence in Healthcare, University of Coimbra, Coimbra, Portugal

Introduction

Non-small cell lung cancer remains the leading cause of cancer-related mortality worldwide, with advanced disease historically associated with poor prognosis prior to the advent of immunotherapy. Agents such as pembrolizumab and nivolumab targeting the PD-1/PD-L1 axis have transformed first- and second-line treatment paradigms for eligible patients [1, 2]. These therapies elicit durable responses by reinvigorating T-cell mediated antitumor immunity, yet their benefit is confined to a minority of individuals.

Response rates to anti-PD-1/PD-L1 therapy typically range between 20 % and 40 % in unselected advanced NSCLC cohorts, highlighting the urgent need for improved predictive tools. PD-L1 expression assessed by immunohistochemistry serves as the current standard

biomarker, yet it demonstrates imperfect sensitivity and specificity across adenocarcinoma and squamous subtypes. Tumor mutational burden and microsatellite instability status provide additional but still limited guidance in routine practice [3, 4].

Histopathology whole-slide images encode rich spatial information on tumor morphology, stromal composition, and immune infiltrate that cannot be fully captured by genomic profiling alone. Genomic expression data, in turn, quantify transcriptomic programs related to immune checkpoints, chemokine signaling, and oncogenic drivers that modulate therapeutic sensitivity. The integration of these orthogonal data streams therefore represents a logical next step in biomarker development [5, 6].

This conceptual framework introduces a multimodal fusion network that combines whole-slide histopathology images with genomic expression data to predict immunotherapy response in NSCLC patients. The architecture emphasizes end-to-end learning, attention-based interpretability, and robustness to missing modalities. Subsequent sections delineate the background, high-level design, individual processing modules, and evaluation considerations for clinical translation [6].

An overview of the proposed multimodal fusion architecture and its hierarchical data integration pipeline is illustrated in **Figure 1.**

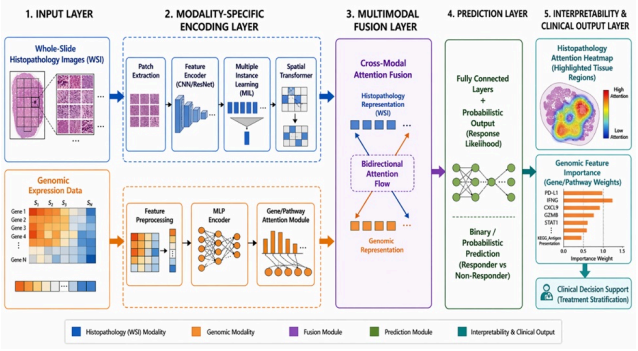

Figure 1. Hierarchical Architecture of a Multimodal Fusion Network Integrating Whole-Slide Histopathology and Genomic Expression Data for Immunotherapy Response Prediction in NSCLC

Background

Immunotherapy in NSCLC

Immune checkpoint inhibitors directed against PD-1 or PD-L1 have revolutionized the therapeutic landscape for advanced NSCLC by blocking inhibitory signals that suppress antitumor T-cell activity. These agents are now routinely employed in both first-line monotherapy and combination regimens with chemotherapy, as well as in the second-line setting following progression on platinum-based therapy. Response assessment relies on RECIST 1.1 criteria with modifications under iRECIST to accommodate pseudoprogression and delayed immune-related responses commonly observed in this modality [7].

Despite these advances, a substantial proportion of patients experience primary or acquired resistance, underscoring the multifactorial nature of immunotherapy outcomes. Factors such as tumor microenvironment composition and systemic immune status interact dynamically to determine clinical benefit. Ongoing research therefore seeks to refine patient selection strategies beyond conventional clinical staging [8].

Current biomarkers

PD-L1 immunohistochemistry using assays such as 22C3, 28-8, or SP263 remains the most widely adopted companion diagnostic for anti-PD-1/PD-L1 therapy in NSCLC, yet cutoffs and scoring systems vary by agent and line of therapy. Tumor mutational burden, typically measured by next-generation sequencing with thresholds around 10 mutations per megabase, has shown complementary predictive value particularly in combination regimens. Microsatellite instability-high status, though rarer in lung cancer, also enriches for responders across several solid tumors [3, 4].

These biomarkers nevertheless suffer from notable limitations including intratumoral heterogeneity, assay variability across platforms, and modest overall predictive accuracy in real-world cohorts. Spatial discordance between primary tumors and metastases further complicates interpretation. Consequently, there is a clear need for integrative approaches that synthesize histopathologic and genomic signals to overcome these shortcomings [9].

Key conceptual differences between unimodal biomarker strategies and the proposed multimodal framework are systematically outlined in **Table 1.**

Table 1. Comparative Theoretical Limitations of Single-Modality Biomarkers versus Multimodal Fusion Approaches in NSCLC Immunotherapy Prediction

Dimension	Histopathology-Only Models	Genomics-Only Models	Multim Fusi Netw
Biological Scope	Captures spatial morphology and immune infiltration	Captures molecular pathways and gene expression	Integra spatial molec determi
Sensitivity to Tumor Heterogeneity	High spatial resolution but limited molecular insight	High molecular depth but lacks spatial context	Simultan models s and mol heteroge
Predictive Robustness	Vulnerable to staining variability and sampling bias	Sensitive to sequencing platform and batch effects	Cross-n redund impro robustn
Interpretability	Visual attention maps highlight tissue regions	Gene-level importance scores identify pathways	Join interpret across t and mol leve
Clinical Translation	Easily accessible but incomplete biomarker	Requires specialized sequencing infrastructure	Levera complem routine molec diagno
Limitation	Misses underlying molecular drivers	Ignores tumor architecture and immune spatial patterns	Requi paired and computa resour

Histopathology as biomarker source

Whole-slide histopathology images contain detailed information on tumor-infiltrating lymphocytes, stromal desmoplasia, and tertiary lymphoid structures that correlate with immunotherapy efficacy. Quantitative assessment of these features can reveal spatial patterns of immune activation not detectable by bulk molecular assays. Necrosis and tumor architecture additionally provide contextual clues regarding treatment sensitivity [10].

Automated deep-learning pipelines have demonstrated the feasibility of extracting such prognostic and predictive signals directly from routine H&E-stained slides. Attention mechanisms within these models can localize biologically relevant regions, thereby linking visual phenotypes to clinical endpoints. This capability positions histopathology as a readily available, cost-effective data source for multimodal modeling [2].

Genomic expression in NSCLC

RNA sequencing profiles in NSCLC reveal immune-related gene signatures, including IFN-γ, CXCL9/10/11, and T-cell receptor repertoire metrics that associate with checkpoint inhibitor response. Driver mutations such as EGFR, KRAS, and STK11 exert distinct effects on the tumor immune microenvironment and can confer primary resistance to immunotherapy. Tumor mutational burden derived from genomic data further stratifies patients when integrated with expression levels [11].

Targeted gene panels and bulk transcriptomic data also capture pathway activation states that modulate antigen presentation and checkpoint molecule expression. These molecular features exhibit complementarity to histopathologic patterns, particularly in distinguishing inflamed from immune-excluded phenotypes. Harnessing both data streams therefore enables a more comprehensive view of the determinants of immunotherapy benefit [12].

Framework Overview

High-level architecture

The proposed multimodal fusion network begins with patch-level feature extraction from whole-slide histopathology images followed by a dedicated WSI encoder that aggregates instance-level representations. Genomic expression data are processed in parallel through a genomic encoder that learns compact embeddings from gene-level inputs. Cross-modal fusion then aligns the two embeddings via attention mechanisms before feeding the joint representation into a final prediction head [6].

This modular design facilitates independent optimization of each modality while enabling information exchange at

multiple stages. The architecture supports both early and intermediate fusion strategies depending on the relative dimensionality and noise characteristics of the inputs. Output layers generate probabilistic estimates of objective response, supporting downstream clinical decision-making [13].

Core assumptions

The framework assumes the availability of paired diagnostic whole-slide H&E images and genomic profiling data, typically obtained from the same diagnostic biopsy or resection specimen. Response labels are derived from RECIST or iRECIST assessments performed in the context of anti-PD-1/PD-L1 therapy. Patient cohorts are restricted to histologically confirmed NSCLC, encompassing both adenocarcinoma and squamous cell carcinoma subtypes [1].

Additional assumptions include standardized slide scanning at clinically routine magnifications and genomic data generated via clinically validated RNA-seq or targeted panels. The model is designed to operate on retrospective or prospective data without requiring real-time tissue processing beyond standard pathology workflows. These conditions ensure feasibility within existing healthcare infrastructures [8].

Design principles

End-to-end multimodal learning is prioritized to allow the network to discover interactions between visual and molecular features that might be missed by hand-crafted rules. Interpretability is embedded through attention maps that surface both patch-level contributions within slides and gene-level importance within expression profiles. Robustness to missing modalities is achieved by incorporating modality dropout during training and fallback unimodal pathways [14].

Scalability and generalizability are further supported by leveraging pre-trained encoders for each data type and by enforcing permutation-invariant operations where appropriate. The overall design aligns with principles of data efficiency and clinical translatability, minimizing reliance on large annotated datasets while maximizing explanatory power [13].

Wsi Processing Module

Patch extraction and encoding

Whole-slide images are tessellated into non-overlapping patches of fixed resolution, typically 256 × 256 or 512 × 512 pixels at 20× or 40× magnification, to manage computational demands. Each patch is passed through a pre-trained feature extractor such as a ResNet backbone or pathology-specific foundation model to obtain low-dimensional embeddings. Dimensionality reduction via principal component analysis or learned projection layers further condenses these representations prior to downstream aggregation [2].

This patch-based workflow preserves local tissue context while enabling parallel processing across gigapixel slides. Normalization steps at both stain and feature levels mitigate batch effects arising from multi-center scanning variability. The resulting patch embeddings serve as the foundational input for subsequent multiple instance learning stages [1].

Multiple Instance Learning (MIL)

Treating each whole-slide image as a bag of unordered patch instances allows multiple instance learning to operate without exhaustive pixel-level annotations. Attention-based MIL assigns learnable weights to individual patches, thereby identifying those most discriminative for immunotherapy response prediction. The final slide-level embedding is computed as a weighted sum of instance features, preserving instance-level interpretability [10].

This paradigm is particularly suited to histopathology because only a small fraction of patches may drive the overall biological signal. Variants incorporating clustering or graph-based relations among instances can further enhance representation quality. Training proceeds with weak supervision from patient-level response labels, promoting data-efficient learning [2].

Spatial transformer

A spatial transformer layer applies self-attention across patch embeddings to model long-range dependencies and tissue architecture within the slide. Positional encodings encode the relative coordinates of patches, allowing the model to capture spatial relationships such as tumor-stroma interfaces or lymphoid aggregates. Multi-head attention mechanisms enable the network to jointly attend to morphological patterns across distant regions [15].

Integration of this transformer component refines the slide-level representation by incorporating contextual information that pure MIL might overlook. Residual connections and layer normalization stabilize training on variable-length bags of patches. The resulting spatially aware embeddings feed directly into the multimodal fusion stage [13].

Genomic Data Processing Module

Input features

Genomic input features are derived from RNA-seq counts or targeted panel data, with emphasis on the most variably expressed genes and pre-defined immune-related gene sets. Mutation status for key NSCLC drivers including TP53, KRAS, EGFR, and STK11 is encoded as binary or one-hot vectors. Additional scalar features such as tumor mutational burden and PD-L1 immunohistochemistry scores are concatenated to enrich the molecular profile [11].

Feature selection or embedding layers reduce the high-dimensional gene space while preserving biologically meaningful signals. Normalization techniques including log-transformation and z-scoring ensure compatibility across sequencing platforms. This curated input vector encapsulates both global transcriptomic programs and specific alterations known to influence immunotherapy outcomes [12].

Genomic encoder

A multilayer perceptron with batch normalization and dropout layers transforms the raw genomic feature vector into a fixed-dimensional embedding suitable for fusion. Optional self-attention over pre-defined gene modules allows the encoder to weigh pathway-level contributions dynamically. Nonlinear activations and residual blocks enhance the model's capacity to capture complex non-linear interactions within the expression data [14].

The encoder is trained jointly with the remainder of the network to ensure that genomic embeddings remain informative for the downstream response prediction task. Regularization strategies prevent overfitting to cohort-specific batch effects. The resulting compact representation encodes molecular drivers in a format directly comparable to the WSI-derived embedding [13].

Multimodal Fusion Module

The architectural components and their respective computational roles within the multimodal framework are detailed in **Table 2.**

Table 2. Structural Decomposition of the Multimodal Fusion Network: Modules, Functions, and Learning Mechanisms

Module	Subcomponent	Input Type	Core Func
WSI Processing	Patch Extraction	Whole-slide images	Conve gigapixel s into manage patche
	Feature Encoder (CNN)	Image patches	Extract vi feature
	Multiple Instance Learning	Patch embeddings	Aggrega informat region
	Spatial Transformer	Patch embeddings + coordinates	Model sp relationsh
Genomic Processing	Feature Preprocessing	RNA-seq / gene panel data	Normalize select feat
	MLP Encoder	Gene expression vectors	Learn nonline molecul representa
	Gene/Pathway Attention	Gene sets	Identify k biologic pathwa
Fusion Module	Cross-Modal Attention	WSI + genomic embeddings	Align modalities learn interacti
Prediction Module	Classification Head	Fused embedding	Predic immunothe respons
Interpretability	Attention Visualization	Model weights	Provid biologic explanati

Fusion strategies

The multimodal fusion module systematically evaluates early, late, and intermediate fusion strategies to integrate embeddings from whole-slide histopathology images and genomic expression data. Early fusion concatenates modality-specific representations at the input level, permitting the network to learn low-level cross-modal interactions directly. Late fusion generates separate unimodal predictions that are subsequently combined through a meta-classifier, offering robustness when one data stream is noisy or incomplete [16, 17]. Intermediate fusion via cross-attention mechanisms allows dynamic alignment of morphological and transcriptomic features across network layers, balancing expressiveness with computational tractability [18]. This hierarchical selection of fusion approaches ensures the framework adapts to the complementary strengths of histopathology and genomics in NSCLC immunotherapy contexts.

Cross-modal attention

Cross-modal attention serves as the core mechanism within the fusion module by treating one modality as the query and the other as key-value pairs to highlight aligned predictive signals. Genomic embeddings can query WSI patch representations to emphasize tissue regions associated with specific molecular pathways, while the reverse direction enables genomic features to contextualize morphological patterns. This bidirectional attention refines the joint latent space without explicit supervision on feature correspondence [19, 20]. The resulting fused representation encodes interactions such as immune infiltrate density modulated by chemokine expression signatures. Such mechanisms promote biologically plausible alignments that enhance overall response prediction fidelity [21].

Attention and Interpretability

Histopathology attention visualization

Histopathology attention visualization generates heatmaps that overlay patch-level importance weights onto the original whole-slide image, revealing regions most influential to the immunotherapy response prediction. These maps typically highlight tumor-stroma interfaces, dense lymphocyte aggregates, and areas of necrosis that correlate with immune activation. Clinicians can inspect the highlighted patches to validate alignment with known histopathological predictors of checkpoint inhibitor benefit [22]. The process maintains full spatial resolution while preserving the weak-supervision paradigm inherent to multiple instance learning. Such visualizations bridge the gap between black-box model outputs and interpretable tissue phenotypes [23].

Genomic feature importance

Genomic feature importance is quantified through attention weights assigned to individual genes or pre-defined pathway modules within the encoder output. High-weight genes frequently map to immune checkpoints, interferon signaling, or resistance-associated drivers such as EGFR and STK11 alterations. These importance scores can be aggregated at the pathway level to provide molecular explanations for predicted response probabilities [24]. The approach facilitates hypothesis generation regarding resistance mechanisms without requiring post-hoc explainability tools. Integration of these weights with histopathology attention maps further supports multi-scale interpretability across data modalities [25].

Prediction and Clinical Decision Support

Response prediction output

The prediction module constitutes the terminal component of the multimodal architecture, transforming fused imaging–genomic representations into clinically interpretable response probabilities. Specifically, the model outputs a continuous likelihood of objective response to anti–PD-1/PD-L1 therapy, operationalized as complete response (CR) or partial response (PR) versus stable disease (SD) or progressive disease (PD), in accordance with RECIST criteria. This formulation aligns the computational output directly with standardized oncologic endpoints, thereby enhancing translational relevance.

To ensure probabilistic interpretability, the final layer employs either sigmoid activation (binary framing) or softmax activation (multi-class extensions), followed by post hoc calibration techniques such as temperature scaling or isotonic regression. These calibration strategies are critical in clinical contexts, where overconfident predictions may lead to suboptimal treatment allocation. The resulting outputs enable stratification of patients into responder and non-responder categories, with tunable

decision thresholds that can be adapted to specific clinical priorities—favoring sensitivity in settings where missing a potential responder carries high risk, or specificity where overtreatment must be minimized [26].

Beyond point estimates, the model incorporates uncertainty quantification through Monte Carlo dropout at inference time. By sampling multiple stochastic forward passes, the system approximates the posterior predictive distribution and derives confidence intervals around response probabilities. This approach captures epistemic uncertainty arising from limited training data or distributional shifts, which is particularly relevant in heterogeneous diseases such as non-small cell lung cancer (NSCLC). The explicit representation of uncertainty allows clinicians to distinguish between high-confidence and ambiguous predictions, facilitating more cautious interpretation in borderline cases.

Importantly, this probabilistic framework extends beyond binary classification by enabling downstream decision-analytic applications, such as risk-adjusted treatment selection, expected utility estimation, and integration into Bayesian clinical decision models [27]. As such, the prediction output is designed not merely as a classifier but as a quantitatively robust input to precision oncology decision-making.

Clinical workflow integration

The proposed system is designed for seamless integration into existing clinical workflows, functioning as an assistive decision-support layer rather than a replacement for established diagnostic modalities. The model ingests routinely acquired diagnostic data, including digitized whole-slide histopathology images from pathology departments and genomic profiling reports generated by molecular laboratories (e.g., next-generation sequencing panels).

Following inference, results are compiled into a structured, clinician-facing report that can be appended to the electronic health record (EHR). This report synthesizes multimodal evidence by highlighting both salient histomorphological features (e.g., tumor-infiltrating lymphocyte density, spatial heterogeneity) and key molecular determinants (e.g., tumor mutational burden, actionable mutations) contributing to the predicted response. Explainability modules—such as attention heatmaps or feature attribution scores—can further enhance interpretability by visually localizing predictive regions within histological slides.

Crucially, the system is positioned to complement, rather than supplant, PD-L1 immunohistochemistry (IHC), which remains a standard biomarker for immunotherapy eligibility. In cases where PD-L1 expression falls within equivocal or borderline ranges, the multimodal model provides orthogonal evidence that may refine therapeutic decisions [28]. This is particularly valuable given the known limitations of PD-L1 as a standalone biomarker, including spatial heterogeneity and assay variability.

Operational deployment is envisioned at the stage of multidisciplinary tumor board discussions, where oncologists, pathologists, radiologists, and molecular specialists collaboratively determine first-line treatment strategies. By presenting an integrated, data-driven assessment of immunotherapy response likelihood, the system supports consensus-building and individualized treatment planning. Importantly, the design prioritizes interoperability with existing digital pathology platforms and clinical information systems, ensuring minimal disruption to standard-of-care processes while enhancing the precision oncology pipeline [29].

Evaluation Strategy

Prediction metrics

Model performance is evaluated using a comprehensive suite of discrimination and calibration metrics tailored to imbalanced clinical datasets. Primary emphasis is placed on the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPRC), which respectively capture global discrimination ability and performance in the context of class imbalance—particularly relevant when responder rates are low.

Complementary metrics include accuracy, sensitivity (recall), specificity, and F1-score, providing a balanced view of classification performance across clinically meaningful dimensions. Calibration is assessed באמצעות the Brier score and reliability diagrams, ensuring that predicted probabilities correspond closely to observed outcome frequencies. Proper calibration is essential for clinical adoption, as decision-making often depends on absolute risk estimates rather than relative rankings.

All metrics are computed at both the aggregate patient level and within clinically relevant subgroups, such as histological subtypes (e.g., adenocarcinoma versus squamous cell carcinoma in NSCLC). This stratified evaluation ensures that model performance is not biased toward dominant subpopulations and supports claims of generalizability across disease heterogeneity [14].

To avoid dependence on arbitrary classification thresholds, threshold-independent analyses (e.g., ROC and PR curves) are emphasized. Where thresholds are required for clinical interpretation, sensitivity analyses are conducted across a range of cutoff values. Additionally, confidence intervals for all reported metrics are estimated באמצעות bootstrapping or cross-validation variance, providing statistical rigor and enabling meaningful comparison with baseline models [13].

Validation protocols

Robust validation is achieved through a multi-tiered strategy encompassing internal cross-validation, external cohort testing, and systematic ablation analyses. Initially, stratified k-fold cross-validation is performed on paired whole-slide image and genomic datasets, ensuring that each fold preserves the distribution of response classes and histological subtypes. This approach mitigates overfitting while maximizing data utilization.

To assess generalizability, the model is subsequently evaluated on independent external datasets collected from multiple institutions, thereby capturing variability in patient demographics, tissue processing protocols, and sequencing platforms. Such multi-center validation is critical for demonstrating real-world applicability and reducing the risk of domain-specific bias.

Ablation studies are conducted to quantify the contribution of individual model components, including imaging-only, genomics-only, and fully fused multimodal configurations. By systematically removing or perturbing modules, these experiments elucidate the incremental value of multimodal integration and identify potential redundancies or bottlenecks in the architecture [6].

Further robustness is examined through sensitivity analyses that simulate missing-modality scenarios using modality dropout techniques. This is particularly important in clinical settings where complete data may not always be available. The model's ability to maintain performance under partial input conditions enhances its practical utility and resilience.

Collectively, these validation protocols establish a rigorous framework for assessing model reliability, interpretability, and translational readiness, thereby supporting its progression toward prospective clinical trials and eventual deployment in precision oncology settings [8].

Conclusion

The multimodal fusion network provides a comprehensive conceptual architecture for integrating whole-slide histopathology images with genomic expression data to predict immunotherapy response in non-small cell lung cancer. By combining multiple instance learning and spatial transformers for histopathology with attention-augmented multilayer perceptrons for genomics, the design captures complementary morphological and molecular determinants of treatment benefit. Cross-modal attention further aligns these data streams to generate interpretable joint representations suited to precision oncology.

Key advantages include the framework's ability to leverage readily available diagnostic specimens, its built-in attention mechanisms for clinical interpretability, and its robustness to partial modality availability through dropout strategies. The modular structure facilitates iterative refinement and extension to additional biomarkers or cancer types while maintaining end-to-end differentiability. These properties position the network as a scalable bridge between computational pathology and molecular oncology.

Limitations encompass the requirement for paired WSI and genomic data at the individual patient level, the computational demands of processing gigapixel slides, and the need for careful harmonization across sequencing platforms and slide scanners. Annotation effort for response labels remains non-trivial, and generalizability across diverse demographic and histologic cohorts requires ongoing validation. Future work must also address ethical considerations surrounding model deployment in diverse healthcare settings.

Implementation of this conceptual framework on public NSCLC immunotherapy cohorts such as those from CPTAC, TCGA, and prospective clinical trials will accelerate its maturation toward real-world application. Collaborative efforts among pathologists, oncologists, and AI researchers are essential to refine and prospectively test

the architecture. Ultimately, such multimodal approaches promise to advance patient stratification and optimize immunotherapy outcomes in advanced non-small cell lung cancer.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 05 Mar 2023 Revised: 29 May 2023 Accepted: 10 Jul 2023
Published online: 20 January 2024

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med. 2018;24(10):1559-67. https://doi.org/10.1038/s41591-018-0177-5.

Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng. 2021;5(6):555-70. https://doi.org/10.1038/s41551-021-00682-w.

Wiesweg M, Mairinger F, Reis H, Goetz M, Kollmeier J, Misch D, et al. Machine learning reveals a PD-L1-independent prediction of response to immunotherapy of non-small cell lung cancer by gene expression context. Eur J Cancer. 2020;140:76-85. https://doi.org/10.1016/j.ejca.2020.08.021.

Niu Y, Wang L, Zhang X, Han Y, Yang C, Wang M, et al. Predicting tumor mutational burden from lung adenocarcinoma histopathological images using deep learning. Front Oncol. 2022;12:927426. https://doi.org/10.3389/fonc.2022.927426.

Dammak S, Cecchini MJ, Breadner D, Ward AD. Using deep learning to predict tumor mutational burden from scans of H&E-stained multicenter slides of lung squamous cell carcinoma. J Med Imaging (Bellingham). 2023;10(1):017502. https://doi.org/10.1117/1.JMI.10.1.017502.

Chen RJ, Lu MY, Wang J, Williamson DFK, Rodig SJ, Lindeman NI, et al. Pathomic fusion: an integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis. IEEE Trans Med Imaging. 2022;41(4):757-70. https://doi.org/10.1109/TMI.2020.3021387.

Li B, Yang L, Zhang H, Li H, Jiang C, Zhang Y, et al. Outcome-supervised deep learning on pathologic whole slide images for survival prediction of immunotherapy in patients with non-small cell lung cancer. Mod Pathol. 2023;36(8):100208. https://doi.org/10.1016/j.modpat.2023.100208.

Park S, Ock CY, Kim H, Pereira S, Park S, Ma M, et al. Artificial intelligence-powered spatial analysis of tumor-infiltrating lymphocytes as complementary biomarker for immune checkpoint inhibition in non-small-cell lung cancer. J Clin Oncol. 2022;40(17):1916-28. https://doi.org/10.1200/JCO.21.02010.

Tian P, He B, Mu W, Liu K, Liu L, Zhan Y, et al. Assessing PD-L1 expression in non-small cell lung cancer and predicting

responses to immune checkpoint inhibitors using deep learning on computed tomography images. Theranostics. 2021;11(5):2098-107. https://doi.org/10.7150/thno.52286.

Rakaee M, Adib E, Ricciuti B, Sholl LM, Shi W, Alchahin A, et al. Association of machine learning-based assessment of tumor-infiltrating lymphocytes on standard histologic images with outcomes of immunotherapy in patients with NSCLC. JAMA Oncol. 2023;9(1):51-60. https://doi.org/10.1001/jamaoncol.2022.4452.

Yang Y, Yang J, Shen L, Chen J, Xia L, Luo Q, et al. A multi-omics-based serial deep learning approach to predict clinical outcomes of single-agent anti-PD-1/PD-L1 immunotherapy in advanced stage non-small-cell lung cancer. Am J Transl Res. 2021;13(2):743-56.

Kong J, Ha D, Lee J, Kim I, Park M, Im SH, et al. Network-based machine learning approach to predict immunotherapy response in cancer patients. Nat Commun. 2022;13(1):3703. https://doi.org/10.1038/s41467-022-31405-6.

Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, et al. Artificial intelligence for multimodal data integration in oncology. Cancer Cell. 2022;40(10):1095-110. https://doi.org/10.1016/j.ccell.2022.09.012.

Steyaert S, Pizurica M, Nagaraj D, Khandelwal P, Hernandez-Boussard T, Gevaert O, et al. Multimodal data fusion for cancer biomarker discovery with deep learning. Nat Mach Intell. 2023;5(4):351-62. https://doi.org/10.1038/s42256-023-00633-5.

Chen RJ, Lu MY, Weng WH, Chen TY, Williamson DFK, Manz T, et al. Multimodal co-attention transformer for survival prediction in gigapixel whole slide images. In: Proc IEEE/CVF Int Conf Comput Vis (ICCV). 2021. p. 4015-25. https://doi.org/10.1109/ICCV48922.2021.00401.

Shamai G, Livne A, Polónia A, Sabo E, Cretu A, Peleg R, et al. Deep learning-based image analysis predicts PD-L1 status from H&E-stained histopathology images in breast cancer. Nat Commun. 2022;13(1):6753. https://doi.org/10.1038/s41467-022-34435-4.

Liu Q, Yao J, Yao L, Chen X, Zhou J, Xiao G, et al. M2 fusion: Bayesian-based multimodal multi-level fusion on colorectal cancer microsatellite instability prediction. In: Int Conf Med Image Comput Comput Assist Interv (MICCAI). Cham: Springer; 2023. p. 125-34. https://doi.org/10.1007/978-3-031-43907-0_13.

Schneider L, Wies C, Krieghoff-Henning EI, Bucher TC, Utikal JS, Schadendorf D, et al. Multimodal integration of image, epigenetic and clinical data to predict BRAF mutation status in melanoma. Eur J Cancer. 2023;183:131-8. https://doi.org/10.1016/j.ejca.2023.02.018.

Xie X, Wang X, Liang Y, Yang J, Wu Y, Zhang T, et al. Evaluating cancer-related biomarkers based on pathological images: a systematic review. Front Oncol. 2021;11:763527. https://doi.org/10.3389/fonc.2021.763527.

Wang C, Ma J, Shao J, Zhang S, Liu Z, Sun H, et al. Predicting EGFR and PD-L1 status in NSCLC patients using multitask AI system based on CT images. Front Immunol. 2022;13:813072. https://doi.org/10.3389/fimmu.2022.813072.

Saad MB, Hong L, Aminu M, Vokes NI, Chen P, Elhalawani H, et al. Predicting benefit from immune checkpoint inhibitors in patients with non-small-cell lung cancer by CT-based ensemble deep learning: a retrospective study. Lancet Digit Health. 2023;5(7):e404-e420. https://doi.org/10.1016/S2589-7500(23)00087-0.

Sadhwani A, Chang HW, Behrooz A, Brown T, Auvigne-Flament I, Patel P, et al. Comparative analysis of machine learning approaches to classify tumor mutation burden in lung adenocarcinoma using histopathology images. Sci Rep. 2021;11(1):16605. https://doi.org/10.1038/s41598-021-95886-4.

Cheng G, Zhang F, Xing Y, Hu X, Zhang H, Sun K, et al. Artificial intelligence-assisted score analysis for predicting the expression of the immunotherapy biomarker PD-L1 in lung cancer. Front Immunol. 2022;13:893198. https://doi.org/10.3389/fimmu.2022.893198.

Lin H, Pan X, Feng Z, Yan L, Hua J, Shao C, et al. Automated whole-slide images assessment of immune infiltration in resected non-small-cell lung cancer: towards better risk-stratification. J Transl Med. 2022;20(1):261. https://doi.org/10.1186/s12967-022-03441-5.

Meng X, Liu Y, Zhang J, Teng F, Xing L, Yu J. PD-1/PD-L1 checkpoint blockades in non-small cell lung cancer: new development and challenges. Cancer Lett. 2017;405:29-37. https://doi.org/10.1016/j.canlet.2017.06.033.

Tan WC, Nerurkar SN, Cai HY, Ng HH, Wu D, Wee YTF, et al. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Commun (Lond). 2020;40(4):135-53. https://doi.org/10.1002/cac2.12023.

Wang S, Yang DM, Rong R, Zhan X, Fujimoto J, Liu H, et al. Artificial intelligence in lung cancer pathology image analysis.

Cancers (Basel). 2019;11(11):1673. https://doi.org/10.3390/cancers11111673.

Liu J, Islam MT, Sang S, Qiu L, Xing L. Biology-aware mutation-based deep learning for outcome prediction of cancer immunotherapy with immune checkpoint inhibitors. NPJ Precis Oncol. 2023;7(1):117. https://doi.org/10.1038/s41698-023-00471-1.

Gao Q, Yang L, Lu M, Jin R, Ye H, Wang X, et al. The artificial intelligence and machine learning in lung cancer immunotherapy. J Hematol Oncol. 2023;16(1):55. https://doi.org/10.1186/s13045-023-01452-7.