TeDyS: Temporal Dynamics with Spatial Context for Longitudinal Progression Prediction

Jeonghyun Hwang*[1], Nur Syamiela Roslan*[2], Jisoo Jeong[1], Sang Myung Woo[3], Thanaporn Viriyasaranon[4](✉), and Jang-Hwan Choi[1,4](✉)

[1] Division of Artificial Intelligence and Software, Ewha Womans University, Seoul, Korea
[2] Division of Mechanical and Biomedical Engineering, Ewha Womans University, Seoul, Korea
[3] Center for Liver and Pancreatobiliary Cancer, National Cancer Center, Goyang, Korea
[4] Department of Artificial Intelligence, Ewha Womans University, Seoul, Korea
{jeonghyunhwang,nursyamiela,thanaporn.v09,choij}@ewha.ac.kr

Abstract. Accurate prediction of disease progression from longitudinal medical imaging is critical for personalized treatment planning, yet remains challenging due to subtle progression cues, scan-to-scan variability, and uneven visit timing. Existing methods often model either temporal consistency across visits or temporal change signals, but rarely capture both in a unified progression framework. We propose TeDyS, a temporal dynamics–conditioned query framework that jointly models short-term and long-term progression patterns under variable inter-visit spacing. TeDyS maintains a set of temporal queries that are conditioned on explicitly decomposed temporal dynamics, including short-term aligned residual changes and long-term baseline-referenced trends. This design enables robust modeling of both fine-grained progression signals and long-term disease trends within a compact latent representation. We evaluate the proposed framework on an in-house longitudinal pancreatic CT cohort and a public longitudinal brain MRI dataset, assessing its performance across different organs and imaging modalities under irregular real-world follow-up. Code will be released at `https://github.com/Ewha-AI/TeDyS`.

Keywords: Longitudinal Imaging · Disease Progression · Temporal Dynamics · Medical Image Analysis

1 Introduction

Longitudinal imaging is commonly acquired during routine follow-up, and clinically meaningful changes often emerge across repeated visits rather than at a single time point [23, 3, 6]. Such serial scans support response assessment and

* Equally contributed.

follow-up management by tracking disease progression: how lesions and surrounding anatomy change across visits over time [24, 3]. Accordingly, progression prediction from multiple longitudinal scans is a key task directly tied to patient management and treatment decision-making [22, 12, 7]. However, reliable progression prediction from longitudinal scans remains challenging, as progression cues are often faint, acquisitions vary across visits, and follow-up intervals are irregular [7, 19, 14]. For progression prediction, both temporal context and meaningful change are essential. Context without meaningful change can miss progression signals, while apparent change without the right context may reflect mismatched regions rather than true progression. Motivated by recent progress in deep learning, automated models have been increasingly explored for progression prediction from serial CT/MRI [21, 7, 4], particularly in cancer and other chronic diseases where progression can be subtle and heterogeneous.

With strong per-scan representations now widely available, existing methods typically adopt one of two practical strategies to incorporate temporal information. A common approach performs explicit comparisons between a limited number of visits, often through pairwise pre–post analysis or difference-based summaries [26, 21]. Many such approaches focus on two time points (e.g., pre- vs. post-treatment) [10, 29, 21], which can capture treatment response but does not naturally extend to multi-visit follow-up, especially when interventions occur between visits. In addition, pairwise comparisons can fragment temporal evolution and limit modeling of cumulative effects or intermediate progression patterns under irregular follow-up [17, 8]. They can also be sensitive to spatial mismatch when lesion appearance, morphology, or extent evolves over time.

An alternative strategy is to model the entire sequence by aggregating visit-wise representations with recurrent, sequence-based, or attention-driven architectures [7, 15, 16]. These models accommodate variable-length follow-up and irregular timing, but temporal change is often encoded implicitly in hidden-state updates or pooled features. As a result, short-term changes between adjacent scans and longer-term progression trends can become entangled, making progression signals harder to distinguish from acquisition-driven variation [4, 20–22].

To address these limitations, we propose **TeDyS**, a **Te**mporal **Dy**namics with **S**patial context framework for longitudinal progression prediction. Following recent advances in query-based temporal state modeling for sequential vision [13, 27, 5, 25, 28], TeDyS models progression via dynamics-conditioned temporal query updates tailored to irregular longitudinal scans. At each visit, spatial features are converted into slot-based latent tokens, and temporal dynamics are explicitly decomposed into two complementary components: *short-term dynamics (STD)*, capturing aligned adjacent-visit residual changes, and *long-term dynamics (LTD)*, modeling baseline-referenced deviations conditioned on elapsed time. The fused dynamics tokens guide query updates across visits, enabling compact accumulation of progression evidence while preserving sensitivity to fine-grained cues and long-range progression trends under irregular follow-up. Compared with conventional temporal aggregation, TeDyS provides (i) an ex-

plicit decomposition of longitudinal dynamics instead of implicit temporal mixing in a generic sequence encoder, and (ii) a query-based temporal state representation that supports structured evidence accumulation with visit-specific spatial context. We evaluate TeDyS on two distinct longitudinal imaging settings: an in-house pancreatic CT cohort and the public OASIS-2 longitudinal brain MRI benchmark [18]. Across these heterogeneous organs and modalities, TeDyS consistently achieves strong performance, with notable gains in recall and AUC, suggesting improved sensitivity to progression-related changes under irregular real-world follow-up.

Our main contributions are as follows:

- We propose TeDyS, a dynamics-conditioned temporal query framework for longitudinal progression prediction under irregular follow-up.
- TeDyS explicitly decomposes longitudinal dynamics into short-term inter-visit changes and long-term baseline-referenced trends.
- We validate TeDyS on in-house pancreatic CT and public OASIS-2 brain MRI, showing robust performance across modality and anatomy.

2 Methods

2.1 Overview

Given a longitudinal sequence of visits $\{I_t\}_{t=1}^{T}$ with elapsed times $\{\tau_t\}$, we maintain a shared set of temporal queries that are updated at each visit using spatial slot tokens X_t and temporal dynamics tokens D_t. The updated queries across visits are aggregated to produce a sequence-level representation $\mathcal{B}$ for prediction, as illustrated in Fig. 1 (a).

Feature Extraction. Spatial features are extracted using frozen nnUNet [9] encoders pretrained for segmentation on PANORAMA [1] (pancreatic CT) and MSD Hippocampus [2] (OASIS-2 MRI). The encoders are used solely for feature extraction and are not fine-tuned during progression prediction training.

For each visit t, we obtain a set of spatial tokens $Z_t \in \mathbb{R}^{S\times d}$ (flattened over spatial locations and projected to width d) from the encoder features. When enabled, we summarize Z_t into a fixed-size set of slot tokens using Slot Attention:

$$X_t = \mathrm{SlotAttn}(Z_t) = \{x_t^i\}_{i=1}^{N} \in \mathbb{R}^{N\times d}, \tag{1}$$

where $\mathrm{SlotAttn}(\cdot)$ iteratively refines N learnable slots via cross-attention to the S spatial tokens.

2.2 Temporal Dynamics

Short-term dynamics. To account for possible permutation of slots across visits, for $t = 2, \ldots, T$ we align tokens between adjacent visits $(t-1 \to t)$ using entropic optimal transport (OT) (Fig. 1(b)). The matching cost is defined as

$$C_{ij} = \left\| x_{t-1}^i - x_t^j \right\|_2^2. \tag{2}$$

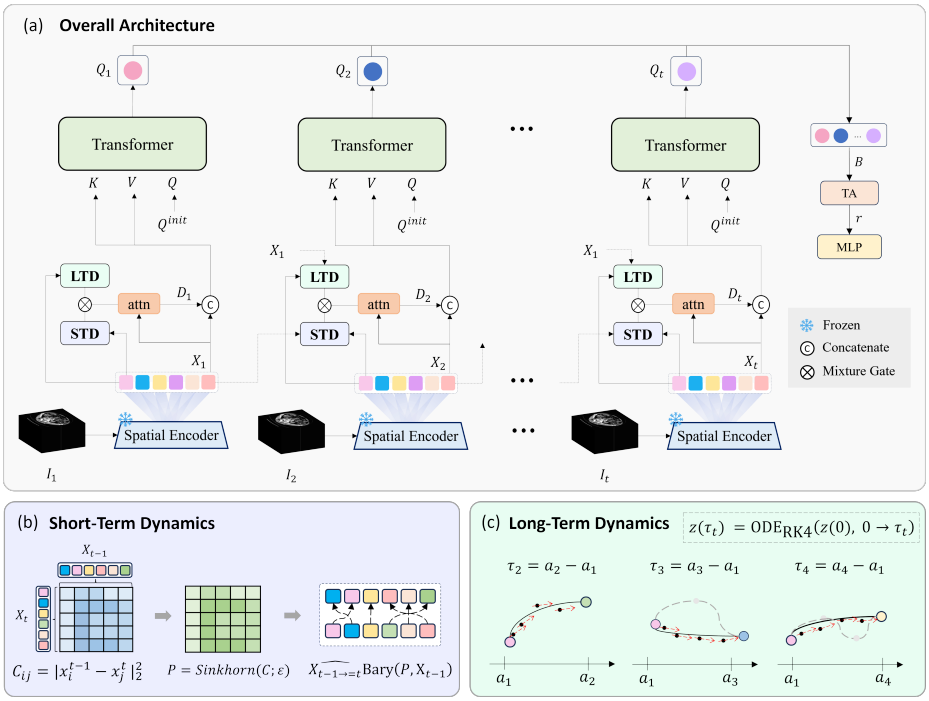

Fig. 1. Overview of TeDyS. (a) Query updates, with (b) adjacent Short-Term Dynamics and (c) baseline-anchored Long-Term Dynamics.

We compute a transport plan P_t via Sinkhorn scaling and obtain aligned previous-visit tokens by barycentric projection:

$$\tilde{x}^j_{t-1\to t} = \frac{\sum_i P_{t,ij}\, x^i_{t-1}}{\sum_i P_{t,ij}}, \quad \tilde{X}_{t-1\to t} \in \mathbb{R}^{N\times d}. \tag{3}$$

The short-term residual is

$$\Delta^{\mathrm{STD}}_t = X_t - \tilde{X}_{t-1\to t}, \quad t \geq 2, \tag{4}$$

and for the baseline visit ($t = 1$), no previous scan is available; we set $\Delta^{\mathrm{STD}}_1 = \mathbf{0}$.

Long-term dynamics. Let $\{a_t\}_{t=1}^T$ denote the acquisition times of visits, with baseline at a_1. We define the elapsed time since baseline as $\tau_t = a_t - a_1$ with $\tau_1 = 0$ (Fig. 1(c)). Given baseline visit tokens X_1, we obtain a patient-level initial state $z(0) = \bar{X}_1$ by mean pooling across the N tokens. We model the baseline-to-current trajectory with a Neural ODE $\frac{dz(\tau)}{d\tau} = f_\theta(z(\tau), \tau)$ and numerically integrate it from $\tau = 0$ to $\tau = \tau_t$ to obtain $z(\tau_t)$. From the resulting state change, we compute a global trend shift

$$\delta_t = W_s\big(z(\tau_t) - z(0)\big).$$

We then form a baseline-referenced token prediction by broadcasting the shift to all tokens:

$$\hat{X}_{t|1} = X_1 + \mathbf{1}\,\delta_t^\top \in \mathbb{R}^{N\times d}, \tag{5}$$

and define the long-term residual as

$$\Delta_t^{\mathrm{LTD}} = X_t - \hat{X}_{t|1}. \tag{6}$$

For the baseline visit ($t = 1$), $\tau_1 = 0$ implies $\hat{X}_{1|1} = X_1$, hence $\Delta_1^{\mathrm{LTD}} = \mathbf{0}$.

Fusion of temporal dynamics. We fuse short- and long-term residuals using a sample-wise gated mixture. Let $\bar{(\cdot)}$ denote mean pooling over tokens, Δt_t the interval between visits $t-1$ and t, $e(\cdot)$ a learned time embedding, and $g(\cdot)$ a learnable mixture gate.:

$$[w_t^{\mathrm{STD}}, w_t^{\mathrm{LTD}}] = \mathrm{softmax}\big(g(\bar{X}_t, \overline{\Delta_t^{\mathrm{STD}}}, \overline{\Delta_t^{\mathrm{LTD}}}, e(\Delta t_t), e(\tau_t))\big), \tag{7}$$

$$\Delta_t^{\mathrm{Mix}} = w_t^{\mathrm{STD}}\Delta_t^{\mathrm{STD}} + w_t^{\mathrm{LTD}}\Delta_t^{\mathrm{LTD}}. \tag{8}$$

$\psi(\cdot)$ denotes a projection network applied before attention, and R is a set of learnable readout tokens shared across visits. We summarize the residual-enhanced tokens into compact dynamics tokens

$$D_t = \mathrm{LN}\big(R + \mathrm{MHA}\big(R, \psi([X_t; \Delta_t^{\mathrm{Mix}}]), \psi([X_t; \Delta_t^{\mathrm{Mix}}])\big)\big) \in \mathbb{R}^{K_d\times d}. \tag{9}$$

2.3 Query Update and Longitudinal Aggregation

We use shared learnable query embeddings $Q^{\mathrm{init}} \in \mathbb{R}^{K\times d}$ for all visits. At each visit,

$$Q_t = \mathrm{Transformer}\big(Q^{\mathrm{init}}, [X_t; D_t]\big) \in \mathbb{R}^{K\times d}. \tag{10}$$

After processing all visits, the updated queries are aggregated via a temporal attention pooling module to produce the final representation $r \in \mathbb{R}^d$ for progression prediction.

3 Experiments

3.1 Datasets

We evaluate the proposed framework on an in-house longitudinal pancreatic CT cohort and a public longitudinal brain MRI dataset to assess cross-organ and cross-modality generalization.

In-house longitudinal pancreatic CT cohort. A retrospective longitudinal cohort of pancreatic cancer cases, provided by the National Cancer Center, Korea, with serial contrast-enhanced abdominal CT examinations acquired between 2014 and 2023 was analyzed. The cohort comprised 113 cases, with a mean of 5.1 examinations per case (range 2–24) and a mean inter-visit interval of 2.3 months. For each case, time-to-progression was discretized into four intervals ($\leq$6, 6–12, 12–18, $\geq$18 months), with class distribution of 24 (21.2%), 33 (29.2%), 18 (15.9%), and 38 (33.6%) cases, respectively. The model takes the truncated pre-event visit sequence as input and predicts the discretized time-to-first-event interval measured from the baseline scan. The endpoint was defined as the earliest occurrence of clinical failure, including radiologic progression, recurrence, or death. Cases without documented failure were assigned to the $\geq$18-month interval only when follow-up exceeded 18 months. The study protocol was approved by the institutional review board with waiver of informed consent.

OASIS-2 longitudinal MRI dataset. We additionally evaluated our method on the OASIS-2 longitudinal brain MRI dataset, which contains 150 subjects with at least two longitudinal scans [18]. A total of 373 scans were included, with a mean of 2.49 scans per subject and a median follow-up of 21.8 months. Progression was defined using the Clinical Dementia Rating (CDR); the first visit with CDR > 0 was considered the progression event. Each subject was assigned a binary label indicating whether progression occurred during follow-up.

3.2 Experimental Setup

All volumes were standardized before model input using resampling, intensity normalization, and automatic foreground cropping [9].

All models were implemented in PyTorch and trained on an NVIDIA RTX A6000 GPU. Models were trained with AdamW for 50 epochs (batch size 8, learning rate 5×10^{-6}, weight decay 1×10^{-2}) using cross-entropy loss. Results are reported with five-fold cross-validation using patient-level non-overlapping splits. For OASIS-2, the five folds contained 29–31 subjects each and preserved the binary progression-label distribution (prog=0/prog=1: 70/80 overall; fold-wise prog=1 counts: 14–19), with the positive class including both baseline-demented subjects and converters during follow-up (14 converters total; fold-wise counts: 1–5). We report Accuracy, Precision, Recall, F1-score, and area under the ROC curve (AUC), summarized as mean $\pm$ standard deviation across folds.

3.3 Results and Analysis

Quantitative Results on the pancreatic CT cohort Table 1 reports the results on the pancreatic CT cohort. TeDyS achieves the best performance across all metrics among the longitudinal baselines. Notably, TeDyS achieves the highest recall (**0.6385**), reflecting its improved ability to capture progression-related signals. This gain is achieved while maintaining competitive precision (0.5676), resulting in the best F1-score (**0.5848**) and AUC (**0.8065**). Given the challenges

Table 1. Quantitative results on the in-house longitudinal pancreatic CT cohort. Best and second-best results are marked in **bold** and underlined, respectively.

Method	Accuracy	Precision	Recall	F1-score	AUC
Lomia-T [21]	0.5043±0.065	0.4144±0.169	0.5000±0.064	0.4145±0.086	0.3812±0.048
T-LSTM [16]	0.5218±0.067	0.5430±0.145	0.5063±0.077	0.4768±0.112	0.7007±0.045
D-LSTM [11]	0.5210±0.032	0.5445±0.171	0.4831±0.039	0.4537±0.070	0.6915±0.086
M-LSTM [15]	0.4949±0.050	0.3550±0.085	0.4130±0.045	0.3550±0.057	0.7068±0.031
TeDyS	**0.6891**±0.068	**0.5676**±0.111	**0.6385**±0.048	**0.5848**±0.082	**0.8065**±0.058

Table 2. Quantitative results on the OASIS-2 longitudinal MRI dataset. Best and second-best results are marked in **bold** and underlined, respectively.

Method	Accuracy	Precision	Recall	F1-score	AUC
LOMIA-T [21]	0.6533±0.029	0.6046±0.158	0.6216±0.071	0.5921±0.113	0.5907±0.024
T-LSTM [16]	0.7199±0.045	0.7346±0.031	0.7075±0.038	0.7032±0.046	0.7389±0.076
D-LSTM [11]	0.7538±0.055	**0.7976**±0.051	0.7336±0.084	0.7220±0.091	0.7200±0.117
M-LSTM [15]	0.7462±0.048	0.7633±0.029	0.7297±0.046	0.7277±0.056	0.6780±0.059
TeDyS	**0.7809**±0.073	0.7966±0.077	**0.7686**±0.065	**0.7707**±0.075	**0.8133**±0.092

of longitudinal pancreatic tumor progression modeling under irregular follow-up intervals, these results highlight the effectiveness of the proposed dynamics-conditioned query formulation in capturing progression-related signals.

Quantitative Results on OASIS-2 Longitudinal MRI Table 2 summarizes results on OASIS-2 (same baselines as Table 1). Despite a modality and anatomical shift from abdominal CT to brain MRI, TeDyS achieves the best overall performance. Notably, improvements in recall (0.7686) and F1-score (0.7707) indicate enhanced sensitivity to progression-related changes while maintaining competitive precision.

ROC Analysis. Figure 2 presents ROC curves on the pancreatic CT cohort and the OASIS-2 longitudinal MRI dataset. TeDyS achieves the highest AUC on both datasets, with consistently strong ROC profiles compared to the longitudinal baselines. These results suggest that the proposed approach is broadly applicable across different organs and imaging modalities, performing well on both abdominal CT and brain MRI.

Ablation Study. Table 3 summarizes component ablations on the in-house pancreatic CT cohort. Both STD and LTD improve the autoregressive query-update baseline; LTD yields the dominant gain (AUC 0.801 vs 0.517), while STD gives a smaller improvement (AUC 0.557). Combining STD+LTD improves Acc/F1 (0.663/0.539), and adding slot-based fusion achieves the best Acc/F1

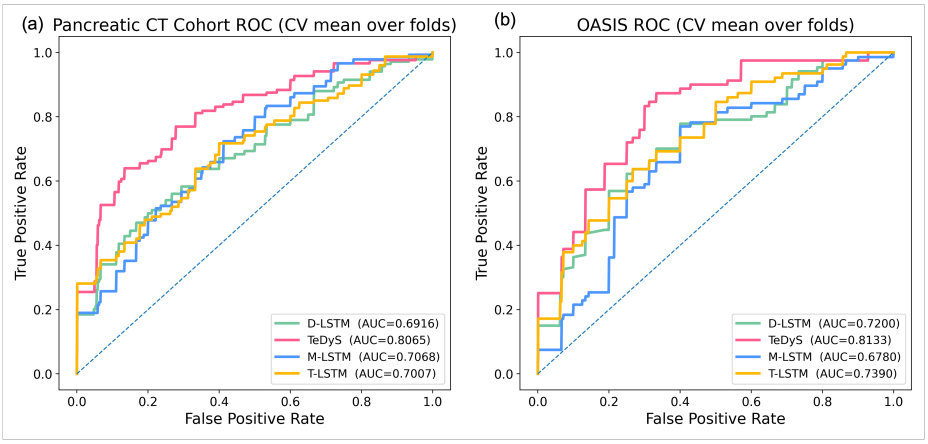

Fig. 2. ROC curves (five-fold cross-validation mean) on (a) the pancreatic CT cohort and (b) the OASIS-2 MRI dataset.

Table 3. Ablation study on the in-house longitudinal pancreatic CT cohort. STD/LTD: short-/long-term dynamics; *Baseline* performs an autoregressive temporal-query update across visits, where queries from the previous visit are used to update those at the next visit. *STD only* and *LTD only* add STD or LTD to the baseline, respectively, and *w/o slot* removes slot attention from our full model.

Method	**Components**			**Results**				
	Slot	STD	LTD	Acc	Prec	Rec	F1	AUC
Baseline	✗	✗	✗	0.4165 ±0.027	0.2073 ±0.060	0.3202 ±0.040	0.2288 ±0.049	0.5165 ±0.075
STD only	✗	✓	✗	0.4960 ±0.079	0.4340 ±0.113	0.4073 ±0.092	0.3602 ±0.108	0.5572 ±0.097
LTD only	✗	✗	✓	0.6535 ±0.071	<u>0.5424</u> ±0.093	0.6099 ±0.051	<u>0.5537</u> ±0.071	<u>0.8012</u> ±0.036
w/o slot	✗	✓	✓	<u>0.6626</u> ±0.070	0.5386 ±0.087	<u>0.6117</u> ±0.053	0.5386 ±0.078	0.7920 ±0.043
Ours	✓	✓	✓	**0.6891** ±0.068	**0.5676** ±0.111	**0.6385** ±0.048	**0.5848** ±0.082	**0.8065** ±0.058

with a slight AUC increase (0.689/0.585/0.807), indicating added benefit when jointly leveraging both dynamics.

4 Conclusion

We presented TeDyS, a dynamics-conditioned query framework for longitudinal progression prediction. By explicitly decomposing temporal change into short-

term residual alignment and long-term baseline-referenced evolution and integrating them through staged fusion and query updates, the design enables stable modeling of longitudinal sequences with non-uniform visit intervals.

Experiments on longitudinal CT and MRI cohorts demonstrate consistent improvements across metrics, with particularly strong recall, underscoring improved sensitivity to progression events. Ablation studies further validate the effectiveness of the proposed temporal modeling strategy.

These results demonstrate that explicitly modeling longitudinal dynamics offers a principled alternative to uniform sequence updates for progression prediction. A remaining limitation is that the OASIS-2 dataset includes a small number of conversion cases, motivating further validation on cohorts with more converters to better assess longitudinal conversion trajectories. Clinically, the proposed framework supports earlier identification of high-risk patients from routine follow-up imaging, enabling more informed risk stratification and treatment planning in longitudinal care.

Acknowledgments. This work was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korean government (MSIT) (Nos. RS-2025-02215813, RS-2025-00520578, and RS-2025-02634603) and by an NRF grant funded by the Korean government (MSIT and MOE) (No. RS-2025-16063688).

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

References

1. Alves, N., Schuurmans, M., Rutkowski, D., Yakar, D., Haldorsen, I., Liedenbaum, M., Molven, A., Vendittelli, P., Litjens, G., Hermans, J., et al.: The panorama study protocol: pancreatic cancer diagnosis-radiologists meet ai. Zenodo (2024)
2. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
3. Aslam, A., Chernyak, V., Tang, A., Miller, F.H., et al.: Ct/mri li-rads 2024 update: treatment response assessment. Radiology **313**(2), e232408 (2024)
4. Chen, Y., Wolterink, J.M., Neve, O.M., Romeijn, S.R., et al.: Vestibular schwannoma growth prediction from longitudinal mri by time-conditioned neural fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 508–518. Springer (2024)
5. Cheng, H.K., Oh, S.W., Price, B., Lee, J.Y., Schwing, A.: Putting the object back into video object segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3151–3161 (2024)
6. Englman, C., Maffei, D., Allen, C., Kirkham, A., et al.: Precise version 2: updated recommendations for reporting prostate magnetic resonance imaging in patients on active surveillance for prostate cancer. European urology **86**(3), 240–255 (2024)
7. Holste, G., Lin, M., Zhou, R., Wang, F., et al.: Harnessing the power of longitudinal medical imaging for eye disease prognosis using transformer-based sequence modeling. NPJ Digital Medicine **7**(1), 216 (2024)

8. Im, J.E., Khalifa, M., Gregory, A.V., Erickson, B.J., Kline, T.L.: A systematic review on the use of registration-based change tracking methods in longitudinal radiological images. Journal of Imaging Informatics in Medicine **38**(4), 2549–2562 (2025)
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
10. Jin, C., Yu, H., Ke, J., Ding, P., et al.: Predicting treatment response from longitudinal images using multi-task deep learning. Nature communications **12**(1), 1851 (2021)
11. Jomeiri, A., Navin, A.H., Shamsi, M.: Longitudinal mri analysis using a hybrid densenet-bilstm method for alzheimer's disease prediction. Behavioural Brain Research **463**, 114900 (2024)
12. Konwer, A., Xu, X., Bae, J., Chen, C., Prasanna, P.: Temporal context matters: Enhancing single image prediction with disease progression representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18824–18835 (2022)
13. Lee, S., Seo, J., Han, K., Choi, M., Im, S.: Cavis: Context-aware video instance segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4507–4517 (2025)
14. Li, T.Z., Xu, K., Gao, R., Tang, Y., et al.: Time-distance vision transformers in lung cancer diagnosis from longitudinal computed tomography. In: Medical Imaging 2023: Image Processing. vol. 12464, pp. 229–238. SPIE (2023)
15. Li, X., Xiao, H., Weng, W., Xu, X., Shi, Y.: Mpbd-lstm: A predictive model for colorectal liver metastases using time series multi-phase contrast-enhanced ct scans. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 379–388. Springer (2023)
16. Liu, X., Wang, M., Aftab, R.: Study on the prediction method of long-term benign and malignant pulmonary lesions based on lstm. Frontiers in Bioengineering and Biotechnology **10**, 791424 (2022)
17. Mao, Y., Zhang, Y., Zou, X., Zheng, Y., et al.: Asymmetric matching in abdominal lymph nodes of follow-up ct scans. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 35–44. Springer (2025)
18. Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults. Journal of cognitive neuroscience **22**(12), 2677–2684 (2010)
19. Montero, I., Sotoudeh-Paima, S., Abadi, E., Samei, E.: Intra-and inter-scanner ct variability and their impact on diagnostic tasks. In: Proceedings of SPIE–the International Society for Optical Engineering. vol. 13405, p. 134054C (2025)
20. Mouadden, N., Laousy, O., Marini, R., Ong, V., et al.: Conditional latent diffusion models for irregularly spaced longitudinal radiological data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 106–115. Springer (2025)
21. Sun, Y., Li, K., Chen, D., Hu, Y., Zhang, S.: Lomia-t: a transformer-based longitudinal medical image analysis framework for predicting treatment response of esophageal cancer. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 426–436. Springer (2024)
22. van Timmeren, J.E., Bussink, J., Koopmans, P., Smeenk, R.J., Monshouwer, R.: Longitudinal image data for outcome modeling. Clinical Oncology **38**, 103610 (2025)

23. Wilson, B.E., Wright, K., Koven, R., Booth, C.M.: Surveillance imaging after curative-intent treatment for cancer: benefits, harms, and evidence. Journal of Clinical Oncology **42**(19), 2245–2249 (2024)
24. Won, S.E., Suh, C.H., Kim, S., Park, H.J., Kim, K.W.: Summary of key points of the response assessment in neuro-oncology (rano) 2.0. Korean Journal of Radiology **25**(5), 407 (2024)
25. Wu, D., Wang, T., Zhang, Y., Zhang, X., Shen, J.: Onlinerefer: A simple online baseline for referring video object segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2761–2770 (2023)
26. Xia, Y., Zhou, J., Xun, X., Zhang, J., et al.: Ct-based multimodal deep learning for non-invasive overall survival prediction in advanced hepatocellular carcinoma patients treated with immunotherapy. Insights into Imaging **15**(1), 214 (2024)
27. Xie, J., Zhong, B., Mo, Z., Zhang, S., et al.: Autoregressive queries for adaptive tracking with spatio-temporal transformers. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19300–19309 (2024)
28. Yang, S., Wang, X., Li, Y., Fang, Y., et al.: Temporally efficient vision transformer for video instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2885–2895 (2022)
29. Ye, G., Wei, Z., Han, C., Wu, G., 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** (2025)