

Original Article

Machine learning-based quantitative pathology reveals intratumoral heterogeneity in ductal carcinoma *in situ* and predicts invasive upstaging

Libo Yang^{1,2*}, Jie Chen^{2*}, Chuanfen Lei⁴, Miao Yang⁵, Pengzhou Cai², Jing Fu⁶, Ting Yin^{1,2}, Peilu Gao¹, Hongyu Wang⁷, Zhiwei Tan⁸, Chen Shu², Li Li², Fei Chen², Leyi Gao^{1,2}, Junsheng Lin¹, Yuhao Yi², Zhang Zhang^{1,3}

¹Department of Pathology, West China Hospital of Sichuan University, Chengdu, Sichuan, China; ²Institute of Clinical Pathology, NHC Key Laboratory of Transplant Engineering and Immunology, West China Hospital, Sichuan University, Chengdu, Sichuan, China; ³Laboratory of Breast Pathology and Artificial Intelligence, West China Hospital, Sichuan University, Chengdu, Sichuan, China; ⁴Department of Pathology, West China Fourth Hospital, Sichuan University, Chengdu, Sichuan, China; ⁵Institute of Systems Science, National University of Singapore, Singapore; ⁶Department of Pathology, Sichuan Provincial People's Hospital, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China; ⁷Department of Pathology, Qilu Hospital of Shandong University, Qingdao, Shandong, China; ⁸Department of Pathology, The Second Affiliated Hospital of Chengdu Medical College, China National Nuclear Corporation 416 Hospital, Chengdu, Sichuan, China. *Equal contributors.

Received June 1, 2026; Accepted September 23, 2026; Epub September 25, 2026; Published September 30, 2026

Abstract: This study aimed to develop a machine learning framework for objective ductal carcinoma in situ (DCIS) grading and to investigate intratumoral heterogeneity as a predictive biomarker. The grading framework was developed using 267 DCIS resections and validated on two external test sets (n=116 and n=69). A single-cell grading model graded individual nuclei using quantitative measures of nuclear size and pleomorphism. An integrated grading model then combined single-cell with architectural features in a support vector machine classifier. Separately, an invasion prediction model was trained on 288 intermediate- and high-grade DCIS cases and validated on 54 cases to predict upstaging to invasive carcinoma >1 mm. Compared to low-grade nuclei, high-grade nuclei were significantly larger and less uniformly oriented (all P<0.001). The single-cell grading model achieved F1 scores of 0.76, 0.62, and 0.80 for low-, intermediate-, and high-grade nuclei, respectively. In one external test set, the integrated grading model achieved F1 scores of 0.83, 0.71, and 0.72 for low-, intermediate-, and high-grade DCIS, respectively. On 54 intermediate- and high-grade DCIS biopsy cases, combining clinical variables with quantitative heterogeneity metrics improved the prediction of upstaging to invasive carcinoma >1 mm over the clinical-only model, increasing the area under the receiver operating characteristic curve from 0.704 to 0.827. These findings identify quantitatively measured intratumoral heterogeneity as a biologically relevant feature that may explain diagnostic ambiguity in borderline cases and help predict extensive invasive upstaging.

Keywords: Ductal carcinoma *in situ*, grading, machine learning, intratumoral heterogeneity

Introduction

Breast ductal carcinoma *in situ* (DCIS) maintains a high incidence with a trend toward younger patients [1]. Even after lumpectomy, patients remain at risk of ipsilateral breast events, including invasive and *in situ* recurrences, and breast cancer-related mortality [2]. The mortality also increases during the first decade after diagnosis [3]. Effective risk stratification

is important for optimal DCIS management. Current clinical prediction models, including Van Nuys Prognostic Index [4], Memorial Sloan Kettering Nomogram [5], and prognostic score [6], combine patient age, tumor size, nuclear grade, and additional clinicopathological variables. Of these factors, nuclear grade is unique in requiring microscopic pathological assessment and is a strong prognostic indicator [7], with low-grade cases associated with better

outcomes than high-grade cases. However, interobserver concordance in histological grading remains low, with only 55% agreement in one multicenter study [8]. This variability may be due to discordant interpretation criteria and poorly defined boundaries between grades, compounded by intratumoral nuclear heterogeneity.

Intratumoral heterogeneity is a recognized feature of DCIS and reflects divergent clonal evolution [9]. Allred et al. found mixtures of cells with different nuclear grades within individual lesions, with 9.2% DCIS cases showing different distribution of low-, intermediate-, and high-grade [10]. Because manual assessment is subjective, the true heterogeneity prevalence may be higher. Hayward et al. reported greater nuclear diversity in high-grade than low-grade DCIS using nuclear size measurements [11]. This approach did not incorporate other critical histologic features for grading, such as necrosis, chromatin patterns, and nucleolar prominence.

In this study, we developed and validated machine learning algorithms for objective and quantitative assessment of DCIS grade. The framework combines single-cell morphology with architectural features to generate a comprehensive case-level classification. We then applied the same quantitative framework to measure intratumoral heterogeneity and assess whether heterogeneity in biopsy specimens predicts invasive carcinoma in the subsequent excision. This approach aims to improve the reproducibility of DCIS grading and provide a quantitative basis for risk stratification. The entire process of this study is shown in **Figure 1**.

Materials and methods

Patient selection

This retrospective study was conducted in accordance with the Declaration of Helsinki and received approval from the Institutional Review Board of West China Hospital (IRS No. 2025-1389). All patient data were fully anonymized before analysis, and the requirement for informed consent was waived. We identified surgical resection specimens diagnosed as DCIS at West China Hospital between January 2015 and December 2021. Three experienced pathologists independently reviewed each case. A total of 267 cases with concordant

grade assignments comprised training set A. Of these, 50 were used for single-cell feature extraction and development of single-cell grading model. The remaining 217 cases were used to develop and internally validate the integrated grading model.

Two independent cohorts were included for external validation. External test set A comprised 116 DCIS biopsy cases diagnosed at West China Hospital between January 2015 and December 2022 for which the pathologists agreed on the grade. External test set B consisted of 69 concordantly graded DCIS biopsy cases from Sichuan Provincial People's Hospital collected between December 2022 and December 2024.

To develop a model for predicting upstaging to invasive carcinoma, we assembled training set B, which consisted of 288 intermediate- and high-grade DCIS biopsy cases from West China Hospital between January 2015 and December 2022 and for which complete surgical excision data were available. External test set C comprised 54 intermediate- and high-grade DCIS biopsy cases diagnosed at West China Hospital between January 2023 and December 2023 each with corresponding excision data.

For all enrolled cases, one representative hematoxylin and eosin (H&E)-stained slide was digitized as a whole-slide image (WSI) for computational analysis. Slides were scanned using a KFBio scanner (KF-PRO-400-HI, Konfoong Bioinformation Tech Co. Ltd., Zhejiang, China). The case-selection process is shown in [Supplementary Figure 1](#).

Image analysis

Overview: The computational pathology pipeline comprised three stages: nuclear segmentation, single-cell feature extraction, and architectural feature extraction. To account for differences in magnification among slides and scanners, image measurements were calibrated using the diameter of annotated erythrocytes as an internal reference.

Nuclei were segmented only within manually annotated neoplastic ductal structures. HoverNet model [12] was used to identify cells in the annotated DCIS ducts, and only cells classified as neoplastic were retained for subsequent model development. The segmentation model had been pretrained and validated using the

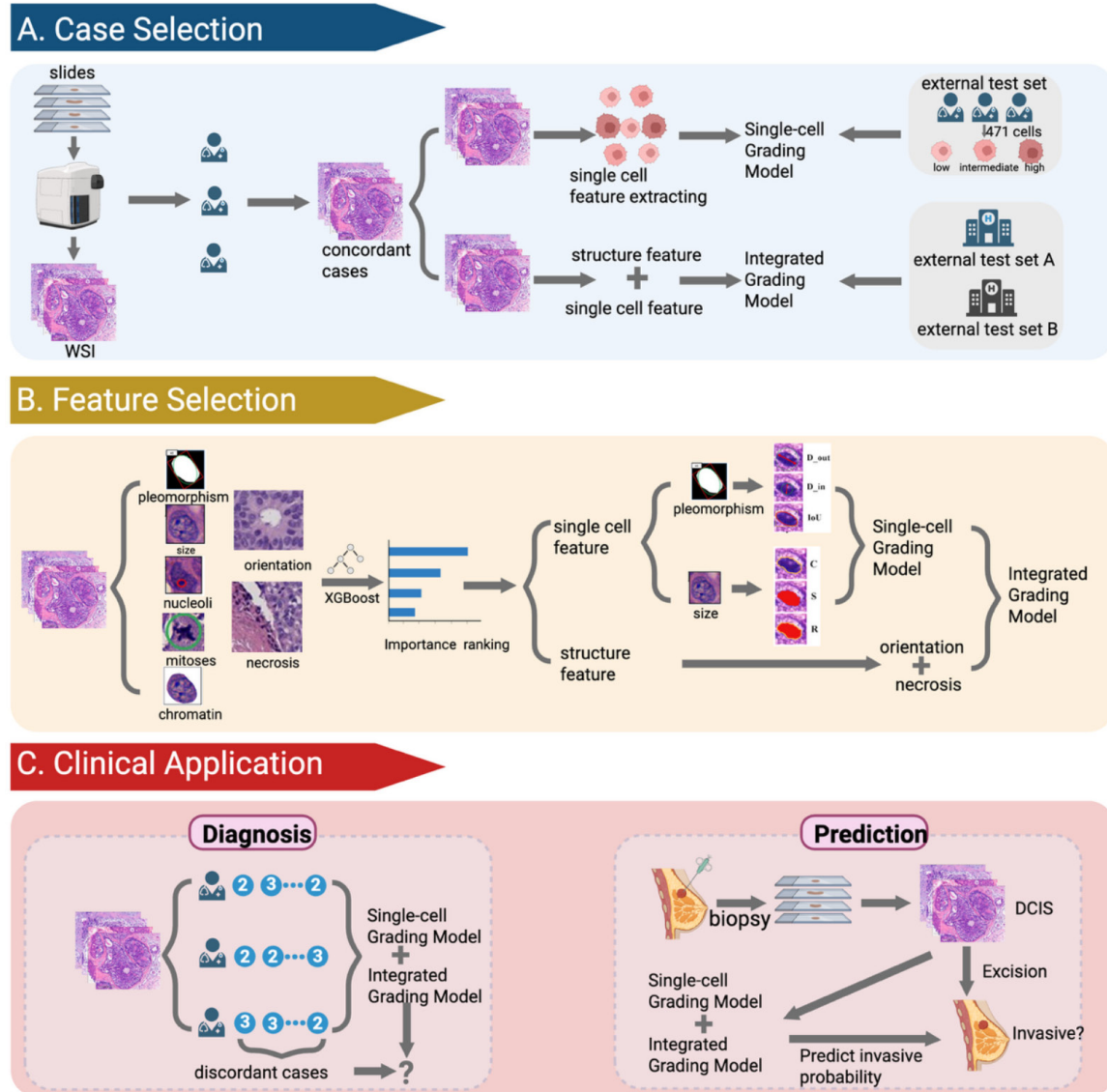


Figure 1. The entire process of this study.

publicly available PanNuke dataset (<https://jgamper.github.io/PanNukeDataset/>).

Single-cell feature extraction: According to College of American Pathologists (CAP) recommendations for DCIS grading, we extracted a comprehensive set of quantitative features corresponding to nuclear grade is evaluated by pleomorphism, size, chromatin, nucleoli, mitoses, and orientation. Candidate single-cell features for pleomorphism, size, chromatin, nucleoli and mitotic activity were calculated as follows. Segmentation accuracy was assessed using the ellipse-overlap ratio (intersection over union [IoU]) between the nuclear contour and a fitted ellipse. Nuclear pleomorphism was

represented by two measures of shape regularity: inscribed circle diameter (D_{in}), defined as the diameter of the largest circle contained within the nuclear contour; and the circumscribed-circle diameter (D_{out}), defined as the diameter of the smallest circle enclosing the contour (Supplementary Figure 2). Nuclear size is described by circumference (C), calculated as the sum of Euclidean distances between adjacent contour points; surface area (S), calculated from the contour coordinates using Green's theorem; and the area-to-squared-circumference ratio (R). Chromatin feature was quantified as nuclear roughness, defined as the standard deviation of pixel intensities within each nucleus: $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$ (Supple-

mentary Figure 3). Nucleoli were detected within segmented nuclei using the Hough transform. The nucleoli feature of each case was calculated by the ratio of cells with detected nucleoli to total number of segmented nuclei (Supplementary Figure 4). Mitotic figures were detected in each region using Tencent MiYing (version 1.0; <https://tencentmiying.com/official/product/pathology>) (Supplementary Figure 5), and the total count was converted to mitoses per 10 high-power fields.

Architectural features extraction: In addition to single-cell features, we quantified necrosis and cell orientation within each annotated DCIS duct. The CAP recommendations discuss necrosis in relation to mammographic calcifications, although necrosis is not itself a nuclear-grade criterion. Necrosis of each case was labeled by pathologists on the WSI. Cell orientation was determined by using principal component analysis to identify the principal axis of each cell contour and then calculating orientation entropy from the spatial relationships among cells, as implemented in the Cell Orientation Entropy (COE) method [13].

Grading model development and validation: All candidate features were put into statistical confusion matrices by machine learning model. XGBoost feature importance was ranked by the number of times each feature was used to split samples across the trees. Nuclear size, nuclear shape regularity, necrosis, and cell orientation were the four most informative features. Necrosis and cell orientation were therefore retained as the principal architectural features.

The selected features fell into two categories: single-cell morphological features (nuclear size and nuclear shape regularity) and architectural features (necrosis and cell orientation). We used them to build two classifiers. The single-cell grading model relied exclusively on morphological features to grade individual cells, whereas the integrated grading model combined aggregated single-cell features with architectural features to classify the lesion as a whole.

The single-cell grading model used a stacked ensemble in which XGBoost and random forest classifiers served as base learners and logistic regression as the meta-learner. XGBoost was

trained with a multiclass softmax objective. The random forest used the Gini impurity criterion and balanced class weights, with individual trees constructed from bootstrap samples and randomly selected feature subsets. Predicted class probabilities from the two base learners were supplied as meta-features to the logistic regression model. Hyperparameters were selected by grid search with threefold cross-validation.

The integrated grading model used a support vector machine (SVM) classifier with a radial basis function kernel to accommodate nonlinear class boundaries. The classifier identified the hyperplane that maximized the margin between classes. Grid search with fivefold cross-validation was used to optimize the regularization parameter (C), kernel coefficient (γ), and class weights. **Figure 2** shows the complete nuclear modeling pipeline.

Development of the Invasion prediction model: We next examined whether nuclear features from the DCIS core-needle biopsy specimens could predict extensive invasive carcinoma measuring >1 mm in the subsequent excision. A separate prediction model was developed from biopsy cohorts (training set B and external test set C). For each biopsy, we collected clinical features, including age, tumor size, ER status, HER2 IHC score, and tumor grade, and quantitative pathological features from the image-analysis pipeline. The latter included the proportions of cells classified by the single-cell model as low, intermediate, or high grade, together with necrosis and cell orientation. Using Training Set B, we compared a clinical-only model with a fused model combining both clinical and quantitative pathological features. Logistic regression and XGBoost were evaluated for each feature set, with fivefold cross-validation used for training and optimization. The best-performing models were then independently assessed in the external test set C. The primary endpoint was extensive invasive carcinoma in the excision.

Statistical analysis

Grading performance was assessed using accuracy $[(TP+TN)/(TP+TN+FP+FN)]$, precision $[TP/(TP+FP)]$, recall $[TP/(TP+FN)]$, and F1-score $[2TP/(2TP+FP+FN)]$, where TP, TN, FP, and FN represent true positives, true negatives, false

Nuclear grading in DCIS

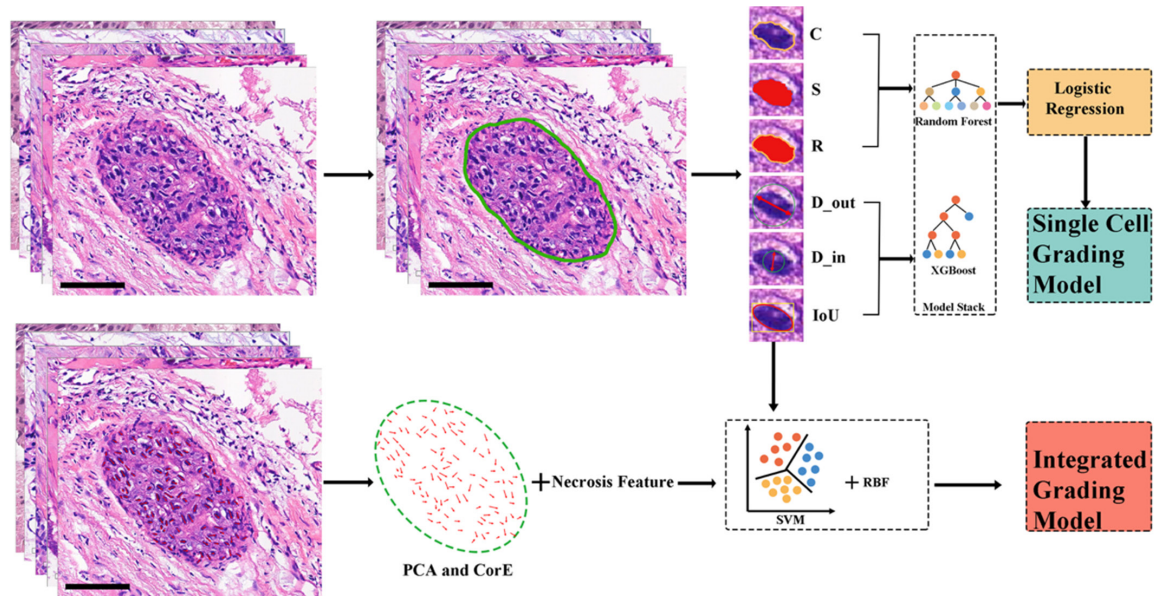


Figure 2. Illustration of model building. Abbreviations: C, circumference; S, surface area; R, ratio of the area to the square of the circumference; IoU, merging ratio; D_in, inscribed circle diameter; D_out, outscribed circle diameter; CORe, Cell Orientation Entropy; SVM, Support vector machine; RBF, radial basis function; PCA, Principal Component Analysis.

positives, and false negatives respectively. For continuous variables compared across multiple groups, analysis of variance (ANOVA) was used to test for differences in group means, followed by the least significant difference (LSD) procedure for pairwise post hoc comparisons. ANOVA and LSD analyses were performed in SPSS, version 16.0 (SPSS Inc., Chicago, IL, USA). Receiver operating characteristic curves for the prediction models were generated and compared using a paired DeLong test in the *pROC* package. To assess the quality of the probabilistic predictions of binary events, Brier score was calculated by *DescTools* package. All packages were run in RStudio, version 1.4.1103 (RStudio, Inc., Boston, MA, USA).

Results

Quantitative single-cell signatures in DCIS cases

A total of 19,800 neoplastic nuclei were segmented and analyzed from the 50 surgical resection cases in the single-cell training set, with 6600 nuclei from each of low-, intermediate-, and high-grade categories. Two groups of features (nuclear size and shape regularity) were modeled as described above. Low-grade

nuclei had significantly smaller circumference and surface areas than both intermediate- and high-grade nuclei (**Figure 3A, 3B**; both $P < 0.001$). Conversely, the compactness ratio (area/perimeter²) was significantly lower in high-grade than in low-grade nuclei (**Figure 3C, 3D**; $P < 0.001$). The combination of inscribed- and circumscribed-circle diameters also distinguished low- from high-grade cells (**Figure 3F**).

The stacked ensemble model was optimized using grid search with threefold cross-validation and then evaluated in an independent test set of 471 nuclei annotated by pathologists. Against these reference annotations, the model achieved F1 scores of 0.76, 0.62, and 0.80 for low-, intermediate-, and high-grade cells, respectively.

Integrated grading model

For case-level grading, we combined aggregated single-cell measurements with necrosis and cell orientation entropy in 217 resection specimens. These cases were divided into an internal training set ($n=80$), validation set ($n=21$), and test set ($n=116$) (**Supplementary Table 1**). High-grade lesions showed a loss of cell orientation (**Supplementary Figure 6**; $P < 0.001$). An SVM classifier was trained on the combined

Nuclear grading in DCIS

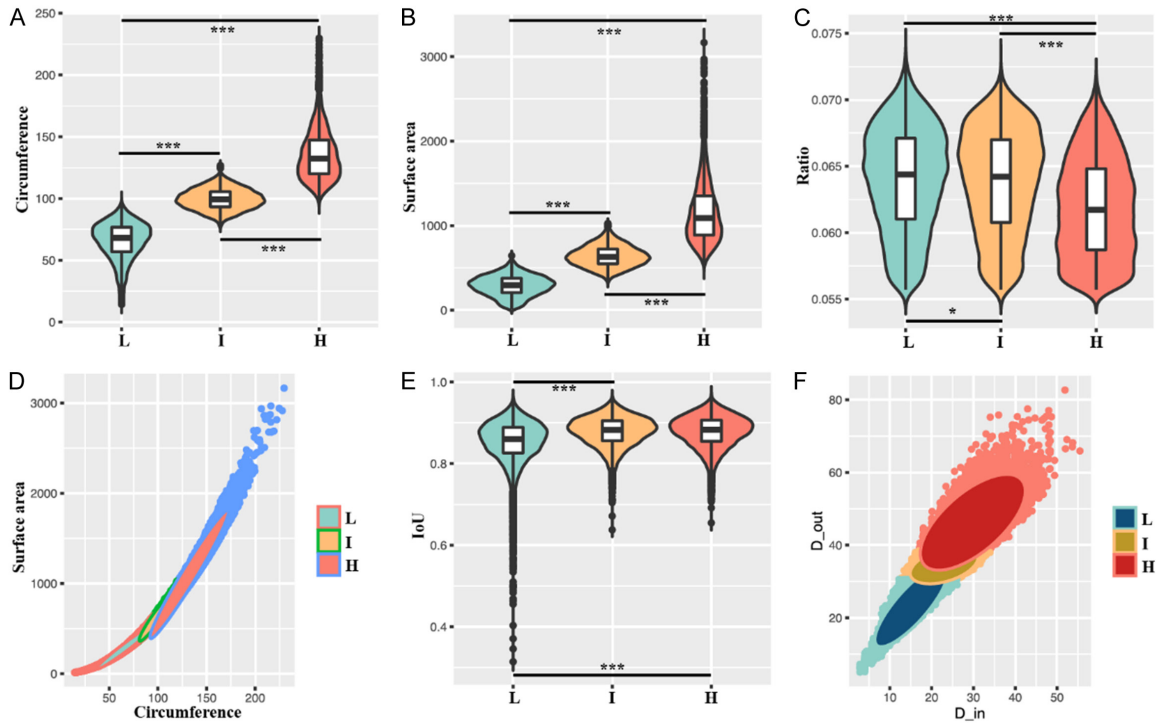


Figure 3. Single-cell features of low-, intermediate-, and high-grade cells. A. Circumference; B. Surface area; C. Ratio of the surface area to the square of the circumference; D. Scatter plot with circumference on the X-axis and surface area on the Y-axis; E. Merging ratio; F. Scatter plot with inscribed circle diameter (D_{in}) on the X-axis and outscribed circle diameter (D_{out}) on the Y-axis. Abbreviations: L, low-grade cells; I, intermediate-grade cells; H, high-grade cells; IoU, merging ratio; D_{in} , inscribed circle diameter; D_{out} , outscribed circle diameter. * $P < 0.050$; ** $P < 0.010$; *** $P < 0.001$. Only comparisons with statistically significant P values are labeled in the figure.

single-cell and whole-duct features (Supplementary Table 2). In the internal test set, F1 scores were 0.79, 0.82 and 0.76 for low-, intermediate-, and high-grade DCIS, respectively. The corresponding scores were 0.63, 0.74, and 0.82 in external test set A and 0.83, 0.71, and 0.72 in external test set B.

For discrimination of high-grade from low-/intermediate-grade DCIS, the area under the curve (AUCs) were 0.874 in internal test set, 0.836 in external test set A, and 0.811 in external test set B (Supplementary Figure 7). For discrimination of low-grade from intermediate-/high-grade DCIS, the corresponding AUCs were 0.874 in internal test set, 0.786 in external test set A, and 0.901 in external test set B, respectively (Supplementary Figure 7).

Quantitative analysis of intratumoral heterogeneity

Using the validated single-cell grading model, we quantified the proportions of low-, intermediate-, and high-grade cells in each case and

compared them with the grades assigned by pathologists and by the model (Figures 4 and 5). All three cell grades occurred in varying proportions across lesions of each case-level grade (Figure 5). Among cases with discordant pathologist- and model-assigned grades, low-grade cases contained a high proportion of low-grade cells (mean $\pm$ SD: $66.8\% \pm 11.9\%$; range: 48.2%-86.4%), whereas high-grade cases were enriched for high-grade cells (mean $\pm$ SD: $56.6\% \pm 10.7\%$; range 25.2%-79.8%). Intermediate-grade cases contained $44.8\% \pm 10.0\%$ low-grade cells, significantly fewer than in low-grade cases ($P < 0.001$), and $35.8\% \pm 11.5\%$ high-grade cells, significantly fewer than in high-grade cases ($P < 0.001$).

Analysis of cases with discordant pathologist assessments

We then investigated whether the models could provide an objective classification when pathologists disagreed. The three pathologists failed to reach a consensus grade in 36 resection cases. All disagreements involved adjacent cat-

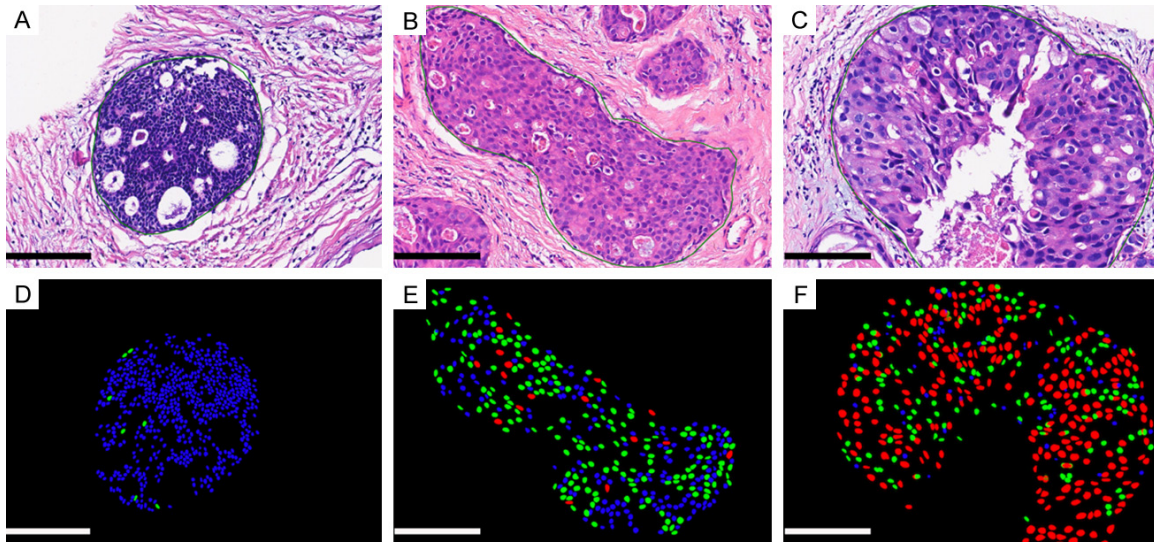


Figure 4. Single-cell grading annotation in each labeled case. (A) low-grade DCIS; (B) intermediate-grade DCIS; (C) high-grade DCIS. (D) single-cell grading annotation in low-grade DCIS (A); (E) single-cell grading annotation in intermediate-grade DCIS (B); (F) single-cell grading annotation in high-grade DCIS (C). (A-F): The scale bar in the figure represents 125 μ m. (D-F): Blues dots refer to low-grade cells, green dots refer to intermediate-grade cells, red dots refer to high-grade cells.

egories: low versus intermediate grade ($n=17$) or intermediate versus high grade ($n=19$). The integrated case-level model classified 3 of these cases as low grade, 16 as intermediate grade, and 17 as high grade (**Figure 6B**). To examine the cellular basis of these classifications, we analyzed the single-cell composition of each lesion (**Figure 6A**). In the 17 cases disputed between low and intermediate grade, the mean proportion of low-grade nuclei laid between the corresponding means for consensus low- and consensus intermediate-grade cases in the external validation cohorts (**Figure 6D**). Similarly, in the 19 cases disputed between intermediate and high grade, the mean proportion of high-grade nuclei laid between the means for consensus intermediate- and consensus high-grade cases (**Figure 6F**).

Intratumoral heterogeneity in biopsy specimens predicts extensive invasion in the subsequent excision

We assessed whether quantitative features derived from biopsy specimens could predict invasive carcinoma measuring >1 mm in the subsequent excision. The clinical information of enrolled DCIS biopsy cases is presented in [Supplementary Table 3](#). In training set B, 90 of 288 cases (31.3%) were upstaged to extensive

invasive carcinoma in the excision specimen. In external test set C, 18 of 54 cases (33.3%) were upstaged. Logistic regression and XGBoost were used to develop a clinical-only model and a fused model that combined clinical variables with single-cell features.

In the training set, the clinical-only model showed modest discrimination, with an AUC of 0.657, sensitivity of 0.571, and specificity of 0.689. The fused model performed better on the same data, with an AUC of 0.976 (**Figure 7A**), significantly higher than that of clinical-only model ($P<0.001$). Its sensitivity and specificity were 1.000 and 0.911, respectively. The Brier score was 0.234 for the clinical-only model and 0.093 for the fused model.

Performance was then assessed in the independent external test set C. The clinical-only model yielded an AUC of 0.704, a sensitivity of 0.556, and a specificity of 0.722. The fused model achieved an AUC of 0.827, a sensitivity of 0.861, and a specificity of 0.833 (**Figure 7B**). Although the difference in AUC between the two models was not significant ($P=0.133$), the fused model showed numerically higher sensitivity and specificity (**Figure 7B**). The Brier score was 0.228 for the clinical-only model and 0.168 for the fused model.

Nuclear grading in DCIS

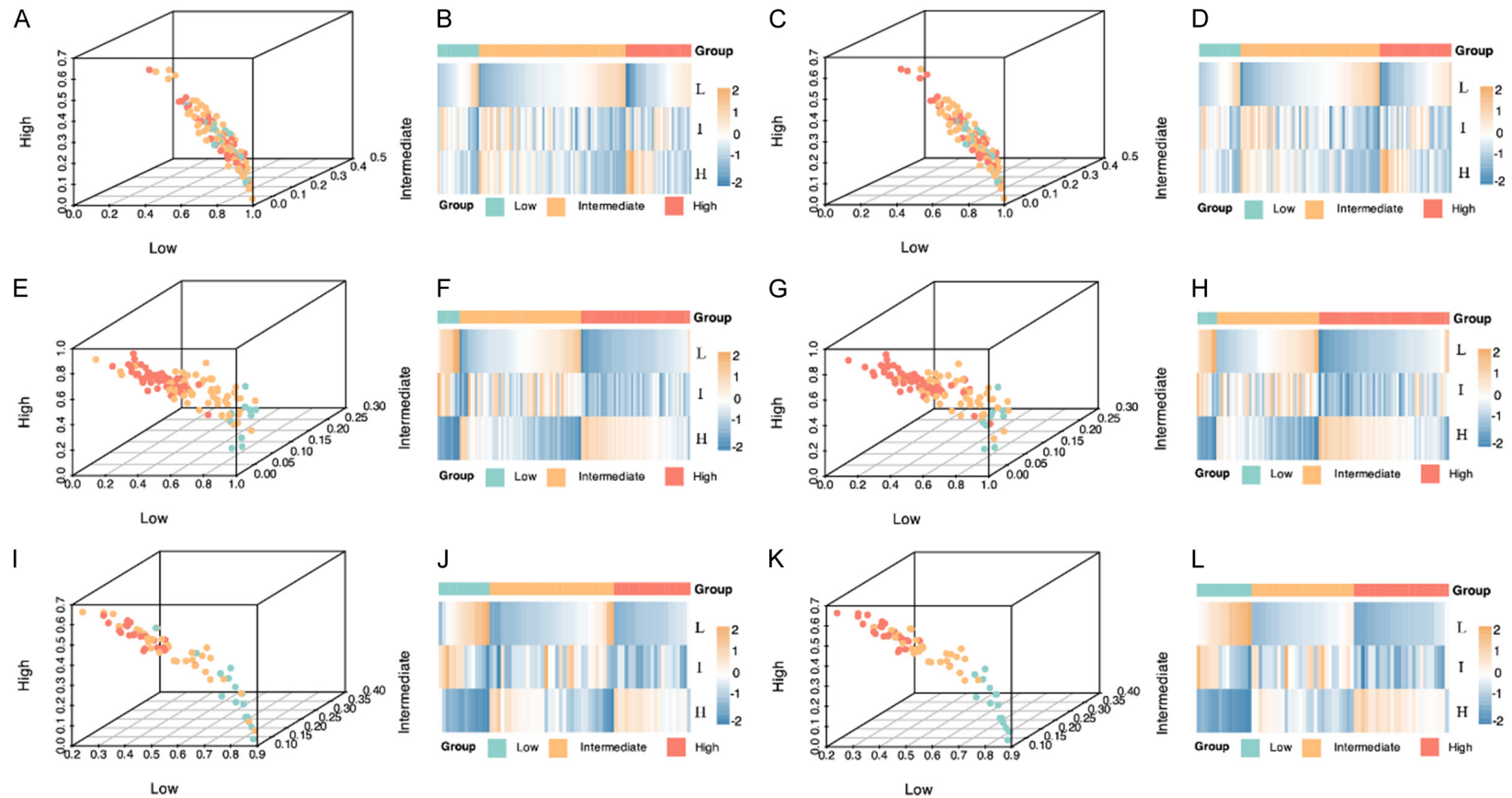


Figure 5. Distribution of low-, intermediate-, and high-grade cells in different sets. (A-D) internal test set; (E-H) external test set A; (I-L) external test set B. (A, E and I) 3D scatter plot shows distribution of low-, intermediate-, and high-grade cells in cases graded by pathologists; (B, F and J) heatmap shows distribution of low-, intermediate-, and high-grade cells in cases graded by the integrated grading model; (C, G and K) 3D scatter plot shows distribution of low-, intermediate-, and high-grade cells in cases graded by the integrated grading model. (A, C, E, G, I and K): x-axis refers to percentage of low-grade cells in each case, y-axis refers to percentage of intermediate-grade cells in each case, and z-axis refers to percentage of high-grade cells in each case; green spot refers to low-grade cases, yellow spot refers to intermediate-grade cases, red spot refers to high-grade cases. Abbreviations: L, low-grade cells; I, intermediate-grade cells; H, high-grade cells.

Nuclear grading in DCIS

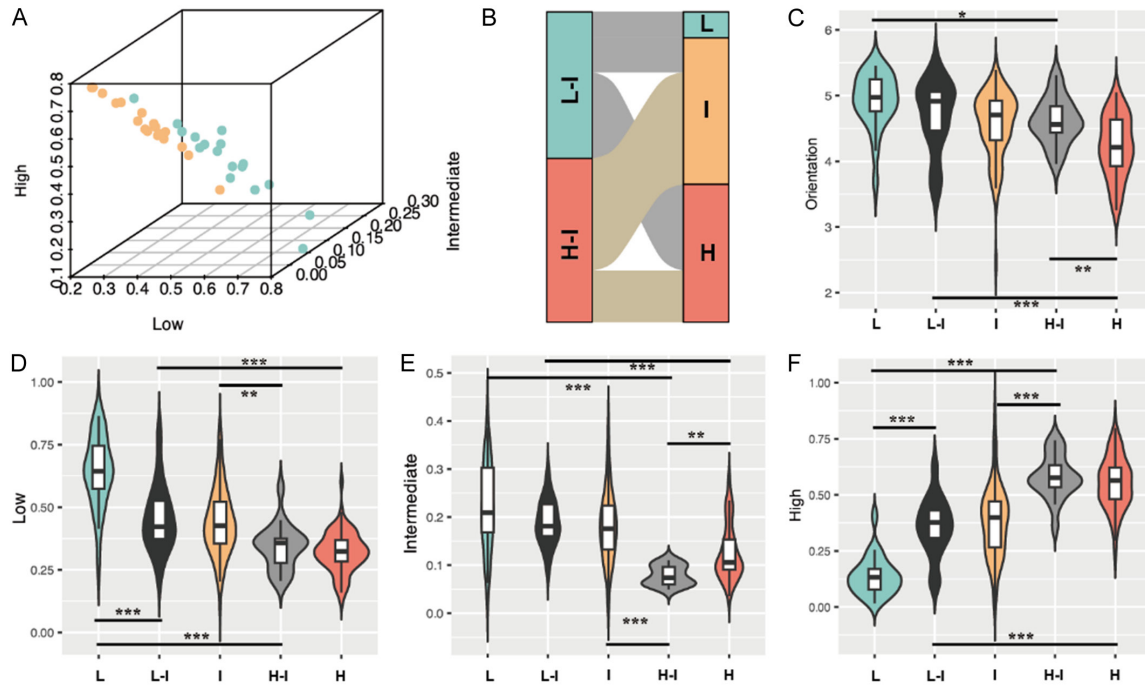


Figure 6. Intratumoral heterogeneity analyses in discordant cases. A. 3D scatter plot showing the distribution of low-, intermediate-, and high-grade cells in cases with discordant pathologist grades, categorized into low versus intermediate (green spots) or intermediate versus high (yellow spots) grades. B. A Sankey diagram shows the discrepancies in case grading by the integrated grading model. The left side shows the pathologist-assigned grades, while the right side shows the model-assigned grades. C. Orientation score comparison between discordant and concordant cases. D. Percentage of low-grade cells in discordant versus concordant cases. E. Percentage of intermediate-grade cells in discordant versus concordant cases. F. Percentage of high-grade cells in discordant versus concordant cases. L, I and H refers to low, intermediate, and high-grade cases in consistent cases, respectively; L-I and H-I refers to discordant results between low and intermediate and between high and intermediate, respectively. *P<0.050; **P<0.010; ***P<0.001. Only comparisons with statistically significant P values are labeled in the figure.

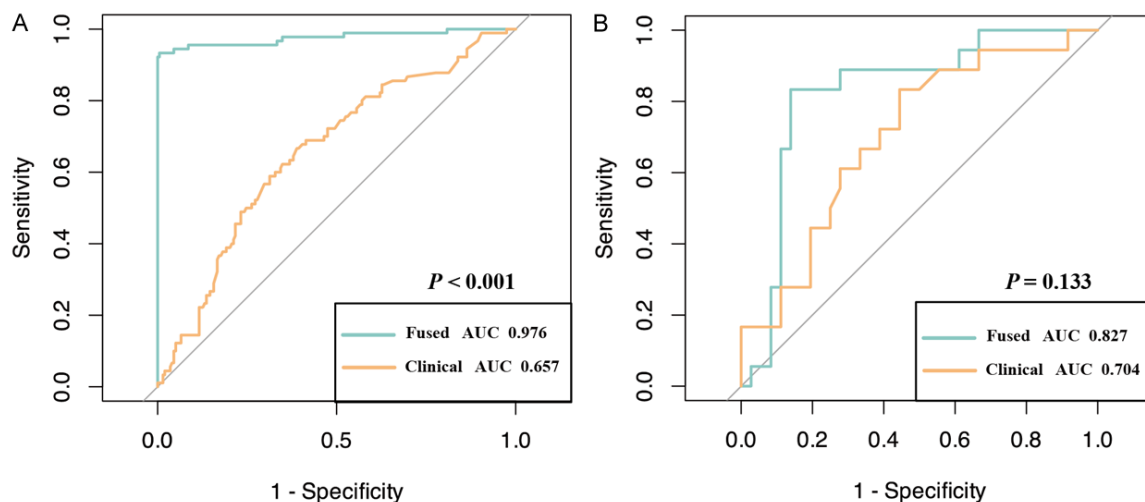


Figure 7. ROC curves of extensive invasion prediction models. A. ROC curves of Clinical model and fused model in training set B. ROC curves of Clinical model and fused model in external test set C. ROC, receiver operating characteristic; AUC, area under the ROC curve.

Discussion

We developed and validated a multilevel machine learning framework to address the persistent problem of interobserver variability in DCIS grading. Quantitative single-cell morphometry distinguished low- from high-grade nuclei and provided an objective cellular basis for grading. Combining these measurements with architectural features enabled accurate, reproducible classification of DCIS as low, intermediate, or high grade across internal and external cohorts. The single-cell model also quantified intratumoral heterogeneity, helping to explain why pathologists disagree on borderline lesions. Finally, heterogeneity measured in biopsy specimens was informative for predicting upstaging to invasive carcinoma in the subsequent excision.

Several institutions have proposed systems for grading DCIS, including recommendations from the CAP [14], Van Nuys [15], and Holland [16]. Nuclear size and pleomorphism are central to all of these systems. Nuclear size is the only grading feature that is readily measured objectively, yet current guidelines do not specify quantitative thresholds for low-, intermediate-, or high-grade nuclei [17]. The remaining criteria are also interpreted subjectively, limiting consistency among pathologists and institutions.

The machine learning approach was designed to operationalize the grading process described in the CAP recommendations [14]. To establish a reliable reference standard, the training cohorts included only cases for which three pathologists independently assigned the same grade. Segmentation and analysis at the level of individual nuclei enabled a granular description of DCIS, while erythrocytes provided an internal scale for reproducible size measurements. Nuclear size differed significantly across grade categories, consistent with findings reported by Hayward et al. [11], and increased progressively from low- to high-grade disease. Pleomorphism also distinguished the grades: low-grade cells had regular nuclear contours, whereas high-grade cells had larger, more irregular nuclei (**Figure 3F**). Although IoU values varied among low-grade cells, the mean IoU values exceeded 0.8 across all three nuclear-grade groups (**Figure 3E**), indicating that the nuclear contours were generally similar to their fitted ellipses. These observations agree with the

established role of pleomorphism in grading [18, 19] and with descriptions of large, irregular nuclei in high-grade DCIS [20]. The model translates these qualitative descriptors into quantitative features.

Guidelines also consider chromatin pattern, nucleolar prominence, and mitotic activity when grading DCIS. In our preliminary feature-importance analysis, however, these variables contributed substantially less to grade discrimination than nuclear size and pleomorphism. One likely explanation is their sensitivity to pre-analytical variation, including tissue processing, fixation time, and staining inconsistency [21]. This can lead to less reliable feature extraction in a computational setting. We therefore excluded them from the final single-cell grading model to improve robustness and generalizability. In a rigorously annotated nuclear test set, the model achieved F1 scores of 0.76 for low-grade and 0.80 for high-grade cells. Performance was lower for intermediate-grade cells (F1=0.62), possibly because this category lacks clear biological and morphological boundaries. This finding underscores the value of quantitative approaches for difficult, intermediate cases.

The integrated model extended the single-cell analysis by incorporating necrosis [16, 18, 19] and cell orientation [14, 18, 19, 22]. Cellular orientation, also termed polarity, describes how cells are arranged relative to the ductal lumen. Several grading recommendations [14, 18, 19, 22] describe preserved polarity as a feature of low-grade DCIS and loss of polarity as a feature of high-grade disease, but the thresholds separating categories remain subjective. Li et al. [23] employed the CORe method to quantify cell orientation and found an association with DCIS Oncotype DX scores. Our CORe analysis likewise identified cell orientation as an informative grading feature: low-grade lesions had higher CORe values and were clearly separated from intermediate- and high-grade cases ([Supplementary Figure 6](#)). The integrated model performed consistently in internal and external test sets. By combining cellular and architectural information, it more closely approximates the lesion-level assessment made by a pathologist.

The single-cell model also enabled direct measurement of intratumoral heterogeneity, which

may reflect the distinct molecular routes through which DCIS develops. Low-grade DCIS arising from normal ductal epithelium is associated with chromosomal alterations such as 1q gain and 16q loss and with hormone receptor-positive disease [24]. High-grade DCIS, by contrast, is associated with changes such as 1p loss and 8q gain and with hormone receptor-negative disease [24]. The coexistence of different nuclear grades within one lesion may therefore reflect distinct cells of origin or divergent evolutionary way toward invasion [25]. Our model captures this variation as grade-specific cell proportions at the duct level. How such heterogeneity should influence diagnosis and whether grading should follow the highest-grade focus or the predominant population remains unsettled. Allred et al. [10] reported approximately half of DCIS cases exhibit varying degrees of grade heterogeneity, but manual quantification is challenging. The present method offers a digital-pathology approach to studying this feature and complements techniques such as multiplex immunofluorescence [26].

In our study, the cell-grade distributions by single-cell grading model supported the case-level classifications made by pathologists: low-grade lesions were enriched for low-grade nuclei, whereas high-grade lesions were enriched for high-grade nuclei. Intermediate-grade lesions had a more mixed composition, with substantial low- and high-grade populations but no component as dominant as in the corresponding low- or high-grade lesions. This method offers a quantitative explanation for the poor interobserver agreement in intermediate-grade DCIS. The analysis of disputed cases reached a similar conclusion. Cases debated as low versus intermediate grade had cell distributions between those of consensus low- and intermediate-grade lesions (**Figure 6D-F**). Similarly, cases debated as intermediate versus high grade showed an intermediate degree of orientation loss between those of consensus intermediate and high-grade lesions (**Figure 6C**). Therefore, diagnostic disagreement may reflect genuine biological and morphological continuity rather than observer subjectivity alone. The machine learning framework provides an objective way to locate individual lesions along that continuous spectrum.

Underestimation of invasion in DCIS diagnosed by core-needle biopsy remains a clinical chal-

lenge. Approximately 10% to 50% biopsy-diagnosed DCIS cases are upstaged to invasive carcinoma after excision [27-29], with higher rates reported for high-grade disease [30]. Although the clinical implications of microinvasion (≤ 1 mm) versus pure DCIS remain debated due to minimal prognostic differences [31, 32]. The presence of extensive invasive carcinoma is associated with poorer outcomes [33] and requires a different therapeutic strategy. We therefore focused on predicting invasive carcinoma >1 mm in biopsy specimens classified as intermediate- or high-grade DCIS. In the external test set, the clinical-only model had modest predictive power (AUC, 0.704). Adding the proportions of low-, intermediate-, and high-grade cells in the model increased the AUC to 0.827. This finding suggests that the quantitative heterogeneity in the initial biopsy tissue contains information about invasive disease in the subsequent excision. This application of the grading model could refine preoperative risk stratification and inform surgical planning for patients with DCIS.

Conclusions

We developed and validated a machine learning framework for objective and reproducible grading of DCIS. The framework translates qualitative histopathological criteria into quantitative measurements and achieved robust performance across independent cohorts. The model is based on a simplified version of CAP criteria and gives an objective quantification method for grading. It also identifies intratumoral heterogeneity as a biologically relevant feature that helps explain diagnostic uncertainty in borderline lesions and may predict upstaging to extensive invasive carcinoma (>1 mm). Given the risk of progression to invasive disease [34] and the increasing emphasis on precise management of low-risk DCIS [35], this approach may improve risk stratification and support clinical decision-making.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No. 82401894).

Disclosure of conflict of interest

None.

Abbreviations

DCIS, ductal carcinoma *in situ*; H&E, hematoxylin and eosin; WSI, whole-slide image; CAP, College of American Pathologists; CORE, Cell Orientation Entropy; SVM, support vector machine; RBF, radial basis function; AUC, Area Under the Curve.

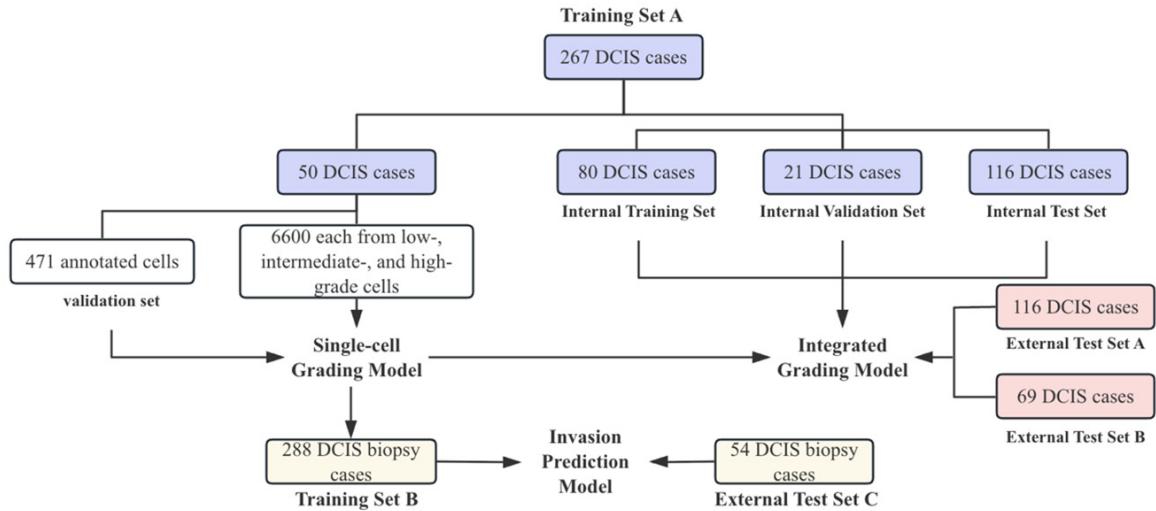
Address correspondence to: Dr. Zhang Zhang, Department of Pathology, West China Hospital of Sichuan University, Guo Xue Xiang No. 37, Chengdu, Sichuan, China. E-mail: zhangzhang@scu.edu.cn; Dr. Yuhao Yi, Institute of Clinical Pathology, NHC Key Laboratory of Transplant Engineering and Immunology, West China Hospital, Sichuan University, Guo Xue Xiang No. 37, Chengdu, Sichuan, China. E-mail: yuhaoyi@scu.edu.cn

References

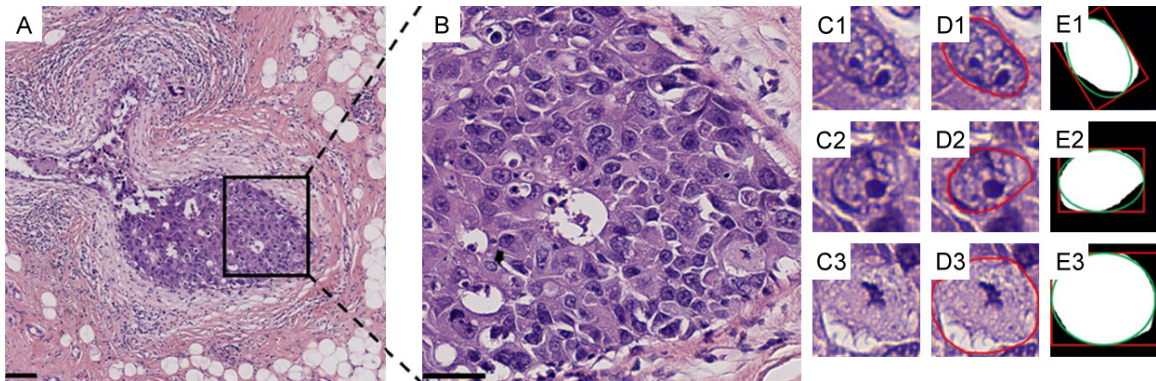
- [1] Ryser MD, Hendrix LH, Worni M, Liu Y, Hyslop T and Hwang ES. Incidence of ductal carcinoma in situ in the United States, 2000-2014. *Cancer Epidemiol Biomarkers Prev* 2019; 28: 1316-1323.
- [2] Yang L, Lu D, Lai Y, Shen M, Yu Q, Lei T, Pu T and Bu H. Prognostic score-based stratification analysis reveals universal benefits of radiotherapy on lowering the risk of ipsilateral breast event for ductal carcinoma in situ patients with different risk levels. *Ann Surg Oncol* 2021; 28: 975-984.
- [3] Giannakeas V, Sopik V and Narod SA. Association of a diagnosis of ductal carcinoma in situ with death from breast cancer. *JAMA Netw Open* 2020; 3: e2017124.
- [4] Boland GP, Chan KC, Knox WF, Roberts SA and Bundred NJ. Value of the Van Nuys Prognostic Index in prediction of recurrence of ductal carcinoma in situ after breast-conserving surgery. *Br J Surg* 2003; 90: 426-432.
- [5] Rudloff U, Jacks LM, Goldberg JI, Wynveen CA, Brogi E, Patil S and Van Zee KJ. Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ. *J Clin Oncol* 2010; 28: 3762-3769.
- [6] Smith GL, Smith BD and Haffty BG. Rationalization and regionalization of treatment for ductal carcinoma in situ of the breast. *Int J Radiat Oncol Biol Phys* 2006; 65: 1397-1403.
- [7] Sagara Y, Freedman RA, Vaz-Luis I, Mallory MA, Wong SM, Aydogan F, DeSantis S, Barry WT and Golshan M. Patient prognostic score and associations with survival improvement offered by radiotherapy after breast-conserving surgery for ductal carcinoma in situ: a population-based longitudinal cohort study. *J Clin Oncol* 2016; 34: 1190-1196.
- [8] Thomas J, Hanby A, Pinder S, Pirrie S, Rea D, Gaunt C, Young J and Francis A. Abstract P3-17-06: LORIS trial of active monitoring for DCIS: how does the online pathology eligibility review process work? *Cancer Research* 2017; 77: P3-17-06-P13-17-06.
- [9] Wilson GM, Dinh P, Pathmanathan N and Graham JD. Ductal carcinoma in situ: molecular changes accompanying disease progression. *J Mammary Gland Biol Neoplasia* 2022; 27: 101-131.
- [10] Allred DC, Wu Y, Mao S, Nagtegaal ID, Lee S, Perou CM, Mohsin SK, O'Connell P, Tsimelzon A and Medina D. Ductal carcinoma in situ and the emergence of diversity during breast cancer evolution. *Clin Cancer Res* 2008; 14: 370-378.
- [11] Hayward MK, Louise Jones J, Hall A, King L, Ironside AJ, Nelson AC, Shelley Hwang E and Weaver VM. Derivation of a nuclear heterogeneity image index to grade DCIS. *Comput Struct Biotechnol J* 2020; 18: 4063-4070.
- [12] Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT and Rajpoot N. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* 2019; 58: 101563.
- [13] Lee G, Ali S, Veltri R, Epstein JI, Christudass C and Madabhushi A. Cell orientation entropy (CORE): predicting biochemical recurrence from prostate cancer tissue microarrays. *Med Image Comput Comput Assist Interv* 2013; 16: 396-403.
- [14] Lester SC, Bose S, Chen YY, Connolly JL, de Baca ME, Fitzgibbons PL, Hayes DF, Kleer C, O'Malley FP, Page DL, Smith BL, Weaver DL and Winer E; Members of the Cancer Committee, College of American Pathologists. Protocol for the examination of specimens from patients with ductal carcinoma in situ of the breast. *Arch Pathol Lab Med* 2009; 133: 15-25.
- [15] Silverstein MJ, Poller DN, Waisman JR, Colburn WJ, Barth A, Gierson ED, Lewinsky B, Gama-gami P and Slamon DJ. Prognostic classification of breast ductal carcinoma-in-situ. *Lancet* 1995; 345: 1154-1157.
- [16] Holland R, Peterse JL, Millis RR, Eusebi V, Faverly D, van de Vijver MJ and Zafrani B. Ductal carcinoma in situ: a proposal for a new classification. *Semin Diagn Pathol* 1994; 11: 167-180.
- [17] Lei CF, Chen J, Li FL, Yang LB, Yang HY, Lian YX, Bu H. The value on computer-aided quantitative analysis in improving the reproducibility of histological grade in ductal carcinoma in situ of breast. *Chinese Journal of Clinical and Experimental Pathology* 2022; 38: 1204-9.
- [18] Ellis IO, Sam-Ai S, Anderson N, et al. Pathology reporting of breast disease in surgical excision

- specimens incorporating the dataset for histological reporting of breast cancer. The Royal College of Pathologists 2016.
- [19] Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, Boecker W, Borisch B, Bussolati G, Connolly CE, Cserni G, Decker T, Dervan P, Drijkoningen M, Ellis IO, Elston CW, Eusebi V, Faverly D, Heikkila P, Holland R, Kerner H, Kulka J, Jacquemier J, Lacerda M, Martinez-Penuela J, De Miguel C, Nordgren H, Peterse JL, Rank F, Regitnig P, Reiner A, Sapino A, Sigal-Zafrani B, Tanous and AM TS, Zozaya E, Wells CA; EC Working Group on Breast Screening Pathology. Quality assurance guidelines for pathology. In: Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L, editors. European guidelines for quality assurance in breast cancer screening and diagnosis, 4th edn. European Commission, Luxemburg; 2006. pp. 219-311.
 - [20] Brown JP and Pinder SE. Ductal carcinoma in situ: current morphological and molecular subtypes. *Diagnostic Histopathology* 2012; 18: 112-118.
 - [21] College of American Pathologists. H&E Troubleshooting Table. <https://documents.cap.org/documents/h-and-e-troubleshooting-guide.pdf>.
 - [22] Consensus Conference on the classification of ductal carcinoma in situ. The consensus conference committee. *Cancer* 1997; 80: 1798-1802.
 - [23] Li H, Whitney J, Bera K, Gilmore H, Thorat MA, Badve S and Madabhushi A. Quantitative nuclear histomorphometric features are predictive of Oncotype DX risk categories in ductal carcinoma in situ: preliminary findings. *Breast Cancer Res* 2019; 21: 114.
 - [24] Gong Y, Kandil D and Khan A. Molecular Pathology of Pre-Invasive Ductal Carcinoma. In: Khan A, Ellis IO, Hanby AM, Cosar EF, Rakha EA, Kandil D, editors. *Precision Molecular Pathology of Breast Cancer*. New York, NY: Springer New York; 2015. pp. 79-93.
 - [25] Hophan SL, Odnokoz O, Liu H, Luo Y, Khan S, Gradishar W, Zhou Z, Badve S, Torres MA and Wan Y. Ductal carcinoma in situ of breast: from molecular etiology to therapeutic management. *Endocrinology* 2022; 163: bqac027.
 - [26] Gerdes MJ, Gokmen-Polar Y, Sui Y, Pang AS, LaPlante N, Harris AL, Tan PH, Ginty F and Badve SS. Single-cell heterogeneity in ductal carcinoma in situ of breast. *Mod Pathol* 2018; 31: 406-417.
 - [27] Shaaban AM, Hilton B, Clements K, Provenzano E, Cheung S, Wallis MG, Sawyer E, Thomas JS, Hanby AM, Pinder SE and Thompson AM; Sloane Project Steering Committee. Pathological features of 11,337 patients with primary ductal carcinoma in situ (DCIS) and subsequent events: results from the UK Sloane Project. *Br J Cancer* 2021; 124: 1009-1017.
 - [28] Brennan ME, Turner RM, Ciatto S, Marinovich ML, French JR, Macaskill P and Houssami N. Ductal carcinoma in situ at core-needle biopsy: meta-analysis of underestimation and predictors of invasive breast cancer. *Radiology* 2011; 260: 119-128.
 - [29] Sheaffer WW, Gray RJ, Wasif N, Stucky CC, Cronin PA, Kosiorek HE, Basu A, Pizzitola VJ, Patel B, Giurescu ME, Lorans R, McCullough AE, Ocal IT and Pockaj BA. Predictive factors of upstaging DCIS to invasive carcinoma in BCT vs mastectomy. *Am J Surg* 2019; 217: 1025-1029.
 - [30] Huo L, Sneige N, Hunt KK, Albarracin CT, Lopez A and Resetskova E. Predictors of invasion in patients with core-needle biopsy-diagnosed ductal carcinoma in situ and recommendations for a selective approach to sentinel lymph node biopsy in ductal carcinoma in situ. *Cancer* 2006; 107: 1760-1768.
 - [31] Shaaban AM, Hilton B, Clements K, Dodwell D, Sharma N, Kirwan C, Sawyer E, Maxwell A, Wallis M, Stobart H, Mylvaganam S, Litherland J, Brace-McDonnell S, Dulson-Cox J, Kearins O, Provenzano E, Ellis IO, Pinder SE and Thompson AM. The presentation, management and outcome of patients with ductal carcinoma in situ (DCIS) with microinvasion (invasion ≤ 1 mm in size)-results from the UK Sloane Project. *Br J Cancer* 2022; 127: 2125-2132.
 - [32] Canas-Marques R and Schnitt SJ. Ductal carcinoma in situ with and without microinvasion: is there a clinically meaningful difference in outcome? *Br J Cancer* 2023; 128: 713-714.
 - [33] Liu G, Kong X, Dai Q, Cheng H, Wang J, Gao J and Wang Y. Clinical features and prognoses of patients with breast cancer who underwent surgery. *JAMA Netw Open* 2023; 6: e2331078.
 - [34] Sung H, Jiang C, Jatoi I and Jemal A. Trends in breast cancer death rates from ductal carcinoma in situ versus invasive cancer in the United States. *Breast Cancer Res* 2025; 27: 164.
 - [35] Hwang ES, Hyslop T, Lynch T, Ryser MD, Weiss A, Wolf A, Norris K, Witten M, Grimm L, Schnitt S, Badve S, Factor R, Frank E, Collyar D, Basila D, Pinto D, Watson MA, West R, Davies L, Donovan JL, Shimada A, Li Y, Li Y, Bennett AV, Rosenberg S, Marks J, Winer E, Boisvert M, Giuliano A, Larson KE, Yost K, McAuliffe PF, Krie A, Tamirisa N, Carey LA, Thompson AM and Partridge AH; COMET Study Investigators. Active monitoring with or without endocrine therapy for low-risk ductal carcinoma in situ: the COMET randomized clinical trial. *JAMA* 2025; 333: 972-80.

Nuclear grading in DCIS

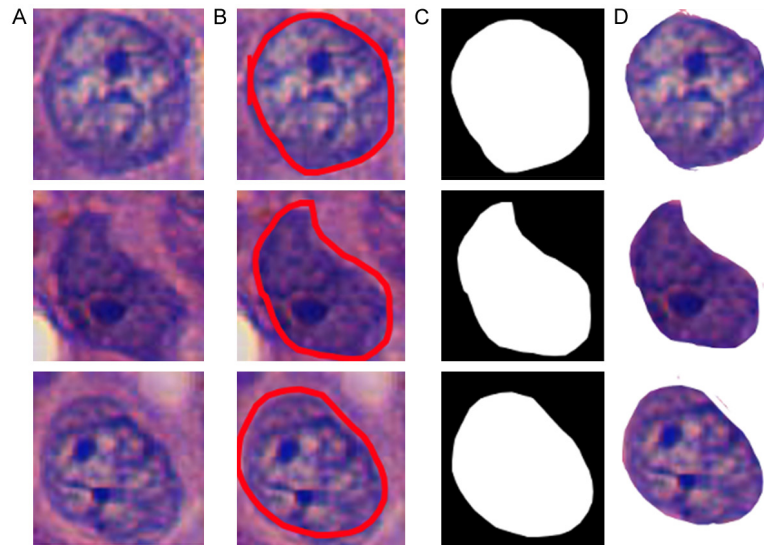


Supplementary Figure 1. Case selection in this study.

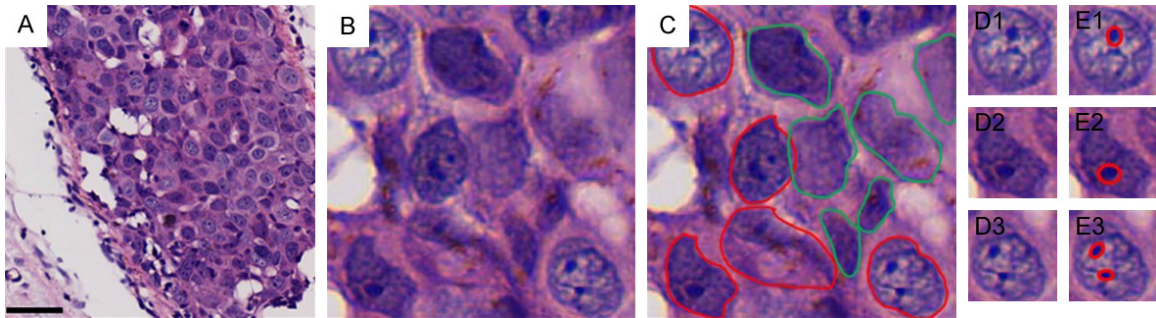


Supplementary Figure 2. Pleomorphism. (A) Original image; (B) A patch from the original image; (C1-C3) A nucleus extracted from b; (D1-D3) Outlines of nuclei in (C1-C3); (E1-E3) Fit the minimum bounding rectangle based on the instance segmentation results of the nuclei (the blank regions E1-E3). The bounding rectangle is indicated by the red lines within (E1-E3). Next, the inscribed ellipse within this bounding rectangle (shown in green in E1-E3) is used as a reference. The scale bar in the (A) represents 200 μm; in (B) represents 50 μm.

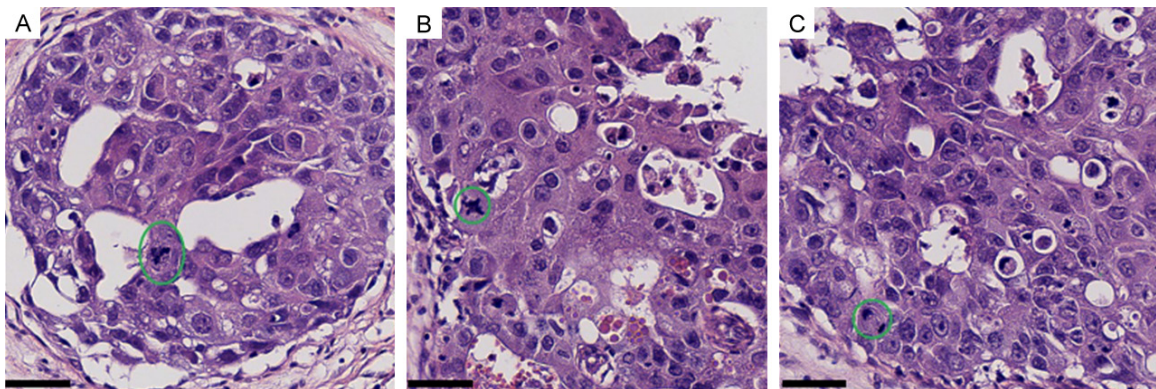
Nuclear grading in DCIS



Supplementary Figure 3. Chromatin features. A. The tumor cell nuclei in the original image. B. The nuclear contours segmented by the model. C. The corresponding mask is generated based on the contours. D. The tumor cell nuclei extracted using the mask. The variance of chromatin uniformity within the nuclei can be calculated by applying the relevant formula to the extracted nuclei.



Supplementary Figure 4. Nucleoli detection. (A) Original image. (B) Amplification of a. (C) The red contour lines in c indicate cells in which nucleoli were detected, while the green contour lines represent cells in which nucleoli were not detected. (D1-D3) The nuclei from the original image based on the results of nuclear instance segmentation. (E1-E3) Red ellipses highlight the nucleoli within the nuclei, which were detected via the Hough Transform algorithm. The scale bar in the (A) represents 100 μm .

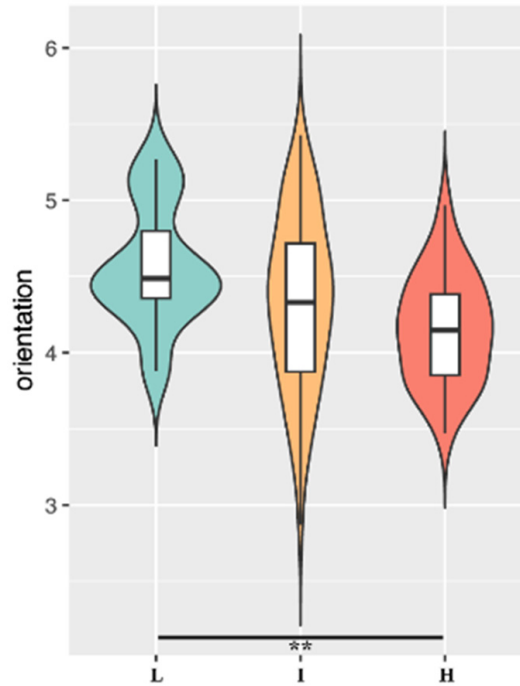


Supplementary Figure 5. Mitoses. A-C. The green circles indicate the mitotic figures. The scale bars represent 25 μm .

Nuclear grading in DCIS

Supplementary Table 1. The numbers of low-, intermediate- and high-grade DCIS cases in each data-set

Dataset	N	Low grade		Intermediate grade		High grade	
		N	%	N	%	N	%
Internal Training Set	80	16	20.0	42	52.5	22	27.5
Internal Validation Set	21	6	28.6	8	38.1	7	33.3
Internal Test Set	116	19	16.4	67	57.8	30	25.9
External Test Set A	116	10	8.6	56	48.3	50	43.1
External Test Set B	69	14	20.3	34	49.3	21	30.4

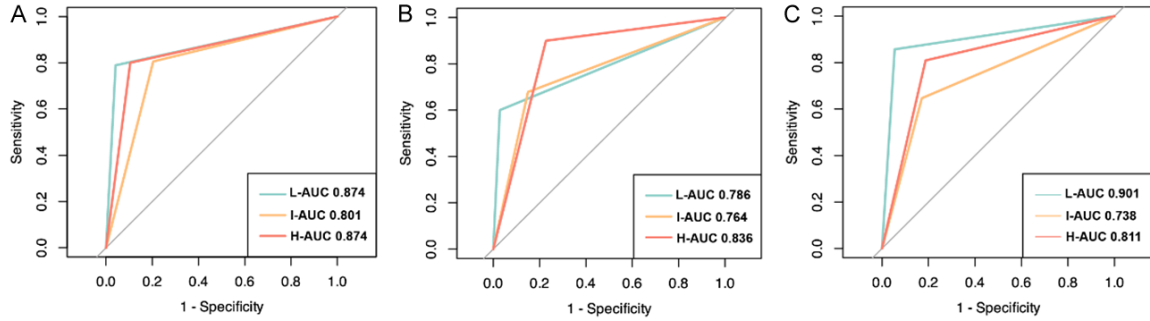


Supplementary Figure 6. Cell orientation in low-, intermediate-, and high-grade DCIS. Abbreviations: L, low-grade cells; I, intermediate-grade cells; H, high-grade cells. ** $P < 0.010$. Only comparisons with statistically significant P values are labeled in the figure.

Supplementary Table 2. Integrated grading model evaluation in different sets

Dataset	Grade	Precision	Recall	F1 score
Internal test set	Low	0.79	0.79	0.79
	Intermediate	0.81	0.84	0.82
	High	0.80	0.73	0.76
External test A	Low	0.67	0.60	0.63
	Intermediate	0.81	0.68	0.74
	High	0.75	0.90	0.82
External test B	Low	0.80	0.86	0.83
	Intermediate	0.79	0.65	0.71
	High	0.65	0.81	0.72

Nuclear grading in DCIS



Supplementary Figure 7. ROC curves for integrated grading model in different sets. A. ROC curves for evaluation of low-, intermediate-, and high-grade DCIS in test set. B. ROC curves for evaluation of low-, intermediate-, and high-grade DCIS in external test set A. C. ROC curves for evaluation of low-, intermediate-, and high-grade DCIS in external test set B. Abbreviations: L, low-grade cells; I, intermediate-grade cells; H, high-grade cells; ROC, receiver operator characteristic; AUC, areas under the ROC curve.

Supplementary Table 3. Clinical features of DCIS biopsy cases in training set B and external test set C

Features	Training set B (N, %)	External test set C (N, %)
Age (years)		
< 40	39 (13.5)	11 (20.4)
40-60	219 (76.0)	38 (70.4)
> 60	30 (10.4)	5 (9.3)
ER status*		
Positive	144 (50.0)	25 (46.3)
Negative	144 (50.0)	29 (53.7)
HER2 IHC score		
0	40 (13.9)	2 (3.7)
1+	47 (16.3)	4 (7.4)
2+	43 (14.9)	16 (29.6)
3+	158 (54.9)	32 (59.3)
Grade		
Intermediate	85 (29.5)	18 (33.3)
High	203 (70.5)	36 (66.7)
Tumor size (cm)		
< 1.6	41 (14.2)	5 (9.3)
1.6-4.0	191 (66.3)	32 (59.3)
> 4.0	56 (19.4)	17 (31.5)
Extensive invasion [#]		
No	198 (68.8)	36 (66.7)
Yes	90 (31.3)	18 (33.3)

*ER status: Positive status means $\geq 1\%$ positive nuclei staining. [#]Extensive invasion: Invasive carcinoma measuring > 1 mm in the final excision specimen.