



Voxel-level radiomics and deep learning for predicting pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant immunotherapy and chemotherapy

Zhen Zhang ^{1,2}, Tianchen Luo,³ Meng Yan,^{2,4} Haixia Shen,¹ Kaiyi Tao,¹ Jian Zeng,¹ Jingping Yuan,⁵ Min Fang,¹ Jian Zheng,⁴ Inigo Bermejo,⁶ Andre Dekker,² Dirk De Ruyscher,² Leonard Wee,² Wencheng Zhang ⁴, Youhua Jiang,¹ Yongling Ji^{1,7}

To cite: Zhang Z, Luo T, Yan M, *et al.* Voxel-level radiomics and deep learning for predicting pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant immunotherapy and chemotherapy. *Journal for ImmunoTherapy of Cancer* 2025;13:e011149. doi:10.1136/jitc-2024-011149

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/jitc-2024-011149>).

ZZ and TL contributed equally.

ZZ and TL are joint first authors.

Accepted 04 March 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to

Yongling Ji; jiyl@zjcc.org.cn

Youhua Jiang;
jiangyh@zjcc.org.cn

ABSTRACT

Background Accurate prediction of pathologic complete response (pCR) following neoadjuvant immunotherapy combined with chemotherapy (nICT) is crucial for tailoring patient care in esophageal squamous cell carcinoma (ESCC). This study aimed to develop and validate a deep learning model using a novel voxel-level radiomics approach to predict pCR based on preoperative CT images.

Methods In this multicenter, retrospective study, 741 patients with ESCC who underwent nICT followed by radical esophagectomy were enrolled from three institutions. Patients from one center were divided into a training set (469 patients) and an internal validation set (118 patients) while the data from the other two centers was used as external validation sets (120 and 34 patients, respectively). The deep learning model, Vision-Mamba, integrated voxel-level radiomics feature maps and CT images for pCR prediction. Additionally, other commonly used deep learning models, including 3D-ResNet and Vision Transformer, as well as traditional radiomics methods, were developed for comparison. Model performance was evaluated using accuracy, area under the curve (AUC), sensitivity, specificity, and prognostic stratification capabilities. The SHapley Additive exPlanations analysis was employed to interpret the model's predictions.

Results The Vision-Mamba model demonstrated robust predictive performance in the training set (accuracy: 0.89, AUC: 0.91, sensitivity: 0.82, specificity: 0.92) and validation sets (accuracy: 0.83–0.91, AUC: 0.83–0.92, sensitivity: 0.73–0.94, specificity: 0.84–1.0). The model outperformed other deep learning models and traditional radiomics methods. The model's ability to stratify patients into high and low-risk groups was validated, showing superior prognostic stratification compared with traditional methods. SHAP provided quantitative and visual model interpretation.

Conclusions We present a voxel-level radiomics-based deep learning model to predict pCR to neoadjuvant immunotherapy combined with chemotherapy based on pretreatment diagnostic CT images with high accuracy and robustness. This model could provide a promising tool for individualized management of patients with ESCC.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Neoadjuvant immunotherapy combined with chemotherapy (nICT) is a promising treatment for esophageal squamous cell carcinoma (ESCC), but accurate prediction of pathologic complete response (pCR) remains challenging. Traditional biomarkers for predicting pCR have limited value, and the role of radiomics in predicting treatment outcomes has been studied primarily in neoadjuvant chemoradiotherapy. Deep learning models have shown potential in medical imaging, but their application in nICT remains limited.

WHAT THIS STUDY ADDS

⇒ This study presents a novel deep learning model, Vision-Mamba, which integrates voxel-level radiomics and CT images to predict pCR in patients with ESCC following nICT. The model outperforms other deep learning models and traditional radiomics methods, demonstrating high accuracy, sensitivity, and specificity across multiple validation cohorts. By combining voxel-level radiomics, the model provides a more detailed and robust prediction of tumor response, offering a promising tool for personalized treatment strategies.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study highlights the potential of voxel-level radiomics combined with deep learning for improving clinical decision-making in ESCC treatment. The model's ability to predict pCR could guide clinicians in selecting candidates for organ-preserving strategies like the watch-and-wait approach, reducing unnecessary surgeries and improving patient quality of life. Future research may focus on validating this model in larger, prospective trials and exploring its integration with other predictive biomarkers, potentially influencing clinical practice guidelines for the treatment of ESCC.

INTRODUCTION

Esophageal squamous cell carcinomas (ESCC) are neoplasms arising from the squamous epithelium and are responsible for over 90% of all esophageal cancers in Asia.¹ Neoadjuvant immunotherapy combined with chemotherapy (nICT) has been established as a promising treatment for locally advanced ESCC, supported by several clinical trials that have demonstrated its acceptable safety and efficacy.^{2–5} Compared with the current standard of care—neoadjuvant chemoradiotherapy (nCRT)—nICT studies achieved comparable R0 resection rates (80.5–98%)^{3,6} and pathologic complete response (pCR) rates (39.2–50%).^{3,6} nICT not only seems to offer the potential for a better long-term prognosis than nCRT,⁷ but it is also a highly recommended option for patients at high risk of radiotherapy complications or those reluctant to undergo radiotherapy.^{8,9} Additionally, 40–50% of patients experience postoperative complications, with major complications occurring in about 10% of patients,^{6,8,10} however the short- and long-term outcomes are similar between planned and salvage esophagectomies.¹¹ Therefore, for patients who will most likely attain pCR, an organ-preserving and function-preserving strategy, known as watch-and-wait, may be considered, where active surveillance and surgery are performed as needed. This underscores the necessity for precise methods to assess responses to nICT, enabling the personalization of treatment plans.

At present, clinically predictive biomarkers for nICT pathological responses are lacking. The most thoroughly investigated biomarkers—microsatellite instability, programmed cell death ligand-1 (PD-L1) expression, and tumor mutational burden^{6,12,13}—are expensive to measure, but offer limited predictive value.^{14,15} Radiomics, on the other hand, which involves the quantitative extraction of features of whole tumors in situ from non-invasive clinical imaging, has demonstrated potential for predicting responses in nCRT.¹⁶ Nonetheless, its applicability in nICT prediction is still in its infancy and requires extensive research.

Previous studies inform us about the known advantages and shortcomings of radiomics. One benefit is the ability to encode tumor phenotype using predefined mathematical formulas that can be later analyzed by machine learning or statistical methods to pinpoint features that predict clinical outcomes.¹⁷ This allows for the efficient distillation of useful quantitative features from a limited sample size¹⁸ and leads to reasonably explainable models. One significant limitation is that each feature is represented by a single value per patient, potentially missing intricate details. End-to-end deep learning approaches facilitate holistic processing from input to output, enabling the model to directly learn relevant features from the input data, and are especially useful when used in conjunction with large and diverse data sets. Nonetheless, if dealing with small sample sizes, deep learning models are at high risk of overfitting to training data, thus lacking generalizability. In practice, one tries to

overcome such limitation by applying data augmentation such as translation, rotation, and scaling¹⁹ of the already-existing images. Introducing external prior knowledge as supplementary training data may also help alleviate these issues, thereby improving the model's performance with small sample size.^{20–23} Radiomics feature maps spatially represent tumor characteristics across the entire tumor volume, preserving spatial heterogeneity and distribution of features that might be lost when reducing them to single summary values. This approach, which we term “voxel-level radiomics”, provides a more nuanced view of the tumor's internal structure and enhances the utility of clinical CT images by addressing the limitations of traditional radiomics, which often oversimplify complex tumor characteristics.

In this study, we proposed a novel “voxel-level radiomics” approach and hypothesized that combining voxel-level radiomics feature maps and CT images with a “3D Vision-based Mamba architecture” could effectively predict primary tumor response after nICT (figure 1A). Due to the effectiveness of the bidirectional state space model in managing long sequences and strong spatial correlations present in medical images, we chose the Vision-Mamba architecture for our current work.²⁴ This study also compared conventional region-based radiomics methods, voxel-level radiomics alone, and CT images alone in constructing deep learning models, as well as evaluating the performance of well-known deep learning models. Furthermore, the SHapley Additive exPlanations (SHAP) analysis was used to enhance the interpretability of the models.

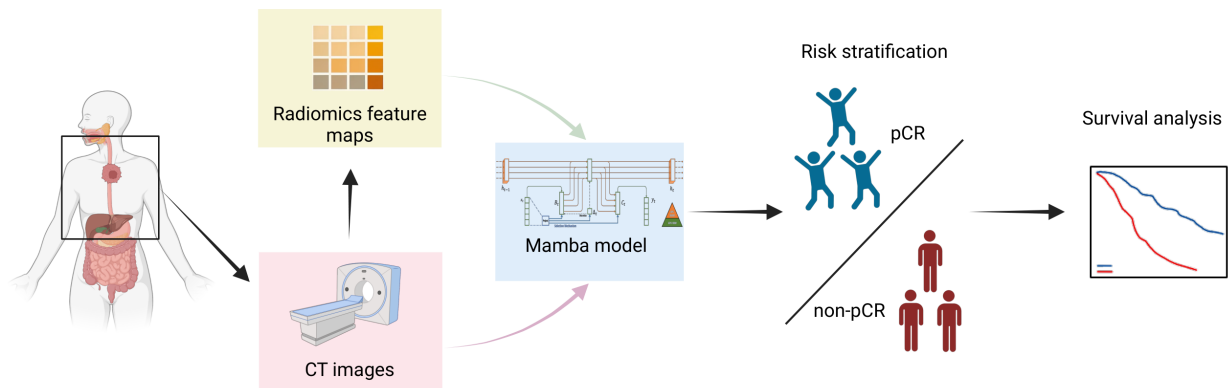
METHODS

Patient enrollment

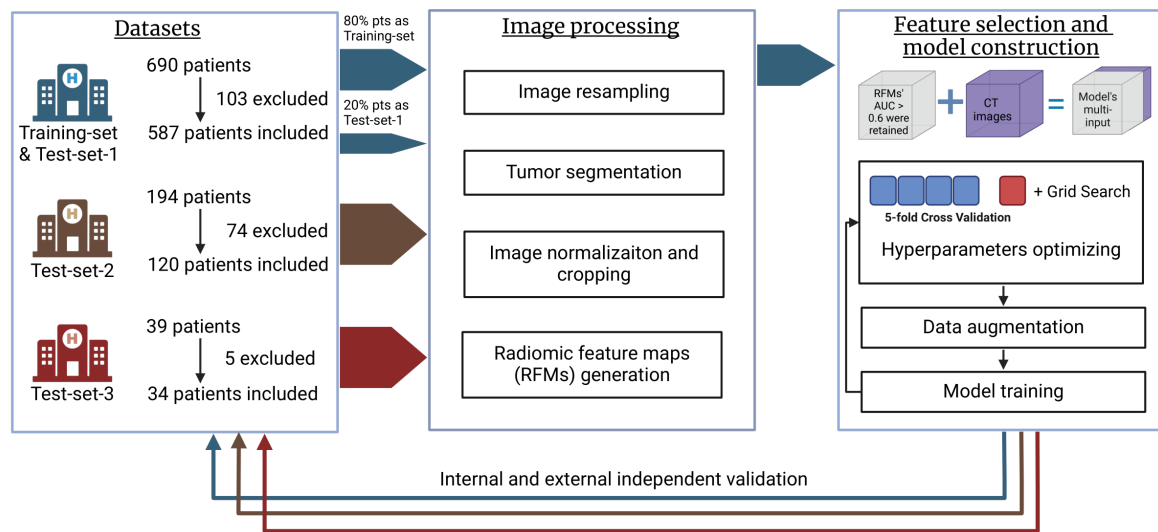
Due to the retrospective nature of the study, the requirement for informed consent was waived by the Institutional Review Board. The study comprised three cohorts of patients who underwent immunochemotherapy prior to radical esophagectomy at three different hospitals: (1) Zhejiang Cancer Hospital, with data from July 2019 to July 2023, which was randomly divided into a training set (80% of patients) and a holdout test set (test-set-1, 20% of patients); (2) Tianjin Medical University Cancer Institute, with patients treated between June 2020 and February 2022 as an independent test set (test-set-2); (3) Renmin Hospital of Wuhan University, with patients treated from July 2020 to September 2023 as a second independent test set (test-set-3). All patients were pathologically confirmed as primary ESCC and had obtained contrast-enhanced chest CT scans within 14 days prior to their neoadjuvant therapy, followed by radical esophagectomy and complete postoperative pathological assessment. Detailed inclusion and exclusion criteria are provided in online supplemental method A.

Treatment protocol and pathological evaluation

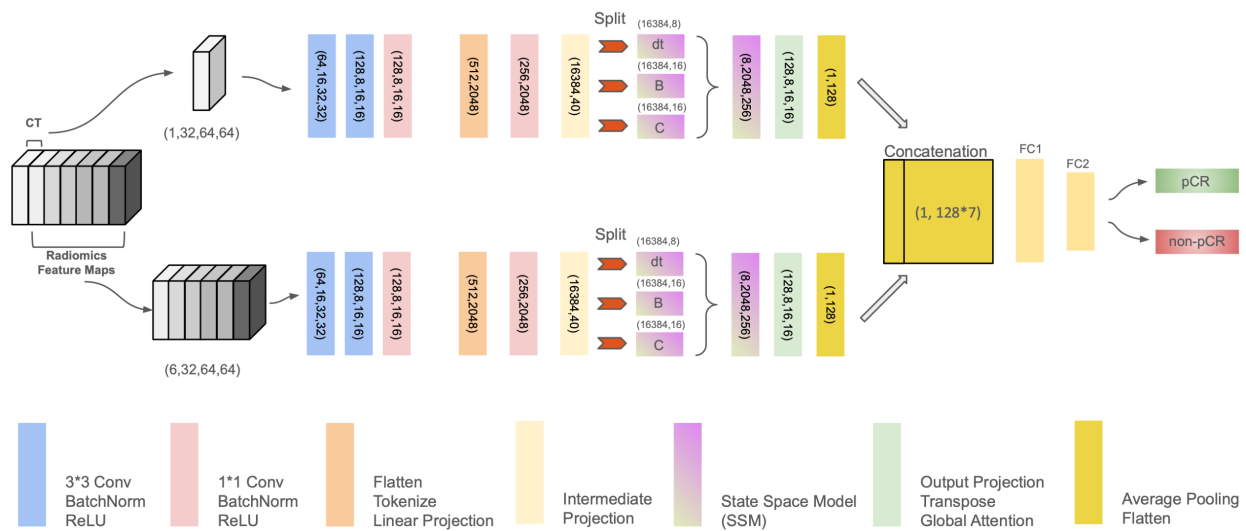
Patients were scheduled to receive at least one cycle of neoadjuvant immunotherapy, starting concurrently with



A



B



C

Figure 1 Study pipeline. (A) The process involved extracting radiomics feature maps from preoperative CT scans and combining them with CT images to predict pathologic complete response (pCR vs non-pCR) using the Vision-Mamba model. Additionally, the model's ability to stratify prognosis was explored. (B) Data from three different hospitals were included in the study. The images underwent processing, segmentation, feature extraction, feature selection, and model building. The performance of the model was then validated using independent validation sets. (C) The Vision-Mamba model architecture includes separate convolutional layers for CT images and shared convolutional layers for all radiomics feature maps. After initial processing with these convolutional layers and activation functions, the data is passed through state space model layers. The outputs are then concatenated and fed into fully connected layers to predict pCR or non-pCR. AUC, area under the curve.

chemotherapy. The immunotherapy involved standard doses (200mg every 3 weeks per cycle) of programmed cell death protein 1 or PD-L1 monoclonal antibodies (sintilimab, camrelizumab, tislelizumab, envafolimab, durvalumab, pembrolizumab, or nivolumab). Chemotherapy regimens were platinum-based and comprised two drugs with the following specifications: (1) TC (Taxane and Carboplatin) regimen (every 3 weeks): one to four cycles of nab-paclitaxel 260 mg/m² (day 1) or paclitaxel 135–175 mg/m² (day 1) + carboplatin area under the curve (AUC) 5 mg/mL/min (day 1) with a 21-day interval; (2) TP (Taxane and Cisplatin) regimen (every 3 weeks): one to four cycles of nab-paclitaxel 260 mg/m² (day 1) or paclitaxel 175 mg/m² (day 1) + cisplatin 75 mg/m² (day 1); (3) other regimens: one to four cycles of nab-paclitaxel 260 mg/m² (day 1) or paclitaxel 175 mg/m² (day 1) + oxaliplatin 130 mg/m² (day 1) with a 21-day interval; one to four cycles of fluorouracil 800–1,000 mg/m² (day 1) or tegafur 40–60 mg/m² (two times per day days 1–14) + cisplatin 75 mg/m² (day 1) or oxaliplatin 130 mg/m² (day 1) with a 21-day interval.

Radical esophagectomy was performed 4–8 weeks after completing the neoadjuvant therapy. The choice of esophagectomy technique—minimally invasive, open surgery, or video-assisted thoracic surgery—was based on the tumor's location and the thoracic surgeon's decision. The surgical approach also included either a two-field or three-field lymphadenectomy.

Pathology specimens obtained from the surgical resections were evaluated by an experienced pathologist and reviewed by a senior pathologist specializing in esophageal cancer. According to the College of American Pathologists Cancer Protocol for Esophageal Carcinoma,²⁵ tumor regression grade (TRG) was classified into four categories: TRG 0 indicated no histologically identifiable cancer cells; TRG 1 represented single cells or rare small groups of cancer cells; TRG 2 represented residual cancer with evident tumor regression but more than single cells or rare small groups of cancer cells; TRG 3 represented extensive residual cancer with no evident tumor regression. pCR was defined as having no viable tumor residual (TRG 0) at the primary tumor site, with TRG 1–3 being classified as non-pCR. This binary endpoint was used to build the prediction model.

Image and radiomics feature processing

The workflow for this section is illustrated in figure 1B. Contrast-enhanced chest CT scans were performed using a range of GE, Siemens, and Philips CT scanners, following standardized scanning protocols and each vendors' default image convolution kernels. Detailed scanning parameters are provided in online supplemental table 1A–C. To standardize spatial resolution across different centers, we resampled the original CT images from all data sets to an isotropic voxel size of 1×1×5 mm (slice thickness 5 mm).

The primary esophageal tumors prior to neoadjuvant therapy were defined as the regions of interest (ROIs) and

manually segmented by two physicians with over 3 years of experience (HS, MY). Tumor boundaries were delineated based on multiple diagnostic modalities, including PET-CT (Positron Emission Tomography - Computed Tomography), esophagograms, and esophagoscopy. These ROIs were further reviewed and manually refined by a senior physician with over 25 years of experience (YJ). Any discrepancies were resolved through collective discussion. The segmentation was performed using 3D Slicer software.²⁶

Subsequently, the CT image intensity was normalized to a range from −110 to 190, a practice derived from clinical experience and standard for deep learning models. The ROI was then cropped from the CT images to isolate the tumor area for analysis, with a padding of 1 voxel added around the edge of the ROI to ensure comprehensive coverage. We employed the “PyRadiomics” Python package²⁵ to extract 90 radiomics feature maps from each ROI. These radiomics feature maps were created by calculating specific radiomic feature values for each voxel within the ROI, thus generating a comprehensive distribution of features. These hand-crafted features, which generally conformed to the Image Biomarker Standardization Initiative (IBSI) guidelines, included 17 intensity features and 73 texture features. Unique exceptions to IBSI have been documented by the software developers, over whom we had no influence.

We performed feature selection by evaluating the predictive power of each individual radiomics feature. Specifically, each of the 90 radiomics feature maps was used as a separate input for training individual Vision-Mamba models, where each model incorporated only one radiomics feature to assess its predictive performance. Features with an area under the receiver operating characteristic greater than 0.6 after model convergence, that is, those demonstrating potential discriminative power, were retained.

Adapting Vision-Mamba architecture for three-dimensional medical image analysis

We modified an existing Vision-Mamba architecture, initially designed for RGB image processing, to accommodate a multi-input three-dimensional (3D) model as depicted in figure 1C. The Mamba model is known for its efficient, hardware-aware designs optimized for long sequence modeling and is particularly adept at visual representation learning.²⁷ It uses bidirectional state space models (SSMs) to effectively model data-dependent global visual contexts and includes positional embeddings for enhanced location-aware recognition. This configuration allows Mamba to process high-resolution images with improved performance and efficiency, circumventing the need for self-attention mechanisms.²⁴

In our adaptation, the input comprises a four-dimensional array (X+1, 32, 64, 64), where X represents the selected voxel-based radiomics feature maps alongside the cropped CT image of the ROI. This array is divided

into $X+1$ individual 3D arrays, which are then separately input into the model.

Each 3D input is subjected to a sequence of transformations including 3D convolutions, batch normalization, and ReLU activation, resulting in consolidated feature maps of shape (128, 8, 16, 16). These maps are then flattened, tokenized, and projected into tensors of shape (256, 2048). Subsequently, these tensors are split into three components: dt (delta time), B (input coupling matrix), and C (output coupling matrix), which are integral to the SSM. The output from the SSM is processed through global attention and average pooling, resulting in feature vectors of dimension (128). These vectors from each input are concatenated to form a comprehensive feature vector of dimension $(128 \times X+1)$.

Finally, this combined feature vector is passed through a fully connected layer. The model employs BCEWithLogitsLoss as its loss function, which combines sigmoid activation with binary cross-entropy loss, and includes a `pos_weight` parameter to manage class imbalance.

Vision-Mamba model construction and training

The configuration of our model includes several key parameters tailored for optimizing performance: the input feature dimensionality (dim) is $X+1$, which includes one original CT image and X selected voxel-based radiomics feature maps, long-range dependency capture (d_state) is set at 16 to balance complexity and efficiency, the depthwise convolution dimensionality (d_conv) is set at 4, and the internal feature dimensionality increase ($expand$) is set at 2. The model processes input hidden states through linear projection followed by depthwise convolution. These features are subsequently divided into dt , B , and C components, sized according to dt_rank and d_state . A selective scan function uses these components for efficient state updates, and an out-projection layer ensures output dimensions match the original inputs.

We trained the model using the Adam optimizer with a batch size of 32. The training data was partitioned into five subsets for fivefold cross-validation, which was used to fine-tune hyperparameters and assess model performance as shown in figure 1B. The initial learning rate was $1e-4$, reduced periodically by a factor of 0.1. To address class imbalance, we implemented BCEWithLogitsLoss with a positive weight of 4.0. Extensive data augmentation techniques were applied, including random horizontal flipping, rotations within ± 60 degrees, scaling between 0.8 and 1.2, and optional elastic deformation to introduce slight distortions. The models were trained and assessed using a single NVIDIA GeForce RTX 4090 GPU, supported by PyTorch V.2.0.0+cu118, CUDA V.11.8, and operated on an Intel Core i7 CPU with 32 GB of RAM. A random seed of 218 ensured consistency across runs. The code is publicly accessible at: https://github.com/Tianchen-Luo/3D_multi_input_Mamba_NEO.

Comparison with 3D-ResNet, vision transformer and classical radiomics models

In addition to the Vision Mamba 3D model, we used two prominent deep learning models in the medical field for comparison: a multi-input 3D-ResNet²⁸ and a multi-input Vision Transformer (ViT)²⁹ model, which used the same radiomics feature maps and initial CT images as our proposed model. The 3D-ResNet model incorporates Basic Block 3D layers, each with 3D convolutions and batch normalization for feature extraction. The ResNet 3D structure consists of an initial 3D convolution layer, followed by four stages of Basic Block 3D layers with spatial downsampling and adaptive average pooling. Each 3D input is independently passed through the shared ResNet-18-based architecture. The features extracted from all inputs are concatenated along the feature dimension and processed through fully connected layers: first a layer with 128 units, followed by a final classification layer.

For the multi-input ViT model, this model processes multiple 3D input volumes by reducing their dimensionality and applying a ViT to each input independently. The multi-input ViT model includes an initial 3D convolutional preprocessing stage to reduce input channels. This stage consists of 3D convolutions, ReLU activations, max pooling, and a 1×1 convolution to reduce channels to 3, matching the ViT input requirements. Each preprocessed input is then fed into a pretrained ViT model. The 3D inputs are reduced to two-dimensional slices via adaptive average pooling, resized to 224×224 dimensions, and normalized. The ViT processes these inputs through its transformer layers. Outputs from the ViT are concatenated along the feature dimension and passed through fully connected layers, reducing dimensionality with ReLU activation and dropout, culminating in a final classification layer to make prediction.

Additionally, the performance of the constructed model was also compared with the Mamba model trained solely on radiomics feature maps (Mamba-Radiomics) or CT images (Mamba-CT). Furthermore, a classical radiomics model, based on a logistic regression model, was constructed for comparative analysis (LR-Radiomics). Detailed information on the construction of these models is provided in online supplemental method B.

Model interpretation using SHAP values

After completing the model training, we used SHAP values to enhance the interpretability of our model.³⁰ SHAP values assign an importance value to each voxel, quantifying their contribution to the model's predictions. The Shapley value for a voxel i is defined as follows:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

where: N is all the voxels in the input images, and S is a subset of voxels excluding $voxel_i$ itself. $f(S)$ is the model prediction using only the subset S of voxels and

$f(S \cup \{i\})$ is the model's output using the voxels in the subset S plus the $voxel_i$.

We computed SHAP values for each voxel in the selected voxel-based radiomics feature maps and CT images to generate their corresponding SHAP value maps. Higher SHAP values indicated regions that positively contributed to the predictions, while lower values indicated less influential regions. Model interpretability is achieved through two primary methods. First, we sum the absolute SHAP values across each voxel in the SHAP value map to rank the importance of radiomics features in predicting pCR. Second, we overlay the SHAP value maps on the original CT images to visually identify the regions most valuable for the model's predictions.

Statistical analysis

Patient characteristics were evaluated using SPSS V.27. Continuous variables were assessed via the Kruskal-Wallis test, while categorical variables were examined using Pearson's χ^2 test or Fisher's exact test. To assess the predictive value of clinical parameters, univariable logistic regression analyses were conducted. Statistical significance was set at a p value of <0.05 for two-tailed tests. The performance of the models was evaluated based on accuracy, AUC, sensitivity, and specificity across all data sets. Survival times were estimated using the Kaplan-Meier method, comparing patients predicted to achieve pCR versus non-pCR, as well as those who actually achieved pCR versus those who did not. Differences in survival outcomes were analyzed using the log-rank test. HRs and 95% CIs were estimated using the Cox proportional hazards model.

Survival analysis was conducted using R software V.4.4.1 with the "survival" package V.3.6.4, and results were visualized using the "survminer" package V.0.4.9. Deep learning models were constructed using Python V.3.10.12.

RESULTS

Patient characteristics

This study included a total of 741 patients from three institutions, as detailed in the patient selection process shown in online supplemental figure 1. The cohort was divided into 469 patients in the training set, 118 in test-set-1 (internal independent validation set), 120 in test-set-2, and 34 in test-set-3. Test-set-2 and 3 served as external validation sets. The distribution of patient characteristics is summarized in table 1. The overall rates of tumor pCR and R0 resection were consistent across the data sets, with a pCR rate of approximately 22% and an R0 resection rate of about 94%. Despite this consistency, variations were observed in age, gender, Eastern Cooperative Oncology Group performance status, tumor location, clinical tumor stage, and immunotherapy regimens across the groups ($p < 0.05$). The use of univariate logistic regression in the training set did not identify any clinical parameters that would be valuable in predicting pCR (online supplemental table 2). However, as shown in online supplemental table 3, female patients and those

who received more than two cycles of nICT treatment were associated with pCR. There were no statistically significant differences in other baseline characteristics between pCR and non-pCR patients.

Evaluation and predictive performance of models

A total of six radiomics features, each with an AUC greater than 0.6, were selected for inclusion in the model, along with the initial CT images (online supplemental table 4). The performance of the Mamba model is detailed in table 2, demonstrating robust predictive capabilities across all data sets, with accuracy ranging from 0.83 to 0.91, AUC from 0.83 to 0.92, sensitivity between 0.73 and 0.94, and specificity between 0.84 and 1.0. Notably, the model maintained favorable predictive accuracy in the external validation sets (test-set-2 and 3), which included different populations and image acquisition parameters.

To further investigate the influence of input modalities and model construction methods, five different models were developed. The predictive performance of these models is provided in online supplemental table 5. In general, the proposed Mamba model exhibited superior performance across all models. The 3D-ResNet and ViT models performed similarly to our proposed model, but they required longer training times. With a batch size of 8 and 469 training samples, the Mamba model completed an epoch 53.18% faster than the ResNet model and 16.49% faster than the ViT model, focusing specifically on the training process. Both the Mamba-Radiomics and Mamba-CT models, as well as the classical LR-Radiomics model, underperformed compared with the proposed model. Specifically, the LR-Radiomics model showed significantly lower AUC values, ranging from 0.52 to 0.63 across the external validation sets, alongside reduced sensitivity and specificity.

Prognostic value of the Vision-Mamba model

The Vision-Mamba model demonstrated significant prognostic stratification capabilities. Kaplan-Meier curves based on whether patients actually achieved pCR (figure 2A) showed clear stratification in the training set, with patients who achieved pCR exhibiting better overall survival (OS). However, this stratification did not reach statistical significance in validation sets. When patients were stratified by the model's predicted pCR status (figure 2B), the differences in prognosis became even more pronounced, although statistical significance was still not achieved in validation sets.

Additionally, patients exhibited the most pronounced prognostic differences when stratified using the median risk score output by the model from the training set (-1.2) as a fixed cut-off value (figure 2C). In test-set-1 and test-set-3, patients with a risk score equal to or greater than -1.2 had significantly better OS compared with those with a risk

Table 1 Patients clinical characteristics across all data sets

Characteristics	Overall (N=741)	Training-set (N=469)	Test-set-1 (N=118)	Test-set-2 (N=120)	Test-set-3 (N=34)	P value
Sex						0.034*
Female	59 (8.0)	30 (6.4)	8 (6.8)	17 (14.2)	4 (11.8)	
Male	682 (92.0)	439 (93.6)	110 (93.2)	103 (85.8)	30 (88.2)	
Age (median (IQR))	65.0 (59.0–69.0)	65.0 (59.0–69.0)	66.5 (61.0–70.0)	62.0 (58.0–66.0)	67.5 (59.0–70.0)	<0.001*
Smoking status						0.836
Never smoked	241 (32.5)	149 (31.8)	42 (35.6)	38 (31.7)	12 (35.3)	
Current or former smoker	500 (67.5)	320 (68.2)	76 (64.4)	82 (68.3)	22 (64.7)	
Drinking status						0.011*
Never drank	222 (30.0)	126 (26.9)	33 (28.0)	50 (41.7)	13 (38.2)	
Current or former drinker	519 (70.0)	343 (73.1)	85 (72.0)	70 (58.3)	21 (61.8)	
ECOG performance status						<0.001*†
0	323 (45.7)	155 (33.0)	65 (55.1)	103 (85.8)	NA	
1	375 (53.0)	306 (65.2)	52 (44.1)	17 (14.2)	NA	
2	9 (1.3)	8 (1.7)	1 (0.8)	0 (0.0)	NA	
Tumor location						0.013*
Upper	93 (12.6)	66 (14.1)	12 (10.2)	9 (7.5)	6 (17.6)	
Middle	400 (54.0)	263 (56.1)	64 (54.2)	63 (52.5)	10 (29.4)	
Lower	248 (33.5)	140 (29.9)	42 (35.6)	48 (40.0)	18 (52.9)	
cT						<0.001*
1	1 (0.0)	0 (0.0)	0 (0.0)	1 (0.8)	0 (0.0)	
2	107 (14.4)	88 (18.7)	16 (13.6)	0 (0.0)	3 (8.8)	
3	586 (79.1)	369 (78.7)	95 (80.5)	97 (80.8)	25 (73.5)	
4	47 (6.3)	12 (2.6)	7 (5.9)	22 (18.3)	6 (17.6)	
cN						0.034*
0	103 (13.9)	57 (12.2)	24 (20.3)	14 (11.7)	8 (23.5)	
1	373 (50.3)	245 (52.2)	57 (48.3)	56 (46.7)	15 (44.1)	
2	246 (33.2)	157 (33.5)	33 (28.0)	48 (40.0)	8 (23.5)	
3	19 (2.6)	10 (2.1)	4 (3.4)	2 (1.7)	3 (8.8)	
cM						0.044*
0	726 (98.0)	460 (98.1)	112 (94.9)	120 (100.0)	34 (100.0)	
1	15 (2.0%)	9 (1.9)	6 (5.1)	0 (0.0)	0 (0.0)	
cTNM stage (AJCC 8th)						<0.001*
I	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	
II	151 (20.4)	100 (21.3)	31 (26.3)	11 (9.2)	9 (26.5)	
III	512 (69.1)	340 (72.5)	71 (60.2)	85 (70.8)	16 (47.1)	
IV	78 (10.5)	29 (6.2)	16 (13.6)	24 (20.0)	9 (26.5)	
Immunotherapy regimen						0.009*
PD-1	699 (94.3)	433 (92.3)	113 (95.8)	120 (100.0)	33 (97.1)	
PD-L1	42 (5.7)	36 (7.7)	5 (4.2)	0 (0.0)	1 (2.9)	
Chemotherapy regimen						<0.001*
T+P	716 (96.6)	466 (99.4)	117 (99.2)	101 (84.2)	32 (94.1)	
Others	25 (3.4)	3 (0.6)	1 (0.8)	19 (15.8)	2 (5.9)	
NICT cycle						<0.001*
≤2	544 (73.4)	390 (83.2)	96 (81.4)	29 (24.2)	29 (85.3)	
>2	197 (26.6)	79 (16.8)	22 (18.6)	91 (75.8)	5 (14.7)	
R0 resection						0.468

Continued

Table 1 Continued

Characteristics	Overall (N=741)	Training-set (N=469)	Test-set-1 (N=118)	Test-set-2 (N=120)	Test-set-3 (N=34)	P value
No	46 (6.2)	29 (6.2)	8 (6.8)	9 (7.5)	0 (0)	0.143
Yes	695 (93.8)	440 (93.8)	110 (93.2)	111 (92.5)	34 (100.0)	
Tumor pCR						0.143
No	579 (78.1)	364 (77.6)	100 (84.7)	92 (76.7)	23 (67.6)	
Yes	162 (21.9)	105 (22.4)	18 (15.3)	28 (23.3)	11 (32.4)	<0.001*
ypT stage						
0	168 (22.7)	105 (22.4)	18 (15.3)	33 (27.5)	12 (35.3)	<0.001*
1	168 (22.7)	113 (24.1)	21 (17.8)	25 (20.8)	9 (26.5)	
2	139 (18.8)	88 (18.8)	21 (17.8)	24 (20.0)	6 (17.6)	
3	263 (35.5)	163 (34.8)	58 (49.2)	37 (30.8)	5 (14.7)	
4	3 (0.4)	0 (0.0)	0 (0.0)	1 (0.8)	2 (5.9)	
ypN stage						0.292
0	421 (56.8)	275 (58.6)	62 (52.5)	62 (51.7)	22 (64.7)	
1	218 (29.4)	128 (27.3)	41 (34.7)	41 (34.2)	8 (23.5)	
2	83 (11.2)	50 (10.7)	13 (11.0)	17 (14.2)	3 (8.8)	
3	19 (2.6)	16 (3.4)	2 (1.7)	0 (0.0)	1 (2.9)	
ypTNM stage (AJCC 8th)						0.078
I	322 (43.5)	216 (46.1)	43 (36.4)	44 (36.7)	19 (55.9)	
II	100 (13.5)	61 (13.0)	19 (16.1)	18 (15.0)	2 (5.9)	
III	300 (40.5)	176 (37.5)	54 (45.8)	58 (48.3)	12 (35.3)	
IV	19 (2.6)	16 (3.4)	2 (1.7)	0 (0.0)	1 (2.9)	
s-LN number (median (IQR))	24.0 (18.0–31.0)	22.0 (17.0–28.0)	25.0 (19.0–35.0)	31.5 (25.0–40.0)	23.0 (14.0–34.0)	<0.001*
Survival time (median (IQR))	672.0 (400.0–983.0)	620.0 (371.0–1007.0)	661.5 (608.0–714.0)	907.5 (580.5–1094.0)	476.5 (355.0–766.0)	<0.001*

Data are n (%), unless otherwise stated.
 *P value below 0.05 was considered statistically significant.
 †P value was calculated comparing the training set, test-set-1 and test-set-2.
 AJCC, American Joint Committee on Cancer; cM, clinical metastasis stage; cN, clinical node stage; cT, clinical tumor stage; cTNM, Clinical Tumor-Node-Metastasis; ECOG, Eastern Cooperative Oncology Group; NICT, neoadjuvant immunochemotherapy; pCR, pathological complete response; PD-1 Inhibitor, programmed cell death protein 1 inhibitor; PD-L1 Inhibitor, programmed cell death ligand 1 inhibitor; s-LN number, surgical lymph node number, defined as the number of lymph nodes were removed from surgery; T+P, paclitaxel in combination with platinum-based chemotherapy; ypN, neoadjuvant pathologic node stage; ypT, neoadjuvant pathologic tumor stage; ypTNM, neoadjuvant pathologic Tumor-Node-Metastasis.

score less than -1.2 (p value <0.05). This stratification method outperformed the others, underscoring the effectiveness of using the risk score to predict patient prognosis.

Univariate Cox regression analysis confirmed that actual pCR status, predicted pCR status, and stratification by a risk score of -1.2 were independent

prognostic factors for OS in patients who underwent nICT (p value <0.001) (online supplemental figure 2).

Model interpretation

After analyzing the specific contributions of input images (figure 3A and online supplemental table 4), we found that CT images contributed the most

Table 2 The performance of the Mamba model

Data set	Accuracy (95% CI)	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Training set	0.91 (0.89 to 0.94)	0.92 (0.90 to 0.95)	0.94 (0.89 to 0.98)	0.91 (0.87 to 0.94)
Test-set-1	0.87 (0.81 to 0.93)	0.83 (0.72 to 0.92)	0.76 (0.68 to 0.94)	0.89 (0.83 to 0.95)
Test-set-2	0.83 (0.77 to 0.90)	0.83 (0.75 to 0.91)	0.82 (0.74 to 0.96)	0.84 (0.77 to 0.91)
Test-set-3	0.91 (0.79 to 1.00)	0.86 (0.73 to 1.00)	0.73 (0.63 to 1.00)	1.00 (1.00 to 1.00)

AUC, area under the curve.

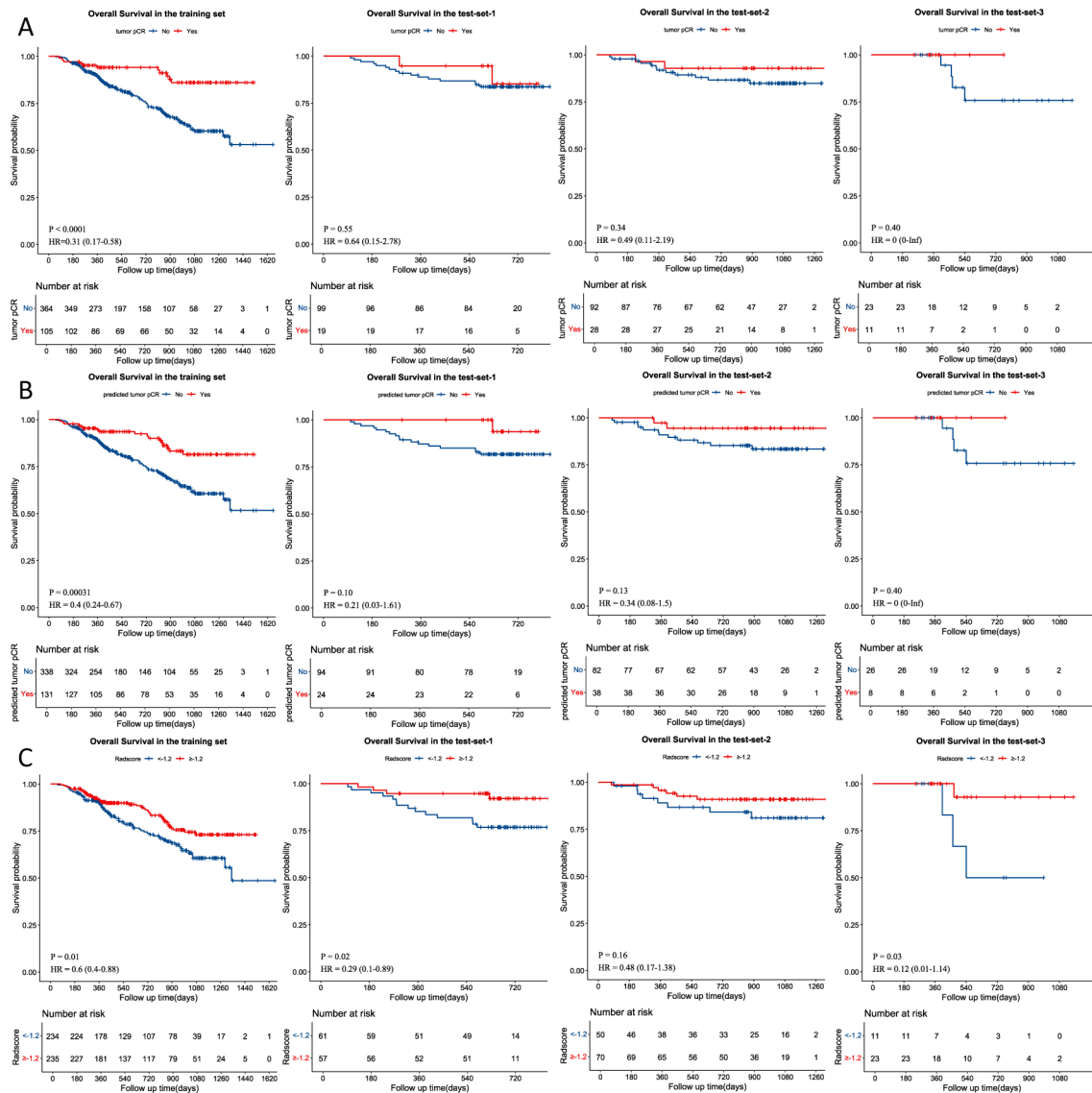


Figure 2 Prognostic stratification performance. (A) Kaplan-Meier (KM) curves for overall survival (OS) stratified by actual pathologic complete response (pCR) status in the training set and three independent validation sets (test-set-1, test-set-2, and test-set-3). (B) KM curves for OS stratified by the model's predicted pCR status. (C) KM curves for OS stratified by the risk scores output by the model, using a median cut-off value of -1.2 from the training set and applying it to the test sets.

(46.13%), followed by two radiomics features: glszm_SmallAreaLowGrayLevelEmphasis (34.97%) and glzm_LargeDependenceHighGrayLevelEmphasis (18.54%). The remaining four radiomics features contributed very little, collectively accounting for only 0.36%.

The SHAP value map provided insights into how the model made predictions for each patient. In figure 3B, CT images were overlaid with the SHAP value map, where darker red areas indicated regions that contributed more significantly to the model's predictions. On examining the SHAP value maps for all patients, we observed that regions with high SHAP values included the tumor necrotic region, the tumor edge region, and some significantly enhanced regions (indicated by the arrows in figure 3B-D).

DISCUSSION

In this study, we developed a deep learning-based model for the early assessment of pCR in patients with ESCC who received neoadjuvant immunotherapy combined with chemotherapy. By integrating voxel-level radiomics feature maps and CT images, our model accurately predicted pCR, achieving favorable AUC, high accuracy, sensitivity, and specificity in one internal independent validation cohort and two external independent validation cohorts. We used a state-of-the-art deep learning method, which demonstrated superior performance compared with other widely-used deep learning methods and conventional radiomics approaches. The integration of voxel-level radiomics and initial CT images into a joint model yielded the best predictive performance. Furthermore, our model effectively stratified patients into high

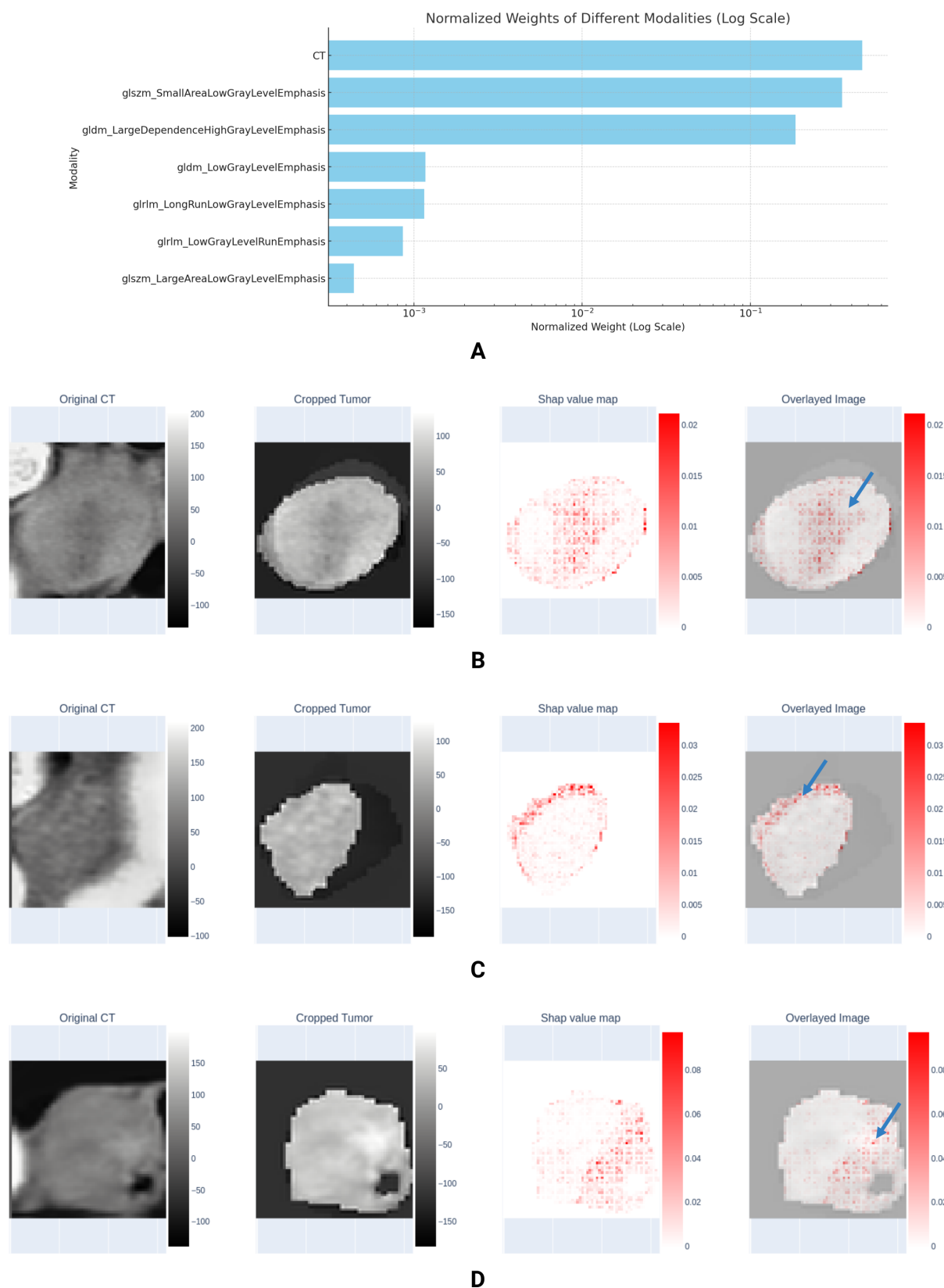


Figure 3 Model interpretation and feature importance. (A) Contributions of different input features to the model's predictions. (B–D) From left to right, each panel shows an original CT image, the cropped tumor region, the SHAP value map, and the overlaid image. The SHAP value maps are overlaid on CT images, with darker red areas indicating regions that contributed more significantly to the model's predictions. (B) The darker red regions, particularly in the tumor necrotic area (indicated by the arrow), highlight areas with a substantial influence on predicting pathologic complete response. (C) The darker red regions in the tumor edge region indicate significant SHAP values contributing to the model's predictions. (D) Further visualizations emphasize the importance of enhanced regions in the model's predictions. SHAP, SHapley Additive exPlanations.

and low-risk groups, and provided an interpretable method to visualize the important regions for predictions. Collectively, our proof-of-concept study demonstrated that the voxel-level radiomics approach, combined with deep learning, can enhance the predictive power of CT images, advancing the goal of precision treatment strategies in ESCC.

Neoadjuvant chemoradiotherapy has long been considered the standard of care for locally advanced esophageal cancer, known as the CROSS (Chemoradiotherapy for Oesophageal Cancer Followed by Surgery Study) protocol.³¹ However, recent studies focusing on the omission of radiotherapy have demonstrated better outcomes for patients compared with the CROSS protocol. For instance, the ESOPEC trial showed that patients treated with perioperative chemotherapy had longer OS (66 months vs 37 months) and comparable pCR rates to the CROSS regimen (19.3% vs 13.5%).³² Similarly, the JCOG 1109 trial revealed that doublet chemotherapy combined with radiotherapy did not significantly improve survival compared with doublet chemotherapy alone.³³ Adding immunotherapy has been shown to further improve outcomes compared with chemotherapy alone. The ESCORT-NEO/NCCES01 trial demonstrated that the addition of immunotherapy increased the pCR rate significantly (28.0% and 15.4% in the immunochemotherapy group vs 4.7% in the chemotherapy group).² In comparing nICT with nCRT, Yang *et al* found that the pCR rates were similar between the nICT and nCRT groups (20.2% vs 29.0%). However, the nICT group experienced fewer adverse events (42.7% vs 55.6%) and had lower postoperative 1-year distant metastasis and recurrence rates.³⁴ Furthermore, Yu *et al* reported that the nICT group had a better 3-year OS rate (91.7% vs 79.8%) and 3-year disease-free survival rate (87.4% vs 72.8%) compared with the nCRT group.⁷ In summary, although the optimal neoadjuvant treatment strategy remains uncertain, current clinical trial evidence and our clinical experience suggest that neoadjuvant immunotherapy combined with chemotherapy holds significant potential. This promising approach underpins the rationale for our study.

Our model has three practical clinical implications. First, for patients who are highly likely to achieve pCR, we may be able to clinically evaluate the feasibility of a watch-and-wait strategy for organ preservation. This approach reduces patient suffering and improves quality of life by avoiding unnecessary surgery. Importantly, by ensuring high specificity, the model helps accurately identify those patients who do not achieve pCR and need timely surgical intervention. This avoids the risk of misclassifying non-pCR patients as pCR, which could delay necessary treatment and compromise long-term outcomes. Second, the model provides accurate risk stratification, allowing for tailored treatment strategies based on patient prognosis. For example, high-risk patients may benefit from more aggressive consolidation therapy, while low-risk patients can avoid overtreatment. Third, our model can serve as

a foundational model that can be fine-tuned for other treatment strategies such as nCRT. It can assist in selecting treatment plans by predicting the likely prognosis of patients undergoing different therapeutic approaches.

To the best of our knowledge, this study represents the largest artificial intelligence (AI)-aided investigation to date predicting pCR in patients with esophageal cancer undergoing nICT, encompassing the highest number of patients. It was also the first to propose a voxel-level radiomics method for constructing a deep learning model. Previous studies, such as Li *et al*'s delta-radiomics approach with 95 patients with ESCC, achieved an AUC of 0.848.³⁵ Similarly, Yang *et al* combined radiomics and hematological features to construct a model with an AUC of 0.934.³⁶ However, these studies were limited by small validation sets (eg, 29 samples) or lacked external validation altogether. In contrast, our study highlighted the poor generalizability and performance of classical radiomics in external validation sets. This limitation is likely due to the high compression of features in classical radiomics, where each feature is represented by a single value. Detailed differences in feature distribution, as observed in radiomics feature maps, are overlooked. Our proposed voxel-level radiomics technique addresses this by analyzing subtle feature distributions and leveraging deep learning models to explore nonlinear relationships and complex interactions between features.

The results demonstrated that our model achieved high accuracy, particularly in specificity, across all test sets (table 2). However, the sensitivity was lower than 0.8 in test-set-1 and test-set-3. This may be attributed to the smaller number of pCR samples compared with non-pCR samples. To address this, efforts should be made to expand the data set or employ data augmentation methods specifically for pCR samples to enhance the model's sensitivity. Despite these sensitivity challenges, the model exhibited consistent performance across different test sets, indicating good generalizability. Nonetheless, future studies could incorporate harmonization methods to further improve the model's generalizability. Regarding prognostic stratification, we found that using the risk scores output by the model provided superior prognostic stratification compared with using the model's pCR predictions or the actual pCR status (figure 2). A subset of non-pCR patients was predicted by the model to achieve pCR. We speculate that these might be patients with a major pathological response, although further statistical analysis is required to validate this hypothesis. While the median cut-off was effective in prognostic stratification, the imbalance between pCR and non-pCR patients in the training cohort could have influenced the model's cut-off value, potentially limiting its generalizability when applied to the validation cohort. Future studies should consider larger data sets and methods to mitigate the impact of data imbalance, such as adjusting the cut-off value or exploring more sophisticated stratification techniques.

We addressed the interpretability of our model from two perspectives. First, we ranked the importance of

input features. The raw CT images, containing all available detailed information, emerged as the most significant feature. Additionally, two radiomics features also held substantial weight. Based on the definitions of these radiomics features, we propose the following hypotheses. The feature `glszm_SmallAreaLowGrayLevelEmphasis` quantifies the prominence of small areas with low gray levels in the image, which is likely representative of necrotic regions within esophageal tumors. This hypothesis was supported by our second interpretability method, the visualization of SHAP value maps, where we observed that necrotic regions significantly influenced the model's predictions. Therefore, we believe that the `glszm_SmallAreaLowGrayLevelEmphasis` feature map may reflect tumor necrosis, contributing greatly to the prediction of pCR. The feature `gldm_LargeDependence-HighGrayLevelEmphasis` quantifies the prominence of large, high-gray-level dependencies, which might represent denser, more continuous regions such as the esophageal wall. Our SHAP value maps also indicated that regions surrounding the esophagus impacted pCR predictions. However, it remains unclear whether this influence is due to the esophageal wall itself or areas adjacent to the tumor. This speculation requires further pathological studies for validation. Moreover, the peritumoral region warrants further analysis in future studies.

In this study, we compared the transformer-based Vision-Mamba model with ViT and a modified ResNet (adapted for 3D data multi-inputs) regarding validation performance, convergence speed, stability, and training efficiency. The Mamba model was selected for its superior performance. It excels in handling long sequence data through SSMs, making it ideal for 3D imaging tasks like CT scans. SSMs process multiple slices as sequences, capturing long-range dependencies and maintaining data continuity. The bidirectional nature of SSMs enhances comprehensive data understanding by processing sequences in both directions. The Mamba model also offers linear time complexity in sequence length, unlike the quadratic complexity in traditional transformers like ViT, resulting in faster processing and lower computational costs. Compared with ResNet, which has deeper networks and more parameters, the Mamba model demonstrated faster training speeds and greater efficiency in handling large 3D data volumes. Specifically, the 3D-ResNet model took twice as long to complete an epoch compared with the Mamba model.

This study had several limitations. First, in this study, the external validation set is concentrated in one country. To validate its generalizability, more cross-country data sets are needed. Additionally, we believe that randomized clinical trials in real-world settings are the best standard for testing AI-based models. However, such trials are currently constrained by ethical and legal issues that need to be addressed. Future rigorous randomized clinical trials should test the model's predictive ability. Second, our study did not incorporate biomarkers previously thought to predict immunotherapy efficacy, such as tumor

mutation burden, PD-L1 expression, and combined positive scores. Due to the financial cost of testing for these biomarkers, a subset of patients in our data set did not have these biomarkers available. Future studies should explore the incorporation of these biomarkers into the modeling process. Third, although we conducted an interpretability analysis of the model and explored it visually, the conclusions drawn are still subjective without solid evidence to prove our conjectures. We believe that future studies should conduct histological analyses to observe and verify the conclusions drawn from radiomics. Additionally, comprehensive and stable biomarkers may be established by combining macroscopic radiology with microscopic pathology. Fourth, the random 8:2 split between the training and validation sets was intended to provide a representative distribution of cases for model training. However, the lower proportion of pCR cases in the internal validation cohort (test-set-1) resulted in reduced sensitivity, with a value of 0.76, which was below the expected threshold. This issue may have arisen due to the relatively small number of pCR cases in the validation set, combined with the inherent variability in pCR rates across different clinical centers. The unequal distribution of pCR and non-pCR cases in the validation set highlights the challenge of ensuring sufficient representation of minority classes in smaller data sets. To address this limitation in future studies, we plan to explore several strategies, such as employing data augmentation techniques to better balance the data set, and incorporating more diverse data sets from multiple centers to more accurately capture the variability in pCR rates across different clinical settings. Finally, while our study has demonstrated promising performance, the relatively low number of pCR cases in the validation sets, may have resulted in an overestimation of the model's predictive performance, as reflected in the AUC values. We acknowledge that the imbalance in the pCR and non-pCR patient distribution could influence the model's sensitivity. In future work, we aim to expand the sample size and collect more pCR cases to improve the model's sensitivity and offer a more balanced evaluation of its performance.

In conclusion, we developed a deep learning model to accurately predict pCR after neoadjuvant immunotherapy combined with chemotherapy in patients with ESCC. This was achieved using a novel voxel-level radiomics approach applied to standard diagnostic CT images obtained before surgery. Our results underscore the potential of this imaging-based biomarker to guide precision treatment decision-making. However, further validation in large prospective trials is necessary to refine these findings and fully assess the clinical utility of our proposed method.

Author affiliations

¹Zhejiang Cancer Hospital, Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, Zhejiang, China

²Department of Radiation Oncology (Maastr), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Center+, Maastricht, The Netherlands

³National University of Singapore, Singapore

⁴Department of Radiation Oncology, Key Laboratory of Cancer Prevention and Therapy, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Tianjin, China

⁵Department of Pathology, Renmin Hospital of Wuhan University, Wuhan, Hubei, China

⁶Hasselt University, Hasselt, Belgium

⁷Zhejiang Key Laboratory of Prevention Diagnosis and Therapy for Gastrointestinal Cancer, Hangzhou, Zhejiang, China

Acknowledgements Figure 1 was created in BioRender. Zhang, Z. (2025) <https://BioRender.com/p00t599>, and Figure 3 was created in BioRender. Zhang, Z. (2025) <https://BioRender.com/r34a875>.

Contributors ZZ: Conceptualization, Methodology, Funding acquisition, Resources, Writing—original draft. TL: Conceptualization, Software, Writing—original draft. MY: Data curation, Methodology, Visualization. HS: Data curation, Methodology. KT: Validation. JZeng: Resources. JY: Resources. MF: Formal analysis. JZheng: Methodology. IB: Supervision, Writing—review and editing. AD: Supervision, Writing—review and editing. DDR: Supervision, Writing—review and editing. LW: Supervision, Writing—review and editing, Methodology. WZ: Project administration, Resources, Writing—review and editing. YJiang: Project administration, Resources, Supervision. YJi: Data curation, Funding acquisition, Project administration, Writing—review and editing. YJi is the guarantor.

Funding This study was funded by the National Natural Science Foundation of China (No. 82303672) and Zhejiang Medical and Health Science and Technology Project (No. 2021KY542).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This retrospective study was approved by the Institutional Review Boards (IRB) of Zhejiang Cancer Hospital (IRB-2023-88).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data are private institutional collections, which may be made available to other researchers upon reasonable request and subject to data sharing agreements – please contact the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Zhen Zhang <http://orcid.org/0000-0001-6335-9529>

Wencheng Zhang <http://orcid.org/0000-0003-3730-5361>

REFERENCES

- Bray F, Laversanne M, Sung H, *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;74:229–63.
- Qin J, Xue L, Hao A, *et al.* Neoadjuvant chemotherapy with or without camrelizumab in resectable esophageal squamous cell carcinoma: the randomized phase 3 ESCORT-NEO/NCCES01 trial. *Nat Med* 2024;30:2549–57.
- Yan X, Duan H, Ni Y, *et al.* Tislelizumab combined with chemotherapy as neoadjuvant therapy for surgically resectable esophageal cancer: A prospective, single-arm, phase II study (TD-NICE). *Int J Surg* 2022;103:106680.
- Sun J-M, Shen L, Shah MA, *et al.* Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study. *Lancet* 2021;398:759–71.
- Janjigian YY, Shitara K, Moehler M, *et al.* First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. *Lancet* 2021;398:27–40.
- Liu J, Yang Y, Liu Z, *et al.* Multicenter, single-arm, phase II trial of camrelizumab and chemotherapy as neoadjuvant treatment for locally advanced esophageal squamous cell carcinoma. *J Immunother Cancer* 2022;10:e004291.
- Yu Y-K, Meng F-Y, Wei X-F, *et al.* Neoadjuvant chemotherapy combined with immunotherapy versus neoadjuvant chemoradiotherapy in patients with locally advanced esophageal squamous cell carcinoma. *J Thorac Cardiovasc Surg* 2024;168:417–28.
- Xu L, Wei X, Li C, *et al.* Pathologic responses and surgical outcomes after neoadjuvant immunochemotherapy versus neoadjuvant chemoradiotherapy in patients with locally advanced esophageal squamous cell carcinoma. *Front Immunol* 2022;13:1052542.
- Xiao X, Yang Y-S, Zeng X-X, *et al.* The comparisons of neoadjuvant chemoimmunotherapy versus chemoradiotherapy for oesophageal squamous cancer. *Eur J Cardiothorac Surg* 2022;62:ezac341.
- Wang H, Tang H, Fang Y, *et al.* Morbidity and Mortality of Patients Who Underwent Minimally Invasive Esophagectomy After Neoadjuvant Chemoradiotherapy vs Neoadjuvant Chemotherapy for Locally Advanced Esophageal Squamous Cell Carcinoma: A Randomized Clinical Trial. *JAMA Surg* 2021;156:444–51.
- Markar S, Gronnier C, Duhamel A, *et al.* Salvage Surgery After Chemoradiotherapy in the Management of Esophageal Cancer: Is It a Viable Therapeutic Option? *J Clin Oncol* 2015;33:3866–73.
- Chen Y, Ren M, Li B, *et al.* Neoadjuvant sintilimab plus chemotherapy for locally advanced resectable esophageal squamous cell carcinoma: a prospective, single-arm, phase II clinical trial (CY-NICE). *J Thorac Dis* 2023;15:6761–75.
- Luchini C, Bibeau F, Ligtenberg MJL, *et al.* ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol* 2019;30:1232–43.
- Doroshov DB, Bhalla S, Beasley MB, *et al.* PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol* 2021;18:345–62.
- Anagnostou V, Bardelli A, Chan TA, *et al.* The status of tumor mutational burden and immunotherapy. *Nat Cancer* 2022;3:652–6.
- Yang Z, Gong J, Li J, *et al.* The gap before real clinical application of imaging-based machine-learning and radiomic models for chemoradiation outcome prediction in esophageal cancer: a systematic review and meta-analysis. *Int J Surg* 2023;109:2451–66.
- Warkentin MT, Al-Sawaihey H, Lam S, *et al.* Radiomics analysis to predict pulmonary nodule malignancy using machine learning approaches. *Thorax* 2024;79:307–15.
- Zhang Z, Wang Z, Yan M, *et al.* Radiomics and Dosiomics Signature From Whole Lung Predicts Radiation Pneumonitis: A Model Development Study With Prospective External Validation and Decision-curve Analysis. *Int J Radiat Oncol Biol Phys* 2023;115:746–58.
- Aharon A, Yair W. Why do deep convolutional networks generalize so poorly to small image transformations. *J Mach Learn Res* 2019;20:1–25.
- Pease M, Arefan D, Barber J, *et al.* Outcome Prediction in Patients with Severe Traumatic Brain Injury Using Deep Learning from Head CT Scans. *Radiology* 2022;304:385–94.
- Talaei Khoei T, Ould Slimane H, Kaabouch N. Deep learning: systematic review, models, challenges, and research directions. *Neural Comput & Applic* 2023;35:23103–24.
- Brigato L, Iocchi L. A close look at deep learning with small data. 2020 25th International Conference on Pattern Recognition (ICPR); Milan, Italy, 2021:2490–7.
- Jiang Z, Zheng T, Liu Y, *et al.* Incorporating Prior Knowledge into Neural Networks through an Implicit Composite Kernel. *arXiv* 2024.
- Zhu L, Liao B, Zhang Q, *et al.* Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. *arXiv* 2024.
- Shi C, Jordan B. Protocol for the examination of specimens from patients with carcinoma of the esophagus. College of American Pathologists Cancer Protocols; 2017.1–17.

- 26 Fedorov A, Beichel R, Kalpathy-Cramer J, *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging* 2012;30:1323–41.
- 27 Gu A, Dao T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv* 2024;31.
- 28 He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA, 2016:770–8.
- 29 Dosovitskiy A, Beyer L, Kolesnikov A, *et al.* An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* 2021;3.
- 30 Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. *arXiv* 2017.
- 31 Eyck BM, van Lanschot JJB, Hulshof MCCM, *et al.* Ten-Year Outcome of Neoadjuvant Chemoradiotherapy Plus Surgery for Esophageal Cancer: The Randomized Controlled CROSS Trial. *J Clin Oncol* 2021;39:1995–2004.
- 32 Hoepfner J, Brunner T, Lordick F, *et al.* Prospective randomized multicenter phase III trial comparing perioperative chemotherapy (FLOT protocol) to neoadjuvant chemoradiation (CROSS protocol) in patients with adenocarcinoma of the esophagus (ESOPEC trial). *JCO* 2024;42:LBA1.
- 33 Kato K, Machida R, Ito Y, *et al.* Doublet chemotherapy, triplet chemotherapy, or doublet chemotherapy combined with radiotherapy as neoadjuvant treatment for locally advanced oesophageal cancer (JCOG1109 NExT): a randomised, controlled, open-label, phase 3 trial. *Lancet* 2024;404:55–66.
- 34 Yang X, Yin H, Zhang S, *et al.* Perioperative outcomes and survival after neoadjuvant immunochemotherapy for locally advanced esophageal squamous cell carcinoma. *J Thorac Cardiovasc Surg* 2025;169:289–300.
- 35 Li K, Li Y, Wang Z, *et al.* Delta-radiomics based on CT predicts pathologic complete response in ESCC treated with neoadjuvant immunochemotherapy and surgery. *Front Oncol* 2023;13:1131883.
- 36 Yang Y, Yi Y, Wang Z, *et al.* A combined nomogram based on radiomics and hematology to predict the pathological complete response of neoadjuvant immunochemotherapy in esophageal squamous cell carcinoma. *BMC Cancer* 2024;24:460.