A Foundation Model Framework for Multi-View MRI Classification of Extramural Vascular Invasion and Mesorectal Fascia Invasion in Rectal Cancer

Yumeng Zhang ¹, Zohaib Salahuddin ¹, Danial Khan¹, Shruti Atul Mali¹, Henry C. Woodruff^{1, 2}, Sina Amirrajab¹, Eduardo Ibor-Crespo ³, Ana Jimenez-Pastor³, Luis Marti-Bonmati ^{4,5}, Philippe Lambin ^{1,2}

Abstract

Background: Accurate MRI-based identification of extramural vascular invasion (EVI) and mesorectal fascia invasion (MFI) is pivotal for risk-stratified management of rectal cancer. However, subjective visual assessment and inter-institutional variability can lead to inconsistent therapeutic decisions beyond mere tumor staging. Reliable detection of EVI directly informs decisions on neoadjuvant therapy and postoperative chemotherapy, while accurate MFI assessment influences surgical planning by predicting circumferential resection margin (CRM) involvement, guiding neoadjuvant chemoradiotherapy (nCRT) to reduce local recurrence risk.

Purpose: To develop and externally evaluate a multi-centre, foundation-model–driven framework that automatically classifies EVI and MFI on axial and sagittal T2-weighted MRI.

Materials and Methods: This retrospective study used 331 pre-treatment rectal-cancer MRI examinations from three European hospitals (GE = 221, Siemens = 106). After TotalSegmentator-guided rectal patch extraction and intensity normalization, a self-supervised frequency-domain harmonization pipeline was trained to minimize scanner-related contrast shifts. Four classifiers were compared: ResNet50, SeResNet, the universal biomedical pretrained transformer (UMedPT) with a lightweight MLP head, and a logistic-regression variant using frozen UMedPT features (UMedPT_LR). The models were trained and validated on 265 cases, and performance was evaluated using a separate test set of 66 cases. Performance was assessed with AUC, balanced accuracy, sensitivity, specificity, and F1 score.

¹ The D-Lab, Department of Precision Medicine, GROW - Research Institute for Oncology and Reproduction, Maastricht University, 6220 MD Maastricht, The Netherlands

² Department of Radiology and Nuclear Medicine, GROW - Research Institute for Oncology and Reproduction, Maastricht University, Medical Center+, 6229 HX Maastricht, The Netherlands

³ Research & Frontiers in AI Department, Quantitative Imaging Biomarkers in Medicine, Quibim SL, Valencia, Spain

⁴ Biomedical Imaging Research Group, La Fe Health Research Institute, Valencia, Spain

⁵ Medical Imaging Department, La Fe University and Polytechnic Hospital, Valencia, Spain

Results: UMedPT_LR achieved the best EVI detection when axial and sagittal features were fused (AUC = 0.82; sensitivity = 0.75; F1 score = 0.73; balanced accuracy = 0.72), surpassing the Chaimeleon Grand-Challenge winner (AUC = 0.74). The highest MFI performance was attained by UMedPT on axial harmonized images (AUC = 0.77; balanced accuracy = 0.72), surpassing the Chaimeleon Grand-Challenge winner (AUC = 0.75). Frequency-domain harmonization improved MFI classification but variably affected EVI performance. Multi-view fusion, that combined axial and sagittal features, consistently enhanced EVI metrics relative to single-plane input. Conventional CNNs (ResNet50, SeResNet) underperformed, especially in F1 score and balanced accuracy.

Conclusion: A foundation-model workflow leveraging harmonization and axial-sagittal feature fusion delivers state-of-the-art, generalizable assessment of key prognostic markers in rectal cancer MRI. Collectively, our results highlight the advantages of foundation model features, image harmonization, and multi-view fusion in improving diagnostic performance in rectal MRI analysis.

Keywords: rectal cancer, MRI, extramural vascular invasion, mesorectal fascia invasion, foundation model, UMedPT, multi-view fusion, deep learning

Introduction

Colorectal cancer is the third most commonly diagnosed malignancy and the second leading cause of cancer-related death worldwide, with rectal cancer comprising approximately one-third of these cases. ¹⁻³ Accurate staging is critical, as five-year survival rates exceed 90% for localized disease but decrease significantly in advanced stages. The introduction of total mesorectal excision (TME) and neoadjuvant chemoradiotherapy (nCRT) has substantially improved outcomes; however, their success hinges on precise staging and the reliable detection of high-risk prognostic markers, such as extramural vascular invasion (EVI) and mesorectal fascia invasion (MFI)^{4,5}.

Detection of EVI strongly influences therapeutic decisions, supporting the use of neoadjuvant chemoradiotherapy even in node-negative (cN0) patients to reduce local recurrence and systemic dissemination risks. Additionally, the presence of EVI may justify postoperative chemotherapy in Stage II patients exhibiting borderline features. Conversely, identification of MFI indicates a high risk of an involved circumferential resection margin (CRM), prompting guidelines to recommend neoadjuvant chemoradiotherapy to downstage the tumor and enhance the likelihood of a clear CRM during surgery.

MRI is recognized as the clinical gold standard for locoregional staging of rectal cancer due to its superior soft-tissue contrast and its ability to assess tumor boundaries, nodal involvement, and proximity to critical structures like the mesorectal fascia⁶. On T2-weighted MRI, EVI typically appears as nodular, beaded, or worm-like signal distortions⁷ reflecting tumor infiltration into vessels beyond the muscularis propria. In contrast, MFI manifests as disruption or blurring of the normally distinct mesorectal fascia, indicating tumor extension toward or into the fascia.

In clinical practice, axial and sagittal MRI views offer complementary perspectives for assessing both EVI and MFI. For EVI, axial views provide clearer visualization of cross-sectional vascular enhancement and extramural protrusions, while sagittal images might depict vascular structures and tumor extension as linear or subtle signals. Regarding MFI, axial T2-weighted images

excel at delineating circumferential tumor spread, measuring the shortest distance to the mesorectal fascia, and assessing the CRM. Sagittal images, despite providing less distinct fascial definition, are particularly beneficial for capturing the vertical extent of tumors and their anatomical relationship to the anal canal and pelvic floor. The integration of these anatomical differences from both axial and sagittal planes is essential for reliable and reproducible tumor staging, particularly in borderline or low rectal lesions.

While experienced radiologists can visually identify these features on T2-weighted MRI, interpretation remains subjective, variable across institutions, and often difficult in borderline cases. These limitations highlight the need for artificial intelligence (AI) -assisted tools to deliver objective and reproducible assessments in support of consistent staging decisions⁸.

In recent years, radiomics and deep learning have emerged as promising tools for extracting high-dimensional imaging features imperceptible to the human eye⁹⁻¹¹. These methods have shown clinical utility in predicting treatment response, nodal metastasis, and survival in rectal cancer patients¹². However, standard deep learning models are sensitive to inter-institutional variability in scanner type and acquisition protocols—an issue that becomes more pronounced in multi-center studies¹³. Harmonization strategies such as ComBat¹⁴ or CycleGAN¹⁵ have been proposed to mitigate these domain shifts, while newer frequency-domain and self-supervised approaches aim to learn modality-invariant representations with minimal supervision¹⁶.

Meanwhile, foundation models, pretrained on large-scale, diverse medical imaging datasets, have recently shown unprecedented generalizability and cross-task transferability. The universal biomedical pretrained model (UMedPT)¹⁷ model, for instance, encodes cross-domain prior knowledge and has outperformed conventional CNNs in a variety of clinical applications, enabling robust performance even with minimal fine-tuning. These capabilities make foundation models particularly suited for scalable and generalizable AI deployment in real-world clinical settings—including the complex task of EVI and MFI classification in rectal cancer.

In this study, we present a novel AI-based framework for automatic classification of EVI and MFI in rectal cancer using both axial and sagittal T2-weighted MRI from multiple centers. We leverage the UMedPT foundation model, as well as, a classification model that applies logistic regression on image features extracted from UMedPT (UMedPT_LR), benchmarking their performance against standard CNNs, including ResNet50 and SeResNet. To mitigate inter-institutional variability, we incorporate a frequency domain self-supervised harmonization strategy. Our goal is to advance the development of robust, generalizable AI systems for rectal cancer staging, with the potential to improve precision risk stratification and support individualized treatment planning in real-world clinical environments.

Methods

Dataset

This retrospective study utilized data from the Chaimeleon grand challenge, a large-scale initiative aimed at developing AI algorithms for cancer diagnosis across multiple tumor types, including lung, breast, prostate, colon and rectum cancers. The program also supports the

establishment of a standardized, privacy-preserving European infrastructure for medical imaging research. In particular, this study included 331 rectal cancer patients from three hospitals – La Fe University Hospital (n=194), ULS Porto (n=113), and CHU Angers (n=24) – utilizing MRI scanners from two main manufacturers – GE (n=221) and Siemens (n=106). Data was processed and securely managed on the Chaimeleon platform, ensuring patient confidentiality. Patient inclusion criteria were: (1) biopsy-proven rectal adenocarcinoma; (2) non-metastasized; (3) availability of pre-treatment multi-view T2-weighted MRI scans (axial and sagittal). Patients lacking baseline MRI scans (n=12) or with incomplete imaging sequences (n=4) were excluded, leaving 331 cases eligible for analysis. As illustrated in **Fig.1.**, the dataset was split into a train_val set (n=265, 80%) and a test set (n=66, 20%). Within the train_val cohort, 80% (n=212) was used for training and the remaining 20% (n=53) for validation. Expert radiologists annotated the presence or absence of EVI and MFI in all cases and manually adjusted to exclude non-tumorous structures. This structured approach facilitated rigorous evaluation and comparison of the proposed foundation model-based AI method for accurate classification of EVI and MFI in rectal cancer.

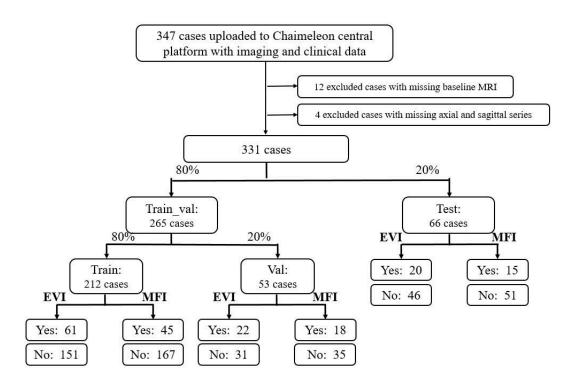


Fig.1. Flowchart of data inclusion, exclusion, and dataset stratification for model development and evaluation

Preprocessing

To prepare the T2-weighted MRI data for classification, all images were first converted from DICOM to NIfTI format. Rectum patches were then extracted through a structured multi-step pipeline. Since no public rectal cancer segmentation dataset was available, we used TotalSegmentator¹⁸ to perform automated anatomical segmentation on pelvic MRI, enabling the

identification of the colon mask. The rectal region was localized primarily based on the colon mask segmented by TotalSegmentator. Additionally, segmentation masks of adjacent anatomical structures, such as the bladder and prostate, were employed as auxiliary references for accurate localization in both axial and sagittal views. A region of interest (ROI) encompassing the rectum was obtained by locating the centroid of the rectal region within the colon mask, followed by center-cropping to a fixed patch size. This ensured consistent coverage and alignment of the rectum across patients, with the ROI centered in the image. The same center-cropping strategy was applied consistently to both axial and sagittal view images.

All volumes were resampled to a uniform voxel spacing to standardize resolution across scans. To account for inter-patient intensity variation, intensity normalization was applied to each image: values were clipped between the 2.5th and 97.5th percentiles, then normalized via Z-score scaling.

Frequency Domain Harmonization

To address inter-institutional variability introduced by divergent acquisition protocols in multi-center MRI datasets, we developed a frequency domain intensity harmonization pipeline, specifically tailored for rectal T2-weighted MRI. This method standardizes image contrast while preserving clinically relevant anatomical features essential for tumor staging and invasion assessment. A curated reference set of ten rectal MRI cases was selected based on consistent acquisition parameters (pulse sequence, TE, TR) and optimal image quality, as assessed by an expert radiologist. These reference scans were used to define the harmonization target space, representing the diversity of tissue-contrast distributions observed across sites.

Each 2D rectal MR slice was first transformed into the frequency domain using Fast Fourier Transform (FFT). Synthetic contrast perturbations were then applied using circular masks centered in the frequency matrix, with varying intensity and spatial frequency content, simulating realistic inter-site contrast differences. The perturbed frequency maps were then reconstructed into the spatial domain via inverse FFT, yielding 200 synthetic variants per slice while preserving anatomical integrity. A custom convolutional autoencoder, equipped with skip connections and trained under self-supervised learning, was then used to reconstruct harmonized images from the contrast-perturbed inputs. The model was optimized using a composite loss combining mean squared error and L2 regularization to ensure fidelity in image reconstruction. This harmonization strategy enables standardized image representation across heterogeneous acquisition protocols, improving the robustness of downstream AI classification tasks.

SeResNet

We employed a modified version of Squeeze-and-Excitation Residual Network-34 (SeResNet-34) as the backbone network for classification, tailored to the anisotropic spatial resolution typical in pelvic MRI. To preserve spatial detail while minimizing feature loss through early downsampling, we customized the network stem to use a convolutional kernel of size $3\times3\times1$ and applied anisotropic strides of (2,2,1) in the first two stages. Max pooling was omitted to retain fine

anatomical structures. LeakyReLU was used in place of ReLU to improve gradient flow in low-contrast regions.

The model comprises five sequential ResidualEncoder stages, with output channel sizes of 32, 64, 128, 256, and 512, respectively. The convolutional kernel size for the first two stages are $3\times3\times1$ and $3\times3\times3$ for the remaining ones. The stride configuration is (2,2,1) for the first two stages and the final stages, followed by (2,2,2) for the remaining stages. Each stage consists of 1, 3, 4, 6, and 3 residual blocks¹⁹, respectively. All residual units are enhanced with squeeze-and-excitation (SE)²⁰ modules, enabling adaptive channel-wise feature recalibration. Following the final stage, a 3D global average pooling layer compresses the spatial dimensions, and the resulting feature vector is passed to a fully connected layer with 512 units. A final sigmoid-activated output layer produces binary predictions for EVI or MFI classification. The full architecture is summarized in **Fig.2**.

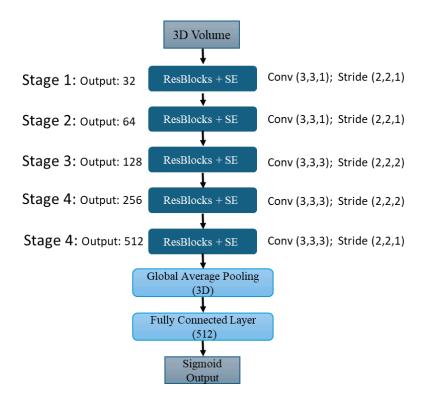


Fig.2. Architecture of the modified SeResNet used for EVI and MFI classification. The model consists of five convolutional stages, each comprising residual blocks enhanced with SE modules. Anisotropic convolution kernels and strides (e.g., $3 \times 3 \times 1$ with stride $2 \times 2 \times 1$) are used in the early stages to accommodate the non-isotropic resolution of pelvic MRI. The final feature map is globally pooled and passed through a 512-unit fully connected layer, followed by a sigmoid output for binary classification. Channel dimensions are progressively expanded across stages, as shown on the left.

Foundation Model

UMedPT

UMedPT¹⁷ is a deep-learning foundation model designed to learn transferable representations across diverse biomedical imaging modalities and tasks. In this study, we adapt UMedPT for the classification of EVI and MFI from multi-slice T2-weighted MRI volumes.

As illustrated in **Fig. 3**, the input 3D volume is first decomposed into a stack of 2D slices. Each slice is processed independently through a shared encoder-squeezer module pretrained within the UMedPT framework. The encoder, based on the Swin Transformer architecture, extracts high-level modality-agnostic features and outputs a hierarchical representation, which is compressed by the squeezer into a 512-dimensional feature vector. Only the encoder and squeezer blocks were retained for feature extraction, as our downstream task does not require the original model's segmentation or localization decoders. To aggregate slice-level features into a volume-level representation, we use a task-specific "grouper" module provided by UMedPT. The grouper computes a learned weighted average of all 512-dimensional slice embeddings within a volume, producing a single accumulated feature vector. This representation is then passed to a lightweight multilayer perceptron (MLP) classifier that outputs binary predictions for EVI or MFI.

During fine-tuning, only the grouper and classifier layers were optimized, while the encoder and squeezer components were frozen to preserve pretrained feature representations. Input images were duplicated across three channels to conform with the encoder's requirements. This architecture enables efficient transfer learning and robust volume-level classification, while minimizing the number of trainable parameters required for domain adaptation.

UMedPT LR

Unlike the MLP-based classifier used in the full UMedPT pipeline, which fine-tunes the grouper and classification head while keeping the shared encoder frozen, UMedPT_LR is a simplified classification model that uses the final 512-dimensional features aggregated by the pretrained UMedPT grouper. In UMedPT_LR, all components of the foundation model, including the encoder, squeezer, and grouper, remain completely frozen. These extracted features are first reduced via Principal Component Analysis (PCA), and then fed into a logistic regression classifier for training.

This streamlined setup offers several practical advantages: it significantly reduces training time, simplifies the overall pipeline, and facilitates rapid experimentation with minimal computational cost. Moreover, the logistic regression model enhances interpretability, as each coefficient corresponds to a contribution of a specific principal component to the prediction—an important consideration for clinical trust and deployment.

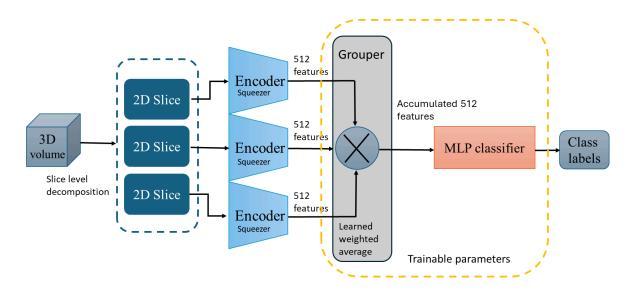


Fig.3. UMedPT architecture for rectal cancer classification using multi-slice MRI. Each 3D MRI volume is decomposed into 2D slices, which are individually encoded and compressed via shared encoders and squeezer modules pretrained in UMedPT. The extracted 512-dimensional features from all slices are aggregated using a learned weighting strategy in the Grouper module. The accumulated representation is then passed to a multi-layer perceptron (MLP) classifier to predict the presence or absence of extramural vascular invasion (EVI) and mesorectal fascia invasion (MFI). During fine-tuning, only the grouper and classifier parameters are updated, while the backbone encoder remains frozen.

Training Strategy

To enhance robustness and generalization, data augmentation was performed using MONAI's framework. Augmentations included random rotation, scaling, shearing, flipping, intensity changes (contrast, gamma, brightness), and MRI-specific artifacts such as Gibbs ringing and bias field inhomogeneity. Patch sizes of $192 \times 192 \times 36$ consistently achieved optimal performance and were used throughout.

SeResNet was optimized using α -balanced focal loss and trained for 200 epochs using SGD with a learning rate of 1×10^{-3} . A warm-up of 10 epochs and ReduceLROnPlateau scheduling were used. UMedPT used a weighted binary cross-entropy loss and AdamW optimizer with a learning rate of 5×10^{-4} . Cosine annealing was used for learning rate decay. Both models were trained with a batch size of 8 using half-precision floating-point arithmetic.

Combination of Axial and Sagittal T2w Images

To fully exploit the complementary information offered by orthogonal imaging planes, we extracted 512-dimensional feature vectors independently from axial and sagittal T2-weighted MR images using the UMedPT foundation model. These features were concatenated to form a single 1024-dimensional representation, integrating both horizontal and vertical anatomical perspectives. To mitigate redundancy and reduce computational complexity, we applied principal component analysis (PCA) to project the concatenated features into a lower-dimensional space. The resulting compressed representations were then fed into a logistic regression model for classification. As illustrated in **Fig.**

4, the axial (a, b) and sagittal (c, d) views provide distinct anatomical context and tumor morphology, suggesting that multi-view fusion may enhance the model's discriminatory capability.

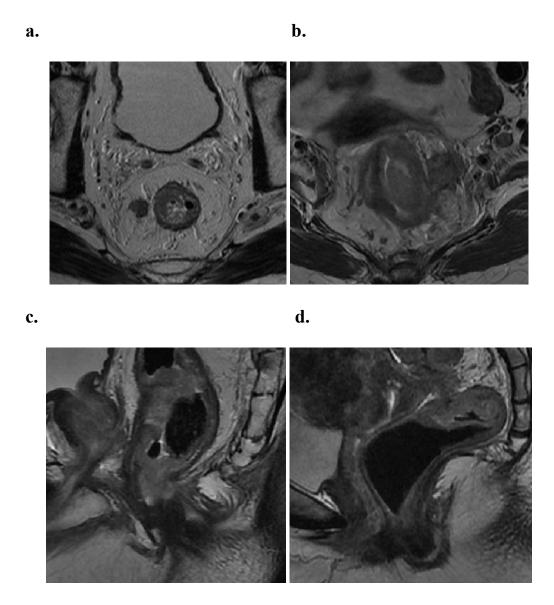


Fig.4. An example image (a) Axial T2w for a patient with EVI+, MFI- (b) Axial T2w for a patient with EVI-, MFI+ (c) Saggital T2w for a patient with EVI+, EVI- (d) Saggital T2w for a patient with EVI-, EVI+

Evaluation Metrics

To comprehensively evaluate model performance on the binary classification task of predicting EVI and MFI, we adopted a set of commonly used metrics consistent with the Chaimeleon challenge guidelines. The primary metric was the AUC score, which reflects the model's ability to discriminate between positive and negative classes across different thresholds. Since AUC alone can be biased in the presence of class imbalance, we also report balanced accuracy, sensitivity, specificity, and F1 score for a more complete assessment. Sensitivity (also known as recall or true positive rate)

and specificity (true negative rate) are defined as equations (1) and (2), respectively. Balanced accuracy accounts for both sensitivity and specificity, offering a fairer evaluation when class distributions are skewed (equation 3). The F1 score provides the harmonic mean of precision and recall (equation 4). To accurately estimate the variability and robustness of these metrics, we employed a bootstrap resampling method with 1000 iterations, computing the 95% confidence intervals (CI) for each metric. These bootstrap-derived confidence intervals offer valuable insight into the precision and reliability of our performance estimates.

Sensitivity (also known as recall or true positive rate) and specificity (true negative rate) are defined as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2}$$

Balanced accuracy accounts for both sensitivity and specificity, offering a fairer evaluation when class distributions are skewed:

$$Balanced\ Accuracy\ =\ \frac{(Sensitivity + Specificity)}{2} \tag{3}$$

The F1 score provides the harmonic mean of precision and recall:

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{4}$$

Results

Patient Characteristics

The characteristics of the patient cohort are summarized in Table 1. The median age of the patients was 65 years (range: 28–89 years), and 60.7% (n=201) were male. Extramural vascular invasion (EVI) was positive in 31.1% (n=103) of cases, while mesorectal fascia invasion (MFI) was positive in 23.6% (n=78). Patients were recruited primarily from three medical centers: LaFe (58.6%, n=194), ULS (34.1%, n=113), and CHU Angers (7.3%, n=24). MRI scans were predominantly obtained using GE scanners (66.8%, n=221), followed by Siemens scanners (32.0%, n=106), with a small minority from other manufacturers (1.2%, n=4). The demographic and clinical characteristics were consistent across the train val, and test cohorts.

Table 1: Patient characteristics

Characteristics	All	Train_Val	Test
Number of Patients	331	265	66
Age (median, range)	65 (28-89)	65 (29-89)	65 (28-87)
Gender (male)	201 (60.7%)	162 (61.1%)	39 (59.1%)
Type of Rectal Cancer			
EVI (+)	103 (31.1%)	83 (31.3%)	20 (30.3%)
MFI (+)	78 (23.6%)	63 (23.8%)	15 (22.7%)
Medical Center			
01: LaFe	194 (58.6%)	158 (59.6%)	36 (54.5%)
03: ULS	113 (34.1%)	87 (32.8%)	26 (39.4%)
06: CHU Angers	24 (7.3%)	20 (7.5%)	4 (6.1%)
Manufacturer			
GE	221 (66.8%)	183 (69.1%)	38 (57.6%)
Siemens	106 (32.0%)	80 (30.2%)	26 (39.4%)
Other	4 (1.2%)	2 (0.7%)	2 (3.0%)

Values in age parentheses are the minimum and maximum. Values in parentheses of other items are the percentages. EVI+: EVI positive. MFI+: MFI positive.

Rectum patch extraction

To ensure consistent localization of the rectal region across patients, we employed a standardized center-cropping approach based on anatomical segmentation using TotalSegmentator. As shown in **Fig. 5**, both axial and sagittal T2-weighted MR images were processed: the rectum was first localized within the colon mask, followed by extraction of a fixed-size patch centered on the rectum. This strategy effectively aligned the rectal region in all cases, enabling robust downstream analysis.

While the colon mask was consistently segmented across all cases, we observed occasional inaccuracies or absence of prostate and bladder segmentation in a small number of examples. Nevertheless, the strong anatomical contrast of rectal structures on T2-weighted images, combined with consistent colon localization, allowed accurate patch extraction even when auxiliary organ masks were suboptimal.

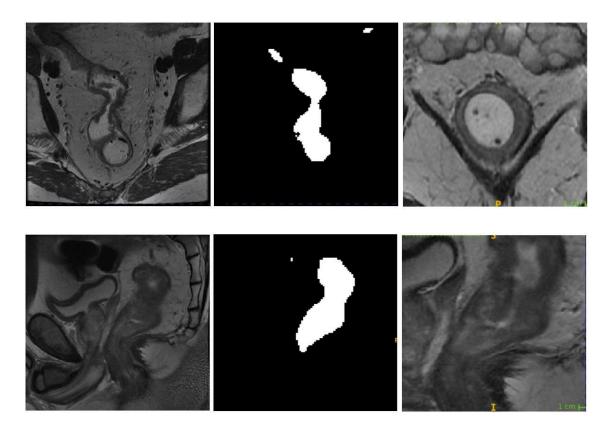


Fig. 5. Rectum patch extraction and anatomical localization using TotalSegmentator on T2-weighted MRI. The first row shows the axial view: left – the original pelvis MR image; middle – the colon mask segmented by TotalSegmentator, from which the rectal region was anatomically localized; right – the rectum patch after center cropping, ensuring the rectum is centered. The second row shows the sagittal view.

Harmonization Results

The frequency domain harmonization pipeline effectively reduced inter-scanner variability and enhanced visual consistency in pelvic MRI datasets, as demonstrated in **Fig. 6.** In both axial and sagittal T2-weighted MR images, harmonization significantly mitigated intensity discrepancies and standardized tissue contrast across different imaging protocols. Anatomical features critical for clinical assessment, such as tumor margins and adjacent soft tissue interfaces, became more visually consistent post-harmonization, facilitating improved feature extraction and reliable downstream AI classification tasks.

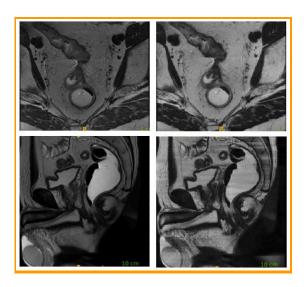


Fig.6. Axial and sagittal pelvic T2-weighted MR images before (left) and after (right) frequency domain harmonization. Images from representative patients highlight reduced variability in image contrast and improved visual consistency post-harmonization, enhancing comparability across heterogeneous acquisition protocols.

Classification Results

We evaluated four deep learning models—Resnet50, SeResnet, UMedPT, and UMedPT_LR—on their ability to classify EVI and MFI using axial and sagittal T2-weighted MR image patches, with and without harmonization.

For overall performance comparison, our models achieved superior results in both EVI and MFI classification tasks compared to the Chaimeleon challenge winners²¹. For EVI, the best performance was achieved using the UMedPT_LR model with combined axial and sagittal original images, yielding an AUC of 0.82, sensitivity of 0.75, F1 score of 0.73, and balanced accuracy of 0.72. In comparison, the challenge winner's model reported lower metrics, with an AUC of 0.74, sensitivity of 0.50, F1 score of 0.50, and balanced accuracy of 0.64, indicating weaker sensitivity and overall classification consistency. For MFI, our best result came from the UMedPT model using axial harmonized images, which achieved an AUC of 0.77, sensitivity of 0.73, specificity of 0.70, F1 score of 0.54, and balanced accuracy of 0.72. This again outperformed the challenge winner's model, which reported an AUC of 0.75, sensitivity of 0.53, specificity of 0.86, F1 score of 0.53, and balanced accuracy of 0.70. These findings demonstrate that our models not only generalize well across orientations and harmonization conditions but also surpass existing state-of-the-art benchmarks in both detection accuracy and robustness.

For EVI classification performance, across all axial original patches (**Table 2**), UMedPT_LR consistently outperformed other models, achieving the highest AUC (0.76) and balanced accuracy (0.70). In contrast, Resnet50 showed poor sensitivity (0.0) despite high specificity (0.98), suggesting weak generalizability. SeResnet and UMedPT also demonstrated moderate performance. On axial harmonized patches (**Table 2**), performance declined for most models. UMedPT_LR's AUC dropped from 0.76 to 0.64, and its F1 score decreased from 0.59 to 0.43, indicating that harmonization may impair EVI classification in axial views. For sagittal original patches (**Table 4**), UMedPT_LR again showed the best results (AUC 0.73, balanced accuracy 0.68), confirming its robustness across

orientations. However, on sagittal harmonized patches (**Table 4**), performance degraded for all models, with both UMedPT and UMedPT LR showing zero sensitivity in some cases.

For MFI classification performance, as shown in **Table 3**, UMedPT_LR outperformed other models on axial original patches for MFI, with a balanced accuracy of 0.71 and AUC of 0.75. On axial harmonized patches, UMedPT achieved the highest AUC (0.77) and balanced accuracy (0.72), suggesting harmonization may benefit MFI detection under certain conditions. Sagittal original patch results (**Table 5**) showed relatively uniform performance across models, though UMedPT_LR again slightly led (AUC 0.62). Harmonization of sagittal patches (**Table 5**) yielded mixed results, with no substantial gain and even declines in F1 scores, especially for UMedPT.

Harmonization yielded no consistent improvement and often degraded performance, particularly for EVI detection. It was beneficial only for MFI classification using UMedPT on axial patches (**Table 3**), where it led to the highest AUC observed across all MFI conditions (0.77).

Table 2 EVI's diagnostic performance of different model on axial T2w Original and Harmonized images

	Axial original patch							
Model	AUC (95% CI)	Sensitivity (95% CI)	* `		Balanced_acc (95% CI)			
Resnet50	0.57 (0.40, 0.70)	0 (0, 0)	0.98 (0.93, 1.0)	0 (0, 0)	0.49 (0.46, 0.50)			
SeResnet	0.57 (0.41, 0.71)	0.70 (0.47, 0.89)	0.44 (0.31, 0.59)	0.48 (0.30, 0.63)	0.57 (0.44, 0.70)			
UMedPT	0.73 (0.57, 0.86)	0.65 (0.43, 0.85)	0.67 (0.53, 0.81)	0.54 (0.35, 0.70)	0.66 (0.53, 0.78)			
UMedPT_ LR	0.76 (0.62, 0.89)	0.60 (0.35, 0.82)	0.80 (0.68, 0.91)	0.59 (0.36, 0.74)	0.70 (0.58, 0.82)			
		Axial harr	nonized patch					
Model	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 (95% CI)	Balanced_acc (95% CI)			
Resnet50	0.54 (0.37, 0.70)	0 (0, 0)	1.0 (1.0, 1.0)	0 (0, 0)	0.50 (0.50, 0.50)			
SeResnet	0.60 (0.45, 0.75)	0.85 (0.67, 1.0)	0.36 (0.22, 0.50)	0.52 (0.36, 0.65)	0.60 (0.49, 0.70)			
UMedPT	0.65 (0.49, 0.80)	0.60 (0.37, 0.80)	0.64 (0.51, 0.79)	0.50 (0.32, 0.67)	0.62 (0.49, 0.75)			
UMedPT_ LR	0.64 (0.47, 0.78)	0.45 (0.22, 0.67)	0.71 (0.58, 0.84)	0.43 (0.21, 0.61)	0.58 (0.45, 0.71)			

Table 3 MFI's diagnostic performance of different model on axial T2w images Original and Harmonized

	Axial original patch							
Model	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 (95% CI)	Balanced_acc (95% CI)			
Resnet50	0.40 (0.24, 0.55)	1.0 (1.0, 1.0)	0 (0,0)	0.38 (0.24, 0.51)	0.50 (0.50, 0.50)			
SeResnet	0.75 (0.61, 0.87)	0.73 (0.50, 0.93)	0.64 (0.50, 0.78)	0.50 (0.30, 0.67)	0.69 (0.55, 0.81)			
UMedPT	0.70 (0.55, 0.84)	0.60 (0.36, 0.83)	0.66 (0.53, 0.78)	0.44 (0.24, 0.62)	0.63 (0.50, 0.78)			
UMedPT_L R	0.75 (0.60, 0.89)	0.60 (0.33, 0.85)	0.82 (0.71, 0.93)	0.55 (0.31, 0.74)	0.71 (0.57, 0.85)			
	Axial harmonized patch							
Model	AUC (95% CI) Sensitivity (95% CI) Specificity (95% F1 (95% CI) Balanced (95% CI)							
Resnet50	0.79 (0.64, 0.90)	0 (0,0)	1.0 (1.0, 1.0)	0 (0,0)	0.50 (0.50, 0.50)			
SeResnet	0.74 (0.58, 0.87)	0.60 (0.35, 0.84)	0.70 (0.57, 0.82)	0.46 (0.26, 0.65)	0.65 (0.52, 0.79)			
UMedPT	0.77 (0.65, 0.89)	0.73 (0.50, 0.94)	0.70 (0.57, 0.82)	0.54 (0.34, 0.71)	0.72 (0.59, 0.85)			
UMedPT_L R	0.72 (0.59, 0.86)	0.40 (0.14, 0.67)	0.90 (0.81, 0.97)	0.46 (0.19, 0.68)	0.65 (0.52, 0.79)			

Table 4 EVI's diagnostic performance of different model on sagittal T2w images Original and Harmonized

Sagittal original patch							
Model	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 (95% CI)	Balanced_acc (95% CI)		
Resnet50	0.55 (0.38, 0.69)	0.05 (0, 0.16)	1.0 (1.0, 1.0)	0.10 (0, 0.28)	0.53 (0.5, 0.58)		
SeResnet	0.52 (0.35, 0.68)	0.55 (0.31, 0.75)	0.58 (0.43, 0.72)	0.44 (0.25, 0.59)	0.56 (0.42, 0.69)		
UMedPT	0.67 (0.53, 0.82)	0.50 (0.28, 0.74)	0.71 (0.58, 0.84)	0.47 (0.26, 0.64)	0.61 (0.47, 0.74)		
UMedPT_ LR	0.73 (0.60, 0.85)	0.60 (0.38, 0.81)	0.76 (0.63, 0.87)	0.56 (0.35, 0.73)	0.68 (0.55, 0.80)		
Sagittal harmonized patch							
Model	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 (95% CI)	Balanced_acc (95% CI)		

Resnet50	0.66 (0.52, 0.80)	0.80 (0.62, 0.95)	0.60 (0.45, 0.73)	0.59 (0.43, 0.73)	0.70 (0.58, 0.81)
SeResnet	0.63 (0.47, 0.79)	0.70 (0.49, 0.91)	0.53 (0.39, 0.68)	0.51 (0.33, 0.67)	0.62 (0.49, 0.74)
UMedPT	0.53 (0.36, 0.69)	0 (0, 0)	1.0 (1.0, 1.0)	0 (0, 0)	0.50 (0.50, 0.50)
UMedPT_ LR	0.55 (0.38, 0.72)	0.35 (0.15, 0.57)	0.80 (0.69, 0.91)	0.39 (0.18, 0.57)	0.58 (0.45, 0.69)

Table 5 MFI's diagnostic performance of different model on sagittal T2w images Original and Harmonized

	Sagittal original patch							
Model	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 (95% CI)	Balanced_acc (95% CI)			
Resnet50	0.45 (0.27, 0.65)	1.0 (1.0, 1.0)	0 (0, 0)	0.38 (0.24, 0.51)	0.50 (0.50, 0.50)			
SeResnet	0.41 (0.27, 0.59)	0.87 (0.67, 1.0)	0.14 (0.06, 0.24)	0.37 (0.22, 0.50)	0.50 (0.39, 0.60)			
UMedPT	0.61 (0.45, 0.78)	0.53 (0.26, 0.79)	0.66 (0.53, 0.80)	0.40 (0.19, 0.59)	0.60 (0.45, 0.74)			
UMedPT_ LR	0.62 (0.46, 0.78)	0.40 (0.14, 0.67)	0.80 (0.68, 0.91)	0.39 (0.16, 0.59)	0.60 (0.47, 0.74)			
		Sagittal harr	nonized patch					
Model	el AUC (95% CI) Sensitivity (95% CI) Specificity (95% F1 (95% CI) Balanced_ac (95% CI)							
Resnet50	0.31 (0.14, 0.51)	1.0 (1.0, 1.0)	0 (0, 0)	0.38 (0.24, 0.51)	0.50 (0.50, 0.50)			
SeResnet	0.73 (0.56, 0.89)	0.73 (0.50, 0.95)	0.70 (0.57, 0.83)	0.54 (0.33, 0.71)	0.72 (0.58, 0.85)			
UMedPT	0.59 (0.44, 0.75)	0.53 (0.29, 0.80)	0.62 (0.48, 0.76)	0.38 (0.18, 0.56)	0.58 (0.43, 0.72)			
UMedPT_ LR	0.60 (0.46, 0.76)	0.27 (0.07, 0.50)	0.76 (0.64, 0.88)	0.26 (0.07, 0.46)	0.51 (0.40, 0.65)			

Table 6 Diagnostic performance of multi-view feature fusion for UMedPT_LR model

feature combination	overall score	auc_evi	auc_mfi	sensitivity _evi	sensitivity _mfi	f1_evi	fl_mfi
axial + sagittal	0.76	0.82	0.71	0.75	0.73	0.73	0.51
axial harmonized + sagittal harmonized	0.59	0.58	0.75	0.35	0.07	0.37	0.13

Combining axial and sagittal views in UMedPT_LR substantially enhanced EVI classification, achieving the highest AUC (0.82), sensitivity (0.75), and F1 score (0.73), as shown in **Table 6**. This was superior to any single-view configuration. However, combining only harmonized views (axial + sagittal) reduced performance (AUC 0.58), suggesting the fusion of original data is more effective.

Receiver operating characteristic (ROC) curves further illustrated these findings. As shown in **Fig.7.** (left), the best EVI classification performance was achieved using UMedPT_LR with both axial and sagittal views. For MFI (**Fig. 7.**, right), UMedPT with axial harmonized patches had the highest curve, consistent with its AUC (0.77).

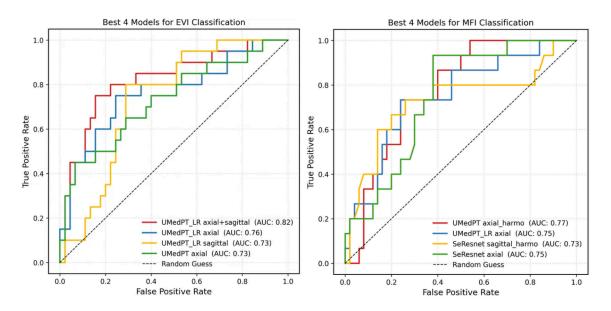


Fig.7. Receiver operating characteristic (ROC) curves of the top four models for classifying extramural vascular invasion (EVI, left) and mesorectal fascia invasion (MFI, right) on the Chaimeleon's testset. For EVI, UMedPT_LR combining axial and sagittal views achieved the highest AUC (0.82). For MFI, the best performance was achieved by UMedPT using axial harmonized patch (AUC: 0.77). "Axial" refers to the original axial patch, and "axial_harmo" indicates the axial harmonized patch.

Fig.8. illustrates one correctly classified (left) and one incorrectly classified (right) example by the UMedPT model on axial T2-weighted MR images. In the left case, both EVI and MFI were correctly predicted with high probabilities (0.76 and 0.79, respectively). A key factor contributing to the correct classification may be the clearer delineation of perirectal tissue planes and a more distinct signal intensity alteration surrounding the rectal wall. Specifically, the left image exhibits a conspicuous disruption of the outer muscularis layer and possible linear vascular signal extending beyond the rectum, which are characteristic imaging signs of EVI. Similarly, the rectal wall abuts or disrupts the mesorectal fascia with intermediate-to-high signal intensity—features suggestive of MFI.

In contrast, the right case was misclassified, with EVI predicted at only 0.46 and MFI at 0.54, despite both being truly positive. This discrepancy could be due to several complicating factors: the image demonstrates irregular signal intensities and possible mucinous or necrotic components within the tumor, which may obscure vascular invasion cues. Moreover, the mesorectal fascia appears

compressed but not clearly interrupted, potentially leading to ambiguity in detecting MFI. The absence of strong linear enhancement or extramural signal disruption may reduce the model's confidence in identifying EVI. Additionally, image artifacts, motion blur, or partial volume effects could further confound the anatomical boundaries, decreasing the model's predictive certainty.

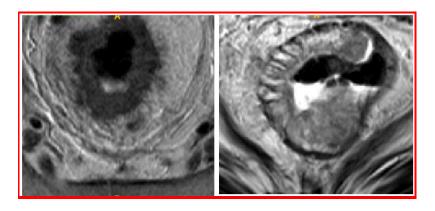


Fig. 8. One example of correct and incorrect predictions from the UMedPT model on axial original T2-weighted MRI images. The left image shows a case labeled as EVI+ and MFI+, with the model predicting a high probability for both conditions (predicted probability: EVI = 0.76, MFI = 0.79). The right image, from another patient also labeled as EVI+ and MFI+, demonstrates a misclassification by the EVI model, with predicted probabilities of EVI = 0.46 and MFI = 0.54, respectively.

Discussion

Our study presents a foundation model-based AI framework that reliably classifies EVI and MFI in multi-center rectal cancer MRI using both axial and sagittal T2-weighted sequences. The UMedPT and UMedPT_LR models demonstrated superior diagnostic performance on most experiments in this study compared to conventional CNNs such as ResNet50 and SeResNet. For example, for EVI best model, UMedPT_LR combined on axial and sagittal original images achieved an AUC of 0.82, and for MFI best model, UMedPT on axial harmonized images achieved an AUC of 0.77. Notably, the best-performing models for EVI and MFI individually outperformed the winning submissions of the Chaimeleon Challenge²¹, which achieves AUC 0.74 for EVI and AUC 0.75 for MFI on testset.

Our comparative experiments further underline the strength of foundation models in both feature extraction and generalizability, especially when applied to harmonized imaging data. This is particularly evident for UMedPT_LR, which showed superior performance even with relatively simple downstream classifiers. These results not only underscore the potential of foundation models in medical imaging, but also provide a promising direction for future lightweight and generalizable diagnostic tools.

Importantly, we systematically evaluated the role of image harmonization across multiple model backbones. For sagittal images, harmonization led to notable performance improvements—raising AUC from approximately 0.5 to 0.6 for EVI classification using ResNet-based models, and from 0.41 to 0.73 for MFI classification using SeResNet. However, harmonization did not universally improve performance, particularly in axial views, suggesting that

residual differences in image acquisition protocols or harmonization accuracy may affect downstream model efficacy.

Analysis of imaging plane-specific features revealed that the performance of sagittal-based models was generally inferior to axial-based models for EVI classification, particularly in foundation model settings. This may be attributed to the typical imaging presentation of EVI, which often involves abnormal vascular structures more clearly visualized in axial views, whereas these features may appear as indistinct points or short lines in sagittal views. In contrast, for MFI classification, although axial images still generally yielded better performance, sagittal views also proved informative. This is likely due to the morphology of MFI: in axial images, MFI often appears as a thickened circular band surrounding the rectum, while in sagittal images, it typically presents as two extended linear structures encasing the rectum, providing complementary anatomical information. Furthermore, sagittal MR sequences contain fewer slices encompassing the rectum compared to axial sequences, which may limit the spatial context available to the model, contributing to the overall reduction in model performance on sagittal inputs.

Moreover, isolating regions of interest is essential to reduce computational complexity and improve AI model performance. By narrowing the model's focus to anatomically and pathologically relevant structures, rectum ROI isolation minimizes background noise and improves diagnostic precision. In this study, TotalSegmentator proved valuable for initial rectum segmentation, allowing consistent ROI extraction across subjects. Additionally, attention-based methods such as MSBC-Net have previously demonstrated that precise localization of rectal anatomy is essential for effective downstream classification, especially in complex pelvic structures²². The availability of publicly annotated datasets could further improve segmentation quality and standardization in future studies.

Finally, we explored a multi-view fusion strategy that combined axial and sagittal features, revealing additive gains in performance. For example, EVI classification improved from AUCs of 0.76 and 0.73 (axial and sagittal, respectively) to 0.82 when fused. For MFI, fusion improved performance to 0.75 compared to 0.72 (axial) and 0.60 (sagittal) separately. This suggests that while each view captures distinct anatomical features, their combination provides complementary information that enhances model robustness.

Collectively, our results highlight the advantages of foundation model features, image harmonization, and multi-view fusion in improving diagnostic performance in rectal MRI analysis. To our knowledge, this is one of the first studies to extensively leverage both axial and sagittal sequences alongside foundation models in this clinical context.

However, this study has several limitations. First, although the data were collected from three different centers, we did not perform a detailed analysis on how data from each center may have influenced the results. The total dataset size was 331, with only 66 cases in the smallest test set, which could make the model results less stable. However, the strong feature extraction ability of foundation models may have helped reduce the impact of limited data. Second, the effects of image harmonization were not consistent across all tasks. While harmonization improved model performance in some settings, it may have also reduced important diagnostic signals in others. More work is needed to improve harmonization methods so that they balance consistency with keeping useful features. Third, EVI and MFI may require different patch sizes or model designs. EVI lesions are often small and vascular, while MFI appears larger and more widespread. Using the same model and patch size for both tasks may limit performance, and task-specific settings might improve

accuracy. Fourth, although UMedPT_LR performed well, model interpretability remains a challenge. Adding explainable AI techniques could improve clinical trust. For example, previous work by Cai et al. ²³ showed that explainable AI tools can maintain strong diagnostic performance while helping users understand model decisions. Finally, we did not compare our model with radiomics-based methods or include clinical data in the analysis. Earlier studies, such as the one by Shin et al. ²⁴, showed that radiomics could outperform radiologists in treatment response assessment, though the results were sensitive to input type. Combining foundation models with radiomics and clinical information may further improve future models. In addition, prospective validation of this algorithm would be valuable—for example, in the context of an in silico trial comparing diagnostic performance across three arms: radiologist alone, AI alone, and radiologist assisted by AI, as previously demonstrated in our prior studies (Sergei et al²⁵; Xian et al²⁶.).

Conclusion

In summary, our study demonstrates that foundation model-based AI frameworks, particularly UMedPT and UMedPT_LR, achieve superior performance in classifying EVI and MFI from multi-center rectal MRI. The integration of axial and sagittal views, harmonized inputs, and rectum patch isolate further enhances diagnostic accuracy. These findings highlight the potential of foundation models in rectal cancer staging and support their future application in clinical workflows.

Grants and Founding

Authors acknowledge financial support from ERC advanced grant (ERC-ADG-2015 n° 694812 - Hypoximmuno),, ERC-2020-PoC: 957565-AUTO.DISTINCT. Authors also acknowledge financial support from the European Union's Horizon research and innovation programme under grant agreement: CHAIMELEON n° 952172 (main contributor), ImmunoSABR n° 733008, EuCanImage n° 952103, TRANSCAN Joint Transnational Call 2016 (JTC2016 CLEARLY n° UM 2017-8295), IMI-OPTIMA n° 101034347, AIDAVA (HORIZON-HLTH-2021-TOOL-06) n°101057062, REALM (HORIZON-HLTH-2022-TOOL-11) n° 101095435, RADIOVAL (HORIZON-HLTH-2021-DISEASE-04-04) n°101057699 and EUCAIM (DIGITAL-2022-CLOUD-AI-02) n°101100633. This study was also supported by the China Scholarship Council grant (202208110055).

Disclosures

Disclosures from the last 36 months within and outside the submitted work: none related to the current manuscript; outside of current manuscript: grants/sponsored research agreements from Radiomics SA, Convert Pharmaceuticals and LivingMed Biotech. He received a presenter fee (in cash or in kind) and/or reimbursement of travel costs/consultancy fee (in cash or in kind) from Radiomics SA, BHV & Roche. PL has shares in the companies Radiomics SA, Convert pharmaceuticals, Comunicare, LivingMed Biotech, BHV and Bactam. PL is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248 and PCT/NL2014/050728), licensed to Radiomics SA; one issued patent on mtDNA (PCT/EP2014/059089), licensed to ptTheragnostic/DNAmito; one non-issued patent on LSRT (PCT/ P126537PC00, US: 17802766), licensed to Varian; three non-patented inventions (softwares) licensed to ptTheragnostic/DNAmito, Radiomics SA and Health Innovation

Ventures and two non-issued, non-licensed patents on Deep Learning-Radiomics (N2024482, N2024889). He confirms that none of the above entities were involved in the preparation of this paper.

References

- Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 74, 229–263 (2024).
- 2. Keller, D. S., Berho, M., Perez, R. O., Wexner, S. D. & Chand, M. The multidisciplinary management of rectal cancer.

 Nat. Rev. Gastroenterol. Hepatol. 17, 414–429 (2020).
- 3. Morgan, E. *et al.* Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* 72, 338–344 (2023).
- van den Broek, J. J., van der Wolf, F. S. W., Heijnen, L. A. & Schreurs, W. H. The prognostic importance of MRI detected extramural vascular invasion (mrEMVI) in locally advanced rectal cancer. *Int. J. Colorectal Dis.* 35, 1849–1854 (2020).
- 5. Paul, S. *et al.* Extramural vascular invasion as an independent prognostic marker in locally advanced rectal cancer: propensity score match pair analysis. *Abdom. Radiol. (NY)* **47**, 3671–3678 (2022).
- Horvat, N., Carlos Tavares Rocha, C., Clemente Oliveira, B., Petkovska, I. & Gollub, M. J. MRI of rectal cancer:
 Tumor staging, imaging techniques, and management. *Radiographics* 39, 367–387 (2019).
- Inoue, A. et al. MRI-detected extramural venous invasion of rectal cancer: Multimodality performance and implications
 at baseline imaging and after neoadjuvant therapy. Insights Imaging 12, 110 (2021).
- 8. Zhang, X.-Y. *et al.* MRI of extramural venous invasion in locally advanced rectal cancer: Relationship to tumor recurrence and overall survival. *Radiology* **289**, 677–685 (2018).
- 9. Lambin, P. *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).
- 10. Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446 (2012).
- 11. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.

 Nat. Commun. 5, 4006 (2014).
- 12. Jiang, X. et al. An MRI Deep learning model predicts outcome in rectal cancer. Radiology 307, e222223 (2023).
- 13. Wichtmann, B. D. *et al.* Are we there yet? The value of deep learning in a multicenter setting for response prediction of locally advanced rectal cancer to neoadjuvant chemoradiotherapy. *Diagnostics (Basel)* **12**, 1601 (2022).
- 14. Pomponio, R. *et al.* Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage* **208**, 116450 (2020).
- 15. Modanwal, G., Vellal, A., Buda, M. & Mazurowski, M. A. MRI image harmonization using cycle-consistent generative

- adversarial network. in *Medical Imaging 2020: Computer-Aided Diagnosis* (eds. Hahn, H. K. & Mazurowski, M. A.) (SPIE, 2020). doi:10.1117/12.2551301.
- 16. Tian, C. *et al.* Deep learning models for preoperative T-stage assessment in rectal cancer using MRI: exploring the impact of rectal filling. *Front. Med. (Lausanne)* **10**, 1326324 (2023).
- 17. Schäfer, R. *et al.* Overcoming data scarcity in biomedical imaging with a foundational multi-task model. *Nat. Comput. Sci.* **4**, 495–509 (2024).
- 18. Akinci D'Antonoli, T. *et al.* TotalSegmentator MRI: Robust sequence-independent segmentation of multiple anatomic structures in MRI. *Radiology* **314**, e241613 (2025).
- 19. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (IEEE, 2016). doi:10.1109/cvpr.2016.90.
- Hu, J., Shen, L., Albanie, S., Sun, G. & Wu, E. Squeeze-and-Excitation Networks. arXiv [cs.CV] (2017) doi:10.48550/ARXIV.1709.01507.
- CHAIMELEON Open Challenges Grand Challenge. grand-challenge.org
 https://chaimeleon.grand-challenge.org/evaluation/e0dc2f29-f8ba-4184-948f-327cc6877982/.
- 22. Meng, P. *et al.* MSBC-Net: Automatic rectal cancer segmentation from MR scans. (2021) doi:10.36227/techrxiv.16577417.v1.
- 23. Cai, L. *et al.* An automated deep learning pipeline for EMVI classification and response prediction of rectal cancer using baseline MRI: a multi-centre study. *NPJ Precis. Oncol.* **8**, 17 (2024).
- 24. Shin, J. *et al.* MRI radiomics model predicts pathologic complete response of rectal cancer following chemoradiotherapy. *Radiology* **303**, 351–358 (2022).
- 25. Primakov, S. P. *et al.* Automated detection and segmentation of non-small cell lung cancer computed tomography images. *Nat. Commun.* **13**, (2022).
- 26. Zhong, X. et al. Methodological explainability evaluation of an interpretable deep learning model for post-hepatectomy liver failure prediction incorporating counterfactual explanations and layerwise relevance propagation: A prospective in silico trial. arXiv [cs.CV] (2024) doi:10.48550/ARXIV.2408.03771.