Leveraging IHC Staining to Prompt HER2 Status Prediction from HE-Stained Histopathology Whole Slide Images

Yuping wang[1], Dongdong Sun[2], Jun Shi[2], Wei Wang[4], Zhiguo Jiang[3], Haibo Wu[4], and Yushan Zheng[1(✉)]

[1] School of Engineering Medicine, Beijing Advanced Innovation Center on Biomedical Engineering, Beihang University, Beijing, 100191, China.
yszheng@buaa.edu.cn
[2] School of Software, Hefei University of Technology, Hefei, 230601, China.
[3] Image Processing Center, School of Astronautics, Beihang University, Beijing, 102206, China.
[4] The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, 230036, China.

Abstract. The development of artificial intelligence has significantly impacted the predictive analysis of molecular biomarkers, which is crucial for targeted cancer therapy. Traditional assessment of HER2 in breast cancer utilizes both Hematoxylin and Eosin (H&E) and Immunohistochemistry (IHC) stained slides. Recent models have sought to predict HER2 status using H&E-stained slides to reduce reliance on the costly and time-consuming IHC staining. However, these models overlook the information from IHC staining. In this paper, we proposes a novel framework that integrates IHC-stained WSIs during the training phase to enhance the HER2 prediction capabilities based on the H&E-stained WSIs. This framework uses IHC-predicted HER2 status as a proxy task, embedding the learned relevant information as prompts into the encoder for H&E slides. Meanwhile, our model only requires H&E slides during inference, which maintains the data-efficiency of the HER2 prediction system. Experimental results show that our method achieves an AUC of 0.860 and a F1 score of 0.652 in the tasks of HER2 0/1+/2+/3+ status grading for breast cancer, which significantly outperforms state-of-the-art models.

Keywords: HER2 · Whole slide image · Multi-modal learning.

1 Introduction

With the rapid development of artificial intelligence, the prediction of molecular biomarkers from pathology whole slide images (WSIs) has gained significant attention due to its implications for targeted therapy. Human Epidermal Growth Factor Receptor 2 (HER2) overexpression serves as a critical biomarker for diagnosing breast cancer [12,10]. In routine breast cancer diagnostics, hematoxylin

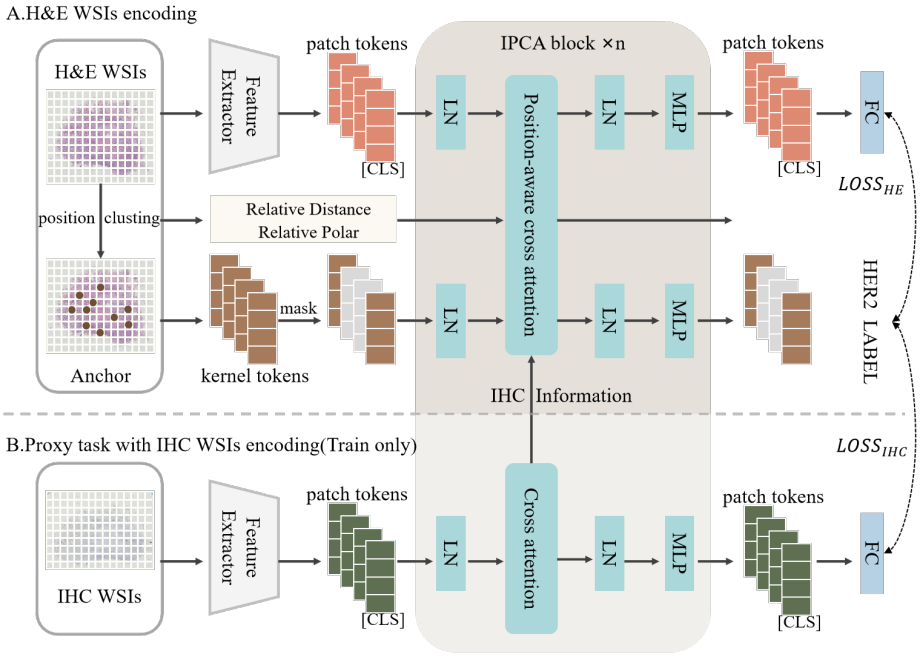

Fig. 1: The overview of the proposed method. (A) illustrates the H&E-stained WSIs encoding process, where the H&E patch tokens are updated by the Position-aware cross attention module with H&E information and through H&E-IHC interaction with IHC information. (B) illustrates the IHC-stained WSIs encoding process, where the IHC patch tokens are updated through cross attention. This process, as a proxy task, is used only during the training phase.

and eosin (H&E) stained slides are initially employed to identify cancerous tissue regions. Following this, Immunohistochemistry (IHC) staining and In Situ Hybridization (ISH) techniques are used to confirm the presence of HER2 overexpression [17]. Accurately predicting HER2 status using WSIs can significantly influence treatment strategies, thereby improve patient outcomes[6].

In early research, manually designed image features coupled with classifiers were employed to predict HER2 status from Immunohistochemistry (IHC) stained slides [3,9]. Recently, deep learning techniques [13,11,20,14] have shown substantial benefits in the analysis of IHC-stained slides, significantly enhancing the accuracy and efficiency of HER2 status prediction. Then, various deep learning frameworks [16,1] have been developed to predict HER2 status using H&E stained slides, aiming to reduce reliance on IHC staining which suffers from longer processing times, higher costs, and increased complexity. Notably, frameworks such as SlideGraph+ [8] and HEAT [2] utilize WSI-level graphs to effectively capture spatial and structural information. These advanced graph-based models significantly improve diagnostic accuracy with H&E-stained slides.

However, these models do not harness the extensive data available from IHC staining, which provides critical insights into HER2 expression. This omission of detailed IHC data during model training could limit the predictive capabilities and accuracy of systems relying solely on H&E-stained slides.

In this paper, we proposed a novel framework for HER2 status prediction based on cooperative training of H&E-stained and IHC-stained WSIs. As illustrated in Fig.1, our framework fully utilizes IHC-stained slides during the training phase to guide the learning of the H&E encoder. This allows more explicit information from the IHC to be embedded into the encoding process of H&E-stained slides. During the inference phase, the framework only requires H&E-stained slides as input and does not depend on IHC slides. Experimental results on a breast cancer dataset demonstrate that our method significantly outperforms baseline models trained solely on H&E-stained slides, with a 2.0% improvement in AUC and an 6.8% improvement in F1 score. The contribution of this paper can be summarized into two key aspects:

1) During the training phase of the H&E encoder, we utilize the prediction of HER2 status from IHC-stained slides as a proxy task. This proxy task allows the model to capitalize on the detailed information specific to HER2 from IHC slides, thereby enhancing its capability to identify similar patterns in H&E stains. In the inference phase, the model exclusively relies on H&E-stained slides, ensuring a high data efficiency.

2) We designed a novel module named IHC-prompted cross-attention (IPCA), as illustrated in Fig.2. The core idea of this module is to establish a set of learnable prompts that are shaped under the dual guidance of both H&E and IHC branch. This design allows the IPCA module to integrate IHC-specific patterns into the H&E encoder, thus improving the model's predictive accuracy and robustness.

2 Methods

2.1 Problem Formulation

The flowchart of the proposed work is illustrated in Fig.1. After WSIs segmentation and patch features extraction, we formulate the features extracted from H&E and IHC slides as $\mathbf{X}_{HE} \in \mathbb{R}^{n_p^h \times d_f}$ and $\mathbf{X}_{IHC} \in \mathbb{R}^{n_p^i \times d_f}$, where n_p^h and n_p^i are the numbers of patches segmented from H&E and IHC slides, and d_f is the feature dimension. Simultaneously, by clustering the position coordinates of the H&E patches, multiple position anchors are extracted and represented as $\mathbf{K}_{HE} \in \mathbb{R}^{n_k^h \times d_f}$, where n_k^h is the number of anchors.

2.2 H&E-stained WSI encoding

Fig.2 illustrates the process of H&E encoding. We utilized the Position-aware Cross-Attention (PACA) module in PAMA [18] to build the encoder for H&E-stained WSIs. PACA can extract the local and global features for WSIs by

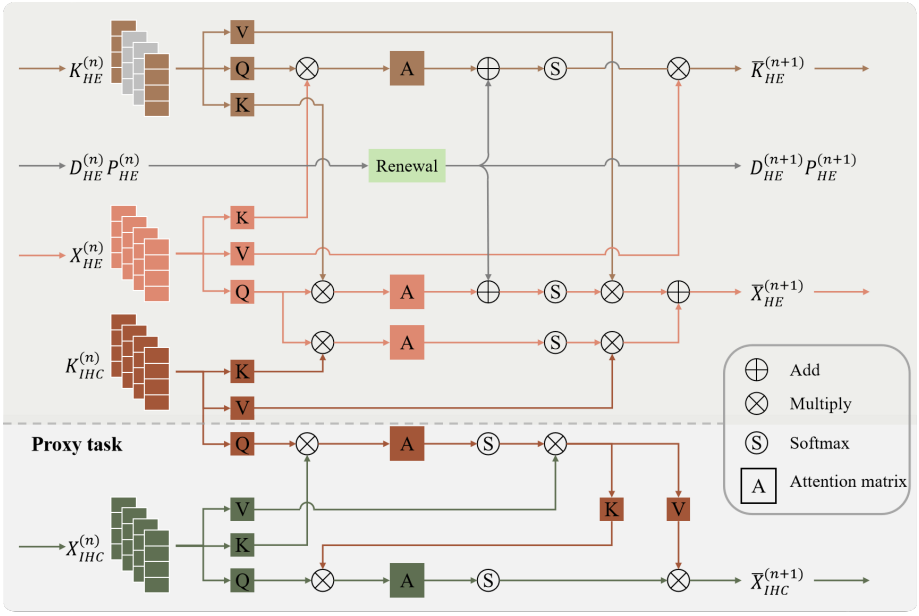

Fig. 2: Detailed View of the proposed IPCA module. The module comprises two main branches: the H&E branch (top) and the IHC branch (bottom) . The two branches interact through a cross-attention mechanism to enhance the predictive capability for HER2 status. Each branch includes transformations for queries, keys, and values, along with an attention matrix and Softmax operation, facilitating feature updating and information transmission.

performing cross-attention operations between anchor tokens and patch tokens. Specifically, a relative spatial distance matrix $\mathbf{D}_{HE} \in \mathbb{N}^{n_k^h \times n_p^h}$ and a relative polar angle matrix $\mathbf{P}_{HE} \in \mathbb{N}^{n_k^h \times n_p^h}$ are calculated between the anchors and patches, to describe the structural information of WSI.

In PACA, the anchors first collect local information from the patches by the formula:

$$\bar{\mathbf{K}}_{HE}^{(n+1)} = \sigma\left(\frac{\mathbf{X}_{HE}^{(n)}\mathbf{W}_q^{(n)} \cdot (\mathbf{X}_{HE}^{(n)}\mathbf{W}_k^{(n)})^\top + \mathbf{\Phi}^{(\mathbf{n})}}{\sqrt{d_e}}\right) \cdot (\mathbf{X}_{HE}^{(n)}\mathbf{W}_v^{(n)}), \tag{1}$$

where $\mathbf{W}_l \in \mathbb{R}^{d_f \times d_e}$ are the learnable parameters for the projection matrices, where $l = q, k, v$ stands for query, key, and value projections, $\bar{\mathbf{K}}_{HE}^{(n+1)}$ is passed to the next block after undergoing layer normalization and MLP operations to obtain $\mathbf{K}_{HE}^{(n+1)}$, and

$$\mathbf{\Phi}^{(n)} = \varphi_d(\mathbf{D}_{HE}^{(n)}) + \varphi_p(\mathbf{P}_{HE}^{(n)}), \tag{2}$$

where φ_d and φ_p are the embedding functions that map the distance and polar angle information into corresponding embedding values.

Symmetrically, each patch token captures the information of all anchors into their own local representations using the equation:

$$\bar{\mathbf{X}}_{HE\leftarrow HE}^{(n)} = \sigma\left(\frac{\mathbf{X}_{HE}^{(n)}\mathbf{W}_q^{(n)} \cdot (\mathbf{K}_{HE}^{(n)}\mathbf{W}_k^{(n)})^\top + \mathbf{\Phi}^{(\mathbf{n})\top}}{\sqrt{d_e}}\right) \cdot (\mathbf{K}_{HE}^{(n)}\mathbf{W}_v^{(n)}). \quad (3)$$

This allows anchor tokens to update patch tokens based on valuable information obtained from the kernels.

2.3 Proxy HER2 prediction task with IHC-stained WSIs

As illustrated in Fig.2, we constructed an additional branch alongside the H&E encoder to serve as a proxy task. This branch processes IHC-stained WSIs and symmetrically predicts the HER2 status. In each IPCA block, a set of prompts, $\mathbf{K}_{IHC} \in \mathbb{R}^{n_k \times d_f}$, are built to describe IHC-related patterns and will be optimized throughout the training process. The module first takes learnable kernel tokens $\mathbf{K}_{IHC}$ and patch feature $\mathbf{X}_{IHC}$ as input. Specifically, *Query* is derived from $\mathbf{K}_{IHC}$, and the *Key* and *Value* are derived from $\mathbf{X}_{IHC}$, all of which are achieved through linear transformations. Afterwards, we get the kernel tokens $\bar{\mathbf{K}}_{IHC}^{(n)}$ with IHC information:

$$\bar{\mathbf{K}}_{IHC}^{(n)} = \sigma\left(\frac{\mathbf{K}_{IHC}^{(n)}\mathbf{W}_q^{(n)} \cdot (\mathbf{X}_{IHC}^{(n)}\mathbf{W}_k^{(n)})^\top}{\sqrt{d_e}}\right) \cdot (\mathbf{X}_{IHC}^{(n)}\mathbf{W}_v^{(n)}). \quad (4)$$

Following this, we utilize $\bar{\mathbf{K}}_{IHC}^{(n)}$ with IHC information to perform another cross-attention with $\mathbf{X}_{IHC}$. In this process, The *Query* is derived from $\mathbf{X}_{IHC}$, and the *Key* and *Value* are derived from $\bar{\mathbf{K}}_{IHC}^{(n)}$. This process can be represented as follows:

$$\bar{\mathbf{X}}_{IHC}^{(n+1)} = \sigma\left(\frac{\mathbf{X}_{IHC}^{(n)}\mathbf{W}_q^{(n)} \cdot (\bar{\mathbf{K}}_{IHC}^{(n)}\mathbf{W}_k^{(n)})^\top}{\sqrt{d_e}}\right) \cdot (\bar{\mathbf{K}}_{IHC}^{(n)}\mathbf{W}_v^{(n)}). \quad (5)$$

As HER2 prediction with IHC-stained WSIs is a proxy task, Eq. 4 and Eq. 5 is only used in the training phase.

2.4 IHC staining information Embedding

Based on Eq. 4 and Eq. 5, $\mathbf{K}_{IHC}^{(n)}$ is expected to learn the prototypes that are crucial to recognize IHC features. Therefore, we can use $\mathbf{K}_{IHC}^{(n)}$ to assist the H&E encoder to extract IHC-related features. To achieve this, we utilize an cross attention, where the *Query* is derived from $\mathbf{X}_{HE}$, while the *Key* and *Value* are derived from $\mathbf{K}_{IHC}$ that contains the IHC information. The result of this cross attention operation is an updated $\bar{\mathbf{X}}_{HE\leftarrow IHC}$, which is then combined with

the $\bar{\mathbf{X}}_{HE\leftarrow HE}$ to produce the final updated $\mathbf{X}_{HE}$. The mathematical description of this process is as follows:

$$\bar{\mathbf{X}}^{(n)}_{HE\leftarrow IHC} = \sigma\left(\frac{\mathbf{X}^{(n)}_{HE}\mathbf{W}^{(n)}_{q}\cdot(\mathbf{K}^{(n)}_{IHC}\mathbf{W}^{(n)}_{k})^{\top}}{\sqrt{d_e}}\right)\cdot(\mathbf{K}^{(n)}_{IHC}\mathbf{W}^{(n)}_{v}), \tag{6}$$

$$\bar{\mathbf{X}}^{(n+1)}_{HE} = \bar{\mathbf{X}}^{(n)}_{HE\leftarrow HE} + \bar{\mathbf{X}}^{(n)}_{HE\leftarrow IHC}. \tag{7}$$

These designs enable the IPCA module to effectively utilize the IHC staining information in H&E WSI encoding even without the input of the IHC slides. It remains data-efficient during inference by relying solely on H&E slides.

2.5 Objective and optimization

Before the model inference, we add [CLS] tokens to the patch tokens of both H&E and IHC slides. After passing through n IPCA blocks, the [CLS] tokens are unfolded and passed through a fully connected layer for HER2 status prediction, categorizing into HER2 0, 1+, 2+, and 3+. Finally, we applied the cross-entropy loss function to both [CLS] tokens for training.

Additionally, to promote the integration of IHC information into the H&E encoder during training, we introduce a kernel masking (KM) strategy for a certain proportion of $\mathbf{K}_{HE}$. This masking strategy encourages the encoder to more effectively integrate IHC information into the H&E WSI encoding.

3 Experiments and Results

3.1 Implementation details

The dataset used in this study is a breast cancer dataset, collected from the First Affiliated Hospital of USTC (University of Science and Technology of China). This dataset comprises 358 cases, with each case including a H&E-stained slide and a paired IHC-stained slide. The HER2 status of each case has been affirmed by expert pathologists. The dataset is split by a ratio of 7:3. Five-fold cross-validation is performed on the training set to determine the model hyperparameters. Then, the determined model was used for testing.

The slides are segmented into 256×256 patches at a 20× magnification using sliding window strategy where the background region of each slide was removed by a threshold. Then, a pre-trained Vision Transformer (ViT-base) from PLIP [4] was applied to extracting features for the patches.

During the training phase, the Adam optimizer [5] is used with a learning rate of 1e-4, employing a cosine decay strategy. The batch size is set to 16. The evaluation metrics include Accuracy (ACC), Macro-AUC, Weighted-AUC, Macro-F1, and Weighted-F1. All experiments are conducted on two GPUs of NVIDIA GeForce RTX 4090.

Table 1: Hyper-parameter Experiments.

Settings	Accuracy	Macro-AUC	Weighted-AUC	Macro-F1	Weighted-F1
$r = 0.75$	0.617±0.003	0.837±0.024	0.847±0.024	0.586±0.034	0.617±0.028
$r = 0.5$	0.616±0.030	0.827±0.027	0.839±0.020	0.589±0.052	0.613±0.047
$r = 0.25$	0.616±0.010	0.833±0.026	0.842±0.021	0.572±0.030	0.603±0.018
$r = 0$	0.600±0.010	0.819±0.034	0.831±0.029	0.564±0.028	0.600±0.019
$n_k = 32$	0.617±0.003	0.837±0.024	0.847±0.024	0.586±0.034	0.610±0.028
$n_k = 64$	0.621±0.003	0.825±0.024	0.832±0.019	0.592±0.028	0.614±0.014
$n_k = 128$	0.623±0.002	0.841±0.022	0.851±0.015	0.568±0.016	0.606±0.014
$n_k = 256$	0.622±0.002	0.847±0.018	0.855±0.010	0.569±0.036	0.604±0.022

3.2 Hyper-parameter settings

We tuned the main hyper-parameters of our model through cross-validation. The results are detailed in Table. 1. Initially, we focused on the kernel masking strategy by tuning the mask-ratio in the range of [0, 0.25, 0.5 0.75]. The results showed that a kernel mask-ratio of 0.75 provided the best performance. Based on this experiment, we finally set the mask-ratio as 0.75. Subsequently, we tuned the number of prompts n_k of the shared $\mathbf{K}_{IHC}$ from 32 to 256. The results show that the performance improves as the number of shared prompts increase, and $n_k = 256$ delivered the best performance. Considering that a larger number of prompts would bring greater computation to the model, we finally set $n_k = 256$.

3.3 Comparison with SOTA methods

Then, we compared our model with six methods, inlcuding CLAM [7], Nystromformer [19], TransMIL [15], SlideGraph+ [8], HEAT [2], and PAMA [18]. These methods are used for training and inference with H&E inputs. The results are presented in Table. 2. Overall, our method demonstrates the best performance across all evaluation metrics. In comparison to the second-best methods, our method demonstrates a performance increase of 1.9% in Macro-AUC and 2.0% in Weighted-AUC. Additionally, it shows improvements of 6.5% in Macro-F1 and 6.8% in Weighted-F1, respectively.

SlideGraph+ and HEAT, both based on graph-based approaches, are considered state-of-the-art methods for predicting HER2 status using H&E inputs. PAMA can more effectively capture the context of WSIs based on its position-aware and anchor cross-attention mechanisms compared to SlideGraph+ and HEAT. It enables PAMA to achieve superior performance through enhanced spatial and semantic representation capabilities. Our proposed method significantly improves upon PAMA by leveraging an anchor structure combined with an IHC-prompted mechanism. We plotted the Macro-average ROC curve for PAMA and our method in Fig. 3. The AUC values for our method are consistently higher than those for PAMA, particularly in predicting HER2 1+ and HER2 2+ status, where we observed increases of 2.3% and 2.0%, respectively.

Table 2: Comparison with the state-of-the-art methods.

Method	Train	Inference	ACC	M-AUC	W-AUC	M-F1	W-F1
CLAM [7]	H&E	H&E	0.572	0.823	0.825	0.543	0.567
Nystromformer [19]	H&E	H&E	0.554	0.806	0.811	0.510	0.536
TransMIL [15]	H&E	H&E	0.576	0.789	0.798	0.545	0.571
SlideGraph+ [8]	H&E	H&E	0.589	0.822	0.825	0.546	0.577
HEAT [2]	H&E	H&E	0.548	0.800	0.803	0.519	0.541
PAMA [18]	H&E	H&E	0.583	0.839	0.840	0.567	0.584
Ours w/o proxy	H&E	H&E	0.601	0.845	0.847	0.578	0.606
Ours	H&E+IHC	H&E	**0.657**	**0.858**	**0.860**	**0.632**	**0.652**

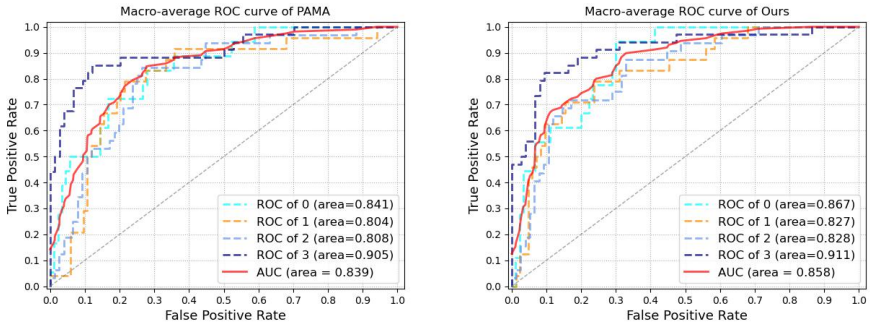

Fig. 3: Comparison of ROC curves between PAMA and our proposed method.

These improvements are crucial, as distinguishing these HER2 statuses using only H&E inputs is challenging due to the subtle morphological features in H&E slides. Our integration of IHC information enhances the discriminative power of the H&E encoder, enabling it to more effectively differentiate these subtle distinctions.

3.4 Ablation Study

We conducted ablation studies to evaluate the impact of embedding IHC information in Table. 2. In *Ours w/o proxy*, we removed the IHC input in both the training and inference stages but retained $\mathbf{K}_{IHC}$ along with the kernel tokens of H&E to update the patch tokens of H&E. Here, $\mathbf{K}_{IHC}$ serves as prompts trained to encapsulate information about H&E, which can be utilized during the inference phase. The results demonstrate that *Ours w/o proxy* outperforms PAMA across all evaluation metrics, indicating the prompt enable leverages the learned information. Furthermore, our method perform better than *Ours w/o proxy* in AUC of 1.3% and F1 of 5.4%, underscoring the importance of embedding IHC information for enhancing the H&E encoder's capabilities.

4 Conclusion

In this paper, we proposed a novel framework for HER2 prediction, which effectively incorporates the information from IHC into the H&E encoder. This model demonstrates data efficiency by utilizing both H&E and IHC inputs during training, while only requiring H&E input during inference. Ablation experiments validate the effective integration of IHC information into the H&E encoder. Our results exhibit superiority over existing HER2 prediction methods. Future work will focus on training the framework using datasets across multiple organs to enhance its generalization ability for HER2 status prediction.

Acknowledgements

This work was partly supported by Beijing Natural Science Foundation (Grant No. 7242270), partly supported by the National Natural Science Foundation of China (Grant No. 62171007, 61901018, and 61906058), partly supported by the Fundamental Research Funds for the Central Universities of China (grant No. YWF-23-Q-1075), and partly supported by Joint Fund for Medical Artificial Intelligence (Grant No. MAI2023C014).

References

1. Anand, D., Kurian, N.C., Dhage, S., Kumar, N., Rane, S., Gann, P.H., Sethi, A.: Deep learning to estimate human epidermal growth factor receptor 2 status from hematoxylin and eosin-stained breast tissue images. Journal of Pathology Informatics **11**(1), 19 (2020)
2. Chan, T.H., Cendra, F.J., Ma, L., Yin, G., Yu, L.: Histopathology whole slide image analysis with heterogeneous graph representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15661–15670 (June 2023)
3. Cordeiro, C.Q., Ioshii, S.O., Alves, J.H., Oliveira, L.F.: An automatic patch-based approach for HER-2 scoring in immunohistochemical breast cancer images using color features. CoRR **abs/1805.05392** (2018)
4. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., Zou, J.: A visual–language foundation model for pathology image analysis using medical twitter. Nature Medicine pp. 1–10 (2023)
5. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization (2017)
6. van der Laak, J., Litjens, G., Ciompi, F.: Deep learning in histopathology: the path to the clinic. Nature Medicine **27**, 775–784 (05 2021)
7. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering **5**(6), 555–570 (2021)
8. Lu, W., Toss, M., Dawood, M., Rakha, E., Rajpoot, N., Minhas, F.: Slidegraph+: Whole slide image level graphs to predict her2 status in breast cancer. Medical Image Analysis **80**, 102486 (2022)
9. Mukundan, R.: Analysis of image feature characteristics for automated scoring of her2 in histology slides. Journal of Imaging **5** (2019)

10. Piccart, M., Lohrisch, C., Di Leo, A., Larsimont, D.: The Predictive Value of HER2 in Breast Cancer. Oncology **61**(Suppl. 2), 73–82 (10 2001)
11. Qaiser, T., Rajpoot, N.M.: Learning where to see: A novel attention model for automated immunohistochemical scoring. CoRR **abs/1903.10762** (2019)
12. Ross, J.S., Slodkowska, E.A., Symmans, W.F., Pusztai, L., Ravdin, P.M., Hortobagyi, G.N.: The HER-2 Receptor and Breast Cancer: Ten Years of Targeted Anti–HER-2 Therapy and Personalized Medicine. The Oncologist **14**(4), 320–368 (04 2009)
13. Saha, M., Chakraborty, C.: Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. IEEE Transactions on Image Processing **27**(5), 2189–2200 (2018)
14. Selcuk, S.Y., Yang, X., Bai, B., Zhang, Y., Li, Y., Aydin, M., Unal, A.F., Gomatam, A., Guo, Z., Angus, D.M., Kolodney, G., Atlan, K., Haran, T.K., Pillar, N., Ozcan, A.: Automated her2 scoring in breast cancer images using deep learning and pyramid sampling (2024)
15. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in Neural Information Processing Systems **34**, 2136–2147 (2021)
16. Wang, J., Zhu, X., Chen, K., Hao, L., Liu, Y.: Hahnet: A convolutional neural network for her2 status classification of breast cancer (2023)
17. Wolff, A.C., Somerfield, M.R., Dowsett, M., Hammond, M.E.H., Hayes, D.F., McShane, L.M., Saphner, T.J., Spears, P.A., Allison, K.H.: Human epidermal growth factor receptor 2 testing in breast cancer: Asco–college of american pathologists guideline update. Journal of Clinical Oncology **41**(22), 3867–3872 (2023), pMID: 37284804
18. Wu, K., Zheng, Y., Shi, J., Xie, F., Jiang, Z.: Position-aware masked autoencoder for histopathology wsi representation learning. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 714–724. Springer Nature Switzerland, Cham (2023)
19. Xiong, Y., Zeng, Z., Chakraborty, R., Tan, M., Fung, G., Li, Y., Singh, V.: Nyströmformer: A nyström-based algorithm for approximating self-attention. Proceedings of the AAAI Conference on Artificial Intelligence **35**, 14138–14148 (05 2021)
20. Yao, Q., Hou, W., Wu, K., Bai, Y., Long, M., Diao, X., Jia, L., Niu, D., Li, X.: Using whole slide gray value map to predict her2 expression and fish status in breast cancer. Cancers **14**(24) (2022)