GMAI-MMBench: A Comprehensive Multimodal Evaluation Benchmark Towards General Medical AI

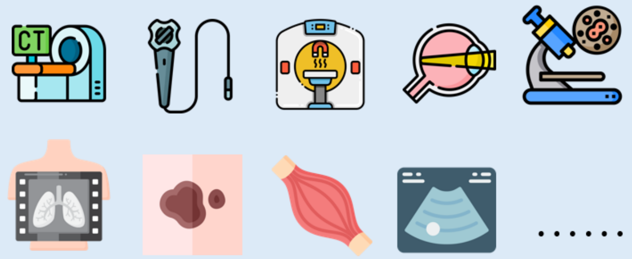
Pengcheng Chen, Jin Ye, Guoan Wang, Yanjun Li, Zhongying Deng, Wei Li, Tianbin Li, Haodong Duan, Ziyan Huang, Yanzhou Su, Benyou Wang, Shaoting Zhang, Bin Fu, Jianfei Cai, Bohan Zhuang, Eric J Seibel, Junjun He, Yu Qiao

NEURAL INFORMATION PROCESSING SYSTEMS

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Data composition

Comprehensive medical knowledge

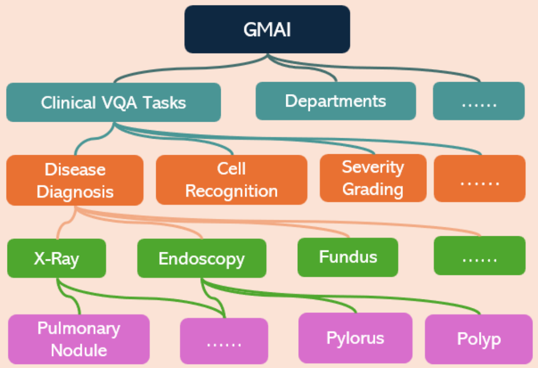


38 Medical image modalities

284 Clinical related datasets

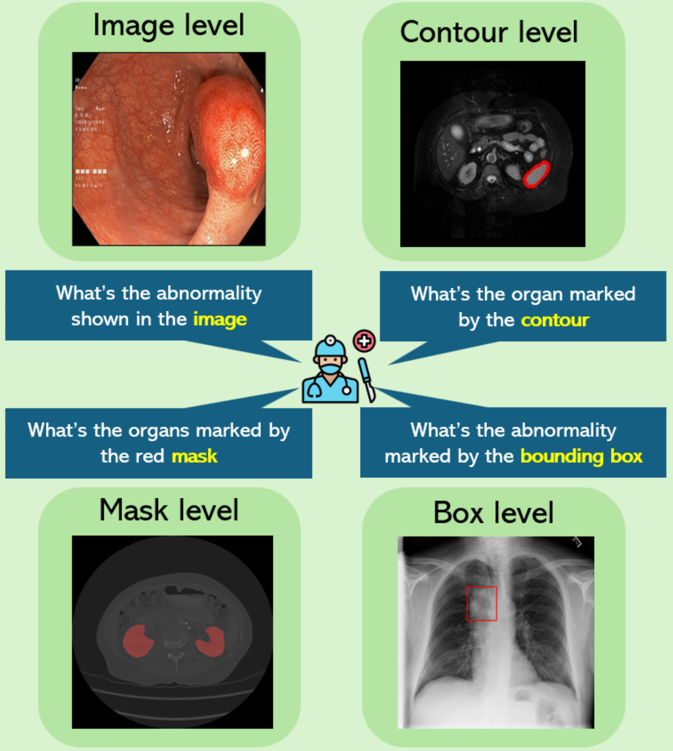
Well-categorized data structure

18 Clinical related tasks
Across 18 departments



Lexical tree structure

Multi-perceptual granularity



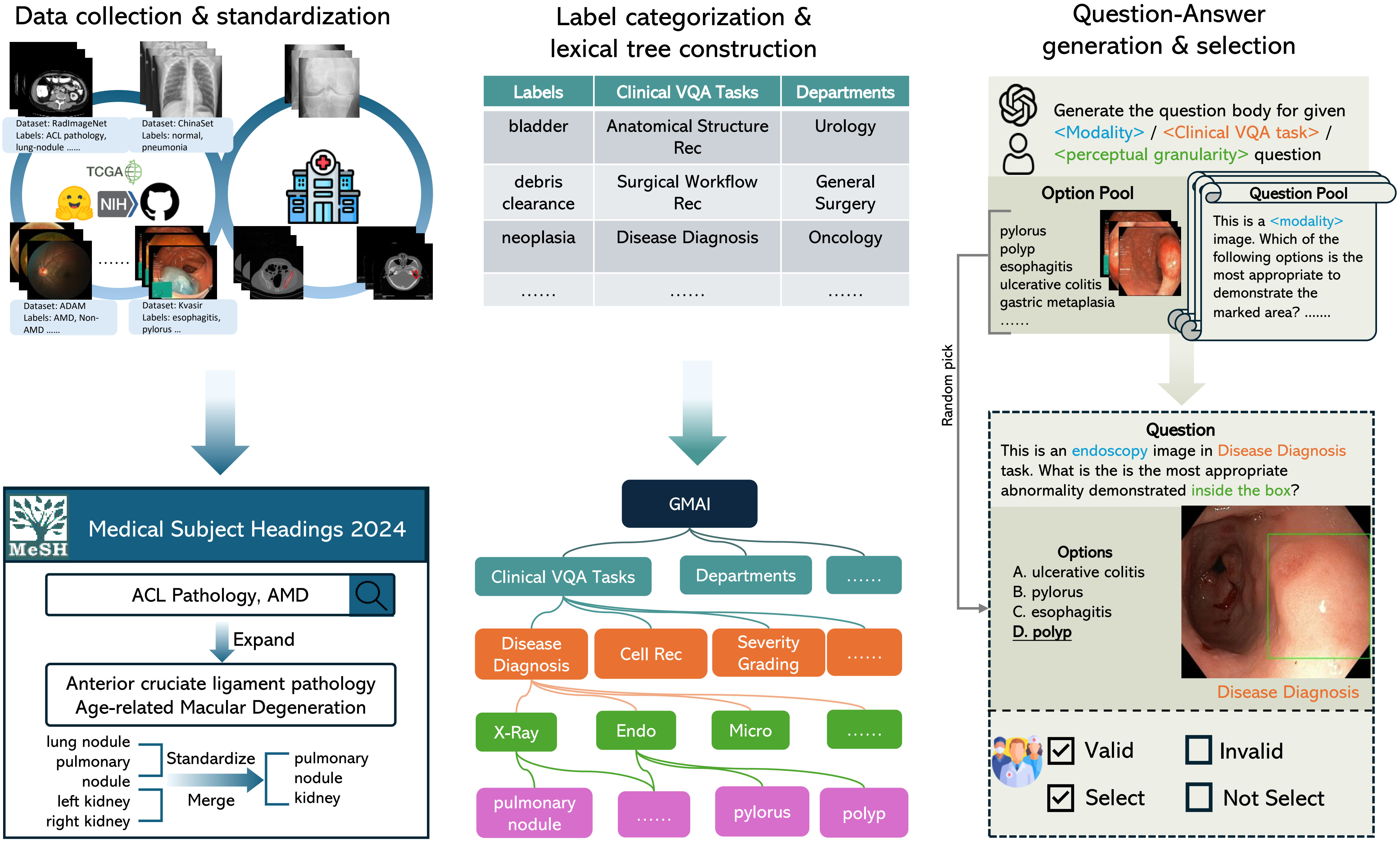
What's the abnormality shown in the image

What's the organ marked by the contour

What's the organs marked by the red mask

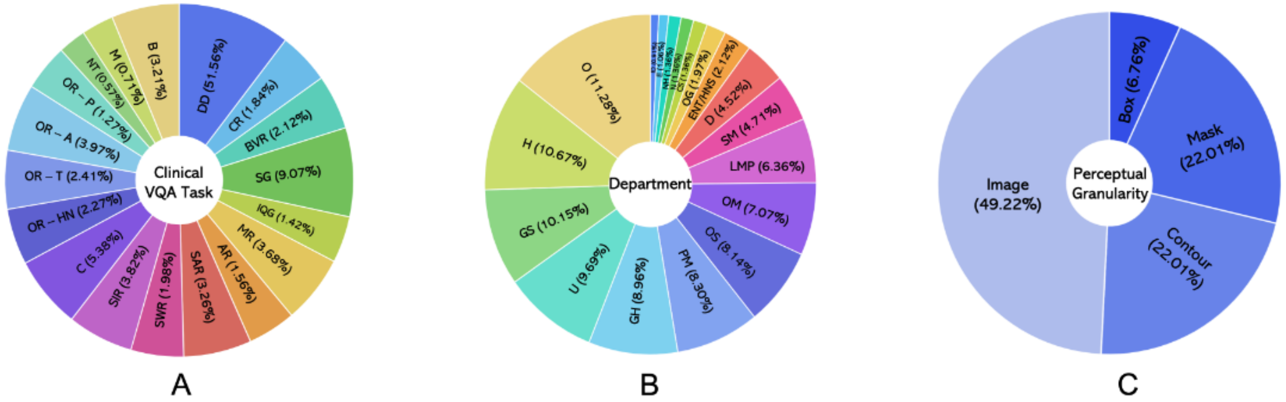
What's the abnormality marked by the bounding box

4 Different perceptual types


Data collection & standardization
Dataset: RadImageNet
Labels: ACL pathology, lung-nodule ......
Dataset: ChinaSet
Labels: normal, pneumonia
TCGA
NIH
......
Dataset: ADAM
Labels: AMD, Non-AMD ......
Dataset: Kvasir
Labels: esophagitis, pylorus ...
Medical Subject Headings 2024
MeSH
ACL Pathology, AMD
Expand
Anterior cruciate ligament pathology
Age-related Macular Degeneration
lung nodule
pulmonary nodule
left kidney
right kidney
Standardize
Merge
pulmonary nodule
kidney
Label categorization & lexical tree construction
GMAI
Clinical VQA Tasks
Departments
......
Disease Diagnosis
Cell Rec
Severity Grading
......
X-Ray
Endo
Micro
......
pulmonary nodule
......
pylorus
polyp
Question-Answer generation & selection
Generate the question body for given <Modality> / <Clinical VQA task> / <perceptual granularity> question
Option Pool
pylorus
polyp
esophagitis
ulcerative colitis
gastric metaplasia
......
Question Pool
This is a <modality> image. Which of the following options is the most appropriate to demonstrate the marked area? ....
Random pick
Question
This is an endoscopy image in Disease Diagnosis task. What is the is the most appropriate abnormality demonstrated inside the box?
Options
A. ulcerative colitis
B. pylorus
C. esophagitis
D. polyp
Disease Diagnosis
☑ Valid
□ Invalid
☑ Select
□ Not Select

Labels	Clinical VQA Tasks	Departments
bladder	Anatomical Structure Rec	Urology
debris clearance	Surgical Workflow Rec	General Surgery
neoplasia	Disease Diagnosis	Oncology
......	......	......

Data distribution



~26000 VQAs

Table 5: Statistics of the clinical VQA tasks and their sub-task abbreviations mentioned in the paper with their corresponding full terms.

Full Name	Abbreviation	Single Choice			Multiple Choice		
		Modalities	Labels	Cases	Modalities	Labels	Cases
Attribute Recognition	AR	5	26	780	1	4	40
Blood Vessels Recognition	BVR	7	15	436	-	-	-
Bone	B	6	22	655	-	-	-
Cell Recognition	CR	4	13	383	1	18	7614
Counting	C	1	38	853	-	-	-
Disease Diagnosis	DD	29	364	10167	3	26	8037
Image Quality Grading	IQG	2	10	300	-	-	-
Microorganism Recognition	MR	3	26	779	-	-	-
Muscle	M	1	5	150	-	-	-
Nervous Tissue	NT	2	4	120	-	-	-
Organ Recognition - Abdomen	OR-A	7	28	838	-	-	-
Organ Recognition - Head and Neck	OR-HN	5	16	480	-	-	-
Organ Recognition - Pelvic	OR-P	6	9	270	-	-	-
Organ Recognition - Thorax	OR-T	9	17	510	-	-	-
Severity Grading	SG	5	64	1678	-	-	-
Surgeon Action Recognition	SAR	1	23	635	-	-	-
Surgical Instrument Recognition	SIR	1	27	790	-	-	-
Surgical Workflow Recognition	SWR	1	14	420	-	-	-

Table 6: Statistics of the departments and their sub-task abbreviations mentioned in the paper with their corresponding full terms.

Full Name	Abbreviation	Single Choice			Multiple Choice		
		Modalities	Labels	Cases	Modalities	Labels	Cases
Cardiovascular Surgery	CS	9	9	270	1	1	424
Dermatology	D	1	30	894	-	-	-
Endocrinology	E	3	7	210	-	-	-
Gastroenterology and Hepatology	GH	7	60	1774	-	-	-
General Surgery	GS	6	68	2009	-	-	-
Hematology	H	6	80	2112	-	-	-
Infectious Diseases	ID	2	7	180	-	-	-
Laboratory Medicine and Pathology	LMP	2	45	1259	1	18	7614
Nephrology and Hypertension	NH	4	9	270	-	-	-
Neurosurgery	N	8	9	270	-	-	-
None (Attributes that do not belong to any department)	N/A	2	15	450	-	-	-
Obstetrics and Gynecology	OG	5	14	389	-	-	-
Oncology (Medical)	OM	20	51	1399	-	-	-
Ophthalmology	O	6	97	2232	2	11	218
Orthopedic Surgery	OS	8	54	1611	-	-	-
Otolaryngology (ENT)/Head and Neck Surgery	ENT/HNS	5	14	420	1	6	1015
Pulmonary Medicine	PM	2	55	1643	1	12	6420
Sports Medicine	SM	3	64	1919	-	-	-
Urology	U	8	33	933	-	-	-

Example

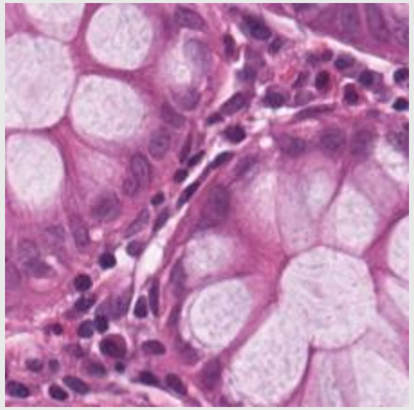
Image level

Question: Determine which option best matches the content displayed in the histology image.

Options:

A. debris
B. lymphocyte
C. normal colonic mucosa
D. smooth muscle

Please select the correct answer from the options above



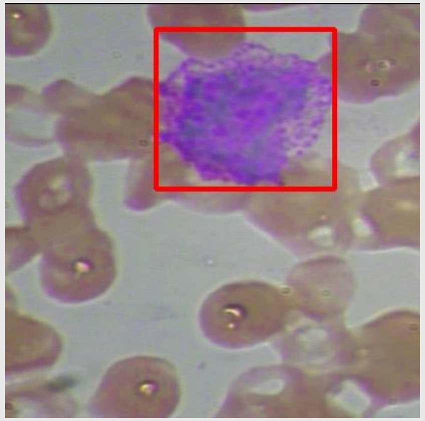
Box level

Question: Observe the microscopy image. Can you identify the target within the outlined box?

Options：

A. red blood cell
B. white blood cell
C. platelet
D. mycobacterium tuberculosis

Please select the correct answer from the options above



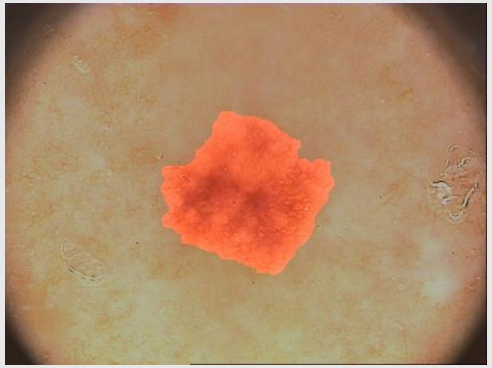
Mask level

Question: Observe the Dermoscopy image. What is the most likely abnormality shown in the highlight area?

Options：

A. pleural effusion
B. esophageal cancer
C. globules skin lesion
D. lung consolidation
E. melanocytic lesions

Please select the correct answer from the options above



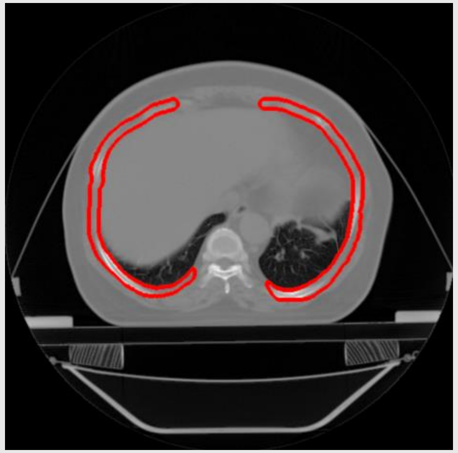
Contour level

Question: Observe the CT image. Can you identify the organ in the highlight area?

Options：

A. spinal cord
B. pulmonary artery
C. chest wall
D. Esophagus

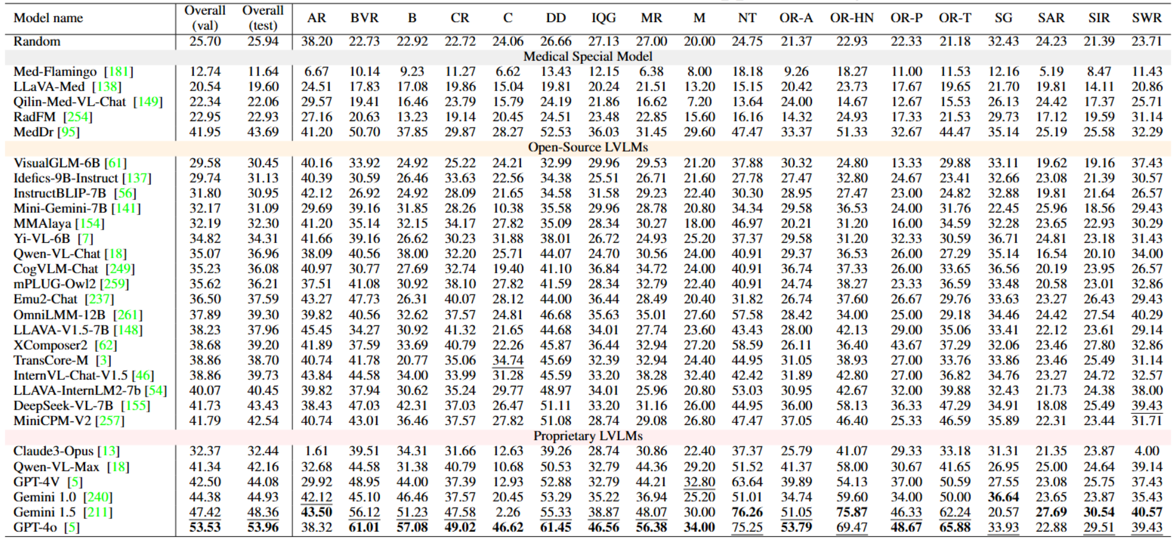
Please select the correct answer from the options above



Evaluation

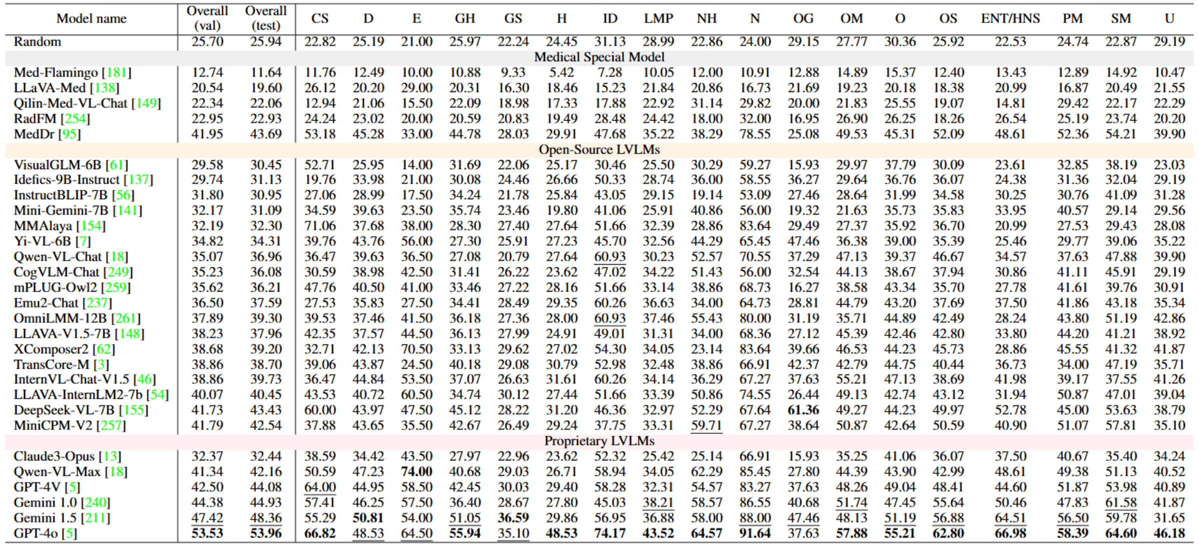
$$\text{ACC} = \frac{n_{\text{correct}}}{n_{\text{questions}}}.$$

ACCs among different VQA tasks



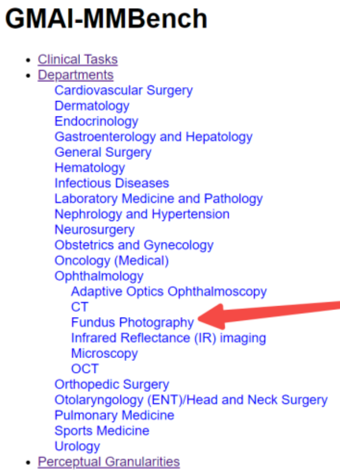
Model name	Overall (val)	Overall (test)	AR	BVR	B	CR	C	DD	IQG	MR	M	NT	OR-A	OR-HN	OR-P	OR-T	SG	SAR	SIR	SWR
Random	25.70	25.94	38.20	22.73	22.92	22.72	24.06	26.66	27.13	27.00	20.00	24.75	21.37	22.93	22.33	21.18	32.43	24.23	21.39	23.71
Medical Special Model																				
Med-Flamingo [181]	12.74	11.64	6.67	10.14	9.23	11.27	6.62	13.43	12.15	6.38	8.00	18.18	9.26	18.27	11.00	11.53	12.16	5.19	8.47	11.43
LLaVA-Med [138]	20.54	19.60	24.51	17.83	17.08	19.86	15.04	19.81	20.24	21.51	13.20	15.15	20.42	23.73	17.67	19.65	21.70	19.81	14.11	20.86
Qilin-Med-VL-Chat [149]	22.34	22.06	29.57	19.41	16.46	23.79	15.79	24.19	21.86	16.62	7.20	13.64	24.00	14.67	12.67	15.53	26.13	24.42	17.37	25.71
RadFM [254]	22.95	22.93	27.16	20.63	13.23	19.14	20.45	24.51	23.48	22.85	15.60	16.16	14.32	24.93	17.33	21.53	29.73	17.12	19.59	31.14
MedDr [95]	41.95	43.69	41.20	50.70	37.85	29.87	28.27	52.53	36.03	31.45	29.60	47.47	33.37	51.33	32.67	44.47	35.14	25.19	25.58	32.29
Open-Source LVLMs																				
VisualGLM-6B [61]	29.58	30.45	40.16	33.92	24.92	25.22	24.21	32.99	29.96	29.53	21.20	37.88	30.32	24.80	13.33	29.88	33.11	19.62	19.16	37.43
Idefics-9B-Instruct [137]	29.74	31.13	40.39	30.59	26.46	33.63	22.56	34.38	25.51	26.71	21.60	27.78	27.47	32.80	24.67	23.41	32.66	23.08	21.39	30.57
InstructBLIP-7B [56]	31.80	30.95	42.12	26.92	24.92	28.09	21.65	34.58	31.58	29.23	22.40	30.30	28.95	27.47	23.00	24.82	32.88	19.81	21.64	26.57
Mini-Gemini-7B [141]	32.17	31.09	29.69	39.16	31.85	28.26	10.38	35.58	29.96	28.78	20.80	34.34	29.58	36.53	24.00	31.76	22.45	25.96	18.56	29.43
MMAlaya [154]	32.19	32.30	41.20	35.14	32.15	34.17	27.82	35.09	28.34	30.27	18.00	46.97	20.21	31.20	16.00	34.59	32.28	23.65	22.93	30.29
Yi-VL-6B [7]	34.82	34.31	41.66	39.16	26.62	30.23	31.88	38.01	26.72	24.93	25.20	37.37	29.58	31.20	32.33	30.59	36.71	24.81	23.18	31.43
Qwen-VL-Chat [18]	35.07	36.96	38.09	40.56	38.00	32.20	25.71	44.07	24.70	30.56	24.00	40.91	29.37	36.53	26.00	27.29	35.14	16.54	20.10	34.00
CogVLM-Chat [249]	35.23	36.08	40.97	30.77	27.69	32.74	19.40	41.10	36.84	34.72	24.00	40.91	36.74	37.33	26.00	33.65	36.56	20.19	23.95	26.57
mPLUG-Owl2 [259]	35.62	36.21	37.51	41.08	30.92	38.10	27.82	41.59	28.34	32.79	22.40	40.91	24.74	38.27	23.33	36.59	33.48	20.58	23.01	32.86
Emu2-Chat [237]	36.50	37.59	43.27	47.73	26.31	40.07	28.12	44.00	36.44	28.49	20.40	31.82	26.74	37.60	26.67	29.76	33.63	23.27	26.43	29.43
OmniLMM-12B [261]	37.89	39.30	39.82	40.56	32.62	37.57	24.81	46.68	35.63	35.01	27.60	57.58	28.42	34.00	25.00	29.18	34.46	24.42	27.54	40.29
LLAVA-V1.5-7B [148]	38.23	37.96	45.45	34.27	30.92	41.32	21.65	44.68	34.01	27.74	23.60	43.43	28.00	42.13	29.00	35.06	33.41	22.12	23.61	29.14
XComposer2 [62]	38.68	39.20	41.89	37.59	33.69	40.79	22.26	45.87	36.44	32.94	27.20	58.59	26.11	36.40	43.67	37.29	32.06	23.46	27.80	32.86
TransCore-M [3]	38.86	38.70	40.74	41.78	20.77	35.06	34.74	45.69	32.39	32.94	24.40	44.95	31.05	38.93	27.00	33.76	33.86	23.46	25.49	31.14
InternVL-Chat-V1.5 [46]	38.86	39.73	43.84	44.58	34.00	33.99	31.28	45.59	33.20	38.28	32.40	42.42	31.89	42.80	27.00	36.82	34.76	23.27	24.72	32.57
LLAVA-InternLM2-7b [54]	40.07	40.45	39.82	37.94	30.62	35.24	29.77	48.97	34.01	25.96	20.80	53.03	30.95	42.67	32.00	39.88	32.43	21.73	24.38	38.00
DeepSeek-VL-7B [155]	41.73	43.43	38.43	47.03	42.31	37.03	26.47	51.11	33.20	31.16	26.00	44.95	36.00	58.13	36.33	47.29	34.91	18.08	25.49	39.43
MiniCPM-V2 [257]	41.79	42.54	40.74	43.01	36.46	37.57	27.82	51.08	28.74	29.08	26.80	47.47	37.05	46.40	25.33	46.59	35.89	22.31	23.44	31.71
Proprietary LVLMs																				
Claude3-Opus [13]	32.37	32.44	1.61	39.51	34.31	31.66	12.63	39.26	28.74	30.86	22.40	37.37	25.79	41.07	29.33	33.18	31.31	21.35	23.87	4.00
Qwen-VL-Max [18]	41.34	42.16	32.68	44.58	31.38	40.79	10.68	50.53	32.79	44.36	29.20	51.52	41.37	58.00	30.67	41.65	26.95	25.00	24.64	39.14
GPT-4V [5]	42.50	44.08	29.92	48.95	44.00	37.39	12.93	52.88	32.79	44.21	32.80	63.64	39.89	54.13	37.00	50.59	27.55	23.08	25.75	37.43
Gemini 1.0 [240]	44.38	44.93	42.12	45.10	46.46	37.57	20.45	53.29	35.22	36.94	25.20	51.01	34.74	59.60	34.00	50.00	**36.64**	23.65	23.87	35.43
Gemini 1.5 [211]	47.42	48.36	**43.50**	56.12	51.23	47.58	2.26	55.33	38.87	48.07	30.00	**76.26**	51.05	**75.87**	46.33	62.24	20.57	**27.69**	**30.54**	**40.57**
GPT-4o [5]	**53.53**	**53.96**	38.32	**61.01**	**57.08**	**49.02**	**46.62**	**61.45**	**46.56**	**56.38**	**34.00**	75.25	**53.79**	69.47	**48.67**	**65.88**	33.93	22.88	29.51	39.43

ACCs among different departments



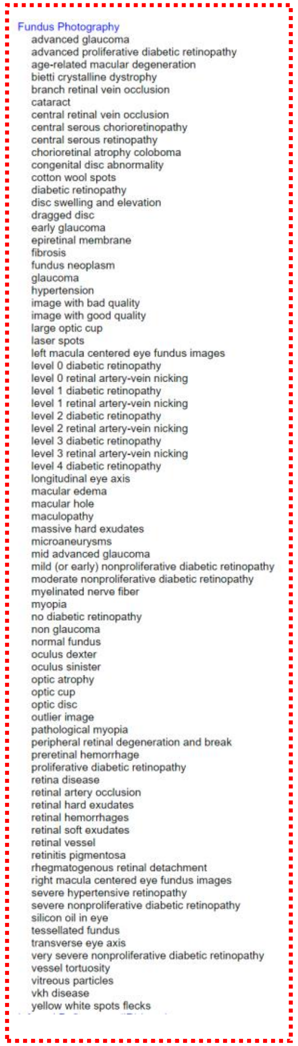
Model name	Overall (val)	Overall (test)	CS	D	E	GH	GS	H	ID	LMP	NH	N	OG	OM	O	OS	ENT/HNS	PM	SM	U
Random	25.70	25.94	22.82	25.19	21.00	25.97	22.24	24.45	31.13	28.99	22.86	24.00	29.15	27.77	30.36	25.92	22.53	24.74	22.87	29.19
Medical Special Model																				
Med-Flamingo [181]	12.74	11.64	11.76	12.49	10.00	10.88	9.33	5.42	7.28	10.05	12.00	10.91	12.88	14.89	15.37	12.40	13.43	12.89	14.92	10.47
LLaVA-Med [138]	20.54	19.60	26.12	20.20	29.00	20.31	16.30	18.46	15.23	21.84	20.86	16.73	21.69	19.23	20.18	18.38	20.99	16.87	20.49	21.55
Qilin-Med-VL-Chat [149]	22.34	22.06	12.94	21.06	15.50	22.09	18.98	17.33	17.88	22.92	31.14	29.82	20.00	21.83	25.55	19.07	14.81	29.42	22.17	22.29
RadFM [254]	22.95	22.93	24.24	23.02	20.00	20.59	20.83	19.49	28.48	24.42	18.00	32.00	16.95	26.90	26.25	18.26	26.54	25.19	23.74	20.20
MedDr [95]	41.95	43.69	53.18	45.28	33.00	44.78	28.03	29.91	47.68	35.22	38.29	78.55	25.08	49.53	45.31	52.09	48.61	52.36	54.21	39.90
Open-Source LVLMs																				
VisualGLM-6B [61]	29.58	30.45	52.71	25.95	14.00	31.69	22.06	25.17	30.46	25.50	30.29	59.27	15.93	29.97	37.79	30.09	23.61	32.85	38.19	23.03
Idefics-9B-Instruct [137]	29.74	31.13	19.76	33.98	21.00	30.08	24.46	26.66	50.33	28.74	36.00	58.55	36.27	29.64	36.76	36.07	24.38	31.36	32.04	29.19
InstructBLIP-7B [56]	31.80	30.95	27.06	28.99	17.50	34.24	21.78	25.84	43.05	29.15	19.14	53.09	27.46	28.64	31.99	34.58	30.25	30.76	41.09	31.28
Mini-Gemini-7B [141]	32.17	31.09	34.59	39.63	23.50	35.74	23.46	19.80	41.06	25.91	40.86	56.00	19.32	21.63	35.73	35.83	33.95	40.57	29.14	29.56
MMAlaya [154]	32.19	32.30	71.06	37.68	38.00	28.30	27.40	27.64	51.66	32.39	28.86	83.64	29.49	27.37	35.92	36.70	20.99	27.53	29.43	28.08
Yi-VL-6B [7]	34.82	34.31	39.76	43.76	56.00	27.30	25.91	27.23	45.70	32.56	44.29	65.45	47.46	36.38	39.00	35.39	25.46	29.77	39.06	35.22
Qwen-VL-Chat [18]	35.07	36.96	36.47	39.63	36.50	27.08	20.79	27.64	60.93	30.23	52.57	70.55	37.29	47.13	39.37	46.67	34.57	37.63	47.88	39.90
CogVLM-Chat [249]	35.23	36.08	30.59	38.98	42.50	31.41	26.22	23.62	47.02	34.22	51.43	56.00	32.54	44.13	38.67	37.94	30.86	41.11	45.91	29.19
mPLUG-Owl2 [259]	35.62	36.21	47.76	40.50	41.00	33.46	27.22	28.16	51.66	33.14	38.86	68.73	16.27	38.58	43.34	35.70	27.78	41.61	39.76	30.91
Emu2-Chat [237]	36.50	37.59	27.53	35.83	27.50	34.41	28.49	29.35	60.26	36.63	34.00	64.73	28.81	44.79	43.20	37.69	37.50	41.86	43.18	35.34
OmniLMM-12B [261]	37.89	39.30	39.53	37.46	41.50	36.18	27.36	28.00	60.93	37.46	55.43	80.00	31.19	35.71	44.89	42.49	28.24	43.80	51.19	42.86
LLAVA-V1.5-7B [148]	38.23	37.96	42.35	37.57	44.50	36.13	27.99	24.91	49.01	31.31	34.00	68.36	27.12	45.39	42.46	42.80	33.80	44.20	41.21	38.92
XComposer2 [62]	38.68	39.20	32.71	42.13	70.50	33.13	29.62	27.02	54.30	34.05	23.14	83.64	39.66	46.53	44.23	45.73	28.86	45.55	41.32	41.87
TransCore-M [3]	38.86	38.70	39.06	43.87	24.50	40.18	29.08	30.79	52.98	32.48	38.86	66.91	42.37	42.79	44.75	40.44	36.73	34.00	47.19	35.71
InternVL-Chat-V1.5 [46]	38.86	39.73	36.47	44.84	53.50	37.07	26.63	31.61	60.26	34.14	36.29	67.27	37.63	55.21	47.13	38.69	41.98	39.17	37.55	41.26
LLAVA-InternLM2-7b [54]	40.07	40.45	43.53	40.72	60.50	34.74	30.12	27.44	51.66	33.39	50.86	74.55	26.44	49.13	42.74	43.12	31.94	50.87	47.01	39.04
DeepSeek-VL-7B [155]	41.73	43.43	60.00	43.97	47.50	45.12	28.22	31.20	46.36	32.97	52.29	67.64	**61.36**	49.27	44.23	49.97	52.78	45.00	53.63	38.79
MiniCPM-V2 [257]	41.79	42.54	37.88	43.65	35.50	42.67	26.49	29.24	37.75	33.31	59.71	67.27	38.64	50.87	42.64	50.59	40.90	51.07	57.81	35.10
Proprietary LVLMs																				
Claude3-Opus [13]	32.37	32.44	38.59	34.42	43.50	27.97	22.96	23.62	52.32	25.42	25.14	66.91	15.93	35.25	41.06	36.07	37.50	40.67	35.40	34.24
Qwen-VL-Max [18]	41.34	42.16	50.59	47.23	**74.00**	40.68	29.03	26.71	58.94	34.05	62.29	85.45	27.80	44.39	43.90	42.99	48.61	49.38	51.13	40.52
GPT-4V [5]	42.50	44.08	64.00	44.95	58.50	42.45	30.03	29.40	58.28	32.31	54.57	83.27	37.63	48.26	49.04	48.41	44.60	51.87	53.98	40.89
Gemini 1.0 [240]	44.38	44.93	57.41	46.25	57.50	36.40	28.67	27.80	45.03	38.21	58.57	86.55	40.68	51.74	47.45	55.64	50.46	47.83	61.58	41.87
Gemini 1.5 [211]	47.42	48.36	55.29	**50.81**	54.00	51.05	**36.59**	29.86	56.95	36.88	58.00	88.00	47.46	48.13	51.19	56.88	64.51	56.50	59.78	31.65
GPT-4o [5]	**53.53**	**53.96**	**66.82**	48.53	64.50	**55.94**	35.10	**48.53**	**74.17**	**43.52**	**64.57**	**91.64**	37.63	**57.88**	**55.21**	**62.80**	**66.98**	**58.39**	**64.60**	**46.18**

Lexical tree



GMAI-MMBench

- Clinical Tasks
- Departments
  - Cardiovascular Surgery
  - Dermatology
  - Endocrinology
  - Gastroenterology and Hepatology
  - General Surgery
  - Hematology
  - Infectious Diseases
  - Laboratory Medicine and Pathology
  - Nephrology and Hypertension
  - Neurosurgery
  - Obstetrics and Gynecology
  - Oncology (Medical)
  - Ophthalmology
    - Adaptive Optics Ophthalmoscopy
    - CT
    - Fundus Photography
    - Infrared Reflectance (IR) imaging
    - Microscopy
    - OCT
  - Orthopedic Surgery
  - Otolaryngology (ENT)/Head and Neck Surgery
  - Pulmonary Medicine
  - Sports Medicine
  - Urology
- Perceptual Granularities



Fundus Photography

- advanced glaucoma
- advanced proliferative diabetic retinopathy
- age-related macular degeneration
- bietti crystalline dystrophy
- branch retinal vein occlusion
- cataract
- central retinal vein occlusion
- central serous chorioretinopathy
- central serous retinopathy
- chorioretinal atrophy coloboma
- congenital disc abnormality
- cotton wool spots
- diabetic retinopathy
- disc swelling and elevation
- dragged disc
- early glaucoma
- epiretinal membrane
- fibrosis
- fundus neoplasm
- glaucoma
- hypertension
- image with bad quality
- image with good quality
- large optic cup
- laser spots
- left macula centered eye fundus images
- level 0 diabetic retinopathy
- level 0 retinal artery-vein nicking
- level 1 diabetic retinopathy
- level 1 retinal artery-vein nicking
- level 2 diabetic retinopathy
- level 2 retinal artery-vein nicking
- level 3 diabetic retinopathy
- level 3 retinal artery-vein nicking
- level 4 diabetic retinopathy
- longitudinal eye axis
- macular edema
- macular hole
- maculopathy
- massive hard exudates
- microaneurysms
- mid advanced glaucoma
- mild (or early) nonproliferative diabetic retinopathy
- moderate nonproliferative diabetic retinopathy
- myelinated nerve fiber
- myopia
- no diabetic retinopathy
- non glaucoma
- normal fundus
- oculus dexter
- oculus sinister
- optic atrophy
- optic cup
- optic disc
- outlier image
- pathological myopia
- peripheral retinal degeneration and break
- preretinal hemorrhage
- proliferative diabetic retinopathy
- retina disease
- retinal artery occlusion
- retinal hard exudates
- retinal hemorrhages
- retinal soft exudates
- retinal vessel
- retinitis pigmentosa
- rhegmatogenous retinal detachment
- right macula centered eye fundus images
- severe hypertensive retinopathy
- severe nonproliferative diabetic retinopathy
- silicon oil in eye
- tessellated fundus
- transverse eye axis
- very severe nonproliferative diabetic retinopathy
- vessel tortuosity
- vitreous particles
- vkh disease
- yellow white spots flecks

Question list

🟥 Keywords from ⬚

Q1: ……
Options : A: … , B: … , C: … , D: advanced glaucoma
answer = advanced glaucoma (hidden, can't be seen by users)

Q2: ……
Options : A: … , B: … , C: advanced proliferative diabetic retinopathy , D: …
answer = advanced proliferative diabetic retinopathy (hidden, can't be seen by users)

Q3: ……
Options : A: age-related macular degeneration, B: … , C: … , D: …
answer = age-related macular degeneration(hidden, can't be seen by users)

Q4: ……
Options : A: bietti crystalline dystrophy, B: … , C: … D: …
answer = bietti crystalline dystrophy (hidden, can't be seen by users)

……

Model list

GPT-4o
Gemini
LLAVA series
Claude3-Opus
InternVL
Yi-VL
Qwen-Max
DeepSeek
……

Model	Accuracy
GPT-4o	…
Claude3-Opus	…
Qwen-Max	…
……	…