Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

Research Article

Performance of a HER2 testing algorithm tailored for urothelial bladder cancer: A Bi-centre study

Aoling Huang [a,1], Yizhi Zhao [b,1], Feng Guan [a], Hongfeng Zhang [c], Bin Luo [a], Ting Xie [a], Shuaijun Chen [c], Xinyue Chen [a], Shuying Ai [a], Xianli Ju [a], Honglin Yan [a], Lin Yang [b,*,2], Jingping Yuan [a,*,2,3]

[a] Department of Pathology, Renmin Hospital of Wuhan University, 238 Jiefang-Road, Wuchang District, Wuhan 430060, PR China
[b] School of Engineering, Westlake University, 600 Dunyu Road, Xihu District, Hangzhou 310030, PR China
[c] Department of Pathology, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430014, PR China

ABSTRACT

Aims: This study aimed to develop an AI algorithm for automated HER2 scoring in urothelial bladder cancer (UBCa) and assess the interobserver agreement using both manual and AI-assisted methods based on breast cancer criteria.
Methods and Results: We utilized 330 slides from two institutions for initial AI development and selected 200 slides for the ring study, involving six pathologists (3 senior, 3 junior). Our AI algorithm achieved high accuracy in two independent tests, with accuracies of 0.94 and 0.92. In the ring study, the AI-assisted method improved both accuracy (0.66 vs 0.94) and consistency (kappa=0.48; 95 % CI, 0.443–0.526 vs kappa=0.87; 95 % CI, 0.852–0.885) compared to manual scoring, especially in HER2-low cases (F1-scores: 0.63 vs 0.92). Additionally, in 62.3 % of heterogeneous HER2-positive cases, the interpretation accuracy significantly improved (0.49 vs 0.93). Pathologists, particularly junior ones, experienced enhanced accuracy and consistency with AI assistance.
Conclusions: This is the first study to provide a quantification algorithm for HER2 scoring in UBCa to assist pathologists in diagnosis. The ring study demonstrated that HER2 scoring based on breast cancer criteria can be effectively applied to UBCa. Furthermore, AI assistance significantly improves the accuracy and consistency of interpretations among pathologists with varying levels of experience, even in heterogeneous cases.

1. Introduction

Bladder cancer (BC) ranks 9th in incidence, with 613,791 new cases, and 13th in mortality [1]. The majority of bladder cancer cases are diagnosed as non-muscle-invasive bladder cancer (NMIBC), while approximately 25 % of patients are diagnosed with muscle-invasive disease, with the vast majority of histological subtypes being urothelial carcinomas of bladder (UBCa) [2–4]. The treatment options such as transurethral resection of bladder tumors, radiotherapy, radical cystectomy and chemotherapy are available [5]. Despite treatment, their efficacy varies, making it challenging to predict outcomes [6,7]. Notably, 31 % to 78 % of NMIBC patients experience tumor recurrence within 5 years, with 1 % to 45 % progressing to muscle-invasive bladder

cancer (MIBC) [2]. Furthermore, up to 50 % of MIBC patients experience distant recurrences, mainly within the first 3 years after treatment, which contributes to a stark 5-year overall survival rate of just 6 % for those with locally advanced metastatic urothelial carcinoma (la/mUC) [8,9]. Therefore, effective management strategies are crucial to mitigate the risk of recurrence, progression, and mortality associated with UBCa. In this regard, emerging therapeutic approaches, including targeted therapies and immunotherapy, offer promising avenues for enhancing treatment efficacy and prolonging patient survival [9,10].

The human epidermal growth factor receptor 2 (HER2) is a tyrosine kinase receptor, exhibiting no ligand but demonstrating a preference for dimerization with the other three family receptors [10]. Dimerization activates tyrosine kinase signaling pathways, promoting cell

---

proliferation, migration, invasion, and survival [11]. Since HER2 was first identified in the mid 1980′s, the comprehension of the biological role of HER2 in tumors, particularly in breast and gastric cancers, has progressively grown [12]. The interpretation of HER2 immunohisto-chemistry (IHC) primarily relies on the integrity of membrane staining and the intensity of that staining, which is categorized into scores of 0, 1 + , 2 + , and 3 + . Over the years, HER2 expression profile and its potential utility as a biomarker in clinical practice has also been explored in UBCa, with observations of its presence in approximately 12–70 % of cases [13]. Furthermore, variable rates of HER2 expression have been observed across different histological subtypes of UBCa. Several studies have found the highest HER2 overexpression in micropapillary carcinomas (56 %−68.6 %), followed by conventional UBCa, with lower expression in cases with squamous differentiation, sarcomatoid carcinomas, and glandular differentiation [14–16]. Anti-HER2 therapy, which includes monoclonal antibodies such as trastuzumab and pertuzumab, as well as tyrosine kinase inhibitors like lapatinib and tucatinib, has been successfully applied to patients with HER2-positive breast cancer [17,18]. However, such anti-HER2 therapy, whether used alone or in combination with conventional chemotherapy as second-line treatments for patients with la/mUC, yielded unsatisfactory results, characterized by low overall response rates [19]. Recently, encouraging results have been demonstrated in UBCa clinical trials by antibody-drug conjugates (ADCs) targeting HER2, which have emerged as a promising strategy in anti-HER2 therapy. Recently, fam-trastuzumab deruxtecan-nxki (T-DXd) was approved by the FDA for the treatment of unresectable or metastatic HER2-positive solid tumors. And the RC48-C005 trial (NCT03507166) demonstrated that Disitamab vedotin (DV, RC48), another ADC, significantly improved objective response rates (ORR) and survival in patients with locally advanced or muscle-invasive urothelial carcinoma, with ORR 51.2 %, median progression-free survival (mPFS) of 6.9 months, and median overall survival (mOS) of 13.9 months [20]. Based on the results, the U.S. Food and Drug Administration (FDA) granted RC48 "Breakthrough Therapy Designation" and RC48 was approved by the China Food and Drug Administration (CFDA) for use in UBCa treatment and was recommended by the Chinese Society of Clinical Oncology (CSCO) guidelines. Furthermore, a preliminary subgroup analysis of the DESTINY-Breast06 trial (NCT04494425) revealed that T-DXd significantly improved outcomes in patients with HER2-low breast cancer (HER2 0–1 +/HER2 1 +) [21]. Similarly, HER2-low status in urothelial carcinoma should also be considered, as results from the RC48-C011 trial (NCT04073602) demonstrate that RC48 provides therapeutic benefits even in cases with relatively low levels of HER2 expression (IHC 1 +) [22]. Therefore, it is crucial to accurately identify HER2-positive UBCa patients, as low levels of HER2 expression are equally important and should not be overlooked. However, unlike in breast cancer, there are no internationally standardized methodologies or interpretation criteria for UBCa, with current assessments largely based on those established for breast cancer [23]. Studies in breast cancer have demonstrated that gene amplification is the primary mechanism driving HER2 overexpression [24,25]. Consequently, the gold standard for HER2-targeted therapy has traditionally been HER2 3 + or HER2 2 + with positive fluorescence in situ hybridization (FISH) results in breast cancer. In contrast, most studies have failed to find a clear correlation between HER2 overexpression and gene amplification in UBCa [26–28]. Additionally, other mechanisms, such as polysomy 17, point mutations, translocations, or transcriptional upregulation [29–31], may also contribute to HER2 protein overexpression. As a result, in some clinical trials [22,32] and in the approved indication of RC48 by CFDA, anti-HER2 therapy for UBCa relies solely on HER2 IHC results without FISH testing. This makes the accuracy of IHC interpretations critical. However, studies have shown that HER2 IHC interpretation is prone to variability, particularly in cases of low HER2 expression, where inconsistencies are even more pronounced [33]. Furthermore, HER2 IHC heterogeneity in UBCa may also contribute to the low consistency in interpretation. Intratumoral HER2 heterogeneity

can manifest as variations in staining intensity or uneven distribution of HER2 expression across tumor cells, with UBCa showing greater heterogeneity [28] than breast cancer [34] and levels similar to those seen in gastric cancer [35]. Thus, there is a pressing need for more objective methods to assist pathologists in interpreting HER2 IHC in UBCa, which would improve the stratification and selection of patients for HER2-targeted therapies.

With the rapid advancement of artificial intelligence (AI) in recent years, digital pathology has seen remarkable progress [36–38]. When it comes to the Automated IHC scoring system, IHC in breast cancer, especially HER2-low breast cancer has been extensively studied [39,40]. Several AI algorithms show promise in overcoming overcome subjectivity and enhance consistency in pathologists' assessments [41]. However, to our knowledge, few studies have focused on automated HER2 algorithms for UBCa, and limited research has been conducted on the agreement of HER2 scoring for UBCa through manual reading based on the 2018 ASCO/CAP breast cancer guidelines [23]. Previously proposed AI algorithms require extensive membrane data for model training, a task widely recognized as highly challenging and costly [42]. Transfer learning leverages pre-trained model parameters from large datasets for new tasks, starting with a current domain model and assuming identical initial parameters. It employs annotated data from another domain to train the model, utilizing features learned from the pre-trained model to accelerate training and enhance task performance [43].

In this study, we initially employed transfer learning to fine-tune the weights of the original HER2 scoring model for breast cancer, facilitating the establishment and validation of an automated scoring model for analyzing HER2 in UBCa. Subsequently, we conducted a two-round reader study to assess inter-pathologist consistency in HER2 scoring for UBCa, following the 2018 ASCO/CAP guidelines [23]. Additionally, we explored the potential of AI in assisting with HER2 interpretation in UBCa, with a particular focus on its utility in evaluating cases with low or heterogeneous HER2 expression.

2. Materials and methods

2.1. Clinical data

We retrospectively collected 438 consecutive cases of invasive UBCa from two institutions, respectively: (a) Renmin Hospital of Wuhan University (RHWU) and (b) The Central Hospital of Wuhan (TCHW). The RHWU cohort included 365 whole slide images (WSIs) of 342 UBCa patients. The TCHW cohort included 104 WSIs of 96 UBCa patients, with multiple WSIs potentially obtained from a single patient. Two pathologists reviewed all slides and excluded cases with factors such as inconspicuous invasive foci, slides with tissue folds, or excessive nuclear or cytoplasmic staining. Ultimately, 330 WSIs were included in the initial model construction study (Fig. 1). For the subsequent ring study, we selected 200 HER2 IHC staining slides, consisting of 95 HER2-negative (HER2 0/1 +) and 105 HER2-positive cases (HER2 2 +/3 +) based on IHC results.

The scoring criteria for HER2 protein expression were interpreted according to the 2018 ASCO/CAP guidelines [23] (0, no staining or membrane staining that is incomplete and faint/barely perceptible and in ≤ 10 % of tumor cells; 1 +, incomplete membrane staining that is faint/barely perceptible and in > 10 % of tumor cells; 2 +, weak to moderate complete membrane staining observed in > 10 % of tumor cells or circumferential membrane staining that is complete, intense, and in < 10 % of tumor cells; and 3 +, circumferential membrane staining that is complete, intense, and in >10 % of tumor cells). HER2 positivity is defined as HER2 2 + and HER2 3 +. Two pathologists independently reviewed the slides and assigned HER2 scores. In cases of inconsistency, a third senior pathologist, with over 15 years of subspecialty experience in urologic pathology, re-evaluated the slides, and an agreement was reached. This consensus served as the gold standard for
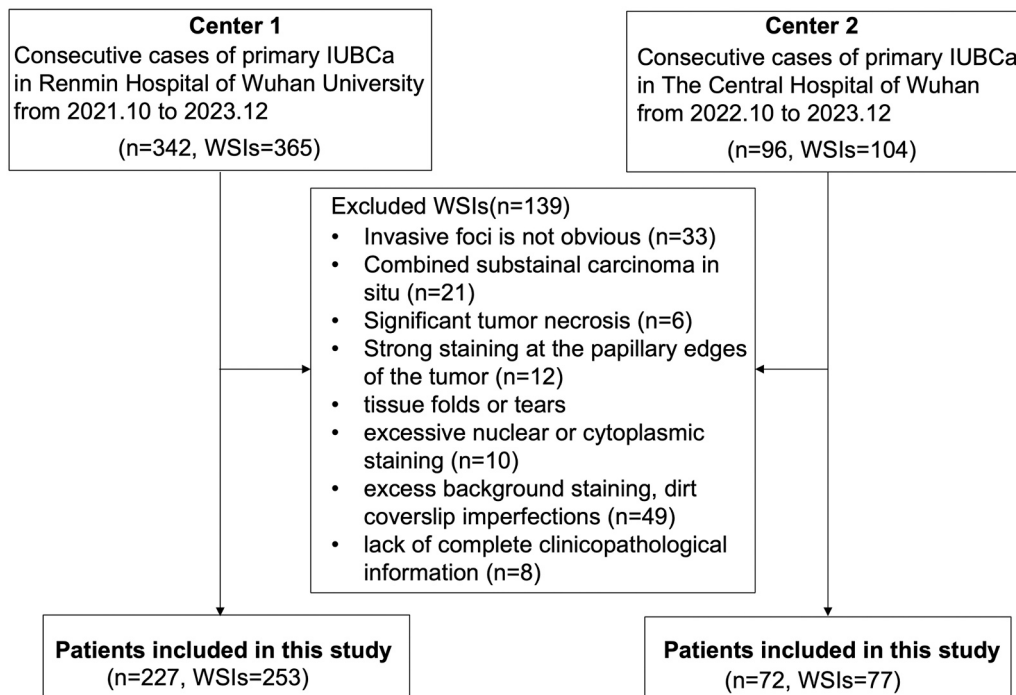
Fig. 1. Flow chart depicting the criteria for patient inclusion and exclusion. IUBCa, invasive urothelial carcinoma of bladder.

the study. HER2 protein expression heterogeneity in this investigation was characterized by membranous staining 2 + and 3 + in 5 %–50 % of tumor cells.

In all samples that were formalin-fixed and paraffin-embedded, IHC for HER2 was conducted using the PATHWAY HER-2/neu (4B5) rabbit monoclonal antibody from Ventana Medical Systems, Inc.(Tucson, AZ). The IHC was conducted with the BenchMark XT automated stainer, also from Ventana Medical Systems. HER2-IHC slides from both cohorts were digitally scanned into WSI format using a KF-PRO-020 digital scanner at × 40 magnification (0.5 μm per pixel).

2.2. Model development

In this study, we employed transfer learning to fine-tune a HER2 model that was previously trained for breast cancer. Despite differences in cancer types, the staining patterns in breast and UBCa are quite similar, with HER2 expression localized to the cell membrane. Given the significantly smaller UBCa dataset, we locked the encoder weights of the pre-trained breast cancer model and trained only the remaining parts. This approach helps mitigate overfitting, thereby enhancing the model's robustness and stability.

During the transfer learning process, pathologists selected 584 patches from 22 representative WSIs from the first RHWU cohort for the training set. Each patch measured 1024×1024 pixels and was meticulously annotated at the cell level, resulting in approximately 100,000 cell-level annotations. These annotations were performed by two pathologists based on consensus, categorizing the cells into six types according to the integrity of the cell membrane and staining intensity: nontumorous cells, negative tumor cells (no staining), weak incomplete membrane staining tumor cells or moderate incomplete membrane staining (1 + cells), weak to moderate complete membrane staining tumor cells (2 + cells), and strong complete membrane staining tumor cells (3 + cells). The remaining 231 WSIs served as the internal validation set to determine performance metrics. Subsequently, the second TCHW cohort of 77 WSIs served as an additional testing cohort to assess the algorithm's robustness. Fig. 2 illustrates the data utilization for training and testing.

The image patches are 1024×1024 pixel RGB images that were

cropped from a 40x magnified WSI. The rationale behind this approach is that processing an entire high-resolution WSI at once would be computationally intractable, given the limitations of existing hardware and software resources. By breaking down the WSI into manageable patches, we can leverage the parallel processing capabilities of modern deep learning frameworks and distribute the computational load across multiple GPU devices or nodes.

Since the HER2 grading is directly based on the statistical counts of different cell types, the scoring strategy adopted in this study was therefore constructed upon cell recognition and classification. In this approach, the cells were directly predicted for their location and type using a point-to-point network (P2PNet) [44] as an approach. This method overcomes the limitations of previous methods that required the probability density maps or pseudo-boxes as learning targets. Instead, it directly received a set of annotated cell points for training and predicted the locations of the cell points during inference. The algorithm used in this study is based on the P2PNet combined with multi-layer feature fusion and a multi-task learning strategy [45]. As a result, this method directly outputs the location and classification of each individual cell point. More specifically, the P2P network provides a set of reference points as a prior. The model then goes through a series of Pyramidal feature aggregation and extraction operations, during which a deform-layer learns the offset of the location of each reference point. Subsequently, the output features from different layers are then fused together and fed into separate multi-layer perceptions for location regression and classification. The model ultimately has two outputs the coordinates of the cell centroids and the cell classification results. In comparison to traditional approaches, this pure point-based multi-task localization method can better locate and parse the individual positions even in dense cell regions. The multi-task learning also provides the model with more cues that can help facilitate cross-task knowledge transfer. Finally, the entire framework of the algorithm is depicted on Fig2.

2.3. Study design

Among the 308 testing slides comprising 231 from the internal test cohort and 77 from the external test cohort, we selected a total of 200
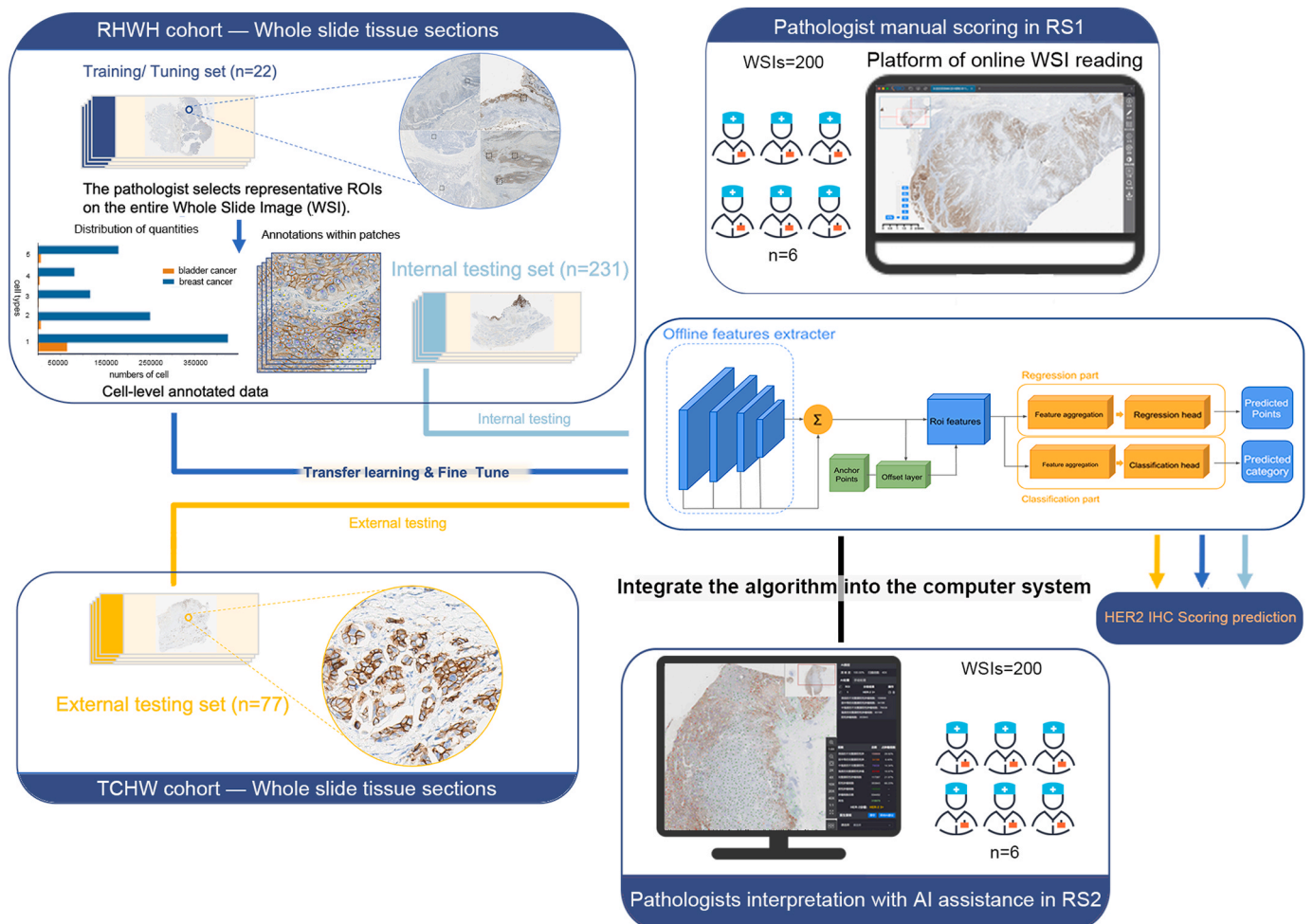
Fig. 2. Study Design Flowchart. The study begins with model construction. Initially, annotation data from regions of interest (ROI) were collected. Using transfer learning with pre-trained weights from a breast cancer model, the model was subsequently trained. The distribution of cell-level annotations for both the breast cancer model and the bladder cancer model was quantitatively depicted, and the network architecture is illustrated. Next, the model's performance was evaluated using two test sets. Additionally, a subset of 200 WSIs was selected for further investigation. In two rounds, RS1 (mannual scoring) and RS2 (AI-assisted scoring), six pathologists independently interpreted the entire set of WSIs.

HER2 IHC slides, consisting of 95 HER2-positive and 105 HER2-negative cases, for our subsequent ring study cohort. For this study, six pathologists from RHWU and TCHW, grouped by their experience level, interpreted these slides. Specifically, junior pathologists with 1–5 years of practice and senior pathologists with 5–10 years of practice, all of whom had experienced in interpreting HER2 IHC sections in routine clinical settings.

First, the pathologists reviewed the 2018 ASCO/CAP guideline [23] for HER2 IHC testing in breast cancer and were trained in use of the AI-assisted device. The research comprised two rounds of ring studies (RS). In the initial round (RS1), the pathologists evaluated the 200 HER2 IHC slides by examining the WSIs, with the ability to zoom in and out at specified magnifications, similar to using a microscope. After a 1-week interval, the second round (RS2) was conducted during which the exact slides were reinterpreted with AI assistance. The AI algorithm was integrated into a computer system that, once WSIs were uploaded, automatically performed patch splitting, cell detection, cell classification, and HER2 IHC scoring prediction. Subsequently, the inference results were then stored in a database and displayed on the front-end screen, assisting pathologists by pre-analyzing HER2 IHC slides and offering computed results as secondary opinions. Finally, during the second ring study, the pathologists finalized the score by incorporating the AI-generated results.

2.4. Statistics

Statistical analyses were conducted with SPSS version 26 and R version 4.2.2. To evaluate the AI algorithm, manual interpretation, and AI-assisted interpretation, various performance metrics were employed, including accuracy, precision, recall, F1-score, and Cohen's kappa. To compare the manual interpretation with AI-assisted interpretation, both the paired t-test and the Wilcoxon rank-sum test were conducted to assess statistical significance. A two-tailed p-value of less than 0.05 was considered statistically significant. Additionally, interobserver agreement was analyzed using Fleiss' kappa, with agreement levels interpreted as follows: 0.01–0.20 indicates slight agreement, 0.21–0.40 reflects fair agreement, 0.41–0.60 signifies moderate agreement, 0.61–0.80 represents substantial agreement, and 0.81–1.00 denotes almost perfect agreement.

3. Results

3.1. Patient cohort and clinicopathological characteristics

The study cohort comprised 330 whole slide images (WSIs) from 299 consecutive cases of primary IUBCa diagnosed at RHWU and TCHW between February 2022 and December 2023. Table 1 presents the clinicopathological details of the cohort. The average age of the patients

Table 1
Clinicopathological features of cases (patients n = 299, WSIs n = 330).

Clinicopathologic features	
Sex (n, %)	
male	240(80.3)
female	59(19.7)
Age, years: mean (range)	69.0(46 −91)
Histological subtypes (n, %)	
Conventional urothelial carcinoma	266 (89.0)
UC with squamous differentiation	14 (4.7)
UC with glandular differentiation	9(3.0)
Micropapillary UC	7(2.3)
Nested UC	2(0.7)
Sarcomatoid UC	1(0.3)
Histological grade (n, %)	
high grade	207(69.2)
low grade	92(30.8)
HER2 IHC score* (n, %)	
0	47(14.2)
1 +	75(22.7)
2 +	101(30.6)
3 +	107(32.4)

* A total of 330 WSIs were included, each with an assigned HER2 score.
UC urothelial carcinoma, IHC immunohistochemistry.

was 69.0 years, ranging from 46 to 91 years. The most prevalent tumor type was conventional urothelial carcinoma (89.0 %), followed by urothelial carcinoma with squamous and glandular differentiation, and others. Additionally, tumors were classified as high grade in 207 cases (69.2 %). Regarding HER2 protein expression, the distribution was as

follows: 0 in 47 slides (14.2 %), 1 + in 75 slides (22.7 %), 2 + in 101 slides (30.6 %), and 3 + in 107 slides (32.4 %). Importantly, scores of 2 + and 3 + were considered HER2 IHC positive, which was observed in 208 slides (63.0 %). Among 299 UBCa patients, the highest HER2 positivity rate was observed in micropapillary carcinoma (6 out of 7, 85.7 %), followed by conventional UBCa (163 out of 266, 61.2 %) and glandular differentiation (5 out of 9, 55.6 %). The positivity rate in cases with squamous differentiation was 6 out of 14 (42.9 %), while all three cases of nested urothelial carcinoma and sarcomatoid urothelial carcinoma were HER2 negative.

3.2. Evaluation of AI quantification algorithm

We began by training an automated algorithm to predict HER2 scores in UBCa. To validate its performance, we tested the model on two independent datasets. As shown in Fig. 3 and Table 2, detailed

Table 2
Classification performance of our AI model in two independent test datasets.

dataset		0	1 +	2 +	3 +
internal test	Accuracy	1.00	0.96	0.91	0.95
	Precision	1.00	0.90	0.90	0.92
	Recall	0.88	0.90	0.90	1.00
	F1-score	0.94	0.90	0.90	0.96
external test	Accuracy	0.88	0.89	0.89	1.00
	Precision	1.00	0.86	0.92	0.98
	Recall	0.93	0.96	0.91	0.95
	F1-score	0.96	0.91	0.91	0.96
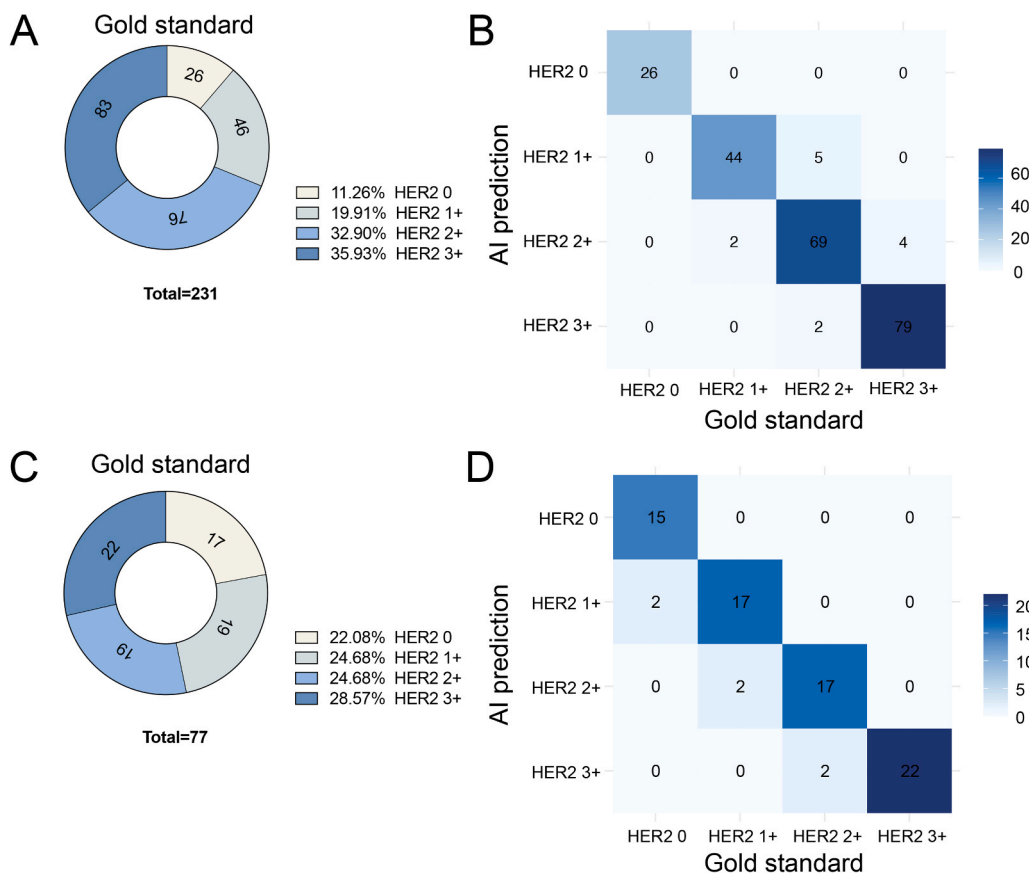
Fig. 3. Ground-truth and validation outcomes at the slide level for the dataset. (A) RHWU dataset Ground-truth: Distribution of HER2 scores, showing 11.26 % HER2 0, 19.91 % HER2 1 + , 32.90 % HER2 2 + , and 35.93 % HER2 3 + (total = 231). (B) Confusion matrix of RHWU dataset: The AI model shows high accuracy, especially for HER2 2 + and 3 + , with some misclassification between HER2 1 + and 2 + . (C) TCHW dataset Ground-truth: Distribution of HER2 scores, with 22.08 % HER2 0, 24.68 % HER2 1 + , 24.68 % HER2 2 + , and 28.57 % HER2 3 + (total = 77). (D) Confusion matrix of TCHW dataset: Similar to the RHWU results, the AI model accurately predicts HER2 2 + and 3 + , with some misclassification between HER2 1 + and 2 + .

validation results are presented. First, 231 WSIs from RHWU were evaluated. On the internal test set from the RHWU cohort, the model achieved an overall accuracy of 0.94, indicating robust predictive capability. Moreover, its accuracy, precision, and F1-score all exceeded 0.9, approaching 1, underscoring its reliability. Then, when tested on the external dataset from TCHW (n = 77), the overall accuracy of the algorithm slightly declined to 0.92. Upon further analysis, we identified that most errors occurred in cases where the cell proportion was close to the 10 % threshold. Additionally, misclassifications of 2 + as 3 + were primarily due to overestimations of strong staining in certain slides.

3.3. Overall ring study results

The interpretation results from all pathologists in RS1 (manual scoring) and RS2 (AI-assisted scoring) are illustrated in Fig. 4. In RS2, referred to as AI-assisted pathologist scoring, pathologists re-evaluated and adjusted the scores of the entire WSI based on AI-generated results. This AI-assisted evaluation in RS2 visually enhanced the accuracy and consistency of the manual evaluation in every HER2 score when measured against the gold standard values.

3.4. Accuracy evaluation in each RS

The accuracy of pathologists' HER2 scoring (RS1), AI-assisted outcomes (RS2) and AI results were compared using confusion matrix (Fig. 5A). AI prediction (RS-AI) represents the HER2 scores generated by evaluating the entire WSI with the aforementioned AI algorithm. As shown in Fig. 5B, the accuracy of RS2 using the AI-assisted method (0.94) was significantly higher than that of RS1 (0.67). The AI-assisted approach reduced the accuracy gap between the gold standard and the pathologists' results by 0.27. Additionally, the accuracy of RS-AI was consistent with the AI-assisted pathologist review in RS2 (0.94).

Next, the accuracy of pathologists' interpretations of HER2 0, HER2 1 + , HER2 2 + , and HER2 3 + tumors were evaluated separately (Fig. 5B). Overall, compared to RS1, the accuracy, recall, and F1 scores for HER2 ratings in RS2 showed varying degrees of improvement. Notably, for HER2 2 + , the F1 scores increased from 0.54 to 0.91, recall improved from 0.53 to 0.89, and precision rose from 0.55 to 0.93. We further investigated the reasons for inconsistent results in RS1. As shown in Fig. 5C, in HER2 0 cases, most errors occurred due to misclassifying HER2 0 as HER2 1 + tumors, while in HER2 2 + cases, most errors stemmed from misclassifying HER2 1 + as HER2 2 + . Importantly, AI-assisted interpretation significantly reduced these errors.

We also examined intratumoral heterogeneity in HER2-positive expressions and evaluated the performance of each HER2 interpretation method with or without heterogeneity. Among HER2-positive cases, 80.0 % (40/50) of HER2 2 + cases exhibited heterogeneity, while 20.0 % (10/50) showed homogeneity. Additionally, in HER2 3 + cases, 47.3 % (26/55) exhibited heterogeneity, while 52.7 % (29/55) showed homogeneity (Fig. 5D). In cases with homogeneous staining, the accuracy of pathologists' review results was 0.85, but it significantly decreased to 0.49 in the presence of heterogeneity (Fig. 5E). AI improved the accuracy under heterogeneous conditions to 0.93, which is comparable to the accuracy under homogeneous conditions (0.96). Thus, the use of AI significantly enhanced the accuracy of HER2-positive expressions with heterogeneity. Fig. 6 showed a case with heterogeneous HER-2 IHC.

3.5. Consistency evaluation of all pathologists in each RS

Heatmaps were used to visualize the changes in consistency of pathologist interpretations between RS1 and RS2 (Fig. 7A-B). In RS1, when pathologists performed manual readings, overall consistency was observed (kappa=0.48; 95 % CI, 0.443–0.526). However, compared to RS1, the AI-assisted evaluation significantly improved the consistency in RS2 (kappa=0.87; 95 % CI, 0.852–0.885). To gain deeper insight, we further analyzed the consistency of the two methods, defining consensus as agreement among at least 5 out of 6 pathologists on the case interpretation (Fig. 7C). The results indicated that in RS1, a minimum of one pathologist interpreted 125 cases as HER2 1 + , with consensus reached in 35 cases (28.0 %). Similarly, 111 cases were labeled as HER2 2 + by a minimum of one pathologist, achieving consensus in 16 cases (14.4 %). In contrast, in RS2, 48 cases were identified as HER2 1 + by a minimum of one pathologist, with 69 cases (69.6 %) reaching consensus. Furthermore, in RS2, 73 cases were designated as HER2 IHC 2 + by a minimum of one pathologist, with 43 cases (58.9 %) reaching consensus. This analysis shows that traditional manual readings showed poorer consistency for HER2 IHC 2 + compared to HER2 IHC 1 + , but both were significantly improved with AI assistance.

Next, the six pathologists were categorized into two groups according to their years of practice. Fig. 7D presents a comparison of accuracy for the two groups in RS1 and RS2. Upon comparing the accuracy between RS1 and RS2, it was evident that both groups of pathologists experienced significant improvements with the AI-assisted method (P < 0.05). The use of AI assistance reduced the accuracy gap between the gold standard and all pathologists, irrespective of their experience levels. Particularly, the greatest improvement in accuracy was observed among junior pathologists, increasing from 0.52 (95 % CI, 0.46–0.58) in
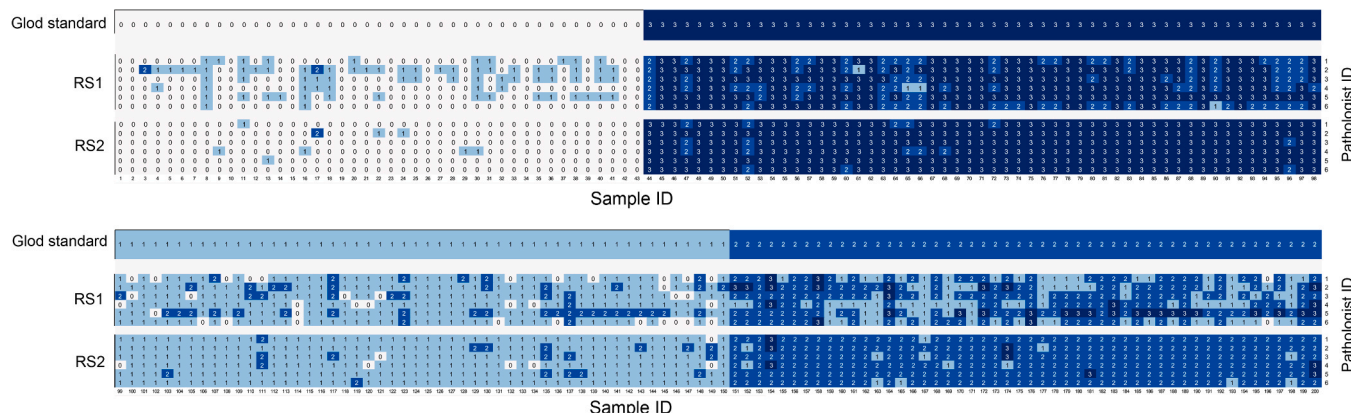

Fig. 4. HER2 interpretation results of 200 cases by six pathologists in two rounds of a ring study. The heatmaps show the HER2 scores assigned by the six pathologists in two rounds (RS1 and RS2), with the gold standard results provided for comparison. Each column represents a sample, and each row corresponds to a pathologist's interpretation. Darker shades indicate higher HER2 scores, ranging from 0 to 3 + . The top panel shows the results from the first round (RS1), while the bottom panel presents results from the second round (RS2). Across both rounds, there is variability in HER2 scoring between pathologists, with the second round demonstrating improved agreement with the gold standard, particularly for higher HER2 scores (2 + and 3 +).
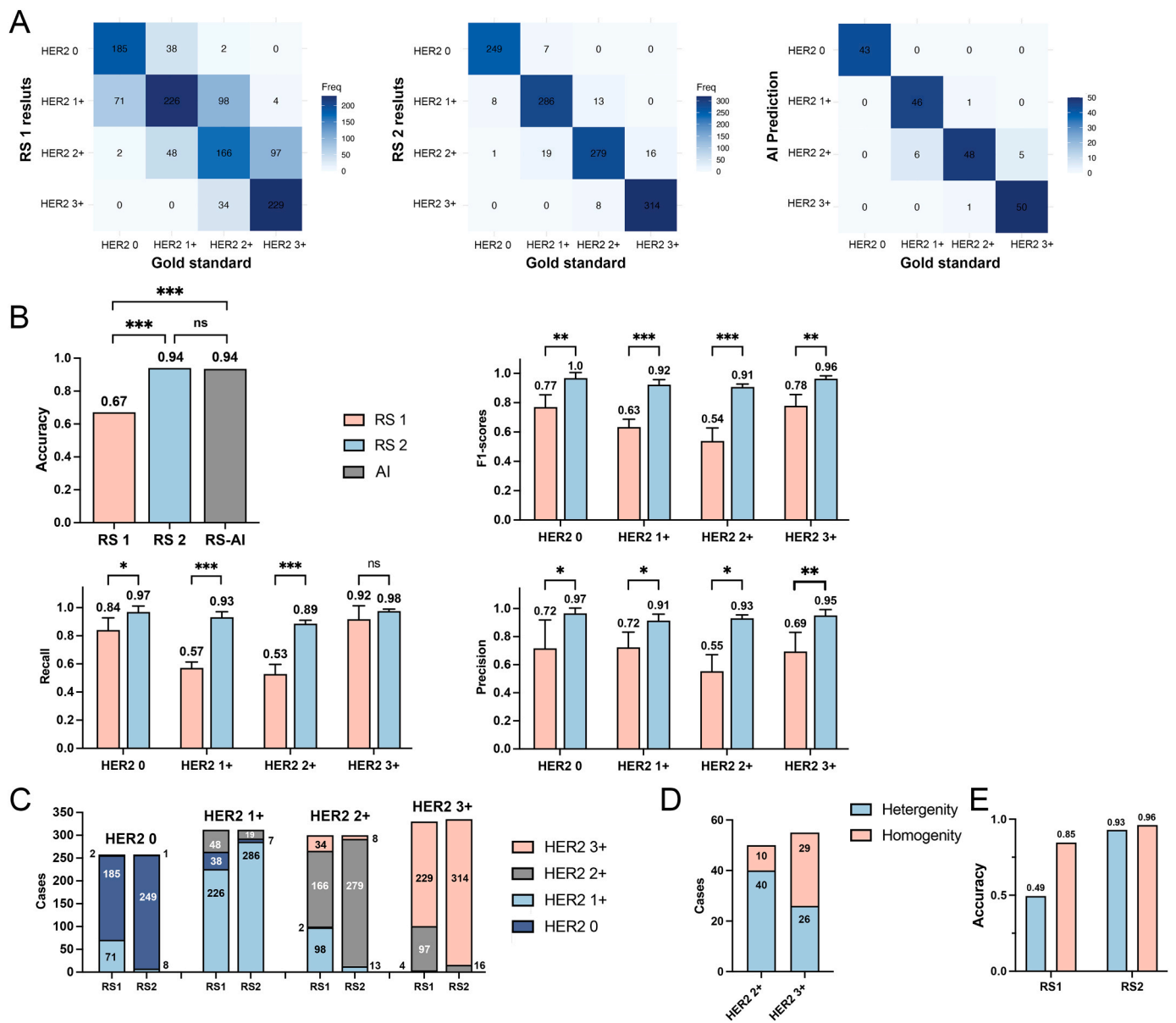
Fig. 5. The accuracy of 2 rounds of ring studies (RSs) for pathologists. (A) Confusion matrix for the two RSs and AI prediction outcomes: The confusion matrices show the accuracy of HER2 scoring by pathologists in RS1 and RS2, as well as the AI model's predictions compared to the gold standard. Overall, RS2 and AI predictions demonstrate improved alignment with the gold standard. (B) Assessment of overall accuracy and statistical comparisons: The bar charts illustrate the overall accuracy, F1-scores, recall, and precision for HER2 0, 1 + , 2 + , and 3 + cases across RS1, RS2, and AI predictions. Statistical tests indicate significant improvements in accuracy and recall between RS1 and RS2 for HER2 0, 1 + , and 2 + cases, while the AI results show comparable performance. Paired t-test and Wilcoxon rank-sum test were conducted to determine significance between groups: ns, P > 0.05; * , P < 0.05; * * , P < 0.01; * ** , P < 0.001. (C) Number of error cases in each grade compared to the gold standard: This chart shows the number of error cases for each HER2 grade. (D) Presence of heterogeneity in HER2 + and HER2 3 + cases: This plot shows the frequency of heterogeneity in HER2 + and HER2 3 + cases. (E) Accuracy in interpreting homogeneous and heterogeneous cases in the two RSs: Accuracy in interpreting homogeneous cases is higher compared to heterogeneous cases in RS1, while RS2 demonstrates comparable accuracy to homogeneous cases.

RS1 to 0.93 (95 % CI, 0.89–0.97) in RS2 (Fig. 7D). The consistency differences between the groups underscore that pathologists generally demonstrated higher consistency in RS2 compared to RS1. Moreover, senior pathologists exhibited greater consistency (Kappa=0.55) in RS1, whereas junior pathologists achieved the highest consistency in RS2 (Kappa=0.87). This significant improvement from 0.45 to 0.87 among junior pathologists represents the most substantial enhancement observed (Fig. 7E). Overall, the acceptance across all pathologists was 0.78, indicating the pathologists' high level of acceptance of the AI results. (Fig. 7F).

4. Discussion

Several clinical trials have shown encouraging results for ADCs in the treatment of HER2-positive UBCa patients [46,47]. In these studies, HER2-positive status is characterized by an IHC score of 2 + or 3 + , regardless of FISH results. For instance, Sheng et al. [46] found that RC48 significantly improved outcomes in HER2-positive la/mUC, with an ORR of 50.5 %. Additionally, Xu et al. [22] reported DV's efficacy in HER2-negative mUC (IHC 0 or IHC 1 +), achieving a mPFS of 5.5 months and an ORR of 26.5 %. Moreover, an ongoing study combining RC48-ADC with toripalimab in HER2-negative la/mUC showed promising ORRs: 83.3 % for IHC 3 + /2 + , 64.3 % for IHC 1 + , and 33.3 %
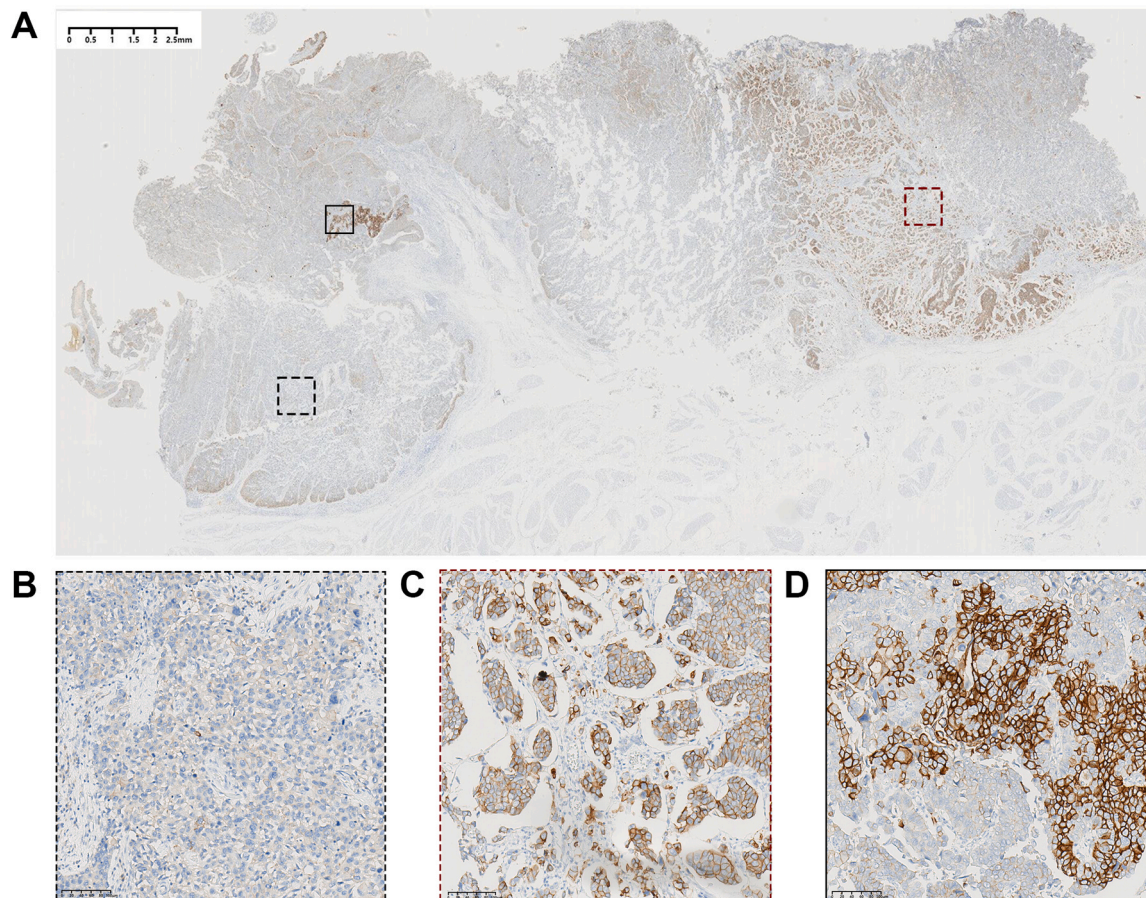
Fig. 6. Urothelial carcinoma of bladder exhibiting heterogeneous HER-2 immunohistochemistry (IHC). (A) Low-resolution view of a tissue section stained for HER-2 IHC. The overall HER-2 status for this case has been classified as HER2 2 + , displaying significant staining heterogeneity. Dotted black line, dotted red line, and solid black line rectangles correspond to the areas shown in (B), (C), and (D), respectively. A scale bar is provided in the figure. (B) The majority of cancer cells display faint HER-2 expression (0–1 +), viewed at 20x magnification. (C) Clusters of tumor cells showing moderately intense, intact HER-2 expression (2 +), viewed at 20x magnification. (D) The majority of tumor cells exhibit strongly positive HER-2 expression (3 +), viewed under 20x magnification.

for IHC 0 [32]. These investigations indicate that ADCs could be an effective treatment option for patients with HER2-positive mUC. Notably, even patients with HER2-low tumors can achieve efficacy rates comparable to those of chemotherapy.

Currently, the determination of HER2 status in UBCa primarily relies on the HER2 IHC method, based on breast cancer criteria, to select candidates for new HER2-targeted therapies. However, a major challenge in employing IHC as the main method for HER2 testing is the intrinsic subjectivity, which leads to considerable intraobserver and interobserver variability. This variability stems not only from the subjective judgments clinical pathologists must make—such as assessing the completeness and intensity of membrane staining and the percentage of positive cells—but also from the inherent heterogeneity of UBCa. Moreover, substantial intratumoral heterogeneity in HER2 protein expression was reported in about 55.5 % of HER2-positive tumors by IHC [28]. Thus, precise pathological evaluation of HER2 status is crucial for identifying patients who are eligible for these advanced treatments.

Our study initially employed transfer learning to train an algorithmic model for automated evaluation of HER2 expression in UBCa. Subsequently, we validated its performance on two independent test sets. In further investigations, we conducted two rounds of studies involving 200 cases to explore the reproductivity of this set of HER2 scoring criteria and the application value of artificial intelligence algorithms in assisting HER2 assessment in UBCa. Our experimental results demonstrate that the HER2 IHC algorithm achieves good accuracy in four-level classifications (0.94; 0.92) among the two independent tests. Moreover, applying this algorithm in subsequent pathologist-assisted scoring

revealed a significant improvement in accuracy and consistency among AI-assisted pathologists in RS2.

To our knowledge, no other study has yet developed a deep learning quantification algorithm for HER2 scoring in UBCa to aid in pathologists' diagnoses. While several AI models have been proposed for breast cancer [48,49], the algorithm we introduced in this study utilizes a point annotation method and transfer learning. We leveraged pre-trained model parameters from extensive datasets and combined them with our annotated data to adapt to our specific task. This strategy was employed to address the challenge of obtaining adequate membrane data necessary for developing the HER2 scoring algorithm in UBCa. Furthermore, our proposed algorithmic model performs cell identification and classification across the entire WSI, calculating the final HER2 score according to established guidelines. This approach is distinct from methods that rely on tissue microarrays or regions of interest [50]. Finally, the AI algorithm demonstrated strong performance across two independent test sets and can be seamlessly integrated into our pathology department's workflow.

Interobserver agreement of HER2 scoring in breast cancer is reported to have kappa values vary from the lowest 0.19 to the highest 0.80 [51, 52], and the interobserver agreement of HER2 scoring in gastric cancers, with reported kappa values between 0.61 and 0.78 using a four-tiered scoring method [53]. In the subsequent reader study, the inter-pathologist agreement in UBCa ranged from 0.39 to 0.63, slightly lower than the aforementioned study results. Several factors may contribute to interobserver variability in UBCa, including the individual pathologist's experience, the lack of specific expertise in HER2 scoring
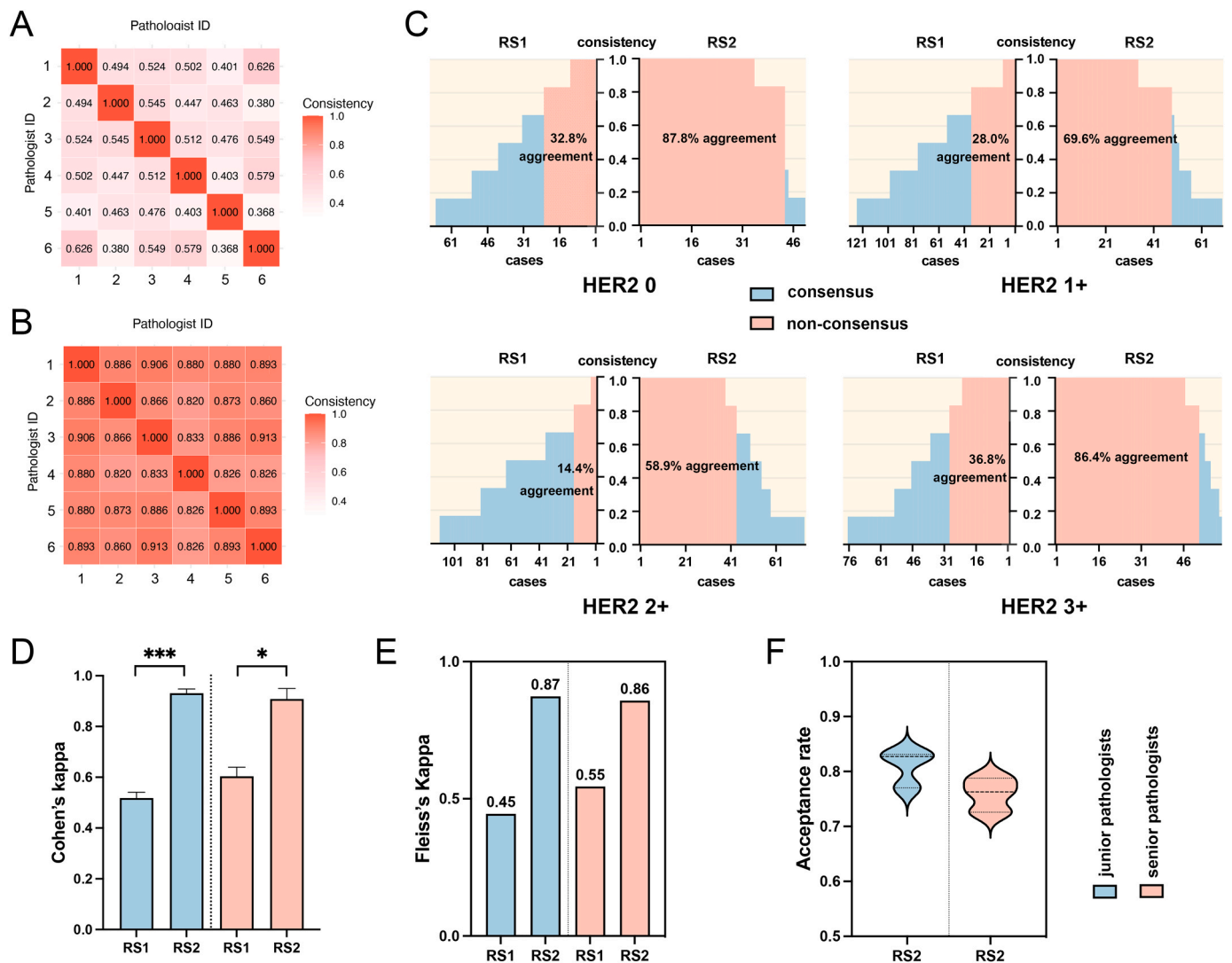
Fig. 7. HER2 scoring agreement in two rounds of ring studies (RSs), along with the accuracy, consistency, and acceptance rate of pathologists at different levels. (A) Total consistency in RS1: The heatmap shows pairwise agreement among pathologists in RS1, with moderate consistency observed across pathologists. (B) Total consistency in RS2: The heatmap shows improved pairwise agreement among pathologists in RS2, with overall consistency higher than in RS1. (C) Agreement for HER2 0, 1 +, 2 +, and 3 + in the two RSs: Bar graphs display the proportion of consensus and non-consensus cases for each HER2 category. Significant improvement in agreement is seen in RS2 for all HER2 categories, particularly for HER2 0 (32.8 % in RS1 vs. 87.8 % in RS2) and HER2 3 + (36.8 % in RS1 vs. 86.4 % in RS2). (D) Accuracy comparison between junior and senior pathologists in the two RSs: Accuracy improved significantly in RS2 for both junior and senior pathologists. Paired t-tests were conducted to determine significance between groups: ns, $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (E) Consistency comparison between junior and senior pathologists: Fleiss's Kappa indicates that both junior and senior pathologists had improved consistency in RS2. (F) Acceptance rate of AI results by pathologists at different levels: Violin plots depict the acceptance rates of AI results between junior and senior pathologists in RS2.

for UBCa, and the tumor's histologic subtype and heterogeneity.

In the RS1, we found that, compared to the gold standard, the accuracy and F1-scores, for HER2 2 + scoring (the threshold for therapeutic efficacy), were all relatively low. This discrepancy may be due to the routine practice in breast cancer HER2 scoring, where equivocal HER2 + results can be further evaluated using FISH or other tests to determine the final outcome. In contrast, for UBCa, pathologists often rely solely on HER2 IHC results to determine the threshold for treatment, leading to a more cautious and stringent selection of HER2-positive cases. In RS2, AI-assisted HER2 2 + scoring demonstrated improved accuracy, recall and precision. This indicates that the AI algorithm can accurately identify HER2 + patients. Additionally, the AI-assisted results showed varying degrees of improvement in distinguishing between HER2 0 and HER2 1 +, highlighting its potential utility in identifying patients with low HER2 expression as well. Furthermore, we observed that pathologists, regardless of their experience levels, all gained advantages from the AI-assisted method. As

expected, the junior pathologist group showed the greatest improvement in accuracy, achieving the highest accuracy in RS2 with the assist of AI. This improvement could be attributed to the junior pathologists' limited experience in HER2 interpretation, whereas senior pathologists may have more confidence in their assessments and therefore may be less inclined to accept AI results. The lower average acceptance among senior pathologists compared to junior pathologists also supports this observation. In conclusion, we found that the AI-assisted method was particularly helpful for pathologists in making decisions about cases where the cell proportion was near the threshold.

According to reports, HER2 heterogeneity in UBCa is more pronounced compared to breast cancer, similar to gastric cancer, and leads to poor agreement in HER2 IHC interpretation [17]. In this study, we explored the potential of AI assistance to enhance HER2 scoring in heterogeneous UBCa cases. The findings validated that AI notably improved the accuracy of HER2 scoring in such cases, achieving levels comparable to homogeneous cases. However, conclusions regarding the

correlation between HER2 heterogeneity and treatment efficacy in UBCa differed [28,54], possibly due to the subjectivity of HER2 IHC evaluation methods. Consequently, AI-based technologies show potential as an approach to deliver more precise and quantitative evaluations.

This study has several limitations. Firstly, the AI algorithm introduced in this research cannot accurately distinguish between carcinoma in situ and invasive carcinoma, potentially leading to misinterpretation in slides containing a mixture of both. Secondly, the algorithm's inability to incorporate a reference for HER2 staining intensity recognition. Furthermore, the gold standard for HER2 scoring used in this research depended on consensus readings from two or three seasoned pathologists, along with subjective definitions and categorizations of heterogeneous cases. This approach retains a degree of subjectivity.

5. Conclusions

This study demonstrated that the feasibility of urothelial carcinoma of the bladder HER2 scoring based on the 2018 ASCO-CAP guidelines. And results underscoring that reproducibility and consistency among pathologists significantly improved with the assist of AI, even in the presence of tumor heterogeneity. Our findings suggest the potential of AI in effectively discerning patients who stand to benefit from the latest ADC therapeutic agents.

CRediT authorship contribution statement

Xinyue Chen: Writing – review & editing, Validation, Resources, Methodology, Investigation, Data curation. **Ting Xie:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Shuaijun Chen:** Resources, Investigation. **Hongfeng Zhang:** Investigation, Data curation. **Bin Luo:** Validation, Resources, Investigation, Data curation. **Jingping Yuan:** Writing – review & editing, Visualization, Validation, Project administration, Conceptualization. **Yizhi Zhao:** Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Feng Guan:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Data curation, Conceptualization. **Honglin Yan:** Writing – review & editing, Validation, Supervision, Project administration, Formal analysis, Conceptualization. **Lin Yang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Aoling Huang:** Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shuying Ai:** Supervision, Resources, Investigation. **Xianli Ju:** Methodology, Investigation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clin 2024;74(3):229–63. https://doi.org/10.3322/caac.21834.

[2] Lobo N, Afferi L, Moschini M, et al. Epidemiology, screening, and prevention of bladder cancer. Eur Urol Oncol 2022;5(6):628–39. https://doi.org/10.1016/j.euo.2022.10.003.

[3] Lopez-Beltran A, Cookson MS, Guercio BJ, Cheng L. Advances in diagnosis and treatment of bladder cancer. Published online February 12, BMJ 2024:e076743. https://doi.org/10.1136/bmj-2023-076743.

[4] Gontero P, Birtle A, Capoun O, et al. European Association of Urology Guidelines on Non–muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ)—A Summary of the 2024 Guidelines Update (Published online) Eur Urol 2024. https://doi.org/10.1016/j.eururo.2024.07.027.

[5] Flaig TW, Spiess PE, Abern M, et al. Bladder cancer, version 3.2024. J Natl Compr Canc Netw 2024;22(4):216–25. https://doi.org/10.6004/jnccn.2024.0024.

[6] Pfail JL, Small AC, Cumarasamy S, Galsky MD. Real world outcomes of patients with bladder cancer. Hematol Oncol Clin North Am 2021;35(3):597–612. https://doi.org/10.1016/j.hoc.2021.01.005.

[7] Vashistha V, Wang H, Mazzone A, et al. Radical cystectomy compared to combined modality treatment for muscle-invasive bladder cancer: a systematic review and meta-analysis. Int J Radiat Oncol Biol Phys 2017;97(5):1002–20. https://doi.org/10.1016/j.ijrobp.2016.11.056.

[8] Alfred Witjes J, Max Bruins H, Carrión A, et al. European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: summary of the 2023 Guidelines. Eur Urol 2024;85(1):17–31. https://doi.org/10.1016/j.eururo.2023.08.016.

[9] Patelli G, Zeppellini A, Spina F, et al. The evolving panorama of HER2-targeted treatments in metastatic urothelial cancer: a systematic review and future perspectives. Cancer Treat Rev 2022;104:102351. https://doi.org/10.1016/j.ctrv.2022.102351.

[10] Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene 2007;26(45):6469–87. https://doi.org/10.1038/sj.onc.1210477.

[11] Hayes DF. HER2 and breast cancer — a phenomenal success story. N Engl J Med 2019;381(13):1284–6. https://doi.org/10.1056/nejmcibr1909386.

[12] Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/ neu oncogene. Science 1987;235(4785):177–82. https://doi.org/10.1126/science.3798106.

[13] Scherrer E, Kang A, Bloudek LM, Koshkin VS. HER2 expression in urothelial carcinoma, a systematic literature review. Front Oncol 2022;12:1011885. https://doi.org/10.3389/fonc.2022.1011885.

[14] Behzatoğlu K, Yörükoğlu K, Demir H, Bal N. Human epidermal growth factor receptor 2 overexpression in micropapillary and other variants of urothelial carcinoma. Eur Urol Focus 2018;4(3):399–404. https://doi.org/10.1016/j.euf.2016.06.007.

[15] Goodman AL, Osunkoya AO. Human epidermal growth factor receptor 2 expression in micropapillary urothelial carcinoma of the bladder: an analysis of 27 cases. Hum Pathol 2016;57:160–4. https://doi.org/10.1016/j.humpath.2016.07.014.

[16] Sanguedolce F, Russo D, Mancini V, et al. Prognostic and therapeutic role of HER2 expression in micropapillary carcinoma of the bladder. Mol Clin Oncol 2019;10(2):205–13. https://doi.org/10.3892/mco.2018.1786.

[17] An J, Peng C, Tang H, Liu X, Peng F. New advances in the research of resistance to neoadjuvant chemotherapy in breast cancer. IJMS 2021;22(17):9644. https://doi.org/10.3390/ijms22179644.

[18] Murthy RK, Loi S, Okines A, et al. Tucatinib, trastuzumab, and capecitabine for HER2-positive metastatic breast cancer. N Engl J Med 2020;382(7):597–609. https://doi.org/10.1056/nejmoa1914609.

[19] Albarrán V, Rosero DI, Chamorro J, et al. Her-2 targeted therapy in advanced urothelial cancer: from monoclonal antibodies to antibody-drug conjugates. Int J Mol Sci 2022;23(20). https://doi.org/10.3390/ijms232012659.

[20] Sheng X, Yan X, Wang L, et al. Open-label, Multicenter, Phase II Study of RC48-ADC, a HER2-Targeting Antibody–Drug Conjugate, in Patients with Locally Advanced or Metastatic Urothelial Carcinoma. Clin Cancer Res 2021;27(1):43–51. https://doi.org/10.1158/1078-0432.ccr-20-2488.

[21] Bardia A, Hu X, Dent R, et al. Trastuzumab Deruxtecan after endocrine therapy in metastatic breast cancer. Published online September 14, N Engl J Med 2024. https://doi.org/10.1056/nejmoa2407086.

[22] Xu H, Sheng X, Zhou L, et al. A phase II study of RC48-ADC in HER2-negative patients with locally advanced or metastatic urothelial carcinoma. JCO 2022;40 (16_suppl). https://doi.org/10.1200/jco.2022.40.16_suppl.4519.

[23] Wolff AC, Somerfield MR, Dowsett M, et al. Human epidermal growth factor receptor 2 testing in breast cancer: ASCO–College of American Pathologists Guideline Update. JCO 2023;41(22):3867–72. https://doi.org/10.1200/jco.22.02864.

[24] Drago JZ, Ferraro E, Abuhadra N, Modi S. Beyond HER2: targeting the ErbB receptor family in breast cancer. Cancer Treat Rev 2022;109:102436. https://doi.org/10.1016/j.ctrv.2022.102436.

[25] Diamantis N, Banerji U. Antibody-drug conjugates—an emerging class of cancer treatment. Br J Cancer 2016;114(4):362–7. https://doi.org/10.1038/bjc.2015.435.

[26] Fleischmann A, Rotzer D, Seiler R, Studer UE, Thalmann GN. Her2 amplification is significantly more frequent in lymph node metastases from urothelial bladder cancer than in the primary tumours. Eur Urol 2011;60(2):350–7. https://doi.org/10.1016/j.eururo.2011.05.035.

[27] Agrawal V, Bharti N, Pandey R. Human epidermal growth factor receptor 2 (HER2) gene amplification in non-muscle invasive urothelial bladder cancers: Identification of patients for targeted therapy. Arab J Urol 2020;18(4):267–72. https://doi.org/10.1080/2090598x.2020.1814183.

[28] Lei H, Ling Y, Yuan P, et al. Assessment of the expression pattern of HER2 and its correlation with HER2-targeting antibody-drug conjugate therapy in urothelial cancer. J Natl Cancer Cent 2023;3(2):121–8. https://doi.org/10.1016/j.jncc.2023.02.003.

[29] Kiss B, Wyatt AW, Douglas J, et al. Her2 alterations in muscle-invasive bladder cancer: patient selection beyond protein expression for targeted therapy. Sci Rep 2017;7(1). https://doi.org/10.1038/srep42713.

[30] Ross JS, Wang K, Al-Rohil RN, et al. Advanced urothelial carcinoma: next-generation sequencing reveals diverse genomic alterations and targets of therapy. Mod Pathol 2014;27(2):271–80. https://doi.org/10.1038/modpathol.2013.135.

[31] Madison RW, Gupta SV, Elamin YY, et al. Urothelial cancer harbours EGFR and HER2 amplifications and exon 20 insertions. BJU Int 2020;125(5):739–46. https://doi.org/10.1111/bju.15006.

[32] Sheng X, Zhou L, Yang K, et al. Disitamab vedotin, a novel humanized anti-HER2 antibody-drug conjugate (ADC), combined with toripalimab in patients with locally advanced or metastatic urothelial carcinoma: An open-label phase 1b/2 study. 4566-4566 JCO 2023;41(16_suppl). https://doi.org/10.1200/jco.2023.41.16_suppl.4566.

[33] Lambein K, Van Bockstal M, Vandemaele L, et al. Distinguishing score 0 from score 1+ in HER2 immunohistochemistry-negative breast cancer. Am J Clin Pathol 2013;140(4):561–6. https://doi.org/10.1309/ajcp4a7ktayhzsoe.

[34] Valenza C, Guidi L, Battaiotto E, et al. Targeting HER2 heterogeneity in breast and gastrointestinal cancers. Trends Cancer Publ Online 2023. https://doi.org/10.1016/j.trecan.2023.11.001.

[35] Zhang H, Wang Y, Wang Y, Wu D, Lin E, Xia Q. Intratumoral and intertumoral heterogeneity of HER2 immunohistochemical expression in gastric cancer. Pathol Res Pr 2020;216(11):153229. https://doi.org/10.1016/j.prp.2020.153229.

[36] Xu Y, Jiang L, Chen W, Huang S, Liu Z, Zhang J. Computer-aided detection and prognosis of colorectal cancer on whole slide images using dual resolution deep learning. J Cancer Res Clin Oncol 2022;149(1):91–101. https://doi.org/10.1007/s00432-022-04435-x.

[37] Bitencourt AGV, Gibbs P, Rossi Saccarelli C, et al. MRI-based machine learning radiomics can predict HER2 expression level and pathologic response after neoadjuvant therapy in HER2 overexpressing breast cancer. EBioMedicine 2020;61:103042. https://doi.org/10.1016/j.ebiom.2020.103042.

[38] Han Z, Lan J, Wang T, et al. A deep learning quantification algorithm for HER2 scoring of gastric cancer. Front Neurosci 2022;16. https://doi.org/10.3389/fnins.2022.877229.

[39] Koopman T, Buikema HJ, Hollema H, Bock GH, Vegt B. What is the added value of digital image analysis of HER 2 immunohistochemistry in breast cancer in clinical practice? A study with multiple platforms. Histopathology 2019;74(6):917–24. https://doi.org/10.1111/his.13812.

[40] Cai J, Zhu C, Chen P, et al. Category Separation For Weakly Supervised Multi-Class Cell Counting. 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). IEEE,; 2022. https://doi.org/10.1109/isbi52829.2022.9761703.

[41] Holten-Rossing H, Møller Talman ML, Kristensson M, Vainer B. Optimizing HER2 assessment in breast cancer: application of automated image analysis. Breast Cancer Res Treat 2015;152(2):367–75. https://doi.org/10.1007/s10549-015-3475-3.

[42] Tuominen VJ, Tolonen TT, Isola J. ImmunoMembrane: a publicly available web application for digital image analysis of HER2 immunohistochemistry. Histopathology 2012;60(5):758–67. https://doi.org/10.1111/j.1365-2559.2011.04142.x.

[43] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22(10):1345–59. https://doi.org/10.1109/tkde.2009.191.

[44] Jomaa M, Mathkour R, Bazi H, Islam Y. MS. end-to-end deep learning fusion of fingerprint and electrocardiogram signals for presentation attack detection. Sensors 2020;20(7). https://doi.org/10.3390/s20072085.

[45] Shui Z, Zhang S, Zhu C, et al. End-to-End Cell Recognition by Point Annotation. In: *Lecture Notes in Computer Science*. Springer Nature Switzerland; 2022. p. 109–18. https://doi.org/10.1007/978-3-031-16440-8_11.

[46] Sheng X, Wang L, He Z, et al. Efficacy and safety of disitamab vedotin in patients with human epidermal growth factor receptor 2-positive locally advanced or metastatic urothelial carcinoma: a combined analysis of two phase II clinical trials. JCO2202912 J Clin Oncol 2023. https://doi.org/10.1200/jco.22.02912.

[47] Chen M, Yao K, Cao M, et al. HER2-targeting antibody-drug conjugate RC48 alone or in combination with immunotherapy for locally advanced or metastatic urothelial carcinoma: a multicenter, real-world study. Cancer Immunol Immunother 2023;72(7):2309–18. https://doi.org/10.1007/s00262-023-03419-1.

[48] Sode M, Thagaard J, Eriksen JO, Laenkholm A. Digital image analysis and assisted reading of the HER2 score display reduced concordance: pitfalls in the categorisation of HER2-low breast cancer. Histopathology 2023;82(6):912–24. https://doi.org/10.1111/his.14877.

[49] Wang CW, Lin KL, Muzakky H, Lin YJ, Chao TK. Weakly supervised bilayer convolutional network in segmentation of HER2 related cells to guide HER2 targeted therapies. Comput Med Imaging Graph 2023;108:102270. https://doi.org/10.1016/j.compmedimag.2023.102270.

[50] Qaiser T, Mukherjee A, Reddy PB C, et al. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. Histopathology 2017;72(2):227–38. https://doi.org/10.1111/his.13333.

[51] Casterá C, Bernet L. HER2 immunohistochemistry inter-observer reproducibility in 205 cases of invasive breast carcinoma additionally tested by ISH. Ann Diagn Pathol 2020;45:151451. https://doi.org/10.1016/j.anndiagpath.2019.151451.

[52] Wu S, Yue M, Zhang J, et al. The role of artificial intelligence in accurate interpretation of HER2 immunohistochemical scores 0 and 1+ in breast cancer. Mod Pathol 2023;36(3):100054. https://doi.org/10.1016/j.modpat.2022.100054.

[53] Koopman T, Louwen M, Hage M, Smits MM, Imholz ALT. Pathologic diagnostics of HER2 positivity in gastroesophageal adenocarcinoma. Am J Clin Pathol 2015;143(2):257–64. https://doi.org/10.1309/ajcpcx69hgddgycq.

[54] Zhou L, Shao Z, Liu Y, et al. HER2 expression associated with clinical characteristics and prognosis of urothelial carcinoma in a Chinese population. Oncol 2023;28(8):e617–24. https://doi.org/10.1093/oncolo/oyad070.