An MRI Foundation Model for Versatile Clinical Applications in Gynecological Cancer via Report Metadata Learning

Shenghua Cheng[1,16,17,+], Hongtao Kang[1,+], Yu Wang[2,+], Wenjun Zheng[1,+], Haoqiang He[3], Peiqi Zhuang[1], Abuduxukuer Aili[2], Lu Yang[2], Longfeng Yang[2], Gengxi Cai[4], Liejing Lu[5], Shuqi Jiang[6], Jiayi Zhuo[7], Yujian Zou[8], Jianpeng Li[8], Zhuangyong Xu[9], Ya Zhang[10], Renbing Zhou[11], Xi Zhong[12], Yu Tan[13], Xian Liu[14], Hui Mai[15], Lan Zhang[1,*], Xiaohua Ban[3,*], Jun Shen[2,18,*], and Xiaohui Duan[2,18,*]

[1]School of Biomedical Engineering, Southern Medical University, Guangzhou, 510515, China.
[2]Department of Radiology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, No. 107 Yanjiang Road West, Guangzhou 510120, Guangdong, China.
[3]Department of Radiology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-Sen University Cancer Center, 651 Dongfeng Road East, Guangzhou, Guangdong 510060, China.
[4]Department of Breast Surgery, The First People's Hospital of Foshan. Foshan 528000, China.
[5]Department of Radiology, The First People's Hospital of Foshan, Foshan 528000, China.
[6]Department of Radiology, The First Affiliated Hospital Guangzhou University of Chinese Medicine, Guangzhou 510405, Guangdong, China.
[7]Department of Radiology, Shenshan Medical Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Shanwei 516621, Guangdong, China.
[8]Department of Radiology, Affiliated Dongguan People's Hospital, Southern Medical University, Dongguan, Guangdong, People's Republic of China.
[9]Department of Radiology, Shantou Central Hospital, No. 114 Waima Road, Shantou 515031, China.
[10]Department of Radiology, The Third Affiliated Hospital of Kunming Medical University, Yunnan Cancer Hospital, No. 519 Kunzhou Road, Xishan District, Kunming 650118,Yunnan, China.
[11]Medical Imaging Center, The First People's Hospital of Kashi, 120 Yingbin Avenue, Kashi City, 844000, Xinjiang, China. 12 Department of Medical Imaging, Guangzhou Institute of Cancer Research, The Affiliated Cancer Hospital, Guangzhou Medical University, Guangzhou, China.
[12]Department of Medical Imaging, Guangzhou Institute of Cancer Research, The Affiliated Cancer Hospital, Guangzhou Medical University, Guangzhou, China.
[13]Department of Radiology, Guangdong Women and Children Hospital, Guangzhou, China.
[14]Department of Radiology, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China.
[15]Department of Radiology, Guangdong Provincial Key Laboratory of Major Obstetric Diseases, Guangdong Provincial Clinical Research Center for Obstetrics and Gynecology, The Third Affiliated Hospital, Guangzhou Medical University, Guangzhou, China.
[16]Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, 510515, China.
[17]Guangdong Province Engineering Laboratory for Medical Imaging and Diagnostic Technology, Southern Medical University Guangzhou, 510515, China.
[18]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Medical Research Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, Guangdong, China.
[*]Corresponding authors: zhang_lan98@163.com (L. Zhang), banxh@sysucc.org.cn (X. Ban), shenjun@mail.sysu.edu.cn (J. Shen), duanxh5@mail.sysu.edu.cn (X. Duan).
[+]These authors contributed equally to this work.

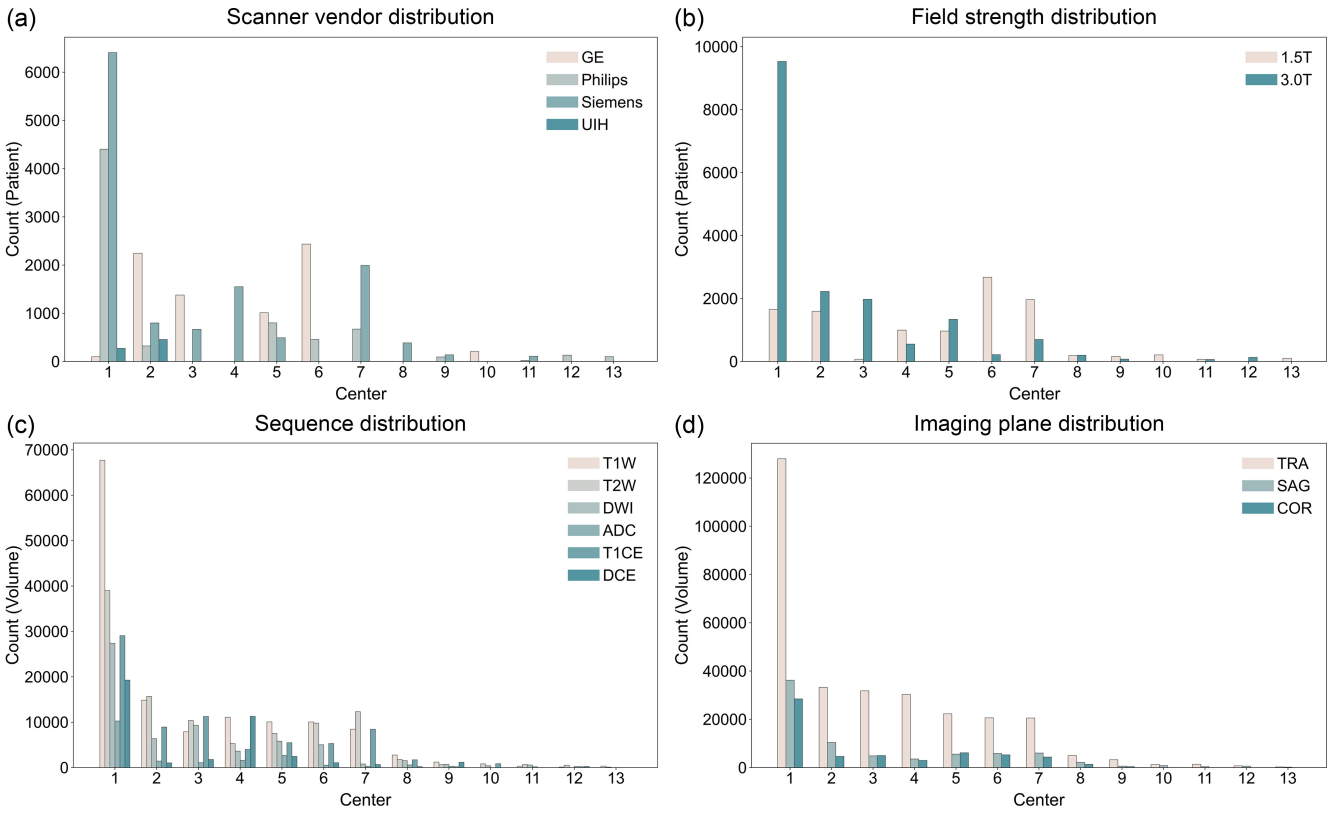

Figure 1. (a) MRI vendors, (b) field strength, (c) sequences, and (d) imaging planes distribution at different medical centers. T1-weighted (T1W), T2-weighted (T2W), T1 contrast-enhanced (T1CE), Diffusion Weighted Imaging (DWI), Apparent Diffusion Coefficient (ADC), and Dynamic Contrast Enhanced (DCE). Transverse(TRA), sagittal(SAG), and coronal(COR).

Table 1. Center information. Our database contains private datasets from 13 medical centers and 2 public datasets.

Alias Name	Center Name	Volume	Patient
Center 1	Sun Yat-sen Memorial Hospital	192663	11422
Center 2	Sun Yat-sen University Cancer Center	48313	4197
Center 3	The First Affiliated Hospital of Guangzhou University of Chinese Medicine	41693	2048
Center 4	Shantou Central Hospital	36883	1621
Center 5	Shenshan Medical Center	34032	2306
Center 6	The First People's Hospital of Foshan	31782	2892
Center 7	Dongguan People's Hospital	30977	2666
Center 8	The First People's Hospital of Kashi	8585	386
Center 9	Yunnan Cancer Hospital	4273	232
Center 10	Guangdong Provincial Maternal and Child Health Hospital	2084	210
Center 11	Guangdong Provincial Hospital of Traditional Chinese Medicine	1766	132
Center 12	The Third Affiliated Hospital of Guangzhou Medical University	1338	132
Center 13	Affiliated Cancer Hospital and Institute of Guangzhou Medical University	461	101
UT_EndoMRI	Public link at: https://zenodo.org/records/13749613	321	146
UMD	Public link at: https://figshare.com/articles/dataset/UMD_zip/23541312	300	300

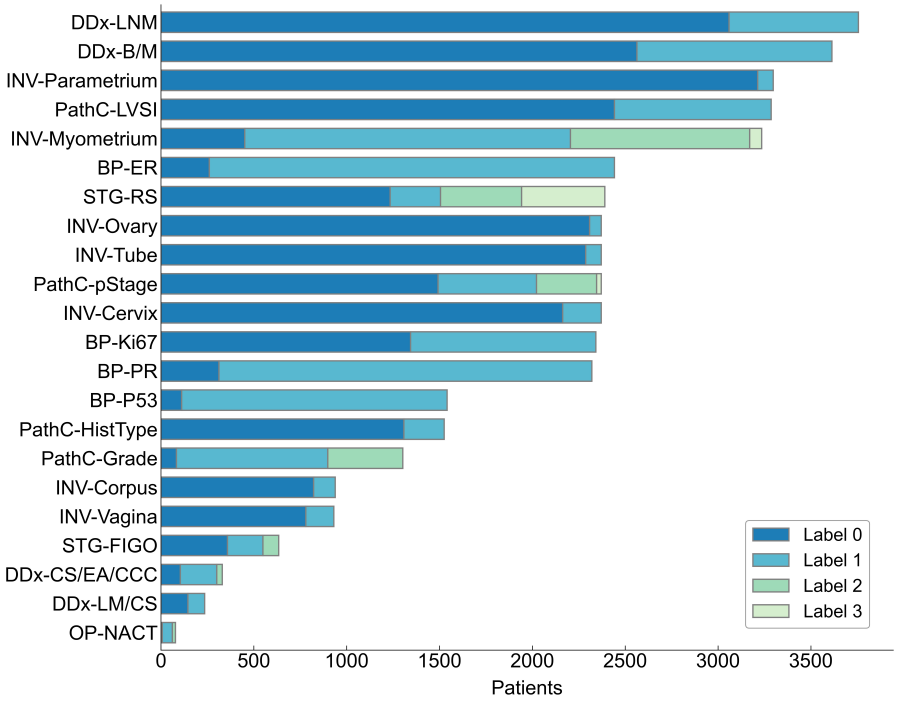

Figure 2. Label distribution in different centers. For binary classification tasks(BP-ER, BP-Ki67, BP-P53, BP-PR, DDx-BM, DDx-LNM, INV-Corpus, INV-Cervix, INV-Myometrium, INV-Ovary, INV-Parametrium, INV-Tube, INV-Vagina, and PathC-LVSI), label 0 represents no metastasis, negative, or benign, while label 1 represents metastasis, positive, or malignant. INV-Myometrium labels 0-3 represent none, less than 1/2, greater than 1/2, and full layer, respectively. PathC-Grade labels 0-2 represent high differentiation, moderate differentiation, and low differentiation, respectively. STG-RS labels 0-3 represent low risk, low-to-medium risk, medium-to-high risk, and high risk, respectively. STG-FIGO labels 0-2 represent FIGO I/II/III, respectively. PathC-HistType labels 0-1 represent squamous cell carcinoma and adenocarcinoma, respectively. PathC-pStage labels 0-3 represent periods I/II/III/IV, respectively. DDx-CS/EA/CCC labels 0-2 represent carcinosarcoma, adenocarcinoma, and clear cell cancer, respectively. DDx-LM/CS labels 0-1 represent uterine leiomyoma and carcinosarcoma, respectively. DDx-LM/CS labels 0-1 represent carcinosarcoma, adenocarcinoma, and clear cell cancer, respectively. DDx-LM/CS labels 0-1 represent uterine fibroids and carcinosarcoma, respectively. OP-NACT labels 0-2 represent complete response, partial response, and stable disease, respectively.

Table 2. This retrospective multi-center study obtained approval from the institutional review committees at all participating centers, and patient written informed consent was waived.

Center Name	Ethics Approval Number
Sun Yat-sen Memorial Hospital	SYSKY-2024-400-01, SYSKY-2024-801-01, SYSKY-2026-025-01
Sun Yat-sen University Cancer Center	SL-G2023-231-01, SL-B2023-348-02
The First Affiliated Hospital of Guangzhou University of Chinese Medicine	K-2025-209
Shantou Central Hospital	2025:026, 2024-113
Shenshan Medical Center	2025-SSKY-044-01, 2025-SSKY-045-01
The First People's Hospital of Foshan	2025-262, 2025-261
Dongguan People's Hospital	KYKT-2025-112, KYKT-2026-026
The First People's Hospital of Kashi	2023:79
Yunnan Cancer Hospital	SLKYCS2024-263
Guangdong Provincial Maternal and Child Health Hospital	202401201
Guangdong Provincial Hospital of Traditional Chinese Medicine	BE2023-146
The Third Affiliated Hospital of Guangzhou Medical University	[2025] No.340
Affiliated Cancer Hospital and Institute of Guangzhou Medical University	2021-SK02

Structured Report Details

The process of transforming unstructured gynecological imaging reports into structured data is achieved through information extraction and standardized categorization (Table 7 and 8). It begins by analyzing free-text descriptions in the original reports (e.g., "significant thickening of the upper endometrial lining with localized soft tissue mass formation") and then maps them to four core modules based on a predefined medical ontology framework: lesion characteristics, extent of invasion, lymph node status, and effusion assessment. Within the lesion module, the system identifies abnormality types (including no abnormality, mass, cystic lesion, nodule, fat-containing lesion, and abnormal signal) across five anatomical regions: cervix, uterus, bilateral adnexa, vagina, and pelvis. The invasion module evaluates involvement in ten key structures, such as the myometrium, junctional zone, and serosal layer. The lymph node module classifies lymph node status across seven regions into three grades, while the effusion module determines the presence of fluid in four anatomical areas. Through this refined structured conversion, the original text is encoded into quantifiable metadata (e.g., pretrain matedata: "01000_1010010000_0000000_0010_1"), which preserves critical clinical information while achieving data standardization. This provides high-quality structured input for subsequent AI model training, clinical decision support, and precision medicine research.

Table 3. The amount of data in the training set, internal test set, and external test set for various tasks. In UT-EndoMRI, some segmentation targets have multiple annotations. The statistics we present here are based on the number of annotations, not the number of images.

Task Categories	Tasks	Training		Internal test		External test	
		Patients	Sequences	Patients	Sequences	Patients	Sequences
Differential Diagnosis	DDx-B/M(OT)	2096	22942	524	5822	253	2749
	DDx-B/M(PG)	592	18027	148	4366		
	DDx-CS/EA/CCC(EC)	264	1196	66	289		
	DDx-LM/CS(EC)	188	1360	47	298		
	DDx-LNM(CC)	586	3627	150	943	189	941
	DDx-LNM(RC)	368	6507	92	1648		
	DDx-LNM(EC)	1716	13084	429	3210	226	1066
Invasion Assessment	INV-Corpus(CC)	600	3699	150	943	188	936
	INV-Cervix(EC)	1716	13084	429	3210	226	1066
	INV-Myometrium(CC)	543	3352	134	846	187	931
	INV-Myometrium(EC)	1716	13084	429	3210	226	1066
	INV-Ovary(EC)	1716	13084	429	3210	226	1066
	INV-Parametrium(CC)	588	3617	150	943	188	936
	INV-Parametrium(EC)	1716	13084	429	3210	226	1066
	INV-Tube(EC)	1716	13084	429	3210	226	1066
	INV-Vagina(CC)	596	3667	147	920	188	936
Staging	STG-FIGO(CC)	404	2543	98	648	132	631
	STG-RS(EC)	1716	13084	429	3210	245	1146
Pathological Characterization	PathC-Grade(CC)	958	8823	226	2044	118	1025
	PathC-HistType(CC)	1120	10562	280	2604	125	1095
	PathC-LVSI(CC)	587	3613	144	906	184	916
	PathC-LVSI(EC)	1716	13084	429	3210	226	1066
	PathC-pStage(EC)	1716	13084	429	3210	226	1066
Biomarker Prediction	BP-ER(Pelvic)	1776	15736	457	3991	209	1188
	BP-Ki67(Pelvic)	1520	13330	446	4055	376	2343
	BP-P53(Pelvic)	1056	7656	277	1958	208	1136
	BP-PR(Pelvic)	1692	14741	430	3711	198	1090
Outcome Prediction	OP-OS(CC)	509	3001	145	917		
	OP-OS(EC)	706	5089	218	4885		
	OP-NACT(CC)	47	857	31	605		
Report Generation	MIL-Report(CC)	6672	98023	261	2991		
	MIL-Report(EC)			406	4638		
	MIL-Report(OT)			135	2051		
	MIL-Report(Pelvic)			866	15093		
Report Generation	R2GenGPT(CC)	6672	98023	261	2991		
	R2GenGPT(EC)			406	4638		
	R2GenGPT(OT)			135	2051		
	R2GenGPT(Pelvic)			866	15093		
Segmentation	ClearTumor	16	88	3	18		
	EndoCancer	330	1040	197	260		
	UMD	240	240	60	60		
	UT-EndoMRI	101	257	45	64		

Table 4. Abbreviation table for various tasks

Abbreviations	Full names
DDx-B/M(OT)	Classification of benign and malignant ovarian tumors.
DDx-B/M(PG)	Classification of benign and malignant parotid gland tumors.
DDx-CS/EA/CCC(EC)	Classification of carcinosarcoma, adenocarcinoma and clear cell carcinoma.
DDx-LM/CS(EC)	Classification of leiomyoma and carcinosarcoma.
DDx-LNM(CC)	Lymph node metastasis prediction in cervical cancer.
DDx-LNM(RC)	Lymph node metastasis prediction in rectal cancer.
DDx-LNM(EC)	Lymph node metastasis prediction in endometrial cancer.
INV-Corpus(CC)	Assessment of uterine corpus invasion in endometrial cancer.
INV-Cervix(EC)	Assessment of cervical invasion in endometrial cancer.
INV-Myometrium(CC)	Assessment of myometrial invasion in cervical cancer.
INV-Myometrium(EC)	Assessment of myometrial invasion in endometrial cancer.
INV-Ovary(EC)	Assessment of ovarian invasion in endometrial cancer.
INV-Parametrium(CC)	Assessment of parametrial invasion in cervical cancer.
INV-Parametrium(EC)	Assessment of parametrial invasion in endometrial cancer.
INV-Tube(EC)	Assessment of fallopian tube invasion in endometrial cancer.
INV-Vagina(CC)	Assessment of vaginal invasion in cervical cancer.
STG-FIGO(CC)	International Federation of Gynecology and Obstetrics(FIGO) staging of cervical cancer.
STG-RS(EC)	Endometrial cancer risk stratification prediction.
PathC-Grade(CC)	Cervical cancer differentiation grade prediction.
PathC-HistType(CC)	Pathological types of cervical cancer.
PathC-LVSI(CC)	Lymphvascular space invasion prediction in cervical cancer.
PathC-LVSI(EC)	Lymphvascular space invasion prediction in endometrial cancer.
PathC-pStage(EC)	Pathological staging of endometrial cancer.
BP-ER(Pelvic)	Prediction of estrogen receptor (ER) in pelvic diseases.
BP-Ki67(Pelvic)	Prediction of Kiel 67 (Ki67) in pelvic diseases.
BP-P53(Pelvic)	Prediction of tumor protein p53 (P53) in pelvic diseases.
BP-PR(Pelvic)	Prediction of progesterone receptor (PR) in pelvic diseases.
OP-OS(CC)	Overall survival (OS) time of cervical cancer patients.
OP-OS(EC)	Overall survival (OS) time of endometrial cancer patients.
OP-NACT(CC)	Prediction of neoadjuvant chemotherapy efficacy in cervical cancery.
MIL-Report(CC)	Report generation via MIL-Report in a cervical cancer patient cohort.
MIL-Report(EC)	Report generation via MIL-Report in a endometrial cancer patient cohort.
MIL-Report(OT)	Report generation via MIL-Report in a ovarian tumor patient cohort.
MIL-Report(Pelvic)	Report generation via MIL-Report in a pelvic disease patient cohort.
R2GenGPT(CC)	Report generation via R2GenGPT in a cervical cancer patient cohort
R2GenGPT(EC)	Report generation via R2GenGPT in a endometrial cancer patient cohort.
R2GenGPT(OT)	Report generation via R2GenGPT in a ovarian tumor patient cohort.
R2GenGPT(Pelvic)	Report generation via R2GenGPT in a pelvic disease patient cohort.
ClearTumor	Clear cell carcinoma lesion segmentation.
EndoCancer	Endometrial cancer lesion segmentation.
UMD	A publicly available uterine segmentation dataset containing four categories: muscular wall, uterine cavity, uterine myoma, and nabothian cyst.
UT-EndoMRI	A publicly available pelvic segmentation dataset containing four categories: uterus, ovary, endometrioma, and cyst.

Table 5. MRI acquisition. T1-weighted (T1W), T1 contrast-enhanced (T1CE); T2-weighted (T2W), Diffusion Weighted Imaging (DWI), Apparent Diffusion Coefficient (ADC), and Dynamic Contrast Enhanced (DCE).

Institution		TR (ms)	TE (ms)	FOV (mm)	Slice Thickness (mm)	Flip Angle
Center 1	T1W	3.09-4.15	1.17-1.9	320-480	3-5	10-90
	T1CE	2.81-600	0.5-12	320-480	3-6	9-150
	T2W	1100-13000	75-130	230-480	3-6	90-160
	DWI	2000-10000	50-115	230-430	3-6	90-180
	ADC	2000-10000	50-115	230-430	3-6	90-180
	DCE	3.6-10	1.15-3.38	250-320	3-5	8-35
Center 2	T1W	3.09-500	1.17-12	220-480	2-7	9-150
	T1CE	3.07-600	0.5-12	250-460	2-7	9-15
	T2W	1100-11300	75-135	200-430	3-7	90-180
	DWI	1650-14500	49-115	240-420	3-6	90-180
	ADC	1650-14500	49-115	240-420	3-6	90-180
	DCE	3.09-6.25	1.17-2.3	260-480	3-5	9-30
Center 3	T1W	500-750	8-20	240-480	3-6	90-180
	T1CE	3.9-5.1	1.2-2.5	250-480	3-5	9-15
	T2W	2500-7500	70-110	240-400	3-6	90-180
	DWI	4200-6500	45-70	220-400	3-5	90
	ADC	4200-6500	45-70	220-400	3-5	90
	DCE	3.16-7.86	1.15-2.36	250-420	3-5	8-15
Center 4	T1W	3.92-6.81	1.39-2.39	300-490	3-5	9-12
	T1CE	2.96-756	1.2-8	200-500	3-7	9-150
	T2W	3000-5600	72-105	240-510	3-7	150-180
	DWI	1900-8500	50-93	280-380	4-7	90-180
	ADC	1900-8500	50-93	280-380	4-7	90-180
	DCE	3.23-5.42	1.18-2.39	250-480	3-7	9-10
Center 5	T1W	100-800	2.25-15	250-400	3-7	70-150
	T1CE	3.6-5.1	1.2-3.1	280-450	3-6	9-150
	T2W	1000-9000	60-130	240-480	3-7	80-180
	DWI	3100-5600	60-90	160-380	3-6	90-180
	ADC	3100-5600	60-90	160-380	3-6	90-180
	DCE	3.6-10.0	1.1-2.0	250-400	3-5	10-35
Center 6	T1W	3.09-500	1.17-12	260-480	3-7	10-150
	T1CE	2.28-400	0.5-13.2	240-480	3-6	10-150
	T2W	1100-11300	75-135	280-510	3-6	90-180
	DWI	820-7500	45-90	280-480	3-7	90-180
	ADC	820-7500	45-90	280-480	3-7	90-180
	DCE	3.52-4.26	1.67-2.1	320-450	4-6	10-15

Table 6. MRI acquisitions. T1-weighted (T1W), T1 contrast-enhanced (T1CE); T2-weighted (T2W), Diffusion Weighted Imaging (DWI), Apparent Diffusion Coefficient (ADC), and Dynamic Contrast Enhanced (DCE)

Institution		TR (ms)	TE (ms)	FOV (mm)	Slice Thickness (mm)	Flip Angle
Center 7	T1W	100-2300	2.25-11	250-400	3-7	10-150
	T1CE	3.7-4.9	1.2-2.4	280-450	3-6	9-10
	T2W	1500-5500	70-120	280-450	3-6	80-180
	DWI	5500-5600	60-65	350-450	3-7	90-180
	ADC	5500-5600	60-65	350-450	3-7	90-180
	DCE	3.66-4.83	1.23-2.35	280-450	3-6	9-10
Center 8	T1W	250-750	2.22-11	250-400	5-6	90-180
	T1CE	3.97-1000	1.23-11	250-450	3-6	10-180
	T2W	1200-4500	51-132	240-480	3-6	90-180
	DWI	3000-8000	51-83	370-420	4-6	90
	ADC	3000-8000	51-83	370-420	4-6	90
	DCE	3.97-9	1.05-2.48	280-400	3-5	4-12
Center 9	T1W	170-877	4.76-10	250-380	5-6	70-150
	T1CE	3.6-5.7	1.2-2.4	240-480	3-6	10-30
	T2W	1500-4534	90-104	250-420	3-6	90-170
	DWI	6400-7500	65-68	220-380	4-6	90
	ADC	6400-7500	65-68	220-380	4-6	90
	DCE	3.6-5.74	1.14-3.38	250-380	3-5	10
Center 10	T1W	3.48	1.5	350-380	4	15
	T1CE	2.8-3.8	1.3-1.7	220-370	4-6	15
	T2W	2941-3200	90-120	260-450	4-6	90
	DWI	3840-5200	76-120	250-370	4-6	90
	ADC	3840-5200	76-120	250-370	4-6	90
Center 11	T1W	455-643	8.6-12.4	280-320	4-6	90
	T1CE	3.97-818	1.29-11	250-430	4-6	90-140
	T2W	3789-5000	85-121	280-450	3-6	140-150
	DWI	3539-6800	80-82	320-390	3-7	90
	ADC	3539-6800	80-82	320-390	3-7	90
	DCE	3.97	1.29	320-360	3	9
Center 12	T1W	490-590	8-10	290-410	5-5	90
	T1CE	2.8-4.88	1-2.39	320-420	3-6	10
	T2W	1250-3000	70-120	280-400	3-5	90
	DWI	1600-6420	50-60	250-420	5-6	90
	ADC	1600-6420	50-60	250-420	5-6	90
Center 13	T1W	492-587	10-10	290-410	5-5	90
	T1CE	2.7-4.9	1.2-3.5	280-420	3-6	10
	T2W	1500-3500	60-130	320-380	3-6	90
	DWI	1250-2026	55-70	380-420	3-6	90
	ADC	1250-2026	55-70	380-420	3-6	90

Table 7. Structured report details.This table demonstrates, through a real-world report example, how to convert a raw report into a structured report, and then into metadata. The metadata will be used for subsequent pre-training and training for downstream tasks.

Structured Dimension	Unstructured Report	Structured Items	Structured Report	Metadata
Part 1: Lesion Status	"The endometrium in the upper uterine segment is markedly thickened, forming a localized soft tissue mass measuring approximately 44mm × 31mm × 39mm."	**Location**: Cervix; Uterus; Bilateral Adnexa; Pelvic; Vagina. **Abnormality (0–5)**: Normality; Mass; Cystic Lesion; Abnormal Signal; Nodule; Fat.	No significant abnormalities were observed in the cervix. A mass is seen in the uterus. No significant abnormalities were observed in the bilateral adnexa or vagina.	**Report generation**: 01000. **Pretrain**:01000.
Part 2: Invasion Status	"Invasion extends to the myometrium at the uterine fundus reaching the serosal surface," "The mass extends distally to the internal cervical os"	**Location**: Myometrium; Junctional; Serosal layer; Vagina; Uterine body; Cervix; Parametrium; Pelvic wall; Bladder; Rectum. **Abnormality (0–1)**: Normality; Abnormality.	Invasion involves the myometrium and serosal layer. No invasion is seen in the vagina, uterine body, pelvic sidewall, bladder, or rectum. Invasion involves the cervix.	**Report generation**: 1010000000. **Pretrain**:1010000000.
Part 3: Lymph Nodes	"No enlarged lymph nodes are seen along the bilateral iliac vessels or in the bilateral inguinal regions"	Location: Pelvic; Along iliac vessels; Para-aortic; Inguinal; Obturator; Presacral; Peritoneal. Abnormality (0–2): Normal; Enlarged lymph node; Small lymph node.	No significant abnormalities are seen in pelvic, iliac vessel, para-aortic, or inguinal lymph nodes.	**Report generation**: 0000000. **Pretrain**:0000000.
Part 4: Effusion Status	"Moderate amount of pelvic effusion."	**Location**: Uterine cavity; Fallopian tubes; Pelvic cavity; Peritoneum. **Abnormality (0–1)**: No effusion, effusion.	No effusion is seen in the uterine cavity or fallopian tubes. Pelvic effusion is present.	**Report generation**: 0010. **Pretrain**:0010.
Imaging Conclusion	"Localized thickening of the upper endometrial lining forming a soft-tissue mass, consistent with endometrial carcinoma, involving the deep myometrium and the internal cervical os, with pelvic effusion."	**Abnormality (0–1)**: No significant abnormality, abnormality.	A mass is seen in the uterus, consistent with endometrial carcinoma, involving the myometrium, serosal layer, and cervix, with pelvic effusion.	**Report generation**: N/A. **Pretrain**:1.

Table 8. Examples of unstructured and structured image reports.

Unstructured Image Reports	Structured Image Reports
Imaging Findings: The uterine volume is slightly enlarged. The endometrium is significantly thickened, with a maximum thickness of approximately 16 mm. It appears isointense on T1-weighted images (T1WI), slightly hyperintense on T2-weighted images (T2WI), and hyperintense on diffusion-weighted imaging (DWI). Mild enhancement is observed on post-contrast scans, with an ill-defined border from the junctional zone. Several cystic lesions exhibiting long T1 and long T2 signals are also noted in the cervix, the largest measuring about 13 mm in length, showing no enhancement. No significant abnormalities are detected in the bilateral adnexa. A small amount of pelvic fluid is present. The urinary bladder is well-distended with clear margins. The rectum appears unremarkable. No significantly enlarged lymph nodes are seen in the bilateral iliac vessel regions or bilateral inguinal areas. The visualized bones show no signs of destruction. **Report Conclusion**: Endometrial thickening, consistent with endometrial carcinoma, suspicious for involvement of the junctional zone. Several cystic foci in the cervix are most likely cysts. A small amount of pelvic fluid is present.	**Imaging Findings**: A cystic lesion is noted in the cervix. A mass is identified within the uterus. No significant abnormalities are detected in the bilateral adnexa. There is involvement of the junctional zone. No invasion is seen into the serosal layer, vagina, uterine body, pelvic sidewall, urinary bladder, or rectum. There is no significant abnormality in the pelvic lymph nodes, lymph nodes adjacent to the iliac vessels, para-aortic lymph nodes, or inguinal lymph nodes. No fluid is seen in the uterine cavity or fallopian tubes. Pelvic fluid is present. **Imaging conclusion**: A uterine mass, consistent with endometrial carcinoma. A cystic lesion is noted in the cervix. There is involvement of the junctional zone. Pelvic fluid is noted. **Report matedata**: 21000000000000000000000010. **Pretraining matedata**: 11000000000000000000000101.
Imaging Findings: The endometrium in the upper uterine segment is markedly thickened, forming a localized soft tissue mass measuring approximately 44mm × 31mm × 39mm. The mass extends distally to the internal os of the cervical canal. It demonstrates an irregular contour and heterogeneous signal intensity: isointense on T1-weighted images, mildly hyperintense on T2-weighted images, with restricted diffusion and heterogeneous enhancement post-contrast. The adjacent junctional zone is obliterated. The mass involves the myometrium from the fundus to the serosal surface. There is right adnexal involvement. No significant abnormality is seen in the left adnexa. The cervix is normal in size and morphology. No obvious lesion is identified in the bilateral parametria. The rectal wall is not thickened. The bladder is poorly distended but with clear margins. No enlarged lymph nodes are seen in the bilateral iliac vessels region or bilateral inguinal areas. The visualized bones show no significant abnormality. **Report Conclusion**: Localized endometrial thickening forming a soft tissue mass in the upper uterine segment, compatible with endometrial carcinoma, involving the deep myometrium and the internal os of the cervical canal, with right adnexal involvement.	**Imaging Findings**:The cervix showed no obvious abnormalities. A mass was seen in the uterus. The bilateral adnexa showed no obvious abnormalities. The myometrium and serosa layer were involved; the vagina, uterine body, cervix, pelvic wall, bladder, and rectum were not invaded. No obvious abnormalities were seen in the pelvic, parailiac, paraaortic, or inguinal lymph nodes. No fluid was seen in the uterine cavity, fallopian tubes, or pelvic cavity. **Imaging conclusion**: A uterine mass is identified, compatible with endometrial carcinoma, involving the myometrium, serosal layer, and cervix. **Report matedata**: 01000101001000000000000000. **Pretraining matedata**: 01000101001000000000000001.

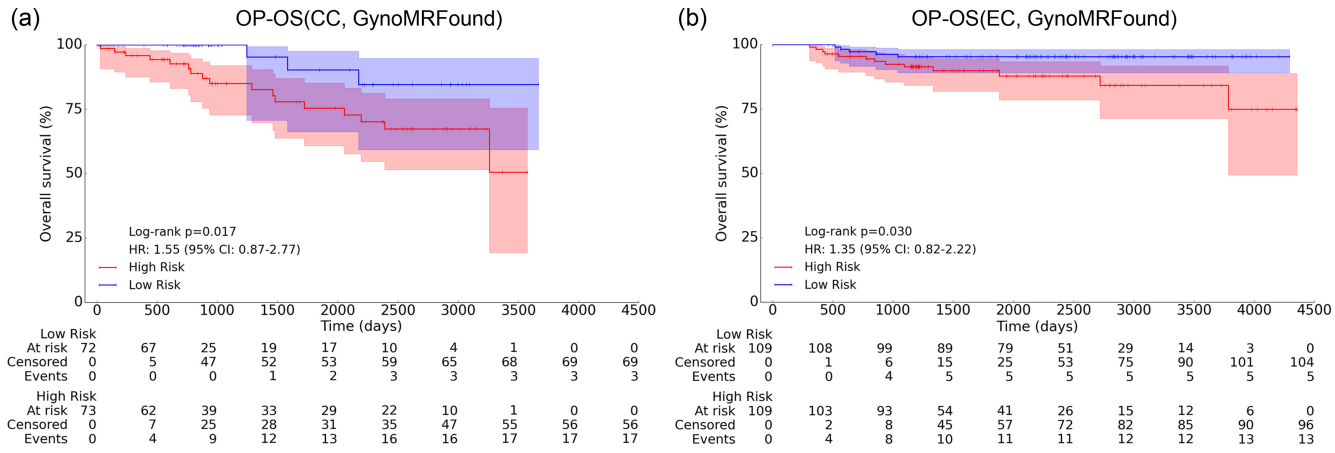

Figure 3. Survival prediction efficacy assessment for patients with cervical cancer and endometrial cancer. (a) Cervical cancer patient cohort. (b) Endometrial cancer patient cohort. The red and blue curves represent the high-risk and low-risk groups, respectively. The table below lists the number of at-risk individuals, the number of events, and the number of censored events at each time point.

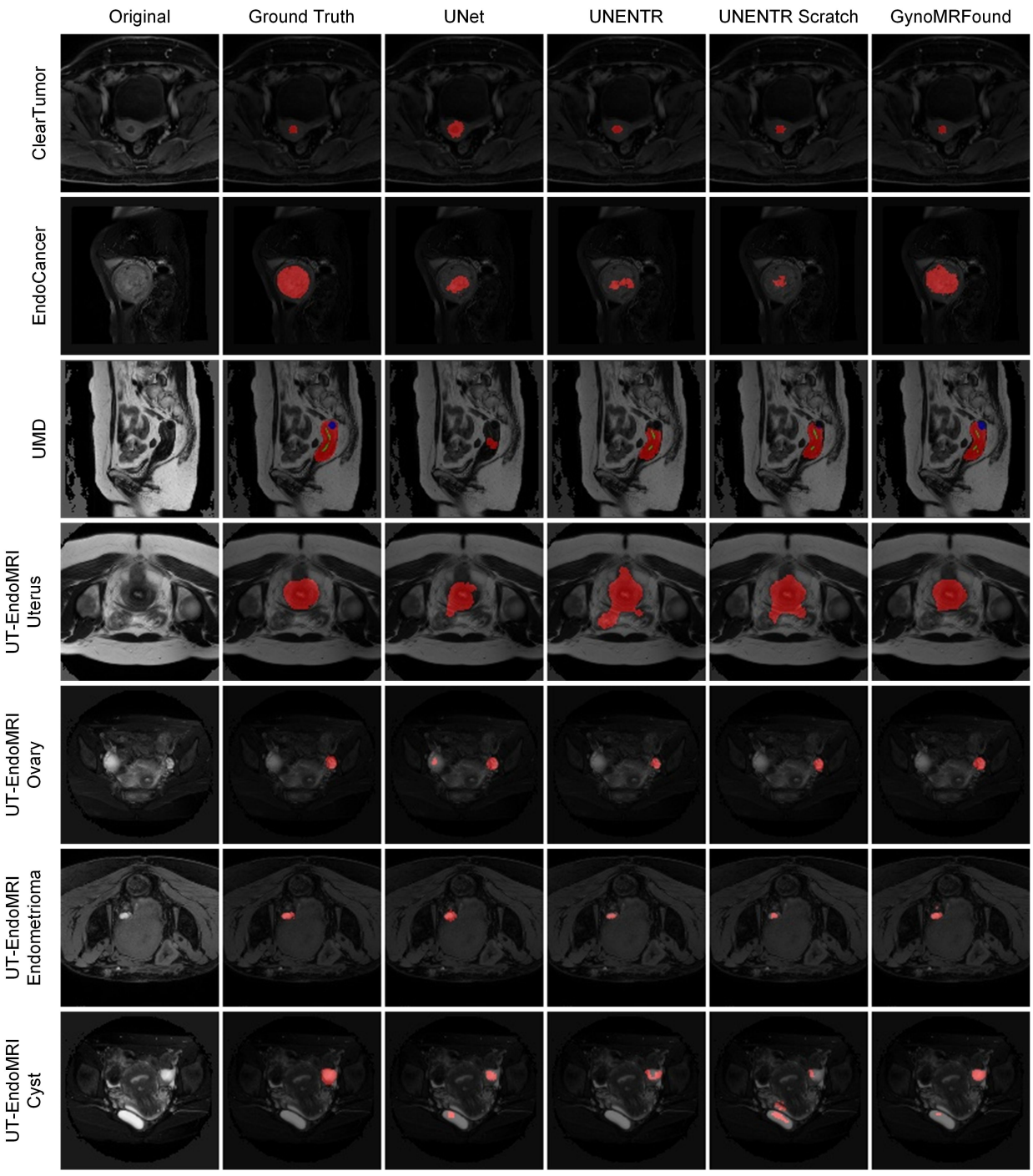

Figure 4. Visual comparison in segmentation tasks.