

#### Supplementary Material 1 (radiomics feature extraction)

There are 274 features based on the original images, 465 features based on LoG-sigma-transformed images, and 744 features based on wavelet-transformed images. All of these features included first-order features, shape-based (3D, 2D) features, grey level co-occurrence matrix (GLCM) features, grey level run length matrix (GLRLM) features, grey level size zone matrix (GLSZM) features, grey level dependence matrix (GLDM) features, and neighbouring grey tone difference matrix (NGTDM) features. For the wavelet filter, each image was filtered with either a low bandpass filter or a high bandpass filter in the x, y, and z directions. For the LoG filter, the sigma values were 5.0, 4.0, 3.0, 2.0 and 1.0 mm.

#### Supplementary Material 2 (feature selection)

The embedded feature selection of the RFC was used to choose predictable features. We selected the features involved in the model leading to the smallest out-of-bag (OOB) error. The importance of each feature could be obtained by summing the Gini impurity decreases for each feature over all trees in the RFC. By computing the RFC scores of importance and deleting the features with small importance, we ranked the remaining features in decreasing order by importance value and selected the main features through the importance threshold.

The optimal importance threshold was obtained by the learning curve of the RFC.

#### Supplementary Material 3 (Results)

We identified the 21 most important MDR predictive biomarkers for model construction based on the optimal threshold, which was set to 0.007 by the learning curve (Fig.S1). These top 21 significant features consist of five categories: 14 first-order features, 3 GLCM features, 1 GLRLM feature, 1 GLDM feature, and 2 GLSZM features. Four first-order features and 1

GLCM feature were obtained from the LoG-sigma-transformed image ( $\sigma = 1, 2, \text{ and } 4$ ).

The remaining 10 first-order features, 2 GLCM features, 1 GLRLM feature, 1 GLDM feature and 2 GLSZM features were obtained from wavelet-transformed images.

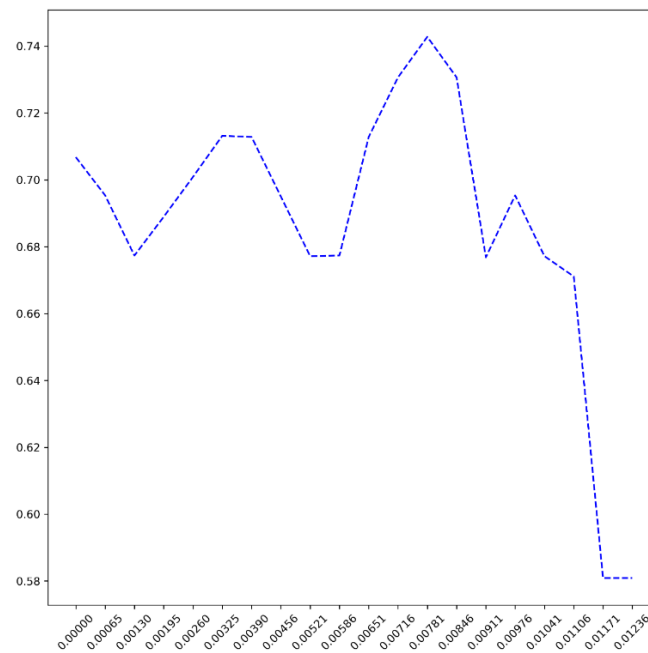


Fig.S1 The learning curve of importance threshold. When the threshold was 0.007, RFC showed the best performance.