



CXPMRG-Bench: Pre-training and Benchmarking for X-ray Medical Report Generation on CheXpert Plus Dataset

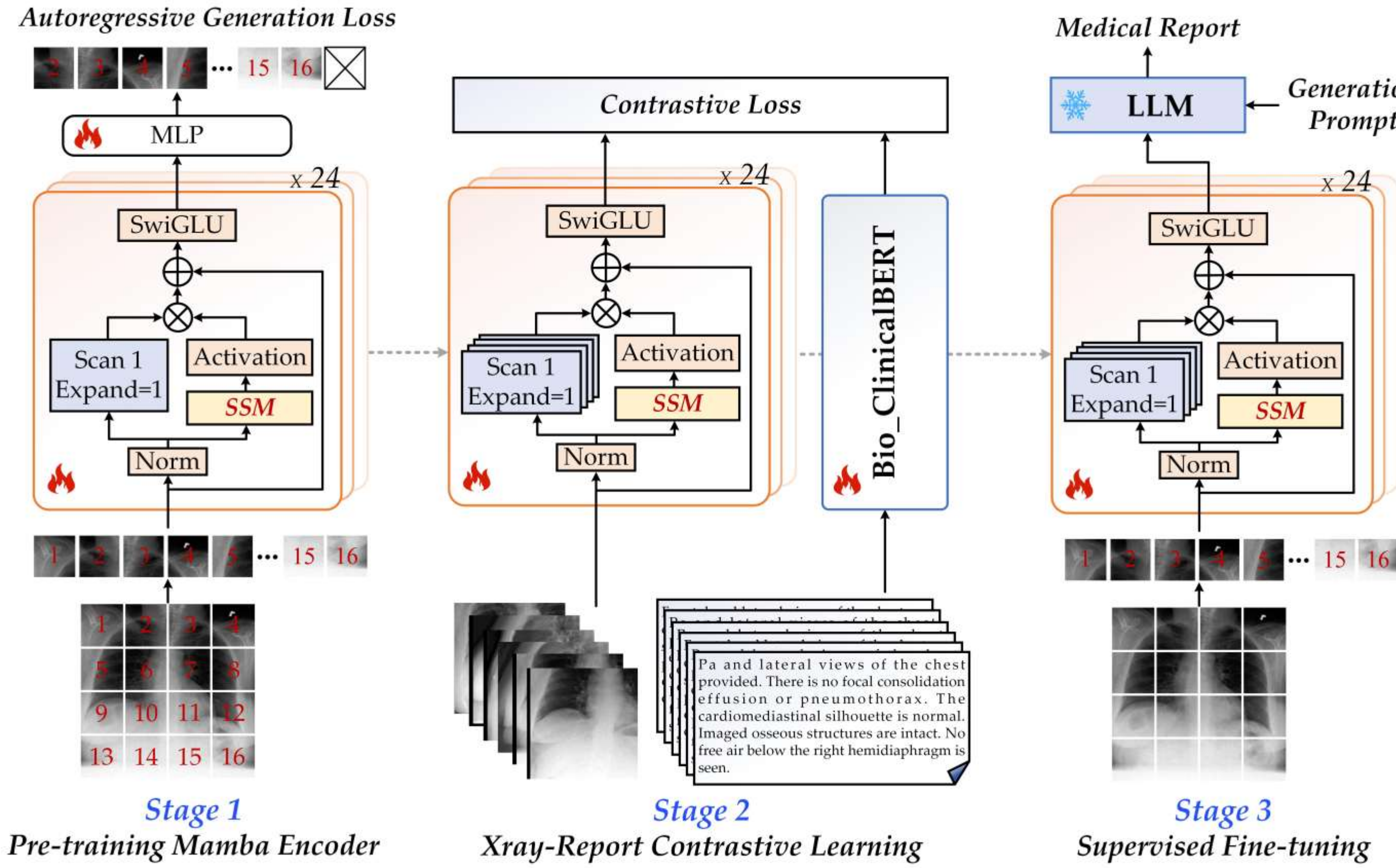


Xiao Wang, Fuling Wang, Yuehang Li, Qingchuan Ma, Shiao Wang, Bo Jiang, Jin Tang

Motivation

- Transformer-based vision backbone models exhibit quadratic time complexity, whereas the Mamba architecture demonstrates linear computational complexity. Current medical pre-training models predominantly employ single-stage frameworks that utilize either image-only datasets or image-report paired data, resulting in inadequate utilization of comprehensive medical data resources.
- The newly released CheXpert Plus, a large-scale thoracic imaging dataset comparable to MIMIC-CXR, holds significant research value. However, its lack of standardized evaluation benchmarks presents challenges for comparative methodological validation in downstream studies.

Proposed MambaXray-VL Pre-training Framework



Our three-stage training—Mamba-based autoregressive pre-training, X-ray image-report contrastive learning, and supervised fine-tuning—maximizes use of standalone X-rays as well as coarse- and high-precision image-report pairs.

Visualization & Conclusions

Visualization of model predictions on MIMIC-CXR. X-ray images and their corresponding ground-truths, along with the output of our model and R2GenGPT model generation reports on the MIMIC-CXR dataset. Matching sentences in our report are highlighted in yellow, R2GenGPT matching sentences are highlighted in cyan, and sentences matching by both models are highlighted in pink.

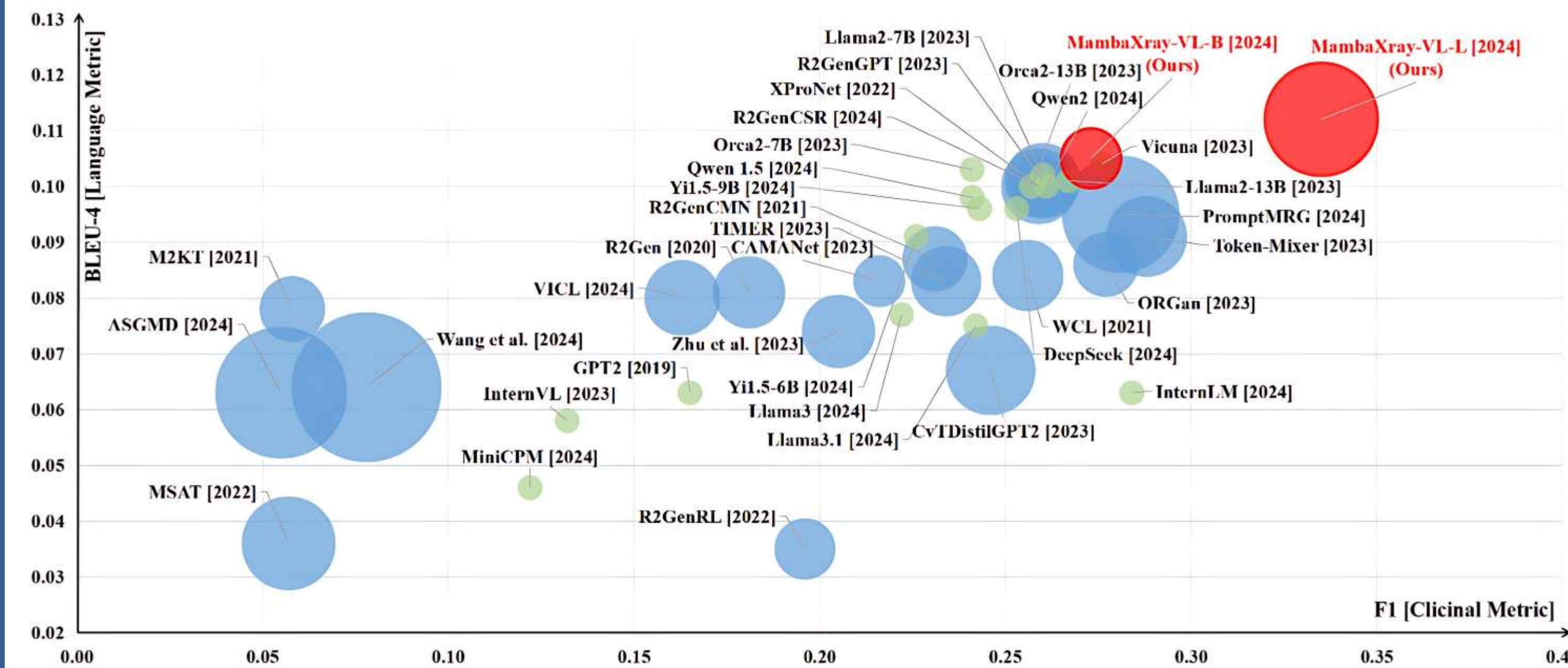
Image	Ground Truth	Ours	R2GenGPT
	Pa and lateral views of the chest provided. Midline sternotomy wires and mediastinal clips are again noted. The previously noted port-a-cath has been removed. The lungs are clear bilaterally without focal consolidation effusion or pneumothorax. Cardiomediastinal silhouette is stable. Bony structures are intact. No free air below the right hemidiaphragm is seen.	Ap upright and lateral views of the chest provided. Midline sternotomy wires and mediastinal clips are again noted. There is no focal consolidation large effusion or pneumothorax. The cardiomediastinal silhouette is stable. Bony structures are intact. No free air below the right hemidiaphragm is seen.	Frontal and lateral views of the chest were obtained. The patient is status post median sternotomy and cabg. The cardiac and mediastinal silhouettes are stable. There is no focal consolidation pleural effusion or pneumothorax. Mild pulmonary vascular congestion is noted. Degenerative changes are seen in the thoracic spine.
	Pa and lateral views of the chest provided. There is no focal consolidation effusion or pneumothorax. The cardiomediastinal silhouette is normal. Imaged osseous structures are intact. No free air below the right hemidiaphragm is seen. Elevation of the right hemidiaphragm is unchanged from chest radiograph.	Pa and lateral views of the chest provided. There is no focal consolidation effusion or pneumothorax. The cardiomediastinal silhouette is normal. Imaged osseous structures are intact. No free air below the right hemidiaphragm is seen. Clips are noted in the right upper quadrant of the abdomen.	Pa and lateral views of the chest provided. There is no focal consolidation effusion or pneumothorax. The cardiomediastinal silhouette is normal. Imaged osseous structures are intact. No free air below the right hemidiaphragm is seen. Surgical clips in the right upper quadrant suggest prior cholecystectomy.
	Frontal and lateral views of the chest were obtained. Dual-lead left-sided pacemaker is again seen with leads extending to the expected positions of the right atrium and right ventricle. The lungs are hyperinflated with flattening of the diaphragms suggesting chronic obstructive pulmonary disease. No pleural effusion or pneumothorax is seen. Slight increased opacity at the right lung base best seen on the front.	Frontal and lateral views of the chest were obtained. Dual-lead left-sided pacemaker is again seen with leads extending to the expected positions of the right atrium and right ventricle. There is no evidence of pneumothorax or pleural effusion. The lungs are hyperinflated with flattening of the diaphragms consistent with chronic obstructive pulmonary disease. Cardiomediastinal silhouette is stable. Bony structures are intact.	Frontal and lateral chest radiographs demonstrate hyper-expanded lungs with flattening of the diaphragms consistent with chronic obstructive pulmonary disease. The cardiomediastinal silhouette is normal. Imaged osseous structures are intact. No free air below the right hemidiaphragm is seen. The cardiac mediastinal and hilar contours are unremarkable. A left-sided pacemaker device is noted with leads terminating in the right atrium and right ventricle.
	As compared to the previous radiograph there is no relevant change. The monitoring and support devices are constant. Low lung volumes. Borderline size of the cardiac silhouette. Mild pulmonary edema. Moderate retrocardiac atelectasis. No evidence of pneumonia.	As compared to the previous radiograph there is no relevant change. The monitoring and support devices are in unchanged position. Low lung volumes with minimal atelectasis at both lung bases. No larger pleural effusions or pneumothorax. Borderline size of the cardiac silhouette. No pulmonary edema. No other parenchymal abnormalities.	In comparison with the study of the monitoring and support devices remain in place. Continued enlargement of the cardiac silhouette with pulmonary vascular congestion and bilateral pleural effusions with compressive atelectasis at the bases. No evidence of acute focal pneumonia or pneumothorax. Central catheters remain in place.

Experiment

Comparison of our model with mainstream models on CheXpert Plus.

Index	Algorithm	Publis	Encoder	Decoder	B4	R	M	C	P	R	F1	Time (min)	Param (M)	Code
#01	R2GenRL [42]	ACL22	Transformer	Transformer	0.035, 0.186, 0.101, 0.012	0.193, 0.229, 0.196	44.33	59.87						URL
#02	XProNet [53]	ECCV22	Transformer	Transformer	0.100, 0.265, 0.146, 0.121	0.314, 0.247, 0.259	6.3	62.35						URL
#03	MSAT [61]	MICCAI22	ViT-B/16	Transformer	0.036, 0.156, 0.066, 0.018	0.044, 0.142, 0.057	5.72	141.10						URL
#04	ORGen [24]	ACL23	CNN	Transformer	0.086, 0.261, 0.135, 0.107	0.288, 0.287, 0.277	46.66	67.50						URL
#05	M2KT [68]	CNSA	CNN	Transformer	0.078, 0.247, 0.101, 0.072	0.044, 0.142, 0.058	22.5	69.07						URL
#06	TIMER [64]	CHIL23	Transformer	Transformer	0.083, 0.254, 0.121, 0.104	0.345, 0.238, 0.234	26.5	79.28						URL
#07	CvT2DistilGPT2 [39]	AIM23	Transformer	GPT2	0.067, 0.238, 0.118, 0.101	0.285, 0.252, 0.246	13.93	128						URL
#08	R2Gen [8]	EMNLP20	Transformer	Transformer	0.081, 0.246, 0.113, 0.077	0.318, 0.200, 0.181	110.05	83.5						URL
#09	R2GenCMN [9]	ACL21	Transformer	Transformer	0.087, 0.256, 0.127, 0.102	0.329, 0.241, 0.231	66.08	67.70						URL
#10	Zhu et al. [76]	MICCAI23	Transformer	Transformer	0.074, 0.235, 0.128, 0.078	0.217, 0.308, 0.205	10.03	85.95						URL
#11	CAMANet [54]	IEEE JBHI23	Swin-Former	Transformer	0.083, 0.249, 0.118, 0.090	0.328, 0.224, 0.216	23.08	43.22						URL
#12	ASGMD [65]	ESWA24	ResNet-101	Transformer	0.063, 0.220, 0.094, 0.044	0.146, 0.108, 0.055	87.37	277.41						URL
#13	Token-Mixer [69]	IEEE TMI23	ResNet-50	Transformer	0.091, 0.261, 0.135, 0.098	0.309, 0.270, 0.288	17.54	104.34						URL
#14	PromptMRG [28]	AAAI24	ResNet-101	Bert	0.095, 0.222, 0.121, 0.044	0.258, 0.265, 0.281	108.45	219.92						URL
#15	R2GenGPT [63]	Meta-Rad.23	Swin-Transformer	Llama2	0.101, 0.266, 0.145, 0.123	0.315, 0.244, 0.260	77.8	90.9						URL
#16	WCL [66]	EMNLP21	Transformer	Transformer	0.084, 0.253, 0.126, 0.103	0.335, 0.259, 0.256	24.08	81.29						URL
#17	R2GenCSR [57]	arXiv24	V-Mamba	Llama2	0.100, 0.265, 0.146, 0.121	0.315, 0.247, 0.259	31.2	91.7						URL
#18	VLCL [7]	arXiv24	Transformer	Transformer	0.080, 0.247, 0.114, 0.072	0.341, 0.175, 0.163	123.71	91.46						URL
#19	Wang et al. [58]	arXiv24	ViT	Llama2	0.064, 0.220, 0.110, 0.059	0.175, 0.099, 0.078	10.82	358.80						URL
#20	MambaXray-VL-B	Ours	MambaXray-VL	Llama2	0.105, 0.267, 0.149, 0.117	0.333, 0.264, 0.273	50.66	57.31						URL
#21	MambaXray-VL-L	Ours	MambaXray-VL	Llama2	0.112, 0.276, 0.157, 0.139	0.377, 0.319, 0.335	55.18	202.32						URL

Benchmark Overview



An overview of Banchmark on the CheXpert Plus dataset, where blue represents the mainstream models, green represents different LLM/VLM, and red represents the model we proposed.

Our method offers three advantages:

- A comprehensive benchmarking framework has been established for the newly released CheXpert Plus dataset, ensuring systematic and equitable performance evaluation.
- We propose a three-stage medical pre-training framework that achieves comprehensive integration of three heterogeneous data modalities.
- The proposed MambaXray-VL architecture demonstrates seamless integration capability as a vision backbone for existing models, achieving significant performance enhancements.