Attention Med3D: A 3D deep learning model based on transfer learning and attention mechanism for predicting mid-treatment chemoradiation response on FDG PET of LA-NSCLC

2026 JULY 19–22 VANCOUVER, BC
JOINT AAPM | COMP MEETING

Chunyan Duan[1], Jing Sun[1], Qiantuo Liu[1], Shijun Chen[1], Xiaojun Wu[1,*], Faisal Yaseen[2], Shouyi Wang[3], Daniel S. Hippe[4], Stephen R. Bowen[2,4]
[1]Tongji University, Shanghai, China, [2]University of Washington, Seattle, WA, USA,
[3]University of Texas, Arlington, TX, USA , [4]Fred Hutchinson Cancer Center, Seattle, WA, USA

INTRODUCTION

❖Accurate assessment of early radiotherapy response in tumors provides crucial guidance for optimizing radiotherapy protocols. FDG PET/CT plays a pivotal role in radiotherapy treatment response evaluation for lung cancer [1, 2].

❖Current radiotherapy response prediction relies on radiomics and 2D deep learning, which underuse tumors' 3D spatial heterogeneity [3]. 3D voxel-level deep learning models are needed to improve prediction.

❖3D deep learning models suffer from overfitting on scarce 3D medical data, which data augmentation and transfer learning such as Med3D pre-trained networks can mitigate [4,5].

❖Since tumors exhibit spatial heterogeneity with region-specific radiotherapy responses, attention mechanisms are needed to adaptively highlight critical tumor regions and improve prediction [6, 7, 8].

REFERENCES

1. S. Sheikhbahaei, E. Mena, A. Yanamadala, S. Reddy, L.B. Solnes, J. Wachsmann, R.M. Subramaniam, The value of FDG PET/CT in treatment response assessment, follow-up, and surveillance of lung cancer, Am. J. Roentgenol. 208(2) (2017) 420-433.
2. Ayeni A, Evbuomwan O. The Role of [18F] FDG PET/CT in Monitoring of Therapy Response in Lung Cancer, Semin. Nucl. Med. 55(2) (2025) 175-189.
3. J. Zhu, B. Xu, T. Fan, S. Ji, K. Gu, J. Ding, H. Lu, J. Ma, Y. Zhou, Sub-regional radiomics combining multichannel 2-dimensional or 3-dimensional deep learning for predicting neoadjuvant chemo-immunotherapy response in esophageal squamous cell carcinoma: a multicenter study, NPJ Precis. Oncol. 9(1) (2025) 248.
4. G. Ayana, K. Dese, A.M. Abagaro, K.C. Jeong, S.-D. Yoon, S.-w. Choe, Multistage transfer learning for medical images, Artif. Intell. Rev. 57(9) (2024) 232.
5. R. Schäfer, T. Nicke, H. Höfener, A. Lange, D. Merhof, F. Feuerhake, V. Schulz, J. Lotz, F. Kiessling, Overcoming data scarcity in biomedical imaging with a foundational multi-task model, Nat. Comput. Sci. 4(7) (2024) 495-509.
6. I. Dagogo-Jack, A.T. Shaw, Tumour heterogeneity and resistance to cancer therapies, Nat. Rev. Clin. Oncol. 15(2) (2018) 81-94.
7. J.P. O'Connor, C.J. Rose, J.C. Waterton, R.A. Carano, G.J. Parker, A. Jackson, Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome, Clin. Cancer Res. 21(2) (2015) 249-257.
8. M. Hassanin, S. Anwar, I. Radwan, F.S. Khan, A. Mian, Visual attention methods in deep learning: An in-depth survey, Inf. Fusion 108 (2024) 102417.

Faisal Yaseen
Department of Radiation Oncology, School of Medicine, University of Washington
1959 North East Pacific Street, Seattle, WA 98195, United States of America
fyaseen@uw.edu
[phone]

METHODS AND MATERIALS

FDG-PET/CT imaging

- **Data**: Twenty-five patients with locally advanced non-small cell lung cancer (LA-NSCLC) underwent baseline FDG-PET/CT imaging (PETpre) and repeat imaging during week 3 (PETmid) of chemoradiotherapy under the FLARE-RT trial (NCT02773238).
- **PETpre**: median 1 week prior, range < 1 month prior to 1st radiotherapy fraction.
- **PETmid**: median 24 Gy in 12 fractions, range 20-26 Gy in 10-13 fractions.
- Standardized acquisition and reconstruction imaging protocol using cross-calibrated PET/CT scanners.
- Primary metabolic tumor volumes (MTV) semi-automatically defined by PET EdgeTM (MIM Software, Cleveland OH).
- Rigid alignment of primary MTV between planning CT, pre-RT PET/CT, mid-RT PET/CT, and RT dose.
- Co-registered PET voxel grids resampled to isotropic 2 mm dimensions.

Data Processing (Figure 1)

I. ROI cropping: The tumor region plus a surrounding margin was extracted from PETpre, PETmid, and Dose, with non-tumor areas masked to zero to focus on the tumor and reduce background interference.

II. Data augmentation: Training samples were expanded via rotation (±30° around x/y/z axes), flipping (along x/y/z axes), and random scaling (0.9-1.1), applied identically to PETpre, PETmid, and Dose.

III. Channel-merge of PETpre and Dose: PETpre and Dose were merged as model inputs and PETmid as the prediction target, with merging performed after augmentation for the training set but directly for the test set.

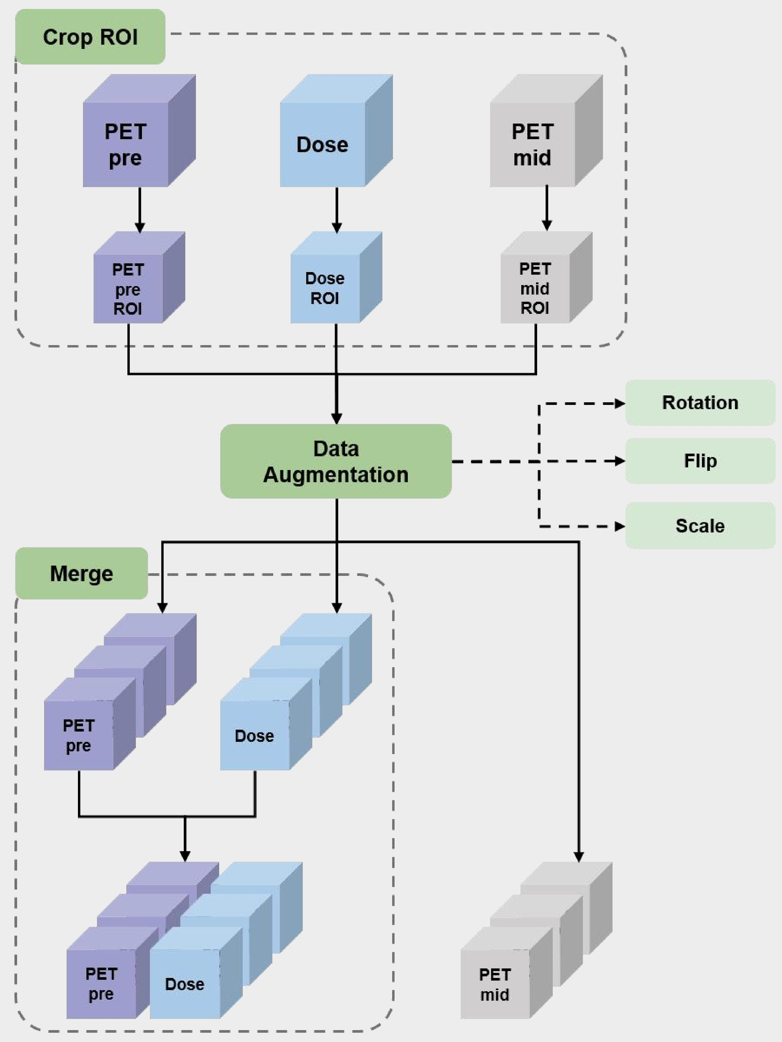

Figure 1. Data Processing. Consists of 3 steps: ROI cropping, data augmentation, channel-merge of PETpre and Dose.

Model Architecture of Attention Med3D (Figure 2)

Attention Med3D used an encoder-decoder architecture. The encoder was initialized via transfer learning from the Med3D pre-trained ResNet50 model, extracting hierarchical features while progressively reducing spatial dimensions. The custom decoder progressively restored spatial dimensions through multi-stage upsampling, with channel and spatial attention refining features at each stage and skip connections preserving multi-scale encoder information for precise prediction. The detailed encoder and decoder structures are shown in Figure 3 and Figure 4.

Comparison Models and Result Validation

- We compared Attention Med3D with several 3D deep learning methods: the widely used 3D U-Net segmentation network, and four 3D ResNet variants (3D ResNet18, 34, 50, and 101), chosen because the Med3D pre-trained model is ResNet-based.
- We employed leave-one-patient-out cross-validation over 25 rounds. In each round, 1 patient served as the test set, 3 patients (one randomly selected from each of the high, medium, and low response groups) formed the validation set, and the remaining 21 patients constituted the training set. To compare model performance, we applied the Friedman test separately for all patients and for each response category.

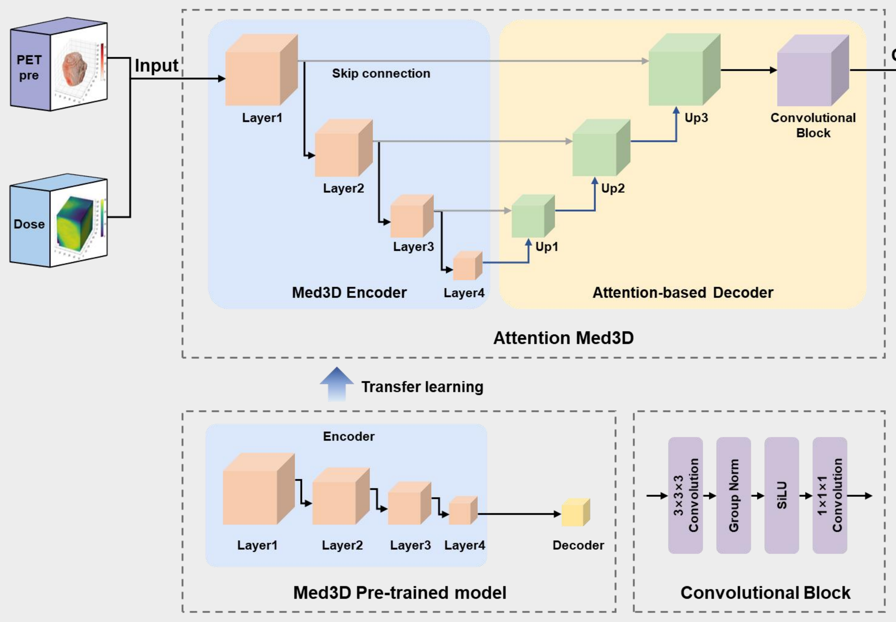

Figure 2. Overall architecture of the Attention Med3D model.

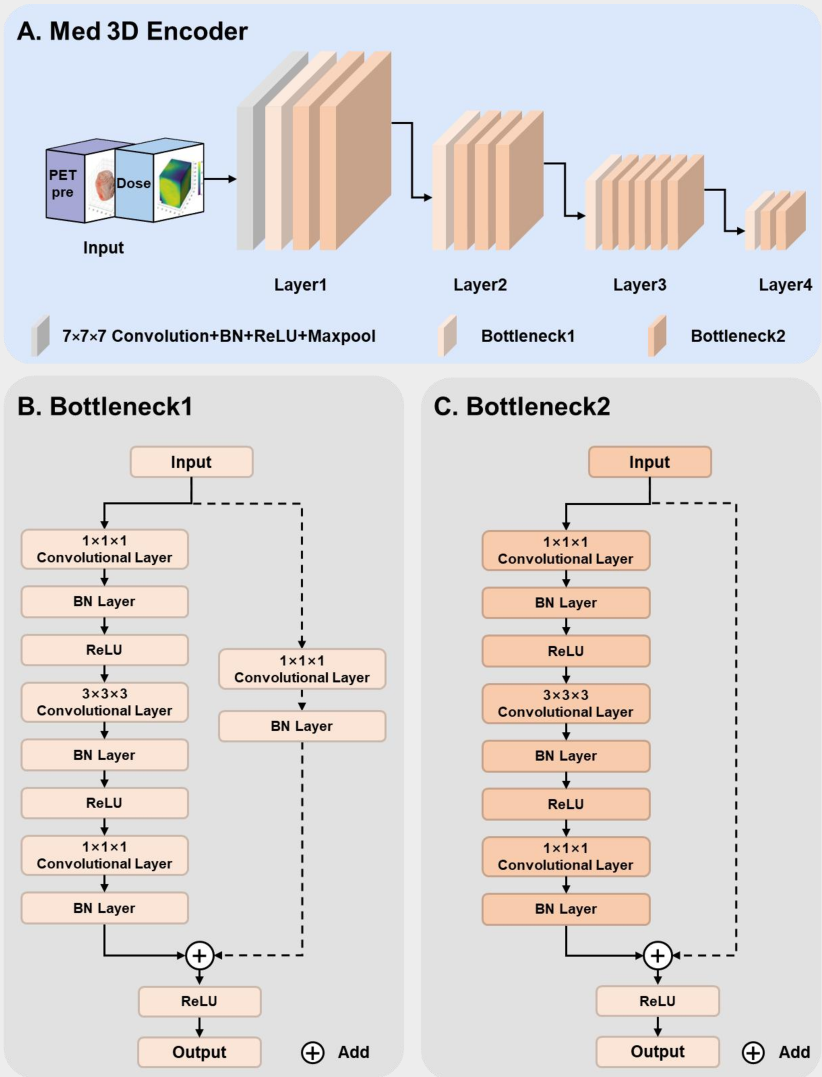

Figure 3. (A) Structure of the Med3D Encoder. The residual block structure is divided into two types: (B) Bottleneck1 and (C) Bottleneck2.

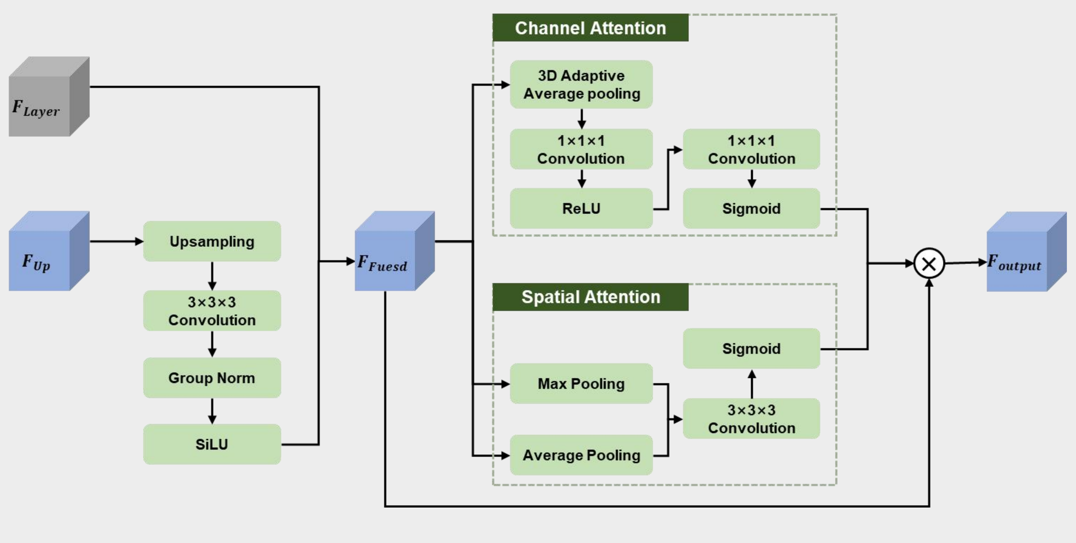

Figure 4. Structure diagram of the attention-based upsampling module (used in Up1, Up2, and Up3).

RESULTS

❖Attention Med3D achieved tumor voxel-level SUV prediction with RMSE of 1.521 and MAE of 1.240. Using the predicted mid-treatment SUV values to calculate the relative SUV change (ΔSUVratio = [SUVpre-SUVmid]/SUVpre), we obtained an RMSE of 0.265 and MAE of 0.215. (Table 1, Table 2)

❖The model numerically outperformed 3D U-Net and 3D ResNet series models for both SUV prediction (RMSE: 1.610-1.810; MAE: 1.353-1.515) and calculated ΔSUVratio (RMSE: 0.298-0.346; MAE: 0.249-0.283), though without statistical significance (Friedman test, $p > 0.05$). (Table 1, Table 2)

❖Figure 5 presents representative voxel-wise PETmid predictions from PETpre across high-, medium-, and low-response groups, with the best and worst cases selected by RMSE. Best-prediction cases showed close visual similarity between predicted and actual PETmid, while worst cases exhibited more noticeable differences. Overall, the predictions' visual appearance was consistent with their RMSE values, with lower-RMSE cases showing closer similarity.

CONCLUSIONS

❖The proposed model explored the application of transfer learning for radiotherapy response prediction and constructed a 3D deep learning framework suitable for limited-sample scenarios.

❖This framework has the potential to help clinicians to optimize individualized radiotherapy protocols for lung cancer patients.

❖**Innovation / Impact**: A 3D deep learning model based on transfer learning and attention mechanisms was developed to achieve accurate voxel-wise tumor SUV prediction in a limited-sample clinical setting. This enables early response prediction from baseline imaging, potentially facilitating timely radiotherapy adaptation.

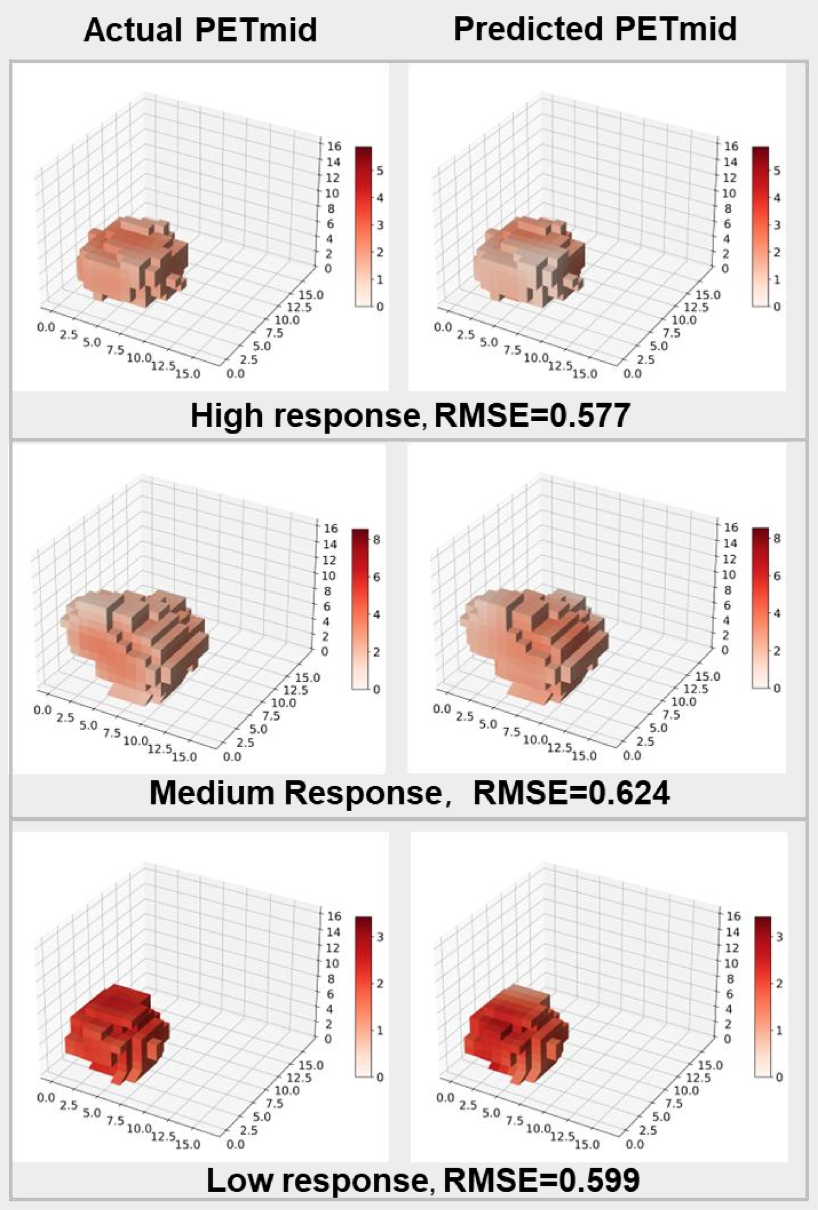

Figure 5. Representative examples of actual mid-treatment PET (Actual PETmid), and Attention Med3D-predicted mid-treatment PET (Predicted PETmid) from 3 patients across different response categories.

Table 1. RMSE of predicted SUV and calculated ΔSUVratio for PETmid image prediction.

	RMSE_SUV				RMSE_ΔSUVratio			
Method	all	high	medium	low	all	high	medium	low
3D U-Net	1.610	1.493	1.787	1.558	0.298	0.194	0.268	0.417
3D ResNet18	1.810	1.955	1.865	1.634	0.336	0.237	0.290	0.464
3D ResNet34	1.670	1.839	1.590	1.591	0.307	0.229	0.246	0.432
3D ResNet50	1.648	1.615	1.814	1.530	0.312	0.193	0.308	0.421
3D ResNet101	1.799	1.565	1.983	1.843	0.346	0.189	0.307	0.520
Attention Med3D	**1.521**	**1.537**	**1.727**	**1.325**	**0.265**	**0.182**	**0.269**	**0.336**
p-value	0.110	0.064	0.221	0.137	0.160	0.073	0.711	0.086

Table 2. MAE of predicted SUV and calculated ΔSUVratio for PETmid image prediction.

	MAE_SUV				MAE_ΔSUVratio			
Method	all	high	medium	low	all	high	medium	low
3D U-Net	1.353	1.205	1.530	1.325	0.249	0.160	0.224	0.349
3D ResNet18	1.515	1.583	1.588	1.390	0.279	0.199	0.238	0.386
3D ResNet34	1.390	1.504	1.319	1.353	0.256	0.192	0.202	0.361
3D ResNet50	1.368	1.306	1.530	1.278	0.254	0.159	0.243	0.349
3D ResNet101	1.508	1.283	1.668	1.565	0.283	0.155	0.256	0.421
Attention Med3D	**1.240**	**1.234**	**1.415**	**1.089**	**0.215**	**0.147**	**0.214**	**0.275**
p-value	0.134	0.121	0.425	0.161	0.110	0.104	0.623	0.120