UNIVERSITAS
GADJAH MADA
MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL
Ahmad Alif Robit Al Hazmi, Moh. Edi Wibowo, S.Kom., M.Kom., Ph.D.;Professor Jia-Ching Wang
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

[1] Emily Alsentzer, John Murphy, William Boag, Wei-Hung Weng, Di Jindi, Tristan Naumann, and Matthew McDermott. Publicly available clinical bert embeddings. In *Proceedings of the 2nd clinical natural language processing workshop*, pages 72–78, 2019.

[2] Ashwini Amin, Prakashini Koteshwara, Siddalingaswamy PC, Stanley Mathew, et al. A systematic literature review on mammography: deep learning techniques for breast cancer detection with global and asian perspectives. *BMC cancer*, 25(1):1–23, 2025.

[3] Marco Cantone, Claudio Marrocco, Francesco Tortorella, and Alessandro Bria. Convolutional networks and transformers for mammography classification: an experimental study. *Sensors*, 23(3):1229, 2023.

[4] Zhenjie Cao, Zhuo Deng, Jie Ma, Jintao Hu, and Lan Ma. Mammovlm: A generative large vision–language model for mammography-related diagnostic assistance. *Information Fusion*, 118:102998, 2025.

[5] Xiao Fang, Yi Lin, Dong Zhang, Kwang-Ting Cheng, and Hao Chen. Aligning medical images with general knowledge from large language models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 57–67. Springer, 2024.

[6] Shantanu Ghosh, Clare B Poynton, Shyam Visweswaran, and Kayhan Batmanghelich. Mammo-clip: A vision language foundation model to enhance data efficiency and robustness in mammography. In *International conference on medical image computing and computer-assisted intervention*, pages 632–642. Springer, 2024.

[7] Chaoqin Huang, Aofan Jiang, Jinghao Feng, Ya Zhang, Xinchao Wang, and Yanfeng Wang. Adapting visual-language models for generalizable anomaly detection in medical images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11375–11385, 2024.

[8] Shih-Cheng Huang, Liyue Shen, Matthew P Lungren, and Serena Yeung. Gloria: A multimodal global-local representation learning framework for label-efficient medical image recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3942–3951, 2021.

[9] Zhi Huang, Federico Bianchi, Mert Yuksekgonul, Thomas J Montine, and James Zou. A visual–language foundation model for pathology image analysis using medical twitter. *Nature medicine*, 29(9):2307–2316, 2023.

[10] Kshitiz Jain, Krithika Rangarajan, and Chetan Arora. Follow the radiologist: Clinically relevant multi-view cues for breast cancer detection from mammograms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 102–112. Springer, 2024.

[11] Alistair EW Johnson, Tom J Pollard, Nathaniel R Greenbaum, Matthew P Lungren, Chih-ying Deng, Yifan Peng, Zhiyong Lu, Roger G Mark, Seth J Berkowitz, and Steven Horng. Mimic-cxr-jpg,

UNIVERSITAS
GADJAH MADA
MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL
Ahmad Alif Robit Al Hazmi, Moh. Edi Wibowo, S.Kom., M.Kom., Ph.D.;Professor Jia-Ching Wang
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

a large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*, 2019.

[12] Muhammad Uzair Khattak, Hanoona Rasheed, Muhammad Maaz, Salman Khan, and Fahad Shahbaz Khan. Maple: Multi-modal prompt learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19113–19122, 2023.

[13] Yangguang Li, Feng Liang, Lichen Zhao, Yufeng Cui, Wanli Ouyang, Jing Shao, Fengwei Yu, and Junjie Yan. Supervision exists everywhere: A data efficient contrastive language-image pre-training paradigm. *arXiv preprint arXiv:2110.05208*, 2021.

[14] Yuhang Liu, Fandong Zhang, Qianyi Zhang, Siwen Wang, Yizhou Wang, and Yizhou Yu. Cross-view correspondence reasoning based on bipartite graph convolutional network for mammogram mass detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3812–3822, 2020.

[15] Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

[16] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[17] William Lotter, Abdul Rahman Diab, Bryan Haslam, Jiye G Kim, Giorgia Grisot, Eric Wu, Kevin Wu, Jorge Onieva Onieva, Yun Boyer, Jerrold L Boxerman, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nature medicine*, 27(2):244–249, 2021.

[18] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg S Corrado, Ara Darzi, et al. International evaluation of an ai system for breast cancer screening. *Nature*, 577(7788):89–94, 2020.

[19] Vishwesh Nath, Wenqi Li, Dong Yang, Andriy Myronenko, Mingxin Zheng, Yao Lu, Zhijian Liu, Hongxu Yin, Yee Man Law, Yucheng Tang, et al. Vila-m3: Enhancing vision-language models with medical expert knowledge. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 14788–14798, 2025.

[20] Hieu T Nguyen, Ha Q Nguyen, Hieu H Pham, Khanh Lam, Linh T Le, Minh Dao, and Van Vu. Vindr-mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Scientific Data*, 10(1):277, 2023.

[21] Thanh-Huy Nguyen, Quang Hien Kha, Thai Ngoc Toan Truong, Ba Thinh Lam, Ba Hung Ngo, Quang Vinh Dinh, and Nguyen Quoc Khanh Le. Towards robust natural-looking mammography lesion synthesis on ipsilateral dual-views breast cancer analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2564–2573, 2023.

[22] American College of Radiology et al. Breast imaging reporting and data system. *BI-RADS*, 2003.

[23] Parita Oza, Urvi Oza, Rajiv Oza, Paawan Sharma, Samir Patel, Pankaj Kumar, and Bakul Gohel. Digital mammography dataset for breast cancer diagnosis research (dmid) with breast mass segmentation analysis. *Biomedical Engineering Letters*, 14(2):317–330, 2024.

[24] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

UNIVERSITAS
GADJAH MADA
MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL
Ahmad Alif Robit Al Hazmi, Moh. Edi Wibowo, S.Kom., M.Kom., Ph.D.;Professor Jia-Ching Wang
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[25] Ji Seung Ryu, Hyunyoung Kang, Yuseong Chu, and Sejung Yang. Vision-language foundation models for medical imaging: a review of current practices and innovations. *Biomedical Engineering Letters*, pages 1–22, 2025.

[26] Jae Won Seo, Young Jae Kim, and Kwang Gi Kim. Leveraging paired mammogram views with deep learning for comprehensive breast cancer detection. *Scientific Reports*, 15(1):4406, 2025.

[27] Chong Su, Zhenghua Gong, and Jie Cao. Multi-view co-occurrence and dual-modality framework for breast cancer classification. *Journal of King Saud University Computer and Information Sciences*, 37(5):78, 2025.

[28] Hongna Tan, Qingxia Wu, Yaping Wu, Bingjie Zheng, Bo Wang, Yan Chen, Lijuan Du, Jing Zhou, Fangfang Fu, Huihui Guo, et al. Mammography-based artificial intelligence for breast cancer detection, diagnosis, and bi-rads categorization using multi-view and multi-level convolutional neural networks. *Insights into Imaging*, 16(1):109, 2025.

[29] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

[30] Yuxing Tang, Zhenjie Cao, Yanbo Zhang, Zhicheng Yang, Zongcheng Ji, Yiwei Wang, Mei Han, Jie Ma, Jing Xiao, and Peng Chang. Leveraging large-scale weakly labeled data for semi-supervised mass detection in mammograms. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3855–3864, 2021.

[31] Rongfang Wang, Jiasheng Chen, Xinlong Zhang, Jing Wang, Hui Liu, Zhiguo Zhou, and Kai Wang. Medpro-dg: Domain-aware masked contrastive prompt learning of institution generalization for outcome prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 413–422. Springer, 2025.

[32] Zifeng Wang, Zhenbang Wu, Dinesh Agarwal, and Jimeng Sun. Medclip: Contrastive learning from unpaired medical images and text. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2022, page 3876, 2022.

[33] Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. Medklip: Medical knowledge enhanced language-image pre-training for x-ray diagnosis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21372–21383, 2023.

[34] Jing Wu, Yuli Wang, Zhusi Zhong, Weihua Liao, Natalia Trayanova, Zhicheng Jiao, and Harrison X Bai. Vision-language foundation model for 3d medical imaging. *npj Artificial Intelligence*, 1(1):17, 2025.

[35] Lewei Yao, Runhui Huang, Lu Hou, Guansong Lu, Minzhe Niu, Hang Xu, Xiaodan Liang, Zhenguo Li, Xin Jiang, and Chunjing Xu. Filip: Fine-grained interactive language-image pre-training. *arXiv preprint arXiv:2111.07783*, 2021.

[36] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917*, 2022.

[37] Xiaohua Zhai, Xiao Wang, Basil Mustafa, Andreas Steiner, Daniel Keysers, Alexander Kolesnikov, and Lucas Beyer. Lit: Zero-shot transfer with locked-image text tuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18123–18133, 2022.

UNIVERSITAS
GADJAH MADA
MAMMOGRAM CLASSIFICATION WITH PROMPT-GUIDED MULTI-VIEW VISION LANGUAGE MODEL
Ahmad Alif Robit Al Hazmi, Moh. Edi Wibowo, S.Kom., M.Kom., Ph.D.;Professor Jia-Ching Wang
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

[38] Shunjie-Fabian Zheng, Hyeonjun Lee, Thijs Kooi, and Ali Diba. Mv-mlm: Bridging multi-view mammography and language for breast cancer diagnosis and risk prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1213–1222, 2025.

[39] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16816–16825, 2022.

[40] Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International Journal of Computer Vision*, 130(9):2337–2348, 2022.