Model	Overall Acc. (%)	Error Category Accuracy (%)						
		AFM	CPE	LEA	LDE	OUG	SPE	VSM
SkyWork$_{R1V3-38B}$	**62.3**	**65.8**	**67.5**	**62.9**	57.7	71.9	51.5	**62.5**
MiniCPM$_{V-4.5}$	58.1	61.1	57.7	61.8	48.5	**77.0**	**59.2**	46.5
Qwen$_{2.5-VL}$	57.0	52.1	61.0	44.9	**67.5**	67.4	58.0	42.4
Intern$_{VL-2.5-78B}$	53.6	53.2	56.1	58.4	63.9	57.8	45.6	41.0
VL-Rethinker$_{7B}$	52.7	48.9	56.9	42.7	60.8	60.7	55.6	38.2
GPT$_5$	52.6	53.7	57.7	50.6	49.5	61.5	43.8	54.2
Eagle$_{2.5-8B}$	49.9	45.3	59.3	50.6	47.9	69.6	44.4	38.2
GPT$_{o3}$	47.9	50.0	56.9	47.2	43.8	51.1	36.7	53.5
MiMo$_{VL-7B-RL-2508}$	47.4	45.3	48.8	46.1	50.0	64.4	35.5	44.4
GLM$_{4.1V-9B-Thinking}$	43.8	43.2	51.2	40.4	45.9	51.1	36.7	38.9
NVILA$_{15B}$	39.4	34.2	31.7	24.7	66.0	42.2	33.1	30.6
MM-Eureka$_{Qwen-32B}$	38.7	31.6	34.1	33.7	55.2	43.7	36.7	30.6
Phi$_{3.5-vision-instruct}$	37.8	34.7	38.2	30.3	38.7	42.2	38.5	40.3
Ovis$_{2.5-9B}$	34.3	29.5	34.1	24.7	39.2	45.2	34.3	29.9
Pixtral$_{12B-2409}$	27.4	27.9	29.3	19.1	33.5	28.1	29.0	19.4
CogVLM$_{2-Llama3-19B}$	22.3	22.6	17.9	12.4	34.0	22.2	20.1	18.8
Kimi$_{VL-A3B-Thinking}$	21.9	23.7	24.4	13.5	27.3	23.0	18.3	18.8
Idefics$_{2-8B}$	17.5	11.1	11.4	6.7	47.4	10.4	16.6	5.6
DeepSeek$_{VL2}$	16.2	17.9	16.3	20.2	8.8	17.8	13.6	22.9
MMaDA$_{8B-MixCoT}$	12.8	17.4	22.8	13.5	8.8	17.0	6.5	6.9
Yi$_{VL-34B}$	12.3	11.6	7.3	4.5	23.7	10.4	11.8	9.0