

RESEARCH

Open Access



Precision HER2: a comprehensive AI system for accurate and consistent evaluation of HER2 expression in invasive breast Cancer

Zhongtang Xiong^{1†}, Kai Liu^{2†}, Shaoyan Liu^{1†}, Jiahao Feng², Jin Wang², Zewen Feng¹, Boan Lai¹, Qingxin Zhang¹, Qingping Jiang^{1*} and Wei Zhang^{1*}

Abstract

Background With the development of novel anti-HER2 targeted drugs, such as ADCs, it has become increasingly important to accurately interpret HER2 expression in breast cancer. Previous studies have demonstrated high intra-observer and inter-observer variabilities in evaluating HER2 staining by human eyes. There exists a strong requirement to develop artificial intelligence (AI) systems to achieve high-precision HER2 expression scoring for better clinical therapy.

Methods In the present study, we collected breast cancer tissue samples and stained consecutive sections with anti-Calponin and anti-HER2 antibodies. High-quality digital images were selected from immunohistochemical slides and interpreted as HER2 3+, 2+, 1+, and 0. AI models were trained and assessed using annotated training and testing sets. The AI model was trained to automatically identify ductal carcinoma in situ (DCIS) by Calponin staining and myoepithelial annotation and filter out DCIS components in HER2-stained slides using image-overlapping techniques. Furthermore, we organized two-phase validation studies. In phase one, pathologists interpreted 112 HER2 whole-slide images (WSIs) without AI assistance, whereas in phase two, pathologists read the same slides using the AI system after a washing period of 2 weeks.

Results Our AI model greatly improved the accuracy of reading (0.902 vs. 0.710). The number of HER2 1+ patients misdiagnosed as HER2 0 was significantly reduced (32/279 vs. 65/279), and they benefitted from ADC drugs. In addition, the AI algorithm improved the intra-group consistency of HER2 readings by pathologists with different years of experience (intra-class correlation coefficient [ICC]: 0.872–0.926 vs. 0.818–0.908), with the improvement most pronounced among junior pathologists (0.885 vs. 0.818).

Conclusions We proposed a high-precision AI system to identify and filter out DCIS components and automatically evaluate HER2 expression in invasive breast cancer.

[†]Zhongtang Xiong, Kai Liu and Shaoyan Liu contributed equally to this work.

*Correspondence:
Qingping Jiang
jiangqp2003@126.com
Wei Zhang
weizhangnfy2015@126.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords HER2, Breast cancer, Artificial intelligence, DCIS, ADCs

Background

Breast cancer is one of the most common malignant tumors in females globally, with high incidence and mortality rates. It is known as “the killer of women’s health” [1, 2]. The expression of human epidermal growth factor receptor 2 (HER2) protein and gene is an essential indicator for deciding the treatment strategy for breast cancer [3, 4]. The expression of HER2 is usually determined by pathologists through immunohistochemistry (IHC) or fluorescence in situ hybridization (FISH) [5, 6]. After the exclusion of in situ components of carcinoma, it is divided into four levels based on the degree and proportion of HER2 membrane staining in invasive cancer, namely, HER2 3+, 2+, 1+, 0 [7]. A combination with FISH further classifies it as HER2-positive (including HER2 3+ or HER2 2+/FISH+) or HER2-negative (including HER2 2+/FISH– or HER2 1+ or 0) expression.

For more than 20 years, only patients with HER2-positive breast cancer have benefited from traditional anti-HER2-targeting drugs such as trastuzumab [8, 9]. Therefore, clinically, HER2 immunohistochemical interpretation is largely based on distinguishing between 2+ and 3+ and between 2+ and 1+, with little attention given to the distinction between 0 and 1+. Recent clinical trials have demonstrated that novel drugs such as antibody-drug conjugates (ADCs) not only have clinical activity against typical HER2-positive breast cancer but also significant clinical activity against tumors with lower or moderate expression of HER2 [10–13]. Thus, patients with low HER2 expression (HER2 1+ or HER2 2+/FISH–) receive clinical attention. The traditional HER2 expression immunohistochemical binary classification can no longer meet the treatment requirements of clinical patients. Therefore, experts propose a more detailed three-category classification of HER2 expression, including HER2 positive, HER2-low, and negative [14]. Several drugs are available for patients with low expression, such as ADCs, which is a newly developed drug with a high price. Therefore, it is essential to accurately interpret HER2 0, 1+, and 2+ expression to enable patients to use the drug economically and reduce the social and economic burden.

Fernandez’s study in 2022 reported that the consistency of HER2 interpretation by pathologists was not high, among which the consistency rate of 0 and 1+ was only 26%, and that of 2+ and 3+ was 58% [15]. This finding was attributed to the following reasons: (1) The interpretation of human eyes is subjective, with great differences between the division of percentage and the division of chromatism. (2) Weak staining of 1+ demands more careful observation under a high-power lens, resulting

in greater inconsistencies between 0 and 1+ interpretation [16]. (3) Because carcinoma in situ cannot be counted, human eye interpretation could interpret carcinoma in situ as resembling invasive carcinoma, or not interpret invasive carcinoma as resembling carcinoma in situ. Therefore, it is important to understand how to accurately judge the immunohistochemical expression of HER2 for proper drug usage. The emergence and progress of artificial intelligence (AI) technology has ushered in a new wave. Computer algorithms can overcome the subjectivity of human eye interpretation by providing automated/semi-automated analysis of pathological digital images [17] and reducing labor intensity. Therefore, we developed a clinically acceptable AI interpretation system based on the limitations of manual interpretation, automatically obtained the refined classification of HER2 expression, and achieved clinical transformation to serve the clinic and assist patients with precision medicine.

Methods

Clinical data

A total of 300 paraffin-embedded specimens from patients with invasive breast cancer of no special type were collected from two medical institutions: the Third Affiliated Hospital of Guangzhou Medical University (TAHGZMU) and the Qianjiang City Central Hospital of Hubei Province (QJCH). The specimens were collected between 2021 and 2023 and included 200 cases from the former institution and 100 cases from the latter. Specimens with unclear invasive lesions and defective slides (such as tissue folds or tears, excessive background staining, dirt, or cover slide defects) were not enrolled in our datasets. Breast cancer biopsy and surgical specimens with good fixation were selected, and all specimens were pathologically confirmed as invasive ductal breast cancer. Two experienced pathologists reviewed the hematoxylin and eosin (H&E) and immunohistochemistry (IHC) slides of all specimens. In our study, 188 samples were selected for AI model learning and training, whereas 112 samples were selected for AI model validation (including 56 biopsy tissue samples and 56 surgical specimens, with 70 from our center and 42 from external institutions). All paraffin samples were prepared as consecutive sections and stained with H&E and IHC (4b5, rabbit monoclonal antibodies; Ventana Medical Systems, Oro Valley, AZ, USA). Two experienced pathologists independently reviewed the slides and interpreted the cases according to the HER2 scoring guidelines [7]. Inconsistent cases were re-evaluated by a third pathologist. When consensus could not be reached, the cases were discussed to

reach an agreement, and consistency scoring was used as the gold standard.

AI system development

All slides were scanned using a three-dimensional (3D) digital scanner (3DHISTECH Ltd., Budapest, HUN), and whole-slide images (WSIs) were obtained. All tissues and data were used after obtaining permission from the hospital's institutional review board. As shown in Fig. 1, two consecutive sections stained with anti-HER2 and anti-Calponin antibodies were input into the HER2 cell detection module and the Calponin tissue segmentation module, respectively. The former was used to detect and grade tumor cells in the HER2 sections, whereas the latter was used to segment the myoepithelial tissue in the Calponin sections. Two 2.5× (magnification) thumbnail images of WSIs were placed in the VALIS [18] image registration module to obtain the transformation matrix for image registration.

The effective area of the section in the HER2 cell detection module was first segmented into multiple 1024×1024 image blocks using a sliding window approach, and the images were processed at 40× magnification to ensure the accuracy of cell detection and grading. All image blocks were input into the pre-trained CSRNet [19] to detect tumor cells and output a tumor cell detection map. Subsequently, the original image blocks and the tumor cell detection map were placed in the tumor cell HER2CIs classification module to classify the tumor cells and obtain their positions and grades in the HER2 section.

The HER2CIs classification module is largely used for the grading of detected tumor cells. The RGB channel

image blocks are converted to HED channel images using color deconvolution, following which different thresholds are used to segment the cell membrane and cytoplasm of different intensities. The completeness and uniformity of cell membrane staining are determined by calculating the connectivity between each cytoplasm and its adjacent cytoplasm in high-intensity cell membrane regions, distinguishing between 3+ and non-3+ HER2. The final grade of the cell is determined by analyzing the strongest staining intensity of the cell membrane surrounding non-3+ HER2 tumor cells.

The effective area of the section in the Calponin tissue segmentation module was segmented into multiple 1024×1024 image blocks using a sliding window approach. The images were processed at 10× magnification due to the large area of the myoepithelial tissue to obtain a myoepithelial tissue segmentation map from all image blocks using the pre-trained HRNet-48 [20]. Afterward, the segmentation map was used to outline the edges of the myoepithelial tissue, and the myoepithelial tissue segmentation was mapped onto the HER2 section using the transformation matrix obtained from the image registration module. Finally, DCIS regions were filtered out based on the myoepithelial tissue area, resulting in the HER2 index of the invasive carcinoma region.

Study design

The IHC of HER2 was interpreted by nine pathologists with different levels of practice experience, including three junior pathologists with 1 to 2 years of experience, three intermediate pathologists with 3 to 5 years of experience, and three senior pathologists with 6 to 10 years of experience. They all had experience interpreting HER2

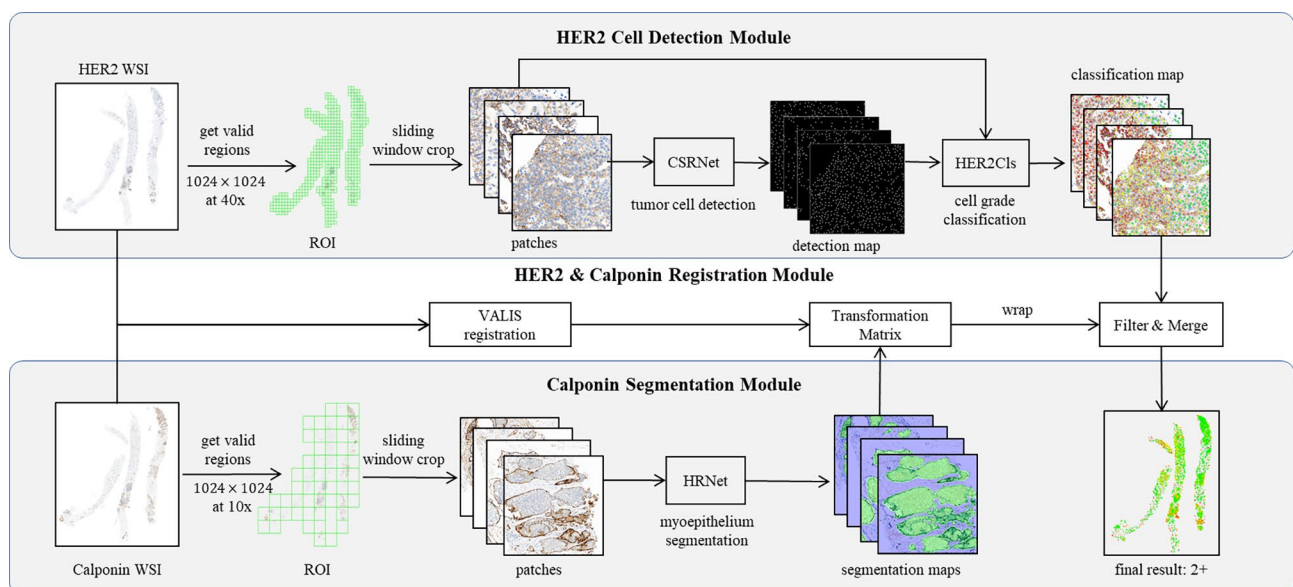


Fig. 1 Framework of the AI detection for HER2 and myoepithelium

IHC in routine clinical practice. Firstly, the pathologists reviewed the 2018 ASCO/CAP guidelines and received training on using AI-assisted equipment. The study was divided into two phases. In phase one, the nine pathologists used WSIs to interpret 112 HER2 IHC slides (56 biopsy specimens and 56 surgical specimens). The pathologists evaluating the WSIs of HER2 individually were called “HER2-pathologists.” After a washout period of 2 weeks, phase two of the study was conducted, in which the pathologists reinterpreted the same slides using the AI-assisted system. The AI system termed “HER2-AI” provided the pathologists with the proportion of different staining scores for HER2 on different slides and displayed a comprehensive recommendation score as a reference for pathologists.

Statistics

Statistical analysis was performed using IBM SPSS Statistics (version 25.0; IBM) and GraphPad Prism 9.01 (GraphPad Software). The accuracy of manual interpretation and the performance of AI-assisted interpretation were evaluated using accuracy, precision, recall, and F1 score. Cohen’s kappa was used to calculate the consistency (accuracy) between the individual readings of pathologists of HER2 results or the results of pathologists evaluating HER2 with AI assistance and the gold standard. Fleiss’ kappa and intra-class correlation coefficient (ICC) were used to evaluate the concordance among observers. Kruskal–Wallis and Wilcoxon rank-sum tests were used to analyze the accuracy differences among the pathologists using different interpretation methods. P -values < 0.05 were considered significant.

Results

Model training and performance

Iterative training of a DL model was performed, followed by evaluating a test set to demonstrate its ability to detect tumor regions. The DL model resulted in high performance on HER2 IHC WSIs. The performance metrics of each model in the module are depicted in Table S1. The accuracy of tumor cell detection was 0.856 in the HER2 cell detection module, with a mean absolute error of 0.0004. The accuracy of cell grade classification was 0.8500, and the F1 score was 0.8200. The segmentation results showed that background-IoU and myoepithelium-IoU were 0.9910 and 0.9100, respectively, in the Calponin segmentation module. The registration results in the HER2 and Calponin registration modules demonstrated a rigid registration error of 0.0830 and a non-rigid registration error of 0.0460. The myoepithelial segmentation model accurately identified the DCIS component in the tumor tissue by annotating and learning from the myoepithelium in Calponin IHC WSI and filtered out the DCIS component in HER2 IHC WSI by image overlapping

method, ultimately obtaining HER2 evaluation grading results in invasive carcinoma in HER2 IHC WSI (Fig. 2). Our DL model exhibited high performance in the test set of both biopsy and resection samples, automatically identifying tumor cells and classifying them with different staining intensities and completeness of HER2 membrane staining. Representative images of IHC sections and their AI results are depicted in Fig. 3.

Overall study interpretation

The results of HER2-pathologists and HER2-AI are presented in Fig. 4 (matrix plot). Compared to the gold standards, the accuracy and consistency of HER2 evaluation with AI assistance were superior to conventional digital slide-based assessment of HER2 grading. HER2-pathologists and HER2-AI exhibited the highest accuracy and consistency in HER2 3+ cases, followed by HER2 0 cases. Compared to HER2-pathologists, HER2-AI significantly decreased the number of cases where HER2 1+ lesions were misinterpreted as HER2 0 lesions. Simultaneously, in HER2-AI, a significant reduction in the number of cases where HER2 2+ lesions were misinterpreted as HER2 1+ lesions was observed.

Accuracy assessment of all pathologists in two phases

The final interpretation result in HER2-AI, referred to as AI-assisted pathologist review, was generated by adjusting the score based on AI results and their perception. As shown in Fig. 5A, the total accuracy of HER2-AI using AI-assisted methods (accuracy: 0.902) was significantly higher compared to HER2-pathologists (accuracy: 0.710). The use of AI-assisted methods reduced the accuracy gap between all pathologists’ results and the gold standard by 0.192. Furthermore, the accuracy of pathologists in interpreting HER2 0, 1+, 2+, and 3+ tumors was evaluated separately. In HER2-pathologists, accuracy (mean values of F1 scores were 0.768, 0.783, 0.716, and 0.764; precision values were 0.786, 0.830, 0.801, and 0.732; and recall values were 0.785, 0.783, 0.746, and 0.780;) for HER2 0, 1+, 2+, and 3+ tumors was relatively poor. The use of HER2-AI improved these values improved to varying degrees. For example, the F1 scores for HER2 0, 1+, 2+, and 3+ increased to 0.959, 0.960, 0.831, and 0.874, respectively (Fig. 5B–D). Cohen’s kappa values were calculated to assess the accuracy of HER2 interpretation when compared with the gold standard in both phases. The accuracy of HER2-AI was significantly higher than that of HER2-pathologists (Fig. 5E).

Next, the influence of different specimen types on HER2 interpretation was compared in the two stages. In the first stage, the accuracy of HER2 interpretation in resected samples was lower than that in biopsy samples. However, the accuracy of HER2 interpretation was significantly improved for both resected specimens and

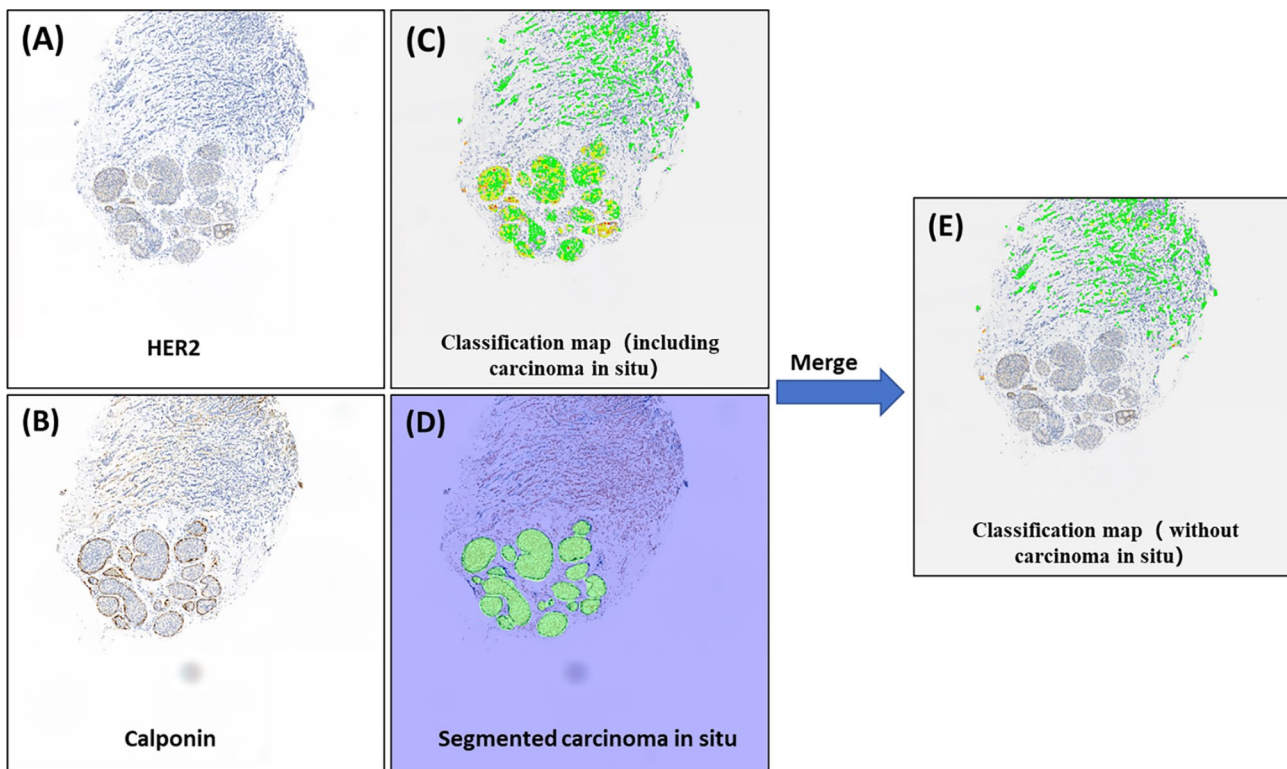


Fig. 2 The flowchart of AI in accurate Interpretation of HER2 in invasive breast cancer. **A, B** Using consecutive sections stained with HER2 and calponin respectively. **C** AI recognizes HER2 staining (including DCIS). **D** AI recognizes DCIS components by calponin staining. **E** Through image overlay technology, AI filters out DCIS component and automatically identifies HER2 expression in invasive cancer components. Green, No staining; Bright yellow-green, Faint/barely perceptible & Incomplete; Orange, Weak to moderate & complete; Red, Intense & complete

biopsy specimens in the second stage, and the difference in HER2 interpretation between different specimen types was reduced (Fig. 5F). Compared to manual interpretation alone, the accuracy of HER2 interpretation was significantly improved using the AI system, both for samples from our institution and from external institutions (Fig. 5G). In addition, nine pathologists were divided into three groups according to their length of practice. Figure 5H shows the comparison of accuracy between the three groups of pathologists in both HER2-pathologists and HER2-AI. In HER2-pathologists, the accuracy of junior and intermediate pathologists varied more than that of senior pathologists. In HER2-AI, junior and intermediate pathologists both benefited from AI-assisted methods, whereas the impact on the accuracy of senior pathologists was negligible. The accuracy gap between pathologists with different experience levels was reduced following the adoption of AI-assisted methods, with the greatest improvement noted in junior pathologists, from 0.539 (95% confidence interval [CI]: 0.535–0.544) in HER2-pathologists to 0.800 (95% CI: 0.784–0.808) in HER2-AI (Fig. 5H).

Consistency assessment in each phase for all pathologists

A heatmap was used to visualize the changes in consistency between HER2-pathologists and HER2-AI (Fig. 6A). AI algorithms significantly improved the overall inter-observer consistency among pathologists. In the first phase, pathologists achieved moderate agreement in HER2 scoring when independently reviewing digital slides (Fleiss Kappa \approx 0.575; 95% CI: 0.575–0.576). In comparison to HER2-pathologists, AI-assisted evaluation in the second phase significantly enhanced the consistency of HER2 interpretation (Fleiss Kappa \approx 0.687; 95% CI: 0.687–0.688) (Fig. 6B). We further analyzed the differences in consistency among pathologists with different years of experience. As shown in Fig. 6C, the consistency among pathologists of different experience levels improved significantly in HER2-AI. In HER2-pathologists, the highest consistency was observed among senior pathologists (ICC \approx 0.908; 95% CI: 0.514–0.626), whereas junior pathologists had the lowest consistency (ICC \approx 0.818; 95% CI: 0.514–0.626). In HER2-AI, the consistency among pathologists of all experience levels notably improved, particularly of those with lower years of experience. For example, the consistency of junior pathologists increased from 0.818 to 0.885, manifesting the greatest advancement.

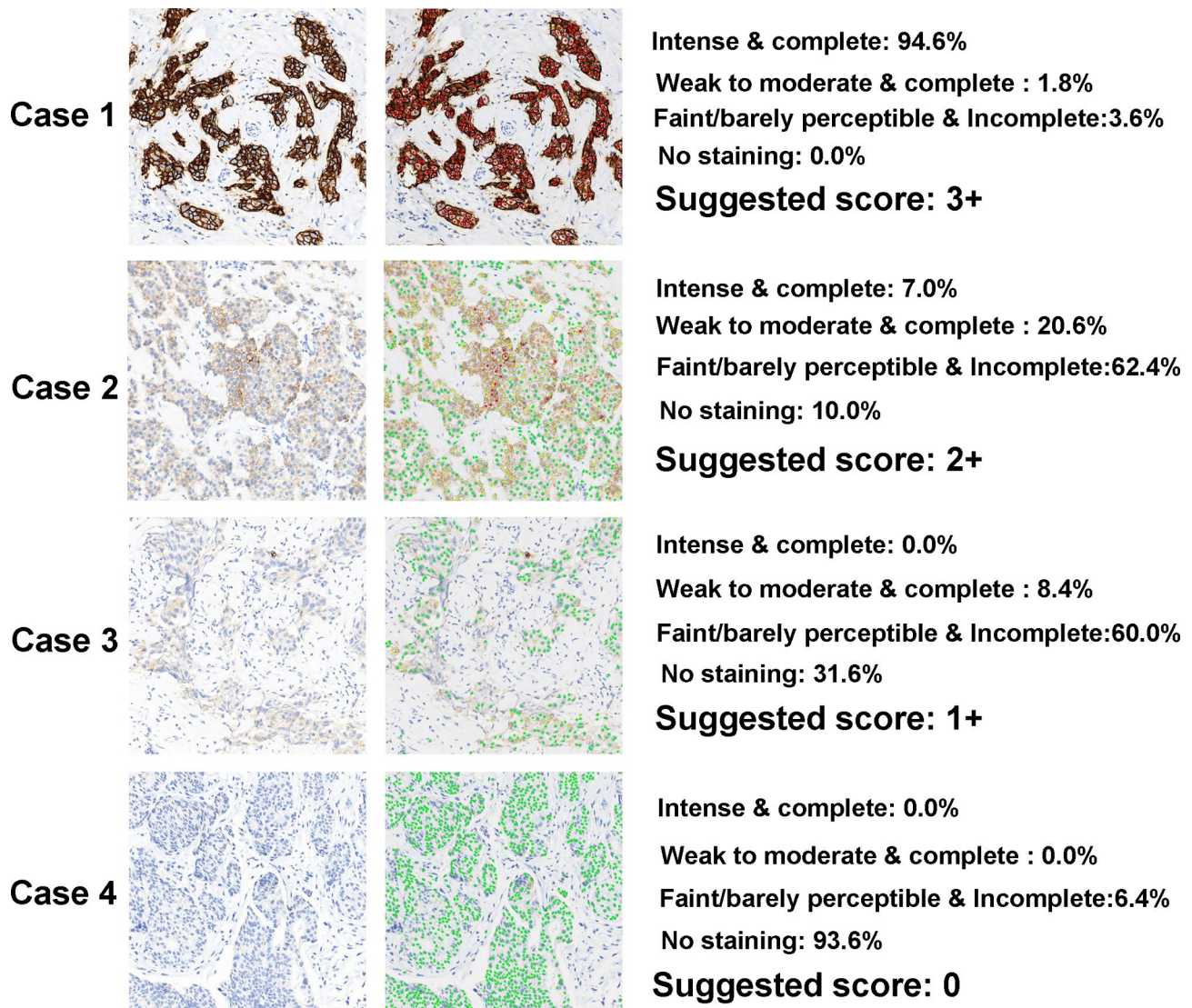


Fig. 3 HER2 images under AI-assisted interpretation. Green, No staining; Bright yellow-green, Faint/barely perceptible & Incomplete; Orange, Weak to moderate & complete; Red, Intense & complete

Discussion

HER2 is an independent prognostic indicator for breast cancer and has been used in selecting treatment plans [21]. Improved treatment strategies for breast cancer, especially the emergence of novel ADC drugs [10, 22], have updated the traditional HER2 expression immunohistochemical classification. Scholars have emphasized the significance of more accurately evaluating and classifying HER2, which demands higher detection requirements for HER2 interpretation. Previous literature reports that pathologists showed a high agreement in HER2 0 and 3+ and less satisfactory agreement in HER2 1+ and HER2 2+ tumors [23]. For instance, Marchio et al. reported that HER2 IHC scores showed poor consistency, particularly in HER2 2+ cases [24], increasing the number of unnecessary FISH examinations and work

intensity. Fernandez et al. recently reported an agreement of only 26% between HER2 0 and 1+ among 18 pathologists [15]. Artificial intelligence can be effective in solving the subjectivity of human eyes in HER2 interpretation. In the present study, we set up rigorous comparison tests between manual interpretation and AI-assisted interpretation. Overall, our DL model displayed excellent performance in accuracy, precision, recall, and F1. Our study demonstrated that the AI model significantly improved the overall accuracy (0.710 vs. 0.901) and consistency of HER2 (Fleiss Kappa: 0.575 vs. 0.687). Consistent with previous studies [23, 25], our study found that pathologists and AI-assisted systems displayed the highest accuracy and consistency in HER2 3+ cases, followed by HER2 0 cases. Compared to human eye interpretation, the AI-assisted system significantly reduced the number

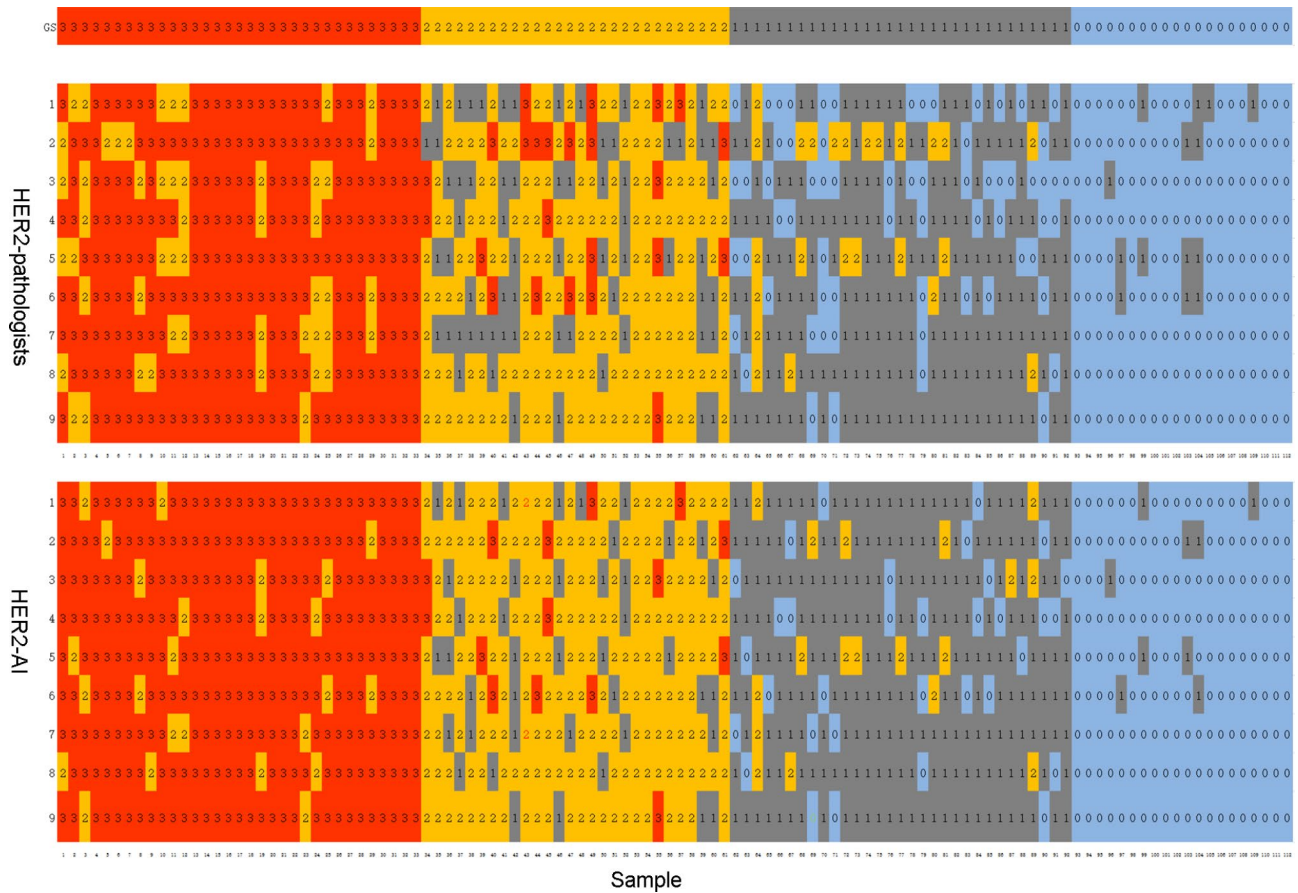


Fig. 4 112 cases were evaluated for HER2 scoring by 9 pathologists in two phases. The numbers represent the corresponding HER2 scores. “GS” represents the gold standard (at the top of the figure)

of cases in which HER2 1+ lesions were misinterpreted as HER2 0 lesions. This allowed the subset of misjudged patients to benefit from novel ADC drugs. The number of cases in which HER2 2+ lesions were misinterpreted as HER2 1+ lesions was significantly reduced in our study. These misdiagnosed patients could benefit from traditional anti-HER2-targeting drugs, such as trastuzumab, reducing their financial burden.

We analyzed the effect of different specimen types (fine needle aspiration vs. surgical resection) on HER2 interpretation results. Compared to small biopsy samples, pathologists displayed higher inter-observer variability to interpret surgical resection specimens during traditional HER2 interpretation. AI assistance reduced the interpretive difference between these two types of specimens, improving the consistency and reproducibility. In addition, we validated samples from two different laboratories. Our AI model displayed high accuracy and consistency in both types of specimens from the two centers. The two comparative verification studies demonstrated that the accuracy of HER2 interpretation improved significantly with the assistance of AI algorithms, particularly for HER2 0 and 1+ cases, irrespective of the

pathologists’ years of experience. This is one of the few studies providing direct evidence that AI assistance can help pathologists better and more accurately diagnose HER2 3+, 2+, 1+, and 0 in breast cancer. In this study, all pathologists with different levels of experience benefited from the AI-assisted approach. Unsurprisingly, the improvement in accuracy and consistency was greatest in the junior pathologist group. AI-assisted improvement was present but not as significant in the intermediate- and high-level pathologist groups. This may be due to the fact that junior pathologists are less experienced in HER2 interpretation and are more likely to accept AI results. However, experienced pathologists, who were more confident in their assessment, showed more resistance to accepting the AI results. The subjective perception of DAB color intensity and width by the human eye, particularly in cases of weak or nearly imperceptible HER2 1+ staining, often leads to poor inter- and intra-observer consistency in HER2 IHC evaluation. The HER2 staining in certain immunohistochemistry sections was faint that some pathologists did not recognize it, yet AI was able to detect it. While AI assistance provided limited benefit to experienced pathologists, it reduced the number of

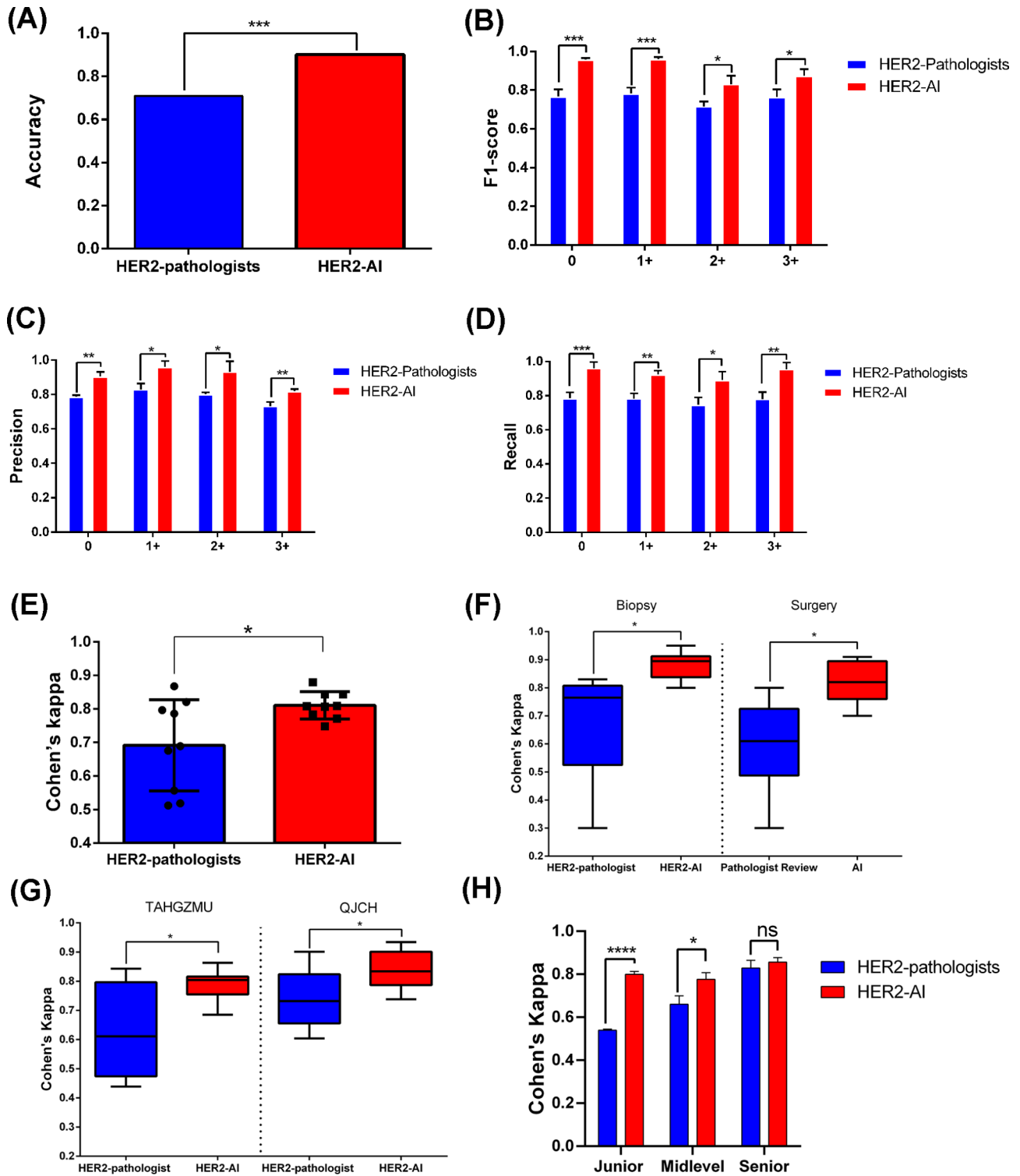


Fig. 5 The accuracy of pathologists was compared in two-phase studies. **A-D** The evaluation of total accuracy and the statistical comparisons of inter-group significance. **E** The accuracy of pathologists significantly improved with the assistance of AI algorithms. **F** Surgical samples and biopsy samples were compared in two-phase experiments. **G** The AI system achieved similar high performance in samples from two centers. **H** The accuracy of junior, intermediate, and senior pathologists was compared in two phases. (ns, $P \geq 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$)

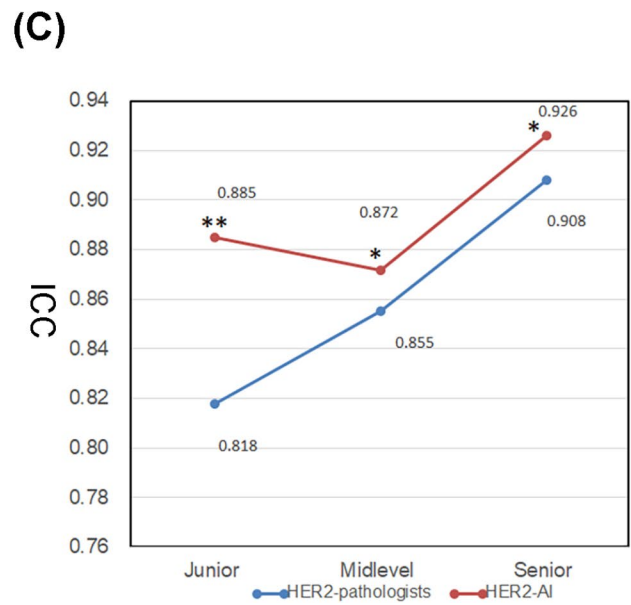
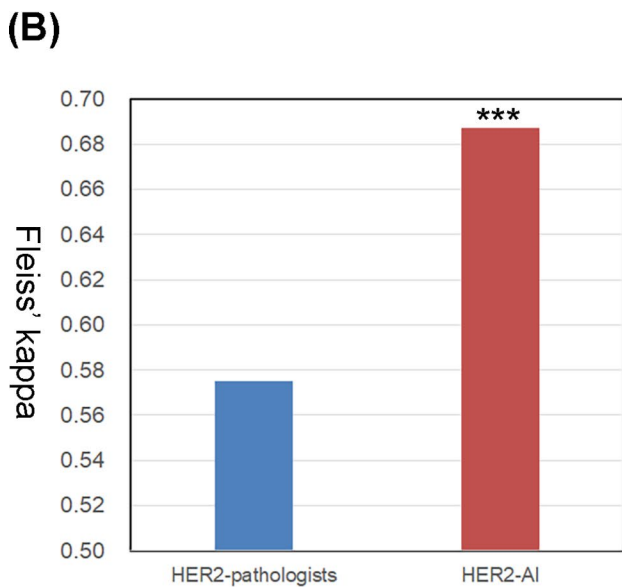
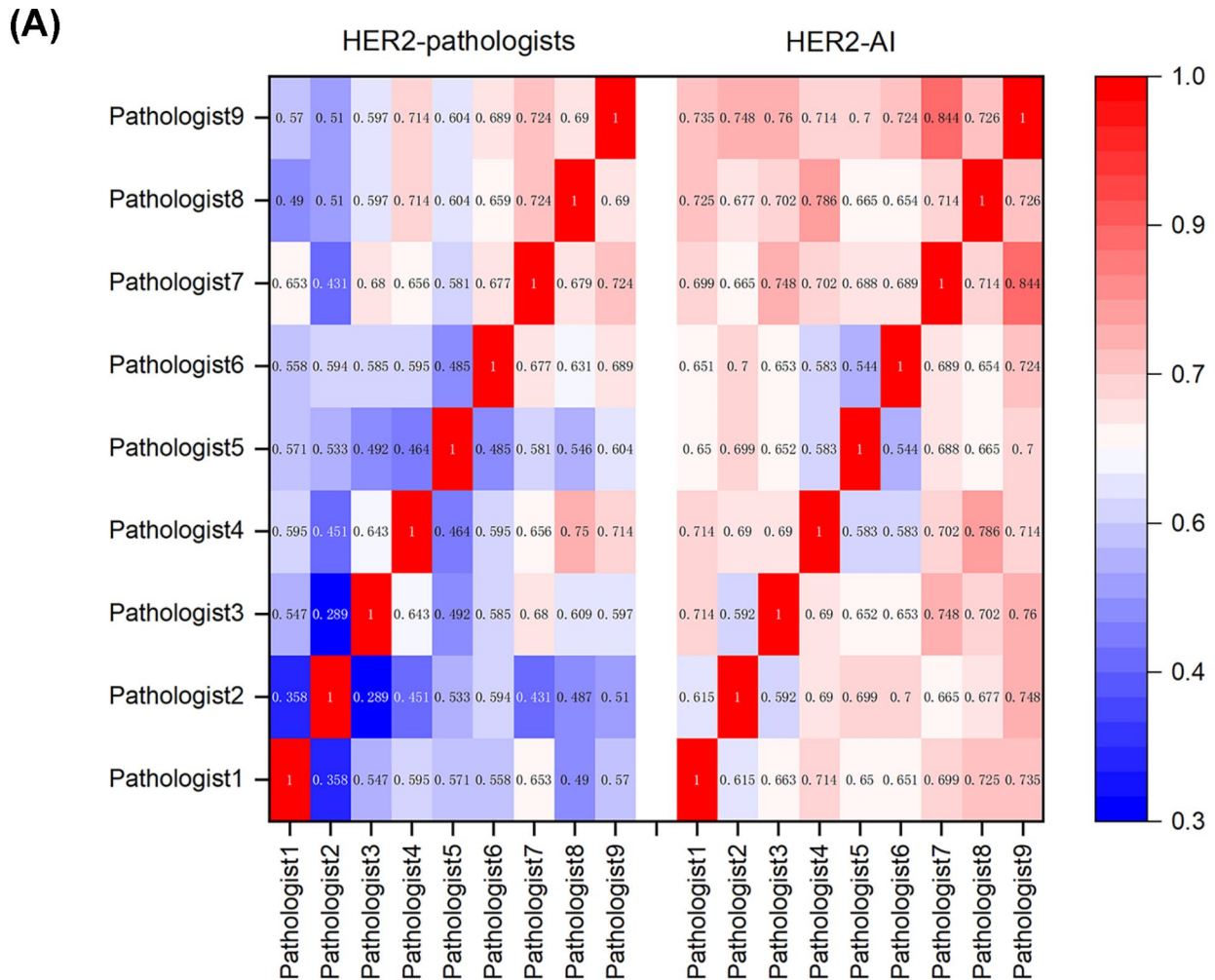


Fig. 6 Consistency of HER2 scoring in two phases. **A** The heatmap shows overall consistency between the two phases. **B** Fleiss' kappa was used to measure the consistency between observers for HER2-pathologists and HER2-AI. **C** Intraobserver concordance of pathologists at different levels between HER2-pathologists (blue) and HER2-AI (Red). (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$)

cases where they misclassified HER2 1+ as 0+, as shown in Fig. 4. Moreover, the human eye exhibits a substantial margin of error when estimating the percentage of HER2 staining, particularly in cases with marked heterogeneous expression. In clinical practice, disagreements among pathologists were more frequent when HER2 expression hovered around the critical 10% threshold. AI-based evaluation techniques provided a more objective and quantitative method, offering pathologists valuable reference data, which could effectively reduce such discrepancies. The role of AI in the context of experienced pathologists was to assist them in revisiting their initial assessments, helping to prevent potential diagnostic errors. Concurrently, the AI algorithm evaluated HER2 staining and its percentage in a relatively objective and reproducible manner. The visualization of the AI model's results, acting as a second opinion, served to bolster the pathologist's diagnostic confidence. In cases where opinions were inconsistent, the pathologist's attention could be directed to the case, and a third-party consultation could be sought if necessary.

A combination of pathologists' expertise and AI models to provide easily perceptible AI results can improve the reliability of pathologist scoring. Recently, computer algorithms, such as AI models, have become an objective and repeatable IHC scoring method for reducing the number of suspicious HER-2 cases [17, 26]. For example, Yue et al. reported that their AI-assisted microscope improved the accuracy of HER2 3+ and 2+ scores, reducing recall of FISH-positive results in IHC 2+ tumors [25]. Similarly, Wu et al.'s study focused specifically on HER2 0 and 1+ interpretation and demonstrated that their AI-assisted system significantly improved the accuracy and consistency of pathologists' HER2 interpretation [27]. A limitation of current AI studies related to HER2 interpretation is their inability to distinguish between carcinoma in situ and invasive carcinoma. The inability to exclude in situ carcinoma components significantly produces biased HER2 interpretation. In clinical work, several specimens contain DCIS components to a greater or lesser extent, it is challenging for the pathologists to distinguish them microscopically. A recent literature review has highlighted the potential for automated assessment of the Ki-67 marker in breast cancer. This could be achieved by filtering out the in situ cancer component in Ki67-stained sections using image alignment techniques [28]. In this study, we attempted to use image overlap techniques to achieve the effect of filtering out in situ carcinoma in breast cancer. Our AI model increased an algorithm to automatically filter out DCIS components by myoepithelial marker Calponin. When the AI system recognizes cancer cells in the specimen, it only provides automatic quantification results and visualization of HER2 expression status in invasive carcinoma components, thereby

greatly improving the accuracy and reliability of AI automatic interpretation of HER2 in practice.

Our AI system is designed to automatically process tumor detection, cell segmentation, and positivity differentiation after pathologists upload the WSIs to the system. The system provides quantitative values as a second opinion for pathologists by pre-reading the HER2 slides. In addition, the AI-assisted system is highly useful for pathologists during HER2 diagnosis. Our AI system marks the HER2 membranous-stained cells with different staining intensities and completeness on the image following segmentation. Some studies have demonstrated that AI-assisted systems, capable of quickly evaluating digital sections and providing visual interpretation results, can significantly enhance efficiency and reduce the workload of pathologists. For instance, the AI-assisted PD-L1 interpretation system developed by Prof. Ling et al. not only improved the diagnostic reproducibility among pathologists but also significantly reduced the time required for interpreting PD-L1 [29]. In our AI model, labelled images of HER2-positive invasive cancer cells could be readily and quickly evaluated by pathologists. Therefore, we speculated that AI-assisted interpretation may enhance the efficiency of pathologists in interpreting HER2 and reduce the workload of pathologists, which is one of the advantages of developing an AI-assisted system. Overall, our findings revealed the potential advantages of AI systems, which could improve accuracy and consistency by providing pre-read HER2 results and visualizing HER2-expressing cells, thereby assisting pathologists in making more reliable diagnoses.

The present study had certain limitations. Firstly, the sample size was relatively limited, comprising only 300 patients and two research centers, with a relatively small dataset. Collecting more datasets from multiple institutions would improve the robustness of the AI model. In addition, more multicenter clinical trials would verify its performance in real-world settings. Secondly, the DL-based tumor detection model occasionally misidentifies non-tumor cells and misclassified HER2-positive cells, generating a certain bias in the HER2 output. Thirdly, we only focused on samples of invasive breast cancer of no special type, and other specific subtypes were not included in the study. Finally, quality control is necessary for the clinical application of AI tools, and it is recommended to integrate human-machine interactions in the future HER2 diagnostic workflow.

In summary, our study extended the application of AI systems to detecting breast cancer, identifying and filtering in situ cancer, and providing better HER2 staining scoring. The AI system displayed high consistency with professional pathologists in HER2 scoring interpretation, especially in HER2-low cases. The pre-read quantitative results and HER2-expressing cell visualization provided

by the AI system could enhance diagnostic reproducibility and efficiency. This study demonstrates that AI-assisted systems could be an effective and valuable tool to overcome the challenges of HER2 assessment in the field of targeted therapy.

Abbreviations

AI	Artificial intelligence
DCIS	Ductal carcinoma in situ. WSIs: whole-slide images
ADC	Antibody-drug conjugate
ICC	Intra-class correlation coefficient
HER2	Human epidermal growth factor receptor 2
H&E	Hematoxylin and eosin
IHC	Immunohistochemistry
FISH	Fluorescence in situ hybridization
3D	Three-dimensional
DL	Deep learning
CSRNet	Cross-Scale Residual Network
VALIS	Virtual Alignment of pathoLogY Image Series

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-024-12980-6>.

Supplementary Material 1

Acknowledgements

We would like to express our gratitude to Dr. Li Zhao, Dr. Yuping Liu, and Dr. Ping Qin for their invaluable assistance in the HER2 interpretation work. We also acknowledge the AI technical support provided by Cells Vision (Guangzhou) Medical Technology Inc., Guangzhou, China. We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript.

Author contributions

A. ZW carried out experiments, analyzed data, and cowrote the manuscript. B. LK completed the development of the AI system, and cowrote the manuscript. C. LSY performed IHC experiments and HER2 quality control. D. XZT carried out data interpretation and conceived the manuscript. FJH and WJ were involved in the development of AI system. E. FZW, LBA, and ZQX are involved in HER2 interpretation. F. JQP provided advice on the structure and critical review of the manuscript. G. All authors reviewed the manuscript and had final approval of the submitted and published versions.

Funding

This work was supported by Beijing Jingjian Pathology Development Foundation (grant number JJDYSG2023-018), and Science and Technology Projects Foundation of Guangzhou city, (grant number 202201020095).

Data availability

All data generated or analyzed during this study are included in article/supplementary materials and are also available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The present study was approved by the Ethics Committee and the Institutional Review Committee of the Third Affiliated Hospital of Guangzhou Medical University (No. 2024129) and was conducted in accordance with the ethical standards of the World Medical Association Declaration of Helsinki. Informed consent was obtained from all individual participants included in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Pathology; Guangdong Provincial Key Laboratory of Major Obstetric Diseases; Guangdong Provincial Clinical Research Center for Obstetrics and Gynecology, The Third Affiliated Hospital, Guangzhou Medical University, NO.63, Duobao Road, Guangzhou 510150, China
²Cells Vision (Guangzhou) Medical Technology Inc, Guangzhou, China

Received: 27 June 2024 / Accepted: 23 September 2024

Published online: 30 September 2024

References

- DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin.* 2017;67(6):439–48.
- Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin.* 2023;73(1):17–48.
- Krystal-Whittemore M, Xu J, Brogi E, Ventura K, Patil S, Ross DS, Dang C, Robson M, Norton L, Morrow M, et al. Pathologic complete response rate according to HER2 detection methods in HER2-positive breast cancer treated with neoadjuvant systemic therapy. *Breast Cancer Res Treat.* 2019;177(1):61–6.
- Fehrenbacher L, Cecchini RS, Geyer CE Jr, Rastogi P, Costantino JP, Atkins JN, Crown JP, Polikoff J, Boileau JF, Provencher L, et al. NSABP B-47/NRG oncology phase III Randomized Trial comparing adjuvant chemotherapy with or without Trastuzumab in high-risk invasive breast Cancer negative for HER2 by FISH and with IHC 1+ or 2. *J Clin Oncol.* 2020;38(5):444–53.
- Wesola M, Jelen M. A comparison of IHC and FISH cytogenetic methods in the evaluation of HER2 status in breast Cancer. *Adv Clin Exp Med.* 2015;24(5):899–903.
- Hwang HC, Gown AM. Evaluation of human epidermal growth factor receptor 2 (HER2) gene status in human breast Cancer formalin-fixed paraffin-embedded (FFPE) tissue specimens by fluorescence in situ hybridization (FISH). *Methods Mol Biol.* 2016;1406:61–70.
- Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, Bilous M, Ellis IO, Fitzgibbons P, Hanna W, et al. Human epidermal growth factor receptor 2 testing in breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol.* 2018;36(20):2105–22.
- Hudis CA. Trastuzumab—mechanism of action and use in clinical practice. *N Engl J Med.* 2007;357(1):39–51.
- Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JM, Bilous M, Fitzgibbons P, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol.* 2013;31(31):3997–4013.
- Nakada T, Sugihara K, Jikoh T, Abe Y, Agatsuma T. The Latest Research and Development into the antibody-drug Conjugate, [fam-] Trastuzumab Deruxtecan (DS-8201a), for HER2 Cancer therapy. *Chem Pharm Bull (Tokyo).* 2019;67(3):173–85.
- Xu Z, Guo D, Jiang Z, Tong R, Jiang P, Bai L, Chen L, Zhu Y, Guo C, Shi J, et al. Novel HER2-Targeting antibody-drug conjugates of Trastuzumab Beyond T-DM1 in breast Cancer: Trastuzumab Deruxtecan (DS-8201a) and (Vic-) Trastuzumab Duocarmazine (SYD985). *Eur J Med Chem.* 2019;183:111682.
- Diaz-Rodriguez E, Gandullo-Sanchez L, Ocana A, Pandiella A. Novel ADCs and strategies to overcome resistance to Anti-HER2 ADCs. *Cancers (Basel)* 2021, 14(1).
- Ferraro E, Drago JZ, Modi S. Implementing antibody-drug conjugates (ADCs) in HER2-positive breast cancer: state of the art and future directions. *Breast Cancer Res.* 2021;23(1):84.
- International Medical Society CA-CA, Breast Cancer Group BoOCMDA. [Consensus on clinical diagnosis and treatment of breast cancer with low expression of human epidermal growth factor receptor 2 (2022 edition)]. *Zhonghua Zhong Liu Za Zhi.* 2022;44(12):1288–95.
- Fernandez AI, Liu M, Bellizzi A, Brock J, Fadare O, Hanley K, Harigopal M, Jorns JM, Kuba MG, Ly A, et al. Examination of low ERBB2 protein expression in breast Cancer tissue. *JAMA Oncol.* 2022;8(4):1–4.
- Lambein K, Van Bockstal M, Vandemaële L, Geenen S, Rottiers I, Nuyts A, Matthyss B, Praet M, Denys H, Libbrecht L. Distinguishing score 0 from score

- 1 + in HER2 immunohistochemistry-negative breast cancer: clinical and pathobiological relevance. *Am J Clin Pathol.* 2013;140(4):561–6.
17. Qaiser T, Mukherjee A, Reddy Pb C, Munugoti SD, Tallam V, Pitkaaho T, Lehtimaki T, Naughton T, Berseth M, Pedraza A, et al. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology.* 2018;72(2):227–38.
 18. Gatenbee CD, Baker AM, Prabhakaran S, Swinyard O, Slebos RJC, Mandal G, Mulholland E, Andor N, Marusyk A, Leedham S, et al. Virtual alignment of pathology image series for multi-gigapixel whole slide images. *Nat Commun.* 2023;14(1):4502.
 19. Li Y, Zhang X, Chen DJL. CSRNNet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes. 2018.
 20. Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M, Wang X, et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans Pattern Anal Mach Intell.* 2021;43(10):3349–64.
 21. Prat A, Baselga J. The role of hormonal therapy in the management of hormonal-receptor-positive breast cancer with co-expression of HER2. *Nat Clin Pract Oncol.* 2008;5(9):531–42.
 22. Doi T, Shitara K, Naito Y, Shimomura A, Fujiwara Y, Yonemori K, Shimizu C, Shimoi T, Kuboki Y, Matsubara N, et al. Safety, pharmacokinetics, and antitumour activity of trastuzumab deruxtecan (DS-8201), a HER2-targeting antibody-drug conjugate, in patients with advanced breast and gastric or gastro-oesophageal tumours: a phase 1 dose-escalation study. *Lancet Oncol.* 2017;18(11):1512–22.
 23. Thomson TA, Hayes MM, Spinelli JJ, Hilland E, Sawrenko C, Phillips D, Dupuis B, Parker RL. HER-2/neu in breast cancer: interobserver variability and performance of immunohistochemistry with 4 antibodies compared with fluorescent in situ hybridization. *Mod Pathol.* 2001;14(11):1079–86.
 24. Modi S, Park H, Murthy RK, Iwata H, Tamura K, Tsurutani J, Moreno-Aspitia A, Doi T, Sagara Y, Redfern C, et al. Antitumor Activity and Safety of Trastuzumab Deruxtecan in patients with HER2-Low-expressing advanced breast Cancer: results from a phase Ib study. *J Clin Oncol.* 2020;38(17):1887–96.
 25. Yue M, Zhang J, Wang X, Yan K, Cai L, Tian K, Niu S, Han X, Yu Y, Huang J, et al. Can AI-assisted microscope facilitate breast HER2 interpretation? A multi-institutional ring study. *Virchows Arch.* 2021;479(3):443–9.
 26. Holten-Rossing H, Moller 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.
 27. Wu S, Yue M, Zhang J, Li X, Li Z, Zhang H, Wang X, Han X, Cai L, Shang 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.
 28. Hida AI, Omanovic D, Pedersen L, Oshiro Y, Ogura T, Nomura T, Kurebayashi J, Kanomata N, Moriya T. Automated assessment of Ki-67 in breast cancer: the utility of digital image analysis using virtual triple staining and whole slide imaging. *Histopathology.* 2020;77(3):471–80.
 29. Wu J, Liu C, Liu X, Sun W, Li L, Gao N, Zhang Y, Yang X, Zhang J, Wang H, et al. Artificial intelligence-assisted system for precision diagnosis of PD-L1 expression in non-small cell lung cancer. *Mod Pathol.* 2022;35(3):403–11.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.