Supplementary Methods S1. Local radiomic feature extraction for visceral adipose tissue heterogeneity score construction

For each patient i and imaging time point $t \in \{\text{pre}, \text{post}\}$, the preprocessed computed tomography enterography (CTE) image and visceral adipose tissue (VAT) mask were denoted by $I_{t,i}$ and $M_{t,i}$, respectively. The VAT region was defined as

$$\Omega_{t,i} = \{x : M_{t,i}(x) = 1\}. \tag{1}$$

For each voxel $x \in \Omega_{t,i}$, a local cubic neighborhood $W(x)$ was centered at x. In the primary implementation, a $3 \times 3 \times 3$ local window was used so that each feature vector represented the immediate local imaging microenvironment around the voxel while preserving spatial resolution.

A total of 104 local radiomic descriptors were extracted from the original CTE image within each local window after excluding diagnostic, nonnumeric, and noninformative outputs. The descriptors covered first-order statistics and texture feature families, including gray-level co-occurrence matrix, gray-level dependence matrix, gray-level run-length matrix, gray-level size-zone matrix, and neighboring gray-tone difference matrix features. The purpose of this local extraction strategy was to represent each VAT voxel by a quantitative vector describing local intensity distribution, gray-level co-occurrence, regional dependence, run-length heterogeneity, zone-size heterogeneity, and neighborhood contrast.

For voxel x, the resulting local feature vector was written as

$$\boldsymbol{\phi}_{t,i}(x) = (\phi_{t,i,1}(x), \phi_{t,i,2}(x), \ldots, \phi_{t,i,104}(x))^{\top} \in \mathbb{R}^{104}. \tag{2}$$

All feature vectors within the VAT mask were stacked to form the voxel-level feature matrix

$$\mathbf{F}_{t,i} = \left[\boldsymbol{\phi}_{t,i}(x_1), \boldsymbol{\phi}_{t,i}(x_2), \ldots, \boldsymbol{\phi}_{t,i}(x_{m_{t,i}})\right]^{\top} \in \mathbb{R}^{m_{t,i} \times 104}, \tag{3}$$

where $m_{t,i} = |\Omega_{t,i}|$ is the number of VAT voxels. Nonfinite feature values were treated as missing values and imputed using the median of the corresponding feature. Feature normalization was then performed before clustering to reduce the influence of different numerical scales across feature families.

The normalized voxel-level matrix $\mathbf{F}_{t,i}$ was clustered by k-means to generate a three-dimensional VAT habitat map. Let K denote the number of clusters and let $C_{t,i,k}$ be the set of voxels assigned to cluster k. The cluster label map was defined as

$$L_{t,i}(x) = k, \qquad x \in C_{t,i,k}, \quad k = 1, \ldots, K. \tag{4}$$

For each cluster, a 26-neighborhood connected-component analysis was performed. The number of connected domains and the volume of the largest connected domain were then used to compute the VAT heterogeneity score described in the main text. This design allows VHS to increase when the VAT habitat map becomes more spatially fragmented, patchy, or disorganized.

Supplementary Table S1 Local radiomic feature families used for VAT habitat construction.

Feature family	Local information represented by the feature family
First-order statistics	Local intensity distribution, including central tendency, dispersion, skewness, kurtosis, entropy, energy, and percentile-based descriptors.
Gray-level co-occurrence matrix	Local pairwise gray-level co-occurrence patterns and second-order texture, including contrast, correlation, homogeneity, dissimilarity, and energy-related descriptors.
Gray-level dependence matrix	Local gray-level dependence and spatial dependence size, reflecting neighborhood-level dependence heterogeneity.
Gray-level run-length matrix	Consecutive run-length patterns of gray levels, capturing fine versus coarse local texture organization.
Gray-level size-zone matrix	Connected zone-size distribution of gray levels, reflecting local zone heterogeneity and spatial aggregation.
Neighboring gray-tone difference matrix	Difference between each gray level and the average neighboring gray tone, representing local contrast and coarseness.

Supplementary Methods S2. Whole-volume VAT radiomics feature engineering and dimensionality reduction

To benchmark the topology-aware VAT heterogeneity score (VHS) framework against conventional whole-volume radiomics, 1,239 VAT radiomic features were extracted from each CTE scan using the whole VAT mask. These features represented intensity, shape, and texture information from the original and filtered images. For each patient, three radiomics feature sets were considered: pretreatment features, post-induction features, and delta features.

Let $\mathbf{r}_{\mathrm{pre},i} \in \mathbb{R}^{1239}$ and $\mathbf{r}_{\mathrm{post},i} \in \mathbb{R}^{1239}$ denote the whole-volume VAT radiomic feature vectors extracted from pretreatment and post-induction CTE scans, respectively. The delta radiomic feature vector was defined as

$$\Delta\mathbf{r}_i = \mathbf{r}_{\mathrm{post},i} - \mathbf{r}_{\mathrm{pre},i}. \tag{5}$$

All feature engineering steps were performed using the training cohort only, and the derived parameters were then applied unchanged to the testing cohort to avoid information leakage.

First, each raw radiomic feature was normalized using parameters estimated from the training cohort. Second, univariable filtering was performed with Student's t test to compare patients with sustained response and those who developed secondary loss of response (SLR). Features with $P < 0.05$ were retained for subsequent analysis. Third, redundancy control was performed using Pearson correlation analysis. For any pair of retained features with an absolute correlation coefficient greater than 0.90, one feature was removed. Priority was given to the feature with the stronger univariable association with SLR or the greater univariable discriminative ability in the training cohort.

Supplementary Algorithm S1 Computation of the VAT heterogeneity score.

Require: Pretreatment image and mask $(I_{\mathrm{pre},i}, M_{\mathrm{pre},i})$; post-induction image and mask $(I_{\mathrm{post},i}, M_{\mathrm{post},i})$; number of clusters K

Ensure: $\mathrm{VHS}_{\mathrm{pre},i}$, $\mathrm{VHS}_{\mathrm{post},i}$, and $\%\Delta\mathrm{VHS}_i$

for $t \in \{\mathrm{pre}, \mathrm{post}\}$ **do**
  Resample $I_{t,i}$ and $M_{t,i}$ to isotropic voxel spacing.
  Binarize the VAT mask and retain the largest connected component.
  Clip VAT voxel intensities to the 1st and 99th percentiles and normalize intensities within VAT.
  for all voxels $x \in \Omega_{t,i}$ **do**
    Extract 104 local radiomic descriptors from the $3 \times 3 \times 3$ neighborhood centered at x.
  end for
  Impute nonfinite local feature values and normalize voxel-level feature vectors.
  Apply k-means clustering to generate the VAT habitat label map $L_{t,i}$.
  for $k = 1$ to K **do**
    Identify spatially connected domains in cluster k using 26-neighborhood connectivity.
    Compute the number of connected domains $n_{t,i,k}$ and the largest connected-domain volume $v_{t,i,k,\max}$.
  end for
  Compute $\mathrm{VHS}_{t,i} = 1 - v_{\mathrm{total},t,i}^{-1} \sum_{k=1}^{K} v_{t,i,k,\max}/n_{t,i,k}$.
end for
Compute $\%\Delta\mathrm{VHS}_i = (\mathrm{VHS}_{\mathrm{pre},i} - \mathrm{VHS}_{\mathrm{post},i})/\mathrm{VHS}_{\mathrm{pre},i} \times 100\%$.

Finally, least absolute shrinkage and selection operator (LASSO) logistic regression was applied to the remaining features to identify a sparse and robust radiomics signature. For a candidate radiomic matrix $\mathbf{Z} \in \mathbb{R}^{n \times q}$ and binary endpoint $y_i \in \{0, 1\}$, the LASSO estimator was defined as

$$(\hat{\beta}_0, \hat{\boldsymbol{\beta}}) = \arg\min_{\beta_0, \boldsymbol{\beta}} \left\{ -\frac{1}{n} \sum_{i=1}^{n} [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] + \lambda \sum_{j=1}^{q} |\beta_j| \right\}, \tag{6}$$

where

$$p_i = \frac{1}{1 + \exp[-(\beta_0 + \boldsymbol{\beta}^\top \mathbf{z}_i)]} \tag{7}$$

and λ is the regularization parameter. The final radiomics score was calculated as a linear combination of the selected standardized features:

$$\mathrm{RadScore}_i = \hat{\beta}_0 + \sum_{j \in \mathcal{S}} \hat{\beta}_j z_{ij}, \tag{8}$$

where $\mathcal{S}$ is the set of features with nonzero coefficients. This pipeline was applied separately to construct $\mathrm{VAT\text{-}RadScore}_{\mathrm{pre}}$, $\mathrm{VAT\text{-}RadScore}_{\mathrm{post}}$, and ΔVAT-RadScore.

Supplementary Algorithm S2 Whole-volume VAT radiomics feature reduction and RadScore construction.

Require: Whole-volume VAT radiomic matrix with 1,239 features; binary SLR endpoint; training and testing cohort labels

Ensure: VAT radiomics score for each patient

1: Extract 1,239 whole-volume VAT radiomic features from each scan.
2: Construct pretreatment, post-induction, and delta radiomic feature sets.
3: Estimate normalization parameters from the training cohort and apply them to both cohorts.
4: In the training cohort, retain features with Student's t test $P < 0.05$.
5: Remove redundant features using Pearson correlation analysis with threshold $|r| > 0.90$.
6: Fit LASSO logistic regression in the training cohort.
7: Define the RadScore as the weighted sum of selected standardized features with nonzero coefficients.
8: Apply the fixed RadScore formula to the testing cohort.

Supplementary Methods S3. Model development, benchmark classifiers, and performance assessment

The primary endpoint was secondary loss of response (SLR), encoded as $y_i = 1$ for SLR and $y_i = 0$ for sustained response. Candidate predictors included baseline clinical variables and the longitudinal imaging descriptor %ΔVHS. Univariable logistic regression was used for initial screening. Variables with $P < 0.20$ were entered into multivariable logistic regression, and variables retained at $P < 0.05$ were used to construct the final nomogram.

The final logistic regression model can be written as

$$\operatorname{logit}(p_i) = \alpha_0 + \sum_{j=1}^{p} \alpha_j x_{ij}, \tag{9}$$

where $p_i = P(y_i = 1 \mid \mathbf{x}_i)$ is the predicted probability of SLR. The clinical-only model used baseline clinical predictors without %ΔVHS. The combined clinical imaging model, namely Nom-%ΔVHS, included %ΔVHS and the retained clinical predictors.

For algorithmic benchmarking, the same retained predictor set was used to train random forest, support vector machine, decision tree, XGBoost, and neural network classifiers. All model training and feature selection procedures were restricted to the training cohort. The independent testing cohort was used only for final model evaluation. The purpose of this benchmark analysis was not to replace the interpretable nomogram, but to examine whether the logistic regression-based model retained competitive performance compared with commonly used nonlinear classifiers.

Discrimination was assessed by the area under the receiver operating characteristic curve. At the selected probability threshold, accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and F1 score were calculated

as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{10}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \qquad \text{Specificity} = \frac{TN}{TN + FP}, \tag{11}$$

$$\text{PPV} = \frac{TP}{TP + FP}, \qquad \text{NPV} = \frac{TN}{TN + FN}, \tag{12}$$

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN}. \tag{13}$$

Calibration was evaluated using calibration curves. Decision curve analysis was used to estimate clinical net benefit across threshold probabilities. Continuous net reclassification improvement and integrated discrimination improvement were calculated to quantify the incremental value of Nom-%ΔVHS relative to comparator models.