TCFD-Net: Temporal-Calibrated Feature Disentanglement Network for Dual-timepoint CT-based Immunotherapy Response Prediction in Advanced Non-small Cell Lung Cancer

Yao Fu[1,2#], Ziwei Gao[3#], Yinuo Lu[1,2#], Weixin Xu[1,2], Mingxin Qi[1,2], Yukun Wu[1,2], Wei Mu[1,2✉]

[1] Key Laboratory of Precision Diagnosis and Treatment Technology and Intelligent Imaging Equipment, School of Engineering Medicine, Beihang University, Beijing, 100190, China
[2] Qingdao Research Institute, Beihang University, Qingdao, 266000, China
[3] School of Biological Science and Medical Engineering, Beihang University, Beijing, 100190, China

Abstract. Longitudinal CT imaging analysis, which captures the spatiotemporal evolution of tumor characteristics across early treatment cycles, holds great promise for predicting immunotherapy response and optimizing personalized treatment strategies in advanced non-small cell lung cancer. However, existing methods that rely on shared-weight encoders often suffer from feature entanglement between temporally stable and treatment-induced dynamic representations, where the dominant static features may suppress subtle dynamic signals that are most predictive of immunotherapy response. To address this challenge, we propose a Temporal-Calibrated Feature Disentanglement Network (TCFD-Net), which explicitly disentangles longitudinal CT features into temporal-invariant and temporal-variant components via a Temporal Invariant-Variant Disentanglement Module, supervised with cross-time mean squared error (MSE) consistency loss, batch distance loss and subspace orthogonality loss to ensure effective disentanglement. Additionally, a Time-Gating Mechanism is designed to adaptively calibrate temporal-variant features based on inter-scan intervals, addressing the irregular scanning protocols common in clinical practice. Experiments on both internal and external testing dataset demonstrate that the proposed method outperforms existing multi-timepoint analysis approaches, achieving AUC improvements of 7.72%-18.84% and accuracy improvements of 10.72%-16.27% on the external testing dataset.

This study was supported by National Natural Science Foundation of China (62571019, 92259205), Shandong Provincial Natural Science Foundation (ZR2024QF010), Science,Technology&Innovation Project of Xiongan New Area (2023XAGG0071), and the fundamental research funds for the central universities.

# Equal contribution. ✉ Corresponding author: weimu@buaa.edu.cn

Our code and 95% Confidence Intervals of our results are available at https://github.com/YFSoaring/TCFD-Net.

Keywords: Longitudinal CT imaging, Temporal-calibrated feature disentanglement network, Immunotherapy response prediction.

1 Introduction

Immune checkpoint inhibitors (ICIs) have revolutionized the treatment landscape for advanced non-small cell lung cancer, significantly improving survival outcomes in this patient population [1-3]. Nevertheless, ICI efficacy remains highly heterogeneous, with only a subset of patients achieving durable clinical benefit [4, 5], underscoring the clinical need for early and accurate identification of treatment responders. Since ICI-induced immune modulation drives continuously evolving tumor responses, baseline or single-timepoint computed tomography (CT) is insufficient to capture these dynamic changes [6, 7]. Therefore, longitudinal CT analysis is essential for capturing subtle spatiotemporal changes that reflect early treatment response and guiding individualized treatment decisions.

Existing longitudinal CT-based approaches have evolved considerably in both feature extraction capacity and temporal fusion strategy. Early methods computed difference or concatenation of hand-crafted radiomic features across time points to capture temporal changes [11, 12], which were subsequently replaced by deep learning-based and foundation model-based representations to enhance feature expressivity and generalizability, while retaining the same fusion strategies [13, 14]. To better capture temporal dynamics, later works introduced more principled modeling objectives, such as contrastive learning to encode treatment-induced tumor changes [15] and multi-task representation learning to enforce temporal continuity of phenotypic trajectories [16]. Yet across all these paradigms, stable features like anatomical structures and treatment-induced dynamic changes remain conflated in the shared feature space, leading to entangled representations where dominant static features may suppress the subtle dynamic signals most predictive of immunotherapy response.

Efforts to disentangle stable and dynamic representations have remained limited. While attempts such as HARM^3-Fusion [17] have introduced latent variable decomposition to partition representations into “inherent” and “variant” components, such approaches rely on implicit regularization rather than explicit disentanglement, limiting their ability to produce discriminative representations for response prediction. This challenge is further exacerbated by irregular follow-up intervals in clinical practice, which introduce interval-dependent confounding into temporal-variant components, thereby complicating the isolation of true treatment-induced dynamic signals.

To address these limitations, a Time-Calibrated Feature Disentanglement Network (TCFD-Net) is proposed for early prediction of immunotherapy response in advanced non-small cell lung cancer based on pre- and post-treatment CT images. The contributions of this work are as follows:

1. We propose a Temporal Invariant-Variant Disentanglement Module that explicitly factorizes longitudinal representations into temporal-invariant and temporal-variant subspaces under joint supervision of cross-time mean squared error (MSE) consistency, batch distance, and subspace orthogonality losses, ensuring the extraction of discriminative treatment-induced dynamic information.

2. We design a Time-Gating Mechanism that adaptively calibrates disentangled temporal-variant representations according to inter-scan intervals, removing interval-dependent confounding to ensure that the temporal-variant subspace faithfully captures treatment-induced dynamic changes rather than acquisition timing artifacts.
3. We conduct multi-center validation of TCFD-Net, demonstrating its superiority over existing longitudinal modeling approaches and consistent generalization across diverse patient cohorts, establishing its clinical potential for early immunotherapy response prediction in advanced non-small cell lung cancer.

2 Methods

2.1 Overall Architecture

The overview of the proposed TCFD-Net framework is depicted in Fig. 1, with two core components: (1) the Temporal Invariant-Variant Disentanglement Module (TIVDM), and (2) the Time-Gating Mechanism (TGM). This model takes a pair of longitudinal CT scans, denoted as the pre-treatment CT image $X_{pre} \in \mathbb{R}^{D\times H\times W}$ and post-treatment CT image $X_{post} \in \mathbb{R}^{D\times H\times W}$, along with their acquisition time interval $\Delta t \in \mathbb{R}^{+}$ as input, and predicts the binary immunotherapy response label. In the TIVDM, the X_{pre} and X_{post} are fed into two structurally separated 3D encoders [18] to extract deep feature representations. Both branches are jointly optimized under classification supervision to learn response-relevant representations. The upper branch is further constrained by a cross-time MSE consistency loss and a batch distance loss to capture temporal-invariant features while preventing representation collapse. Meanwhile, a subspace orthogonality loss encourages the lower branch to encode complementary temporal-variant signals associated with treatment-induced changes. To mitigate the impact of irregular follow-up intervals, the TGM adaptively calibrates the temporal-variant features based on Δt, ensuring accurate modeling of treatment dynamics across variable intervals. Finally, the disentangled temporal-invariant and calibrated temporal-variant representations are subsequently fused and fed into a classifier for treatment response prediction.

2.2 Temporal Invariant-Variant Disentanglement Module

To explicitly decompose the longitudinal CT into temporal-invariant and temporal-variant representations, thereby enhancing the model's ability to capture the evolutionary process related to immunotherapy, the TIVDM is designed.

The Temporal Invariant-Variant Disentanglement Module (TIVDM) uses two structurally identical but parameter-independent 3D Swin Transformers (3D ST) with an embedding dimension of 16 and a depth configuration of [2,2,2,2] to extract temporal-invariant and variant representations, respectively. Given the pre-CT X_{pre}^{i}, and post-CT X_{post}^{i} of the i-th patient, the temporal-invariant features F_{inv}^{i} and temporal-variant

features F_{var}^{i} are extracted by the two 3D encoders ($Encoder_{inv}$, $Encoder_{var}$), respectively, as follows:

$$F_{inv,t}^{i} = Encoder_{inv}(X_t^i), t \in \{pre, post\} \tag{1}$$

$$F_{var,t}^{i} = Encoder_{var}(X_t^i), t \in \{pre, post\} \tag{2}$$

where $F_{inv,t}^{i} \in \mathbb{R}^{d_1}$ represents the extracted features, d_1 represents the feature dimension.

To capture the feature with temporal invariance, we introduce a cross-time MSE consistency loss ($\mathcal{L}_{mse}$) to minimize the discrepancy between pre- and post-treatment representations within the same patient, while simultaneously employing a batch distance loss ($\mathcal{L}_{dist}$) to maximize the feature variance between different patients to prevent representation collapse. It can be formulated as:

$$\mathcal{L}_{mse} = \frac{1}{B}\sum_{i=1}^{B}\left\|F_{inv,pre}^{i} - F_{inv,post}^{i}\right\|_2^2 \tag{3}$$

$$\mathcal{L}_{dist} = \frac{1}{B^2}\sum_{i=1}^{B}\sum_{j=1}^{B}\left(\frac{F_{inv}^{i}\cdot F_{inv}^{j}}{\left\|F_{inv}^{i}\right\|_2\left\|F_{inv}^{j}\right\|_2} - \delta_{ij}\right)^2 \tag{4}$$

where B represents the batch-size, $\|\cdot\|_2^2$ is the square L_2 norm. δ_{ij} is the Kronecker delta which equals 1 if i=j and 0 otherwise.

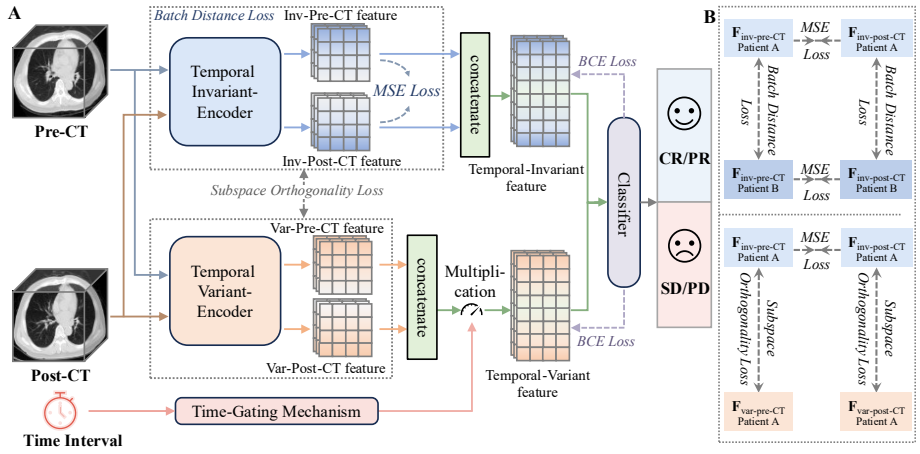

Fig.1. Overview architecture of TCFD-Net framework. **A:** Our framework contains two key components: Temporal Invariant-Variant Disentanglement Module and Time-Gating Mechanism. **B:** Joint optimization using cross-time MSE consistency and Batch Distance losses on temporal-invariant features, together with Orthogonality-constrained subspace separation.

To capture the dynamic representations sensitive to treatment-induced changes, we introduce a subspace orthogonality loss ($\mathcal{L}_{orth}$). This forces the temporal-variant features to be perpendicular to the temporal-invariant features in the embedding space.

Given the batch-normalized feature matrices $\tilde{F}_{inv} \in \mathbb{R}^{B\times 2d_1}$ and $\tilde{F}_{var} \in \mathbb{R}^{B\times 2d_1}$, the loss is defined as the squared Frobenius norm of their cross-correlation matrix:

$$\mathcal{L}_{orth} = \left\| \frac{1}{B-1} \left(\tilde{F}_{inv}\right)^T \tilde{F}_{var} \right\|_F^2 \tag{5}$$

where B represents the batch-size, $\|\cdot\|_F^2$ is the square Frobenius norm.

2.3 Time-Gating Mechanism

To mitigate the observational bias introduced by irregular imaging intervals, we propose the TGM that calibrates the temporal-variant representations based on the time interval.

Given the time interval Δt_i between the pre- and post-CT of the i-th patient, we first project this scalar time value into a high-dimensional gating vector g_i using a two-layer multi-layer perceptron (MLP). The gating vector is computed as:

$$g_i = \sigma_2(W_2 \cdot \sigma_1(W_1 \cdot \Delta t_i + b_1) + b_2) \tag{6}$$

where $W_1 \in \mathbb{R}^{d_2\times 1}$ and $W_2 \in \mathbb{R}^{2d_1\times d_2}$ are learnable weight matrices, $b_1 \in \mathbb{R}^{d_2}$ and $b_2 \in \mathbb{R}^{2d_1}$ are bias terms, σ_1 denotes the ReLU function, and σ_2 denotes the Sigmoid function.

Finally, the time-calibrated variant feature $\bar{F}_{var}^i \in \mathbb{R}^{B\times 2d_1}$, is obtained by modulating the original time-variant feature F_{var}^i via element-wise multiplication:

$$\bar{F}_{var}^i = F_{var}^i \odot g_i \tag{7}$$

where $\odot$ represents the Hadamard product.

3 Experiments

3.1 Datasets and Implementation Details

Data Collection A total of 1133 patients with advanced non-small cell lung cancer receiving immunotherapy between 2016 and 2023 were retrospectively enrolled from two collaborating centers. Treatment responses were categorized per RECIST 1.1 criteria, grouping Complete Response (CR) and Partial Response (PR) as Responders, while Stable Disease (SD) and Progressive Disease (PD) were classified as Non-responders [19].

The primary dataset from Center A (n=763) comprised 444 Responders and 319 Non-responders. It was randomly partitioned into training (n=534), validation (n=76), and internal testing (n=153) datasets at an approximate ratio of 7:1:2. To evaluate the model's generalizability, an independent dataset from Center B (n=370) served as the external testing dataset. This dataset included 211 Responders and 159 Non-responders. The median interval between pre- and post-treatment CT scans was 8.4 weeks (4.6–29.6) and 8.0 weeks (4.6–24.0) in Centers A and B, respectively.

Implementation Details and Evaluation The input to the model is a whole-lung CT volume, and resized to 128×128×128. The model was implemented in PyTorch 1.12 using Python 3.7 and trained on a NVIDIA RTX 3090 GPU. Training was performed using the Adam optimizer with an initial learning rate of 0.0001. A step-wise learning rate scheduler was adopted with a step size of 10 epochs and a decay factor of 0.5. The batch size was set to 48, and the model was trained for 50 epochs. The final model weights were selected based on the lowest validation loss. For response prediction, binary cross-entropy loss ($\mathcal{L}_{cls}$) was used as the primary classification objective. The overall training objective was defined as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{cls} + \lambda_2 \mathcal{L}_{mse} + \lambda_3 \mathcal{L}_{dist} + \lambda_4 \mathcal{L}_{orth} \tag{8}$$

where λ balances all loss components during training. The coefficients λ_1, λ_2, λ_3, λ_4 were set to 0.4, 0.2, 0.2, 0.2. The model performance was evaluated using Area Under the Receiver Operating Characteristic Curve (AUC), Accuracy (ACC), Sensitivity (SEN) and Specificity (SPE) [20]. The cut-off value for binary classification was selected based on the maximum Youden index derived from the training dataset.

3.2 Results

Ablation Study The key components of our model is TIVDM and TGM. To assess the contribution of each component, ablation studies were conducted in the internal and external testing dataset. As show in Table 1, the baseline model achieved AUCs of 0.784 and 0.773, with ACCs of 0.732 and 0.673 on the internal and external testing datasets, respectively. Incorporating TIVDM improved the AUC to 0.815 internally and 0.801 externally, indicating that explicit disentanglement of time-invariant and time-variant representations enhances longitudinal modeling capability. Similarly, introducing TGM led to consistent performance gains (internal AUC = 0.812; external AUC = 0.792), demonstrating the importance of dynamic temporal interval modulation. When both TIVDM and TGM were jointly applied, the model achieved the best performance, with AUCs increasing to 0.848 (internal) and 0.828 (external), and ACCs reaching 0.752 and 0.754, respectively. In terms of SEN and SPE, TIVDM primarily increased SEN, while TGM mainly improved SPE. Their joint integration achieved a better balance between SEN and SPE on both internal and external testing datasets. Compared with the baseline, the full model yielded substantial improvements across both cohorts. These results confirm the complementary and synergistic effects of the proposed modules and their contribution to improved generalization performance.

Effectiveness and Interpretability of TIVDM To further validate the effectiveness of TIVDM, we integrated it into three representative backbone architectures (3D ResNet-18 [21], 3D Vision Transformer (3D ViT) [22], and 3D ST [18]) and compared their performance before and after incorporation. As shown in Table 2, TIVDM consistently improved AUC across all models in both internal and external testing cohorts. Although slight fluctuations in ACC were observed in the internal testing cohort, ACC consistently improved across all models in the external testing cohort, indicating enhanced

cross-center robustness. In addition, TIVDM consistently improved SEN across all baseline architectures and generally maintained competitive SPE.

Table 1. Ablation study on model components with the AUC and ACC metric on internal testing set and external testing set.

3D ST (Baseline)	TIVDM	TGM	Internal Testing Dataset				External Testing Dataset			
			AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE
√			0.784	0.732	0.727	0.736	0.773	0.673	0.761	0.607
√	√		0.815	0.712	0.758	0.678	0.801	0.727	0.837	0.645
√		√	0.812	0.719	0.576	0.828	0.792	0.726	0.648	0.792
√	√	√	0.848	0.752	0.742	0.759	0.828	0.754	0.780	0.735

Table 2. Performance comparison of different backbone architectures with and without TIVDM on internal and external testing sets.

Model	Baseline	TIVDM	Internal Testing Dataset				External Testing Dataset			
			AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE
3D ResNet 18	√		0.781	0.719	0.667	0.759	0.730	0.684	0.692	0.678
3D ResNet 18	√	√	0.801	0.706	0.758	0.667	0.783	0.697	0.723	0.678
3D ViT	√		0.637	0.536	0.561	0.517	0.650	0.592	0.679	0.526
3D ViT	√	√	0.660	0.641	0.576	0.690	0.679	0.649	0.717	0.597
3D ST	√		0.784	0.732	0.727	0.736	0.773	0.673	0.761	0.607
3D ST	√	√	0.815	0.712	0.758	0.678	0.801	0.727	0.837	0.645

To explain this performance improvement, we visualized the feature distribution using t-distributed Stochastic Neighbor Embedding (t-SNE) [23] (Fig. 2). First, regarding subspace orthogonality (Fig. 2A, C), features without disentanglement appear mixed, whereas TIVDM successfully separates representations into two distinct clusters (temporal-invariant vs. temporal-variant), confirming effective subspace disentanglement. Second, regarding temporal dynamics (Fig. 2B, D), within the temporal-invariant subspace, pre- and post-treatment features are highly overlapped, indicating that the learned representations predominantly encode time-invariant components rather than dynamic treatment-induced changes. In contrast, the temporal-variant subspace shows a clear distribution shift between pre- and post-CT features. This confirms that our module can accurately disentangle treatment-induced dynamic and static information explicitly, providing interpretability for the performance improvement.

Comparison with other Methods We compared our method with several representative longitudinal modeling approaches, including LC-NICER [14], M2Fusion [15], TRL [16], and HARM^3 [17] (Table 3). For fair comparison, only the imaging branches of multimodal models (M2Fusion and HARM^3) were adopted, and the deep learning branch of LC-NICER was used for evaluation. As shown in Table 3, our method

achieved the best performance across both internal and external testing cohorts. Specifically, our model attained an AUC of 0.848 on the internal dataset and 0.828 on the external dataset, outperforming the second-best method (M2Fusion) by 2.6% and 5.9%, respectively. Notably, the performance gap was more pronounced in the external cohort, where other methods demonstrated reduced generalization stability, while our approach maintained stable accuracy (ACC=0.754). Moreover, TCFD-Net consistently achieved superior SEN and SPE, reaching SEN values of 0.742 and 0.780 and SPE values of 0.759 and 0.735 in the internal and external cohorts, respectively, indicating more reliable identification. This robust generalization capability confirms that the TIVDM and TGM are more effective in capturing deeply treatment-induced dynamics than traditional weight-sharing or implicit fusion paradigms.

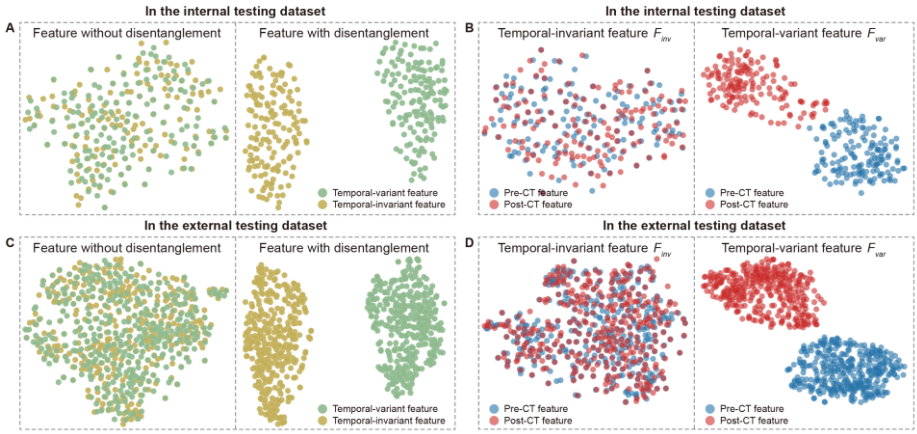

Fig. 2. The t-SNE visualization of feature distributions on internal and external testing datasets. **A, C:** Comparison of features with and without disentanglement in the internal and external testing dataset. **B, D:** Pre- and post-CT features visualized in temporal-invariant and variant in the internal and external testing dataset.

Table 3. Comparison with other methods. *, **, and *** represent $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively, in DeLong tests comparing AUCs with TCFD-Net.

Model	Internal Testing Dataset				External Testing Dataset			
	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE
LC-NICER	0.809	0.726	0.697	0.747	0.730***	0.649	0.679	0.626
M2Fusion	0.822	0.719	0.712	0.724	0.769**	0.681	0.755	0.626
TRL	0.723**	0.667	0.561	0.747	0.697***	0.651	0.579	0.706
HARM^{3}	0.715***	0.647	0.576	0.701	0.725***	0.654	0.679	0.635
TCFD-Net	0.848	0.752	0.742	0.759	0.828	0.754	0.780	0.735

4 Conclusions

In this work, we present a novel longitudinal CT analysis framework for immunotherapy response prediction, termed TCFD-Net. The proposed model explicitly disentangles time-invariant representations from time-variant representations via a TIVDM, structurally isolating stable features from treatment-induced dynamic signals under joint supervisory constraints. This effectively prevents dominant static features from suppressing these subtle temporal changes that are most predictive of immunotherapy response. To further purify the disentangled temporal-variant subspace, a TGM adaptively calibrates the temporal-variant representations according to inter-scan intervals, removing interval-dependent confounding and ensuring that the temporal-variant subspace faithfully captures true biological treatment response rather than acquisition timing artifacts. Multi-center validation demonstrates that TCFD-Net achieves superior and generalizable performance over existing longitudinal modeling approaches, highlighting the clinical value of explicit feature disentanglement coupled with interval-aware calibration for reliable immunotherapy response prediction. The proposed disentanglement-based longitudinal modeling paradigm may offer a generalizable foundation for broader applications in treatment response assessment beyond immunotherapy. Several limitations merit consideration. Performance was evaluated using a single train–validation–test split without resampling. Furthermore, multimodal integration, prospective validation, and generalizability to other tumor types require further study.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Mountzios, G., Remon, J., Hendriks, L.E.L., et al.: Immune-checkpoint inhibition for resectable non-small-cell lung cancer—opportunities and challenges. Nat. Rev. Clin. Oncol. **20**(10), 664–677 (2023)
2. Ponvilawan, B., Bansal, D., Abdel Karim, N., et al.: Long-Term Survival by Number of Immune Checkpoint Inhibitors in PD-L1 – Negative Metastatic NSCLC: A Systematic Review and Meta-Analysis. JAMA Netw. Open **8**(2), e2457357 (2025)
3. Zhou F, Qiao M, Zhou C.: The cutting-edge progress of immune-checkpoint blockade in lung cancer. Cell. Mol. Immunol **18**(2), 279–293 (2021)
4. Reck, M., Frost, N., Peters, S., et al.: Treatment of NSCLC after chemoimmunotherapy—are we making headway? Nat. Rev. Clin. Oncol. **22**(11), 806–830 (2025)
5. Tang, L.B., Peng, Y.L., Chen, J., et al.: Rechallenge with immune-checkpoint inhibitors in patients with advanced-stage lung cancer. Nat. Rev. Clin. Oncol. **22**(8), 546–565 (2025)
6. Sun, L., Bleiberg, B., Hwang, W.T., et al.: Association between duration of immunotherapy and overall survival in advanced non – small cell lung cancer. JAMA Oncol. **9**(8), 1075–1082 (2023)
7. Zhao, Y., He, Y., Wang, W., et al.: Efficacy and safety of immune checkpoint inhibitors for individuals with advanced EGFR-mutated non-small-cell lung cancer who progressed on EGFR tyrosine-kinase inhibitors: a systematic review, meta-analysis, and network meta-analysis. Lancet Oncol. **25**(10), 1347–1356 (2024)

8. Wu, Y., Zhang, W., Liang, X., et al.: Habitat radiomics analysis for progress free survival and immune-related adverse reaction prediction in non-small cell lung cancer treated by immunotherapy. J. Transl. Med. **23**(1), 393 (2025)
9. Saad, M.B., Hong, L., Aminu, M., 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 **5**(7), e404–e420 (2023)
10. Wu, S., Zhan, W., Liu, L., et al.: Pretreatment radiomic biomarker for immunotherapy responder prediction in stage IB – IV NSCLC (LCDigital-IO Study): a multicenter retrospective study. J. Immunother. Cancer **11**(10), e007369 (2023)
11. Gong, J., et al.: A short-term follow-up CT based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer. Oncoimmunology **11**(1), 2028962 (2022)
12. Farina, B., Guerra, A.D.R., Bermejo-Peláez, D., et al.: Integration of longitudinal deep-radiomics and clinical data improves the prediction of durable benefits to anti-PD-1/PD-L1 immunotherapy in advanced NSCLC patients. J. Transl. Med. **21**(1), 174 (2023)
13. Sako, C., Kurland, B.F., Schmidt, T.G., et al.: Deep-learning serial ct prediction of survival in immunotherapy-treated non – small cell lung cancer. JAMA Netw. Open **9**(1), e2555759 (2026)
14. Ye, G., Wei, Z., Han, C., et al.: AI-derived longitudinal and multi-dimensional CT classifier for non-small cell lung cancer to optimize neoadjuvant chemoimmunotherapy decision: a multicentre retrospective study. EClinicalMedicine **89**, 1–12 (2025)
15. Zhang, S., Du, S., Sun, C., et al.: M2Fusion: multi-time multimodal fusion for prediction of pathological complete response in breast cancer. In: Linguraru, M.G., Dou, Q., Feragen, A., et al. (eds.) Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS, vol. 15005, pp. 458–468. Springer, Cham (2024)
16. Janíčková, I., Tan, Y.Y., Helbich, T.H., et al.: Temporal Representation Learning of Phenotype Trajectories for pCR Prediction in Breast Cancer. In: Gee, J.C., Alexander, D.C., Hong, J., et al. (eds.) Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS, vol. 15974, pp. 606–615. Springer, Cham (2025)
17. Wang, J., Liu, X., Gong, Z., et al.: HARM3-Fusion: Hierarchical Attentional Representation Learning of Multi-modal, Multi-timer, Multi-sequence Fusion for Pathological Complete Response Prediction of Head and Neck Squamous Cell Carcinoma. In: Gee, J.C., Alexander, D.C., Hong, J., et al. (eds.) Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS, vol. 15969, pp. 246–255. Springer, Cham (2025)
18. Liu, Z., Lin, Y., Cao, Y., et al.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022. IEEE, Los Alamitos (2021)
19. Eisenhauer E A, Therasse P, Bogaerts J, et al.: New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur. J. Cancer **45**(2), 228–247 (2009)
20. Lin, C., Tan, W., Zhou, Z., et al.: ScreenLungNet: Personalized Long-Term Prediction of Lung Cancer Risk from a Single Low-Dose CT Screening. Radiology **317**(1), e251310 (2025)
21. Zhou, M., Li, M., Cao, Q., et al.: Malignant pleural mesothelioma classification and survival prediction with CT imaging using ResNet. Eur. Radiol. **25**(1), 1–12 (2025)
22. Ranftl, R., Bochkovskiy, A., Koltun, V.: Vision transformers for dense prediction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 12179–12188. IEEE, Los Alamitos (2021)

23. Qiao, J., Jin, J., Wang, D., et al.: A self-conformation-aware pre-training framework for molecular property prediction with substructure interpretability. Nat. Commun. **16**(1), 4382 (2025)