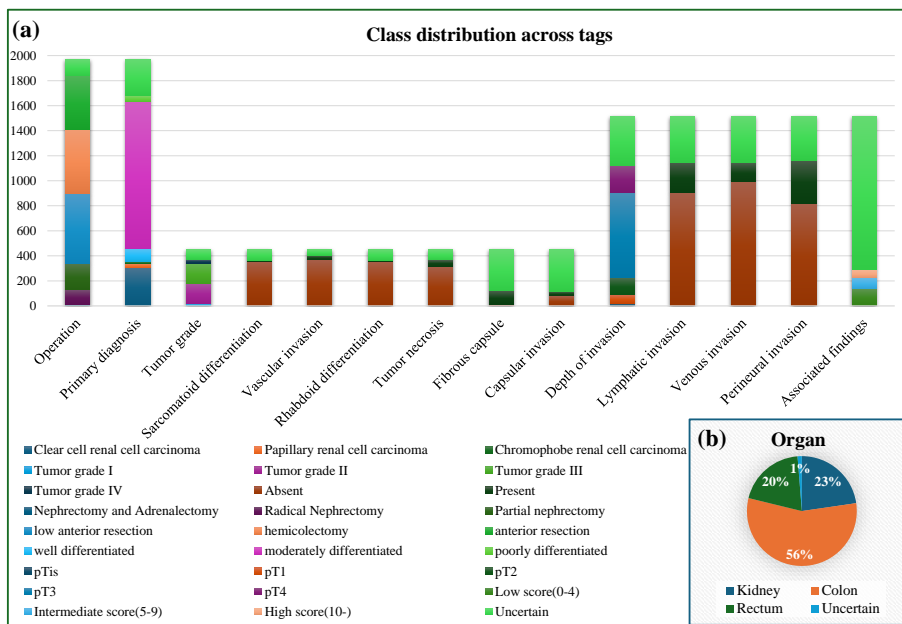


## Supplementary

## Dataset



**Fig. S1.** Dataset composition. (a) Distribution of inner classes for each tag. The x-axis represents the tag names, while the y-axis indicates the number of occurrences for each tag. After selecting inner classes for each tag, we integrated them into one group. Therefore, the output dimension of the tag classifier becomes the size of that group (27: 26 classes + 1 'uncertain' class). (b) Distribution of organs in the dataset. Data labeled as 'uncertain' is extracted from colon or rectum, where there are no keywords related to organs in the reports.

**Table S1.** Implementation details. For training stage-1, we adopt the default settings from HIPT. To evaluate its effectiveness, we employ a classifier consisting of two linear layers followed by a softmax layer. While for stage-2 training, we run for 300 epochs and choose the model with the best tag classification accuracy on the validation set as the representative model. For comparative experiments without a  $Cl_{stag}$ , we select the model with the smallest  $L_{sen}$ .

	stage-1	stage-2
Epoch Number	500	300
Batch Size	64	1
Optimizer	SGD	AdamW
Learning Rate	1e-1	3e-3

(a) Reports (ex.1)		
<p>radical nephrectomy  <b>Primary tumor diagnosis: Papillary renal cell carcinoma</b>            Tumor grade (WHO / ISUP grade): II. Nucleoli are conspicuous and eosinophilic at x400 and visible but not prominent at x100            Sarcomatoid differentiation: Absent            Vascular invasion: Not identified            Rhabdoid differentiation: Absent            Tumor necrosis: Absent</p>		GT
<p>radical nephrectomy  <b>Primary tumor diagnosis: Papillary renal cell carcinoma</b>            Tumor grade (WHO / ISUP grade): II. Nucleoli are conspicuous and eosinophilic at x400 and visible but not prominent at x100            Sarcomatoid differentiation: absent            Vascular invasion: Not identified            Rhabdoid differentiation: absent            Tumor necrosis: absent            Capsular invasion: absent            Fibrous capsule: present (complete)</p>	METEOR: 0.9590	Ours
<p>partial nephrectomy  <b>Primary tumor diagnosis: Clear cell renal cell carcinoma</b>            Tumor grade (WHO / ISUP grade): III. Nucleoli are conspicuous and eosinophilic at x400 and visible but not prominent at x100            Sarcomatoid differentiation: Absent            Vascular invasion: Not identified            Rhabdoid differentiation: Absent            Tumor necrosis: Absent</p>	METEOR: 0.9404	MR-ViT <sub>S</sub> *
<p>low anterior resection  <b>Adenocarcinoma, moderately differentiated</b></p>	METEOR: 0.0105 Missing: tumor grade, sarcomatoid, vascular, rhaboid, necrosis	HIPT <sub>S</sub> *
(b) Reports (ex.2)		
<p>partial nephrectomy  <b>Primary tumor diagnosis: Clear cell renal cell carcinoma, with cystic change</b>            Tumor grade (WHO / ISUP grade): II. Nucleoli are conspicuous and eosinophilic at x400 and visible but not prominent at x100            Sarcomatoid differentiation: absent            Vascular invasion: Not identified            Rhabdoid differentiation: absent            Tumor necrosis: absent            Capsular invasion: absent            Fibrous capsule: present ( complete )</p>		GT
<p>partial nephrectomy  <b>Primary tumor diagnosis: Clear cell renal cell carcinoma, with cystic change</b>            Tumor grade (WHO / ISUP grade): II. Nucleoli are absent or inconspicuous and basophilic at x400            Sarcomatoid differentiation: absent            Vascular invasion: Not identified            Rhabdoid differentiation: absent            Tumor necrosis: absent            Capsular invasion: absent Tissue labeled as "peri-renal fat" (#2), biopsy; No tumor present            Fibrous capsule: present ( complete )</p>	METEOR: 0.8405	Ours
<p>partial nephrectomy  <b>Primary tumor diagnosis: Clear cell renal cell carcinoma</b>            Tumor grade (WHO / ISUP grade): II. Nucleoli are conspicuous and eosinophilic at x400 and visible but not prominent at x100            Sarcomatoid differentiation: Absent            Vascular invasion: Not identified            Rhabdoid differentiation: Absent            Tumor necrosis: Absent</p>	METEOR: 0.7786 Missing: Capsular, Fibrous	MR-ViT <sub>S</sub> *
<p>partial nephrectomy  <b>Primary tumor diagnosis: Clear cell renal cell carcinoma</b>            Tumor grade (WHO / ISUP grade): III. Nucleoli are conspicuous and eosinophilic at x400 and visible but not prominent at x100            Sarcomatoid differentiation: absent            Vascular invasion: Not identified            Rhabdoid differentiation: absent            Tumor necrosis: absent</p>	METEOR: 0.8695 Missing: Capsular, Fibrous	HIPT <sub>S</sub> *
	Match      Mismatch      not in GT	

**Fig. S2.** Additional report generation results for the kidney. (a) Ours provides accurate descriptions for all tags, generating additional descriptions for tags not mentioned in the ground truth report. Despite the high number of incorrect descriptions in MR-ViT<sub>S</sub>\*, our METEOR score is only slightly upper than MR-ViT<sub>S</sub>\*. (b) While our approach provides accurate descriptions for all tags, HIPT<sub>S</sub>\* has one incorrect description and misses descriptions for two tags. Surprisingly, the METEOR score is higher for HIPT<sub>S</sub>\* than Ours. This result suggests that NLG metrics may not absolutely represent the quality of medical report generation.