ORIGINAL ARTICLE

Validation of Whole Slide Imaging Diagnosis for Breast Core Needle Biopsy Specimens at Thammasat University Hospital

Worakit Kaewnopparat, MD[1], Adiluck Pisutpunya, MD[1], Wanwisa Himakhun, MD[1]

[1] Department of Pathology, Faculty of Medicine, Thammasat University, Pathum Thani, Thailand

Background: Whole slide imaging (WSI) is a digital technology developed as an adjunct to, or an alternative modality for, conventional light microscopy (CLM) in pathological diagnosis. However, validation remains essential before implementation, particularly for breast core needle biopsy (CNB), where diagnostic accuracy is critical for patient management.

Objective: To validate the diagnostic concordance, reproducibility, and efficiency of WSI in comparison to CLM for the interpretation of CNB, encompassing histological assessment and biomarker evaluation, specifically estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67.

Materials and Methods: Three observers retrospectively assessed 79 CNB cases and biomarker subgroups using CLM and WSI after a 2-week washout. Agreement among observers was evaluated using Cohen's kappa. The analysis focused on concordance, major and minor discordance, and interpretation time.

Results: Intra-observer agreement demonstrated substantial to almost perfect levels for hematoxylin and eosin (H&E) (κ=0.85 to 0.92) and biomarkers; ER (κ=0.80 to 0.84), PR (κ=0.69 to 1.00), HER2 (κ=0.80 to 0.89), and Ki-67 (κ=0.86 to 0.89), with p<0.001. The concordance rate for H&E was between 81% to 87%, with major discordance occurring in only 1% to 2.5% of cases. Biomarker concordance reached 80% to 90%, with the majority of discordances being minor. WSI interpretation time was comparable to or faster than CLM for most IHC markers, driven by user experience.

Conclusion: WSI demonstrates accuracy, reproducibility, and efficiency in the diagnosis of breast CNB, in accordance with the College of American Pathologists (CAP) validation standards. It is a safe alternative to microscopy and can be used in regular pathology, even in resource-bound settings.

Keywords: Whole slide imaging; Digital pathology; Breast core needle biopsy; Diagnostic concordance

Received 9 December 2025 | Revised 2 February 2026 | Accepted 3 February 2026

J Med Assoc Thai 2026;109(5):353-64

Website: http://www.jmatonline.com

The breast core needle biopsy (CNB) has become the standard procedure for detecting suspicious breast lesions over the last four decades. It is minimally invasive and guided by palpation or imaging, such as mammography or ultrasound. Histopathological evaluation of tissue cores using conventional light microscopy (CLM) is the gold standard for diagnostic confirmation. CNB helps identify benign, premalignant, and malignant breast lesions and guides surgical and treatment decisions. CNB specimens can accurately assess histologic subtype, tumor grade, and predictive biomarkers in malignancies, enabling timely and appropriate oncologic management[1-5].

Automated whole slide imaging (WSI) has quickly become a crucial diagnostic pathology technique. The approach is mostly used for digital picture transmission, second opinion consultation, quality assurance, and teaching. WSI might substitute conventional microscopy as a first-line diagnostic platform due to digital image resolution and technological advancements. To ensure digitized slide diagnostic performance is comparable to glass slides using CLM, WSI must be validated. Several guideline statements regarding the application of WSI for routine diagnostics have been published[6-10]. The College of American Pathologists (CAP) updated its guidelines for the validation of WSI for pathology diagnostic purposes in 2021, recommending a sample set of at least 60 cases for one application [e.g.,

Correspondence to:

Himakhun W.

Department of Pathology, Faculty of Medicine, Thammasat University Hospital, Thammasat University, Pathum Thani 12120, Thailand.

Phone: +66-813468861

Email: hwanwisa@tu.ac.th, ooiill24@gmail.com

How to cite this article:

Kaewnopparat W, Pisutpunya A, Himakhun W. Validation of Whole Slide Imaging Diagnosis for Breast Core Needle Biopsy Specimens at Thammasat University Hospital. J Med Assoc Thai 2026;109:353-64.
DOI: 10.35755/jmedassocthai.2026.5.03887

hematoxylin and eosin (H&E), frozen section, or hematology], diagnostic concordance between digital and glass slides for the same observer, and a washout period of at least two weeks between viewing digital and glass slide[10].

Studies from the University of Nebraska Medical Center and the University of Miami demonstrated high diagnostic reliability of WSI for breast CNB. Using the 2013 CAP guidelines, the Nebraska study reported a WSI-CLM concordance of 97.1%, with intra-observer agreement increasing to 95.4% after a washout period[11]. Similarly, the Miami study showed high intra-observer repeatability with diagnostic performance comparable to conventional glass slide microscopy[12]. In Thailand, a validation study of a low-cost WSI system utilizing a consumer-grade laptop for frozen section lymph node evaluation revealed substantial intra- and inter-observer concordance among three pathologists. There was no significant difference in diagnostic turnaround time between digital and glass slides. However, experienced users exhibited a tendency for expedited diagnosis using WSI[13].

This study seeks to validate WSI for the diagnosis of breast CNB specimens by comparing digital slide interpretations with traditional glass slide diagnoses. This study aimed to determine the diagnostic concordance, identify inconsistencies, and evaluate intra-observer agreement between the two modalities within the authors' institutional context. Validation is crucial prior to the use of WSI for the diagnosis of primary breast lesions in Thailand, guaranteeing that digital pathology offers a precise and secure alternative to the light microscope in this setting.

Materials and Methods

Selection of cases and immunohistochemical analysis

Four hundred forty breast CNB specimens diagnosed at Thammasat University Hospital between January and December 2023 were reviewed retrospectively. Of these, 79 cases representing routine primary breast lesions were included for analysis. All H&E-stained CNB cases were selected consecutively from the routine diagnostic workload to reflect real-world practice and minimize selection bias, with representation across B1-B5 diagnostic categories. The distribution of clinicopathological characteristics across B1-B5 diagnostic categories is summarized in Table 1. Non-primary breast lesions were excluded, as were cases with unavailable histologic materials or slides unsuitable for WSI because of technical issues such as poor staining, cracked slides, or displaced coverslips. For the immunohistochemistry (IHC) component, cases were also selected consecutively. From 286 breast CNB cases diagnosed between January 2023 and December 2024, 20 were selected for each biomarker: estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67, resulting in 80 IHC slides (Figure 1). Tissue sections were trimmed to 5 to 6 μm and placed on glass slides. All IHC staining was carried out on the Ventana BenchMark XT automated platform with the ultraView DABv3 detection system. Primary antibodies included anti-ER (SP1), anti-PR (1E2), anti-HER2/neu (4B5), and anti-Ki-67 (30-9), all of which were Ventana rabbit monoclonal antibodies. Thammasat University Hospital's regular routine diagnostic protocols were followed throughout all procedures, with no study-specific alterations.

Table 1. Patient demographics and clinicopathologic characteristics of the breast CNB cohort included in the H&E validation study (n=79)

Characteristics	Value
Study cohort (H&E CNB cases)	79
Age (years); mean±SD (range)	51.6±13.3 (24 to 86)
Sex (female/male); n (%)	78 (98.7)/1 (1.3)
B category distribution; n (%)	
B1 (Normal tissue)	5 (6.3)
B2 (Benign lesion)	25 (31.6)
B3 (Lesion of uncertain malignant potential)	21 (26.6)
B4 (Suspicious for malignancy)	8 (10.1)
B5 (Malignancy)	20 (25.3)

CNB=core needle biopsy; H&E=hematoxylin and eosin; SD=standard deviation

Conventional light microscope and pathologist diagnosis and scoring

Histological diagnosis and IHC staining assessment were performed using glass slides via CLM with eyepieces with a field number of 22 mm (Nikon Eclipse Si, Tokyo, Japan). Cases were initially reviewed by three observers (WK, WH, and AP), each with varying levels of experience and training in the diagnosis of CLM and WSI (Figure 2). The histology diagnosis was divided into five categories, B1-B5, according to the European guideline for breast biopsy interpretation as follows[14].

B1 (Normal tissue): Specimens in which only normal breast tissue is identified (e.g., breast duct, normal lobule with or without calcification).

B2 (Benign lesion): Non-neoplastic or benign

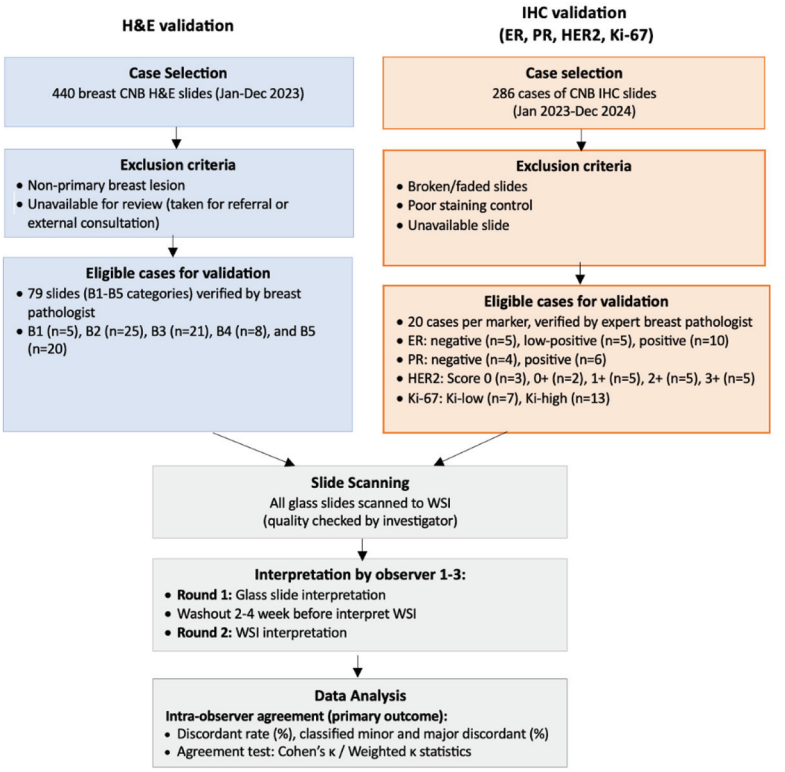

Figure 1. A study flow: Archival breast core biopsy H&E slides (n=440) were screened; after exclusions, 79 B1-B5 cases were selected, quality-verified, scanned to WSI, and interpreted by three observers on glass, then WSI after washout. Intra- and inter-observer agreement, discordance rates, κ statistics, and interpretation time were analyzed. Phase 2 applied the same process to 20 expert-verified IHC cases per biomarker (ER, PR, HER2, Ki-67) stratified by staining categories.

H&E=hematoxylin and eosin; IHC=immunohistochemistry; CNB=core needle biopsy; WSI=whole slide imaging; ER=estrogen receptor; PR=progesterone receptor; HER2=human epidermal growth factor receptor 2; κ=kappa coefficient

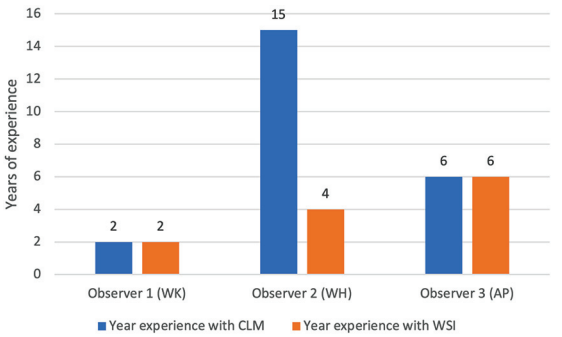

Figure 2. Year experience of Observer 1-3 diagnosis of CLM and WSI.

CLM=conventional light microscopy; WSI=whole slide imaging

pathologic processes that can be specifically diagnosed, such as fibroadenoma, fibrocystic change, sclerosing adenosis, duct ectasia, abscess, and fat necrosis.

B3 (Lesions of uncertain malignant potential): Histologically benign lesions that show architectural/cytologic atypia or carry an increased risk of malignancy and may warrant altered clinical management, including atypical intraductal epithelial proliferation, lobular neoplasia, cellular fibroepithelial lesions/benign phyllodes tumors, papillary lesions (papilloma with or without epithelial atypia), radial scars, and mucocele-like lesions.

B4 (Suspicious for malignancy): Lesions with features suggestive of malignancy but insufficient for a definitive diagnosis due to limitations such as tissue artifact or limited sampling. Examples include low-grade ductal carcinoma in situ (DCIS) or minimal malignant tissue.

B5 (Malignancy): Lesions diagnostic of malignancy, including in situ and invasive lesions, include intermediate- to high-grade DCIS, pleomorphic lobular carcinoma in situ (LCIS), and invasive breast carcinoma of various types.

The specific diagnostic entities selected and represented in each B category in this study are summarized in Table 2.

Table 2. Expanded diagnostic entities corresponding to B1-B5 categories in the breast CNB H&E cohort (n=79)

B category	Expanded specific diagnostic entities (examples in this cohort)	n (%)
B1 (Normal tissue)	Normal breast tissue (ducts/lobules) ± calcification	5 (6.3)
B2 (Benign lesion)	Fibroadenoma	7 (8.9)
	Fibrocystic change	4 (5.1)
	Inflammatory breast lesion (e.g., mastitis/abscess)	3 (3.8)
	Other benign/non-neoplastic lesions (optional, if present): duct ectasia/fat necrosis/sclerosing adenosis	11 (13.9)
B3 (Uncertain malignant potential)	Benign phyllodes tumor/cellular fibroepithelial lesion	12 (15.1)
	Low-grade atypical proliferative lesion (e.g., atypical intraductal epithelial proliferation/ADH-like)	1 (1.2)
	Papillary lesion without definite malignancy (if applicable)	8 (10.1)
	Other B3 entities (optional): lobular neoplasia/radial scar/mucocele-like lesion	-
B4 (Suspicious for malignancy)	Papillary lesion with atypical epithelial proliferation/suspicious atypia	5 (6.32)
	Borderline phyllodes tumor (suspicious)	1 (1.2)
	Other suspicious lesions (optional): suspicious for low-grade DCIS/limited malignant tissue	2 (2.5)
B5 (Malignancy)	Ductal carcinoma in situ (DCIS)	5 (6.3)
	Invasive breast carcinoma of no special type (NST)	12 (15.8)
	Invasive lobular carcinoma (ILC)	1 (1.2)
	Mucinous carcinoma	1 (1.2)
	Other invasive carcinoma or malignancy types (optional, if present)	1 (1.2)
Total		79 (100)

ADH= atypical duct hyperplasia, CNB=core needle biopsy; H&E=hematoxylin and eosin

Interpretation of ER and PR IHC followed the 2018 American Society of Clinical Oncology (ASCO)/CAP guidelines[15], with assessment based on the percentage of tumor cell nuclear staining. PR expression was dichotomized as negative (less than 1%) or positive (1% or more), in accordance with guideline recommendations. For ER, although clinical reporting adheres to the ASCO/CAP three-tier system (negative: less than 1%, low-positive: 1% to 10%, and positive: more than 10%), ER expression in this study was further stratified into four analytical categories (less than 1%, 1% to 10%, 11% to 75%, and more than 75%). This approach, adapted from prior WSI validation studies in breast pathology[16], was used to enhance sensitivity for detecting discordant interpretations between CLM and WSI, including subtle quantitative shifts that do not affect clinical categorization. ER was selected for this expanded stratification because it is the principal predictive biomarker for luminal-type breast carcinoma and the only marker validated in randomized trials to predict benefit from endocrine therapy, whereas PR serves a complementary prognostic role and assists in identifying potential false-negative ER and rare ER-negative/PR-positive tumors, for which evidence guiding endocrine therapy remains limited[17,18].

HER2 status was classified into five categories according to the 2023 ASCO/CAP updated guidelines and the criteria applied in the DESTINY-Breast06 trial[19,20]: 0 (negative, no membrane staining), 0+ (negative, incomplete membrane staining), 1+ (negative), 2+ (equivocal), and 3+ (positive). Ki-67 immunoreactivity was defined by nuclear staining of any intensity. The Ki-67 labeling index was assessed using a global scoring approach, estimating the average percentage of positively stained tumor cell nuclei within invasive tumor areas, in accordance with the recommendations of the International Ki-67 in Breast Cancer Working Group[21].

Slide digitization and diagnosis with WSI

For WSI, H&E, and IHC-stained slides were imaged at a high resolution (0.121 µm/pixel) and 20× magnification using a whole slide scanner (PANNORAMIC Scan II, 3DHISTECH, Budapest, Hungary) with a single z-plane. The MRXS format was used to generate and save digital images, which were then managed with server software (Panoramic Scanner, 3DHISTECH) and retrieved using a file management web interface (CaseViewer, 3DHISTECH). The average scanned image size was 1.2 to 1.95 GB. The quality of scanned digital images was assessed to ensure that they were suitable for analysis and in focus. Different consumer-grade laptop models with high-definition displays (2,732×2,048 pixels) were employed for intra-observer diagnosis to optimize image quality. The glass slides were reviewed by each observer, and the

WSI was performed after a minimum of two weeks of washout.

Definition of intra-observer concordance, minor discordance, and major discordance

Concordance for H&E was defined as the identical diagnostic categorization between WSI-based and CLM-based interpretations. Minor discordant refers to variations within the same benign or malignant category that lack clinical significance, while major discordant involve benign-malignant reversals or significantly different diagnoses within the same category that may influence treatment or prognosis. In the context of ER IHC, concordance was characterized by identical categorical grouping. Minor discordant were restricted to transitions between Group 3 (11% to 75% positive cells) and Group 4 (more than 75%), whereas major discordant encompassed any other categorical alterations that could impact clinical decision-making. HER2 IHC concordance required the same score (0, 1+, 2+, or 3+), minor disparities were confined to score 0 versus 1+ (both HER2-negative), and major discordances included therapeutic category shifts. For PR and Ki-67 IHC, major and minor discordance were not separately defined, as both markers had only two categories. Ki-67 was classified (low for less than 20%) or high (20% or more), and PR as negative (less than 1%) or positive (1% or more). Concordance was defined as agreement within the same category. Any transition between categories was considered clinically significant, indicating discordance.

Statistical analysis

Intra-observer agreement of CLM and WSI was assessed using Cohen's kappa coefficient. The strength of agreement was interpreted according to standard benchmarks: 0.01 to 0.20 (slight), 0.21 to 0.40 (fair), 0.41 to 0.60 (moderate), 0.61 to 0.80 (substantial), and 0.81 to 0.99 (almost perfect)[22]. A two-sided p-value of less than 0.05 was considered statistically significant. All statistical analyses were conducted using IBM SPSS Statistics, version 26.0 (IBM Corp., Armonk, NY, USA).

Ethical approval

All procedures performed in the study were approved by the Human Research Ethics Committee of Thammasat University (Medicine) (COA No. 092/2024). Formal written informed consent was not required with a waiver by the appropriate human ethics committee.

Results

Patients and clinicopathologic characteristics

Seventy-nine H&E-stained CNB cases, consecutively selected from 440 CNB cases accessed between January and December 2023 in primary routine diagnostic practice, were included. The overall mean age was 51.6±13.3 years (range of 24 to 86). The vast majority of patients were female (78 cases, 98.7%), with only one male breast biopsy (1.3%). The distribution of pathological diagnoses across the standard categories B1-B5 is as follows. Five cases (6.3%) were classified as Category 1 (normal breast tissue), 25 cases (31.6%) as Category 2 (e.g., fibroadenoma, fibrocystic changes, and inflammatory breast lesion), 21 cases (26.6%) as Category 3 (e.g., benign phyllodes tumor and low-grade atypical proliferative lesions), 8 cases (10.1%) as Category 4 (e.g., papillary lesion with atypical cells proliferation, or borderline phyllodes tumor), and 20 cases (25.3%) as Category 5 (including DCIS, invasive breast carcinoma of no special type, invasive lobular carcinoma, and mucinous carcinoma). All 20 malignant cases were primary breast carcinomas diagnosed with CNB.

Intra-observer agreement between CLM and WSI

All three observers exhibited substantial intra-observer agreement between CLM and WSI across all assessed parameters. Kappa values for H&E diagnosis ranged from 0.85 to 0.92, reflecting almost perfect agreement. ER interpretation demonstrated substantial to almost perfect agreement (κ=0.80 to 0.84), whereas PR interpretation exhibited substantial to almost agreement (κ=0.69 to 1.00). HER2 scoring exhibited substantial to almost perfect concordance (κ=0.80 to 0.89). The Ki-67 assessment demonstrated almost perfect concordance between CLM and WSI, with kappa values between 0.86 and 0.89. All results demonstrated statistical significance ($p<0.001$) (Table 3, Figure 3).

Intra-observer concordance, minor discordance, and major discordance between CLM and WSI

High concordance rates were noted among all observers for most parameters (Table 4), especially PR and Ki-67, which demonstrated consistent agreement of 95% to 100% with little to no discordance. The interpretation of H&E demonstrated a high level of concordance (81% to 87.3%), with only 1% to 2.5% major discordance.

In the case of ER, concordance varied between 80% and 90%, while significant discordance was

Table 3. Intra-observer agreement between conventional light microscope and whole slide imaging diagnoses

	Intra-observer agreement between CLM and WSI; k-value (95% CI)				
	HE (CLM vs. WSI)	ER (CLM vs. WSI)	PR (CLM vs. WSI)	HER2 (CLM vs. WSI)	Ki-67 (CLM vs. WSI)
Observer 1	0.92 (0.87 to 0.97)	0.84 (0.69 to 0.98)	0.86 (0.59 to 1.00)	0.82 (0.67 to 0.97)	0.86 (0.59 to 1.00)
p-value	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Observer 2	0.85 (0.78 to 0.92)	0.80 (0.64 to 0.95)	1.00 (1.00 to 1.00)	0.89 (0.77 to 1.00)	0.89 (0.69 to 1.00)
p-value	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Observer 3	0.85 (0.78 to 0.93)	0.84 (0.68 to 0.99)	0.69 (0.31 to 1.00)	0.80 (0.66 to 0.94)	0.88 (0.64 to 1.00)
p-value	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001

CLM=conventional light microscope; WSI=whole slide imaging; CI=confidence interval; H&E=hematoxylin and eosin; ER=estrogen receptor; PR=progesterone receptor; HER2=human epidermal growth factor receptor 2; κ=kappa coefficient

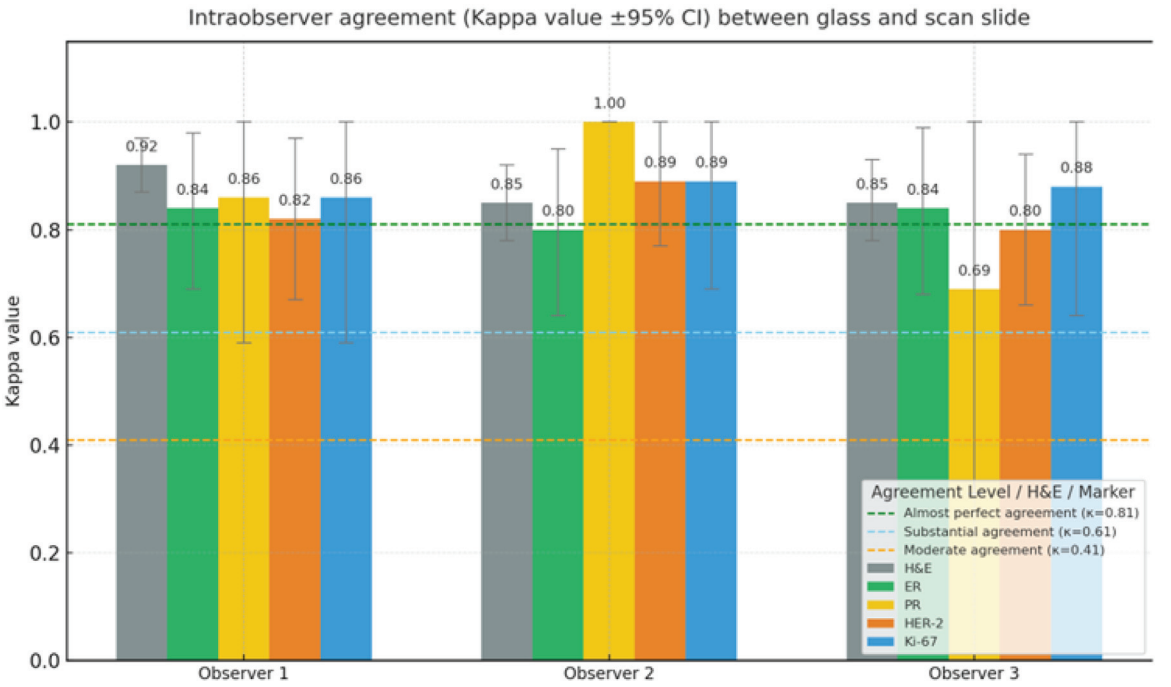

Figure 3. Intra-observer agreement of breast cancer biomarkers expression between CLM and WSI. All kappa coefficients demonstrated significant (p<0.001).

CLM=conventional light microscopy; WSI=whole slide imaging; H&E=hematoxylin and eosin; ER=estrogen receptor; PR=progesterone receptor; HER2=human epidermal growth factor receptor 2; CI=confidence interval; κ=kappa coefficient

observed at a rate of 15% among all observers. HER2 exhibited moderate concordance, ranging from 75% to 85%, alongside minor discordance between 10% and 20%, and major discordance between 5% and 15%.

Most discordant interpretations were minor, while major discordances were infrequent across all markers. Two major H&E discrepancies were found. In the first case, Observer 3 reclassified a lesion from fibroepithelial lesion, which cannot exclude borderline phyllodes tumor (B3) on CLM, to fibroepithelial lesion with occult invasive mammary carcinoma (B5) on WSI, while Observers 1 and 2 consistently diagnosed invasive lobular carcinoma (B5) on both modalities (Figure 4A). In the second case, which had a prior history of treated invasive lobular carcinoma, Observer 1 showed complete concordance, diagnosing benign breast tissue with foreign body granuloma (B2) on both CLM and WSI. In contrast, Observer 2 upgraded the diagnosis from reactive post-treatment changes without residual malignancy (B2) to rare atypical cell clusters suspicious for occult tumor or usual ductal hyperplasia (B4) on WSI, while Observer 3 downgraded the diagnosis from suspicious lymphovascular tumor emboli with foreign body reaction (B4) on CLM to

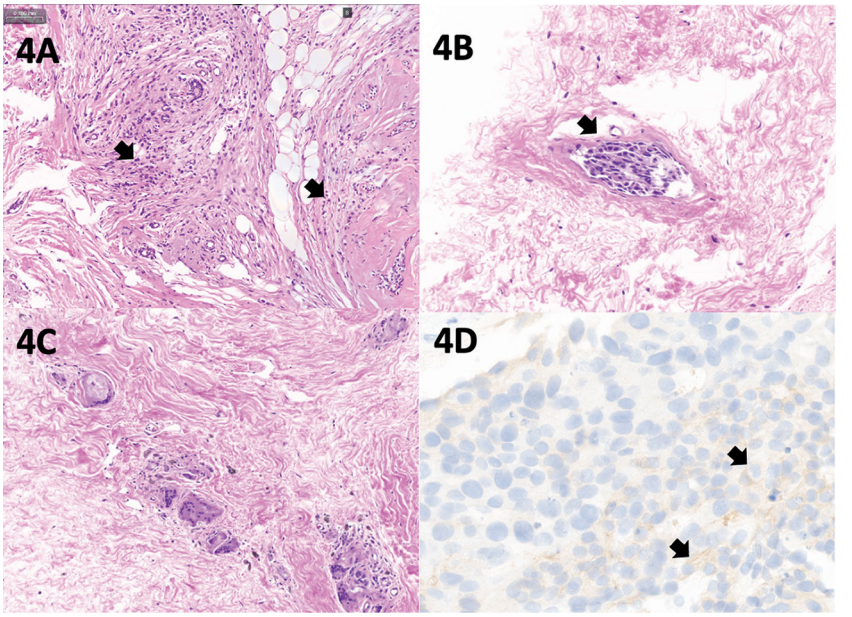

Figure 4. (4A) Representative histologic images from a case interpreted by Observer 3, demonstrating upgrading from a fibroepithelial lesion (B3) on CLM to invasive mammary carcinoma (B5) on WSI. Small occult invasive foci are more readily appreciated on WSI (arrows). (4B, 4C) In a post-treatment case, Observer 2 upgraded the diagnosis from reactive post-treatment changes (B2) on CLM to rare atypical cell clusters (B4) on WSI (arrow). In contrast, Observer 3 downgraded the diagnosis from suspicious lymphovascular tumor emboli with foreign body reaction (B4) on CLM to benign breast tissue with foreign body reaction (B2) on WSI. (4D) HER2 immunohistochemistry shows reclassification from a negative score of 0 (null staining) on CLM to low-level HER2 expression (score 1+) on WSI (arrows).

CLM=conventional light microscopy; WSI=whole slide imaging; HER2=human epidermal growth factor receptor 2

Table 4. Intra-observer concordance, minor discordance, major discordance (H&E, ER, HER2), and unstratified discordance (PR, Ki-67) between CLM and WSI diagnoses

	Concordance n (%)	Minor discordance n (%)	Major discordance n (%)	Discordance n (%)
Observer 1				
H&E	69 (87.3)	10 (12.7)	0 (0.0)	-
ER	16 (80.0)	1 (5.0)	3 (15.0)	-
PR	19 (95.0)	-	-	1 (5.0)
HER2	15 (75.0)	2 (10.0)	3 (15.0)	-
Ki-67	19 (95.0)	-	-	1 (5.0)
Observer 2				
H&E	64 (81.0)	14 (17.7)	1 (1.3)	-
ER	15 (90.0)	2 (10.0)	3 (15.0)	-
PR	20 (100.0)	-	-	0 (0.0)
HER2	17 (85.0)	2 (10.0)	1 (5.0)	-
Ki-67	19 (95.0)	-	-	1 (5.0)
Observer 3				
H&E	64 (81.0)	13 (16.5)	2 (2.5)	-
ER	17 (85.0)	1 (5.0)	3 (15.0)	-
PR	18 (90.0)	-	-	2 (10.0)
HER2	15 (75.0)	4 (20.0)	1 (5.0)	-
Ki-67	19 (95.0)	-	-	1 (5.0)

H&E=hematoxylin and eosin; ER=estrogen receptor; PR=progesterone receptor; HER2=human epidermal growth factor receptor 2; CLM=conventional light microscopy; WSI=whole slide imaging

benign breast tissue with foreign body reaction (B2) on WSI (Figure 4B and 4C). Immunohistochemical discordance was confined to HER2 interpretation, primarily reflecting reclassification from null stain (score of 0) to low-positive (score of 1+) results (Figure 4D).

Inter-observer agreement of CLM and WSI

Using CLM, all observer pairs demonstrated substantial to almost perfect agreement across H&E and all biomarkers (κ=0.67 to 0.88, $p<0.001$). The highest concordance was observed for ER and HER2, while PR and Ki-67 showed slightly lower but still substantial agreement (Table 5). With WSI, H&E, ER, and HER2 maintained similarly high agreement levels (κ=0.71 to 0.86). However, PR and Ki-67 showed rather variable results, particularly between Observer 2 and 3 for PR (κ=0.48), indicating moderate consistency in certain borderline cases on WSI (Table 6).

Time use for CLM and WS

In general, CLM required a longer interpretation duration than WSI among the majority of observers and IHC markers. The most time-consuming task

Table 5. Interobserver agreement of CLM diagnoses

	Interobserver agreement of CLM between the observers k-value (95% CI)		
	Observer 1 vs. 2	Observer 1 vs. 3	Observer 2 vs. 3
H&E	0.77 (0.68 to 0.85)	0.76 (0.66 to 0.85)	0.79 (0.71 to 0.87)
p-value	p<0.001	p<0.001	p<0.001
ER	0.80 (0.64 to 0.95)	0.87 (0.74 to 1.00)	0.83 (0.65 to 1.00)
p-value	p<0.001	p<0.001	p<0.001
PR	0.69 (0.29 to 1.00)	0.86 (0.59 to 1.00)	0.86 (0.59 to 1.00)
p-value	p<0.001	p<0.001	p<0.001
HER-2	0.86 (0.73 to 0.99)	0.79 (0.64 to 0.94)	0.79 (0.62 to 0.95)
p-value	p<0.001	p<0.001	p<0.001
Ki-67	0.67 (0.34 to 1.00)	0.88 (0.64 to 1.00)	0.78 (0.50 to 1.00)
p-value	p<0.001	p<0.001	p<0.001

CLM=conventional light microscope; CI=confidence interval; H&E=hematoxylin and eosin; ER=estrogen receptor; PR=progesterone receptor; HER2=human epidermal growth factor receptor 2; κ=kappa coefficient

Table 6. Interobserver agreement of WSI diagnoses

	Interobserver agreement of WSI between the observers k-value (95% CI)		
	Observer 1 vs. 2	Observer 1 vs. 3	Observer 2 vs. 3
H&E	0.80 (0.71 to 0.89)	0.75 (0.66 to 0.85)	0.78 (0.70 to 0.87)
p-value	p<0.001	p<0.001	p<0.001
ER	0.84 (0.69 to 0.98)	0.80 (0.63 to 0.96)	0.80 (0.64 to 0.96)
p-value	p<0.001	p<0.001	p<0.001
PR	0.86 (0.69 to 1.00)	0.69 (0.31 to 1.00)	0.48 (–0.16 to 0.98)
p-value	p<0.001	p<0.001	p<0.001
HER-2	0.71 (0.54 to 0.88)	0.84 (0.67 to 1.00)	0.76 (0.58 to 0.94)
p-value	p<0.001	p<0.001	p<0.001
Ki-67	0.63 (0.28 to 0.99)	0.86 (0.59 to 1.00)	0.77 (0.46 to 1.00)
p-value	p<0.001	p<0.001	p<0.001

WSI=whole slide imaging; CI=confidence interval; H&E=hematoxylin and eosin; ER=estrogen receptor; PR=progesterone receptor; HER2=human epidermal growth factor receptor 2; κ=kappa coefficient

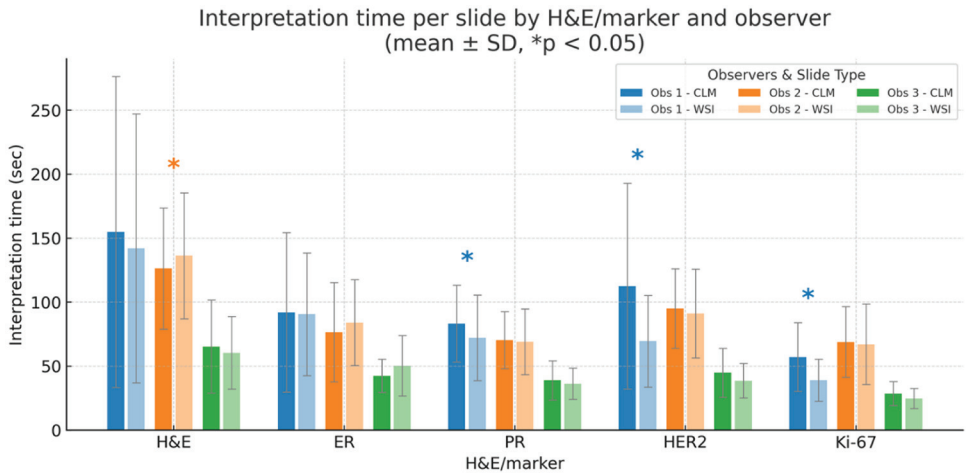

Figure 5. Comparison of interpretation time (second) by HE and IHC markers between CLM and WSI among three observers.

CLM=conventional light microscopy; WSI=whole slide imaging; H&E=hematoxylin and eosin; SD=standard deviation; Obs=observer

for all observers was H&E evaluation. Interpretation times for ER and PR were similar among modalities, with only slight variations. HER2 evaluation revealed the most substantial temporal disparity, with WSI demonstrating a quicker performance for Observer 1 (p<0.05). The interpretation of Ki-67 was the most rapid among all markers, and once more, WSI appeared to minimize interpretation time, particularly for Observer 1.

Despite inter-observer variability, the general trend is that WSI provides comparable or expedited interpretation relative to traditional microscopy, especially for PR, HER2, and Ki-67 assessments (Figure 5).

Discussion

This study demonstrates that WSI can be reliably used for primary diagnosis of breast CNB specimens in a resource-limited tertiary-care setting. Overall, WSI showed strong agreement with CLM across the B1-B5 diagnostic spectrum and for key breast biomarkers (ER, PR, HER2, and Ki-67). Clinically significant discordances were infrequent, supporting the practical safety of adopting WSI for routine breast CNB interpretation when an appropriate local validation process is performed.

With respect to H&E interpretation, diagnostic agreement between WSI and CLM was consistently high across observers. Most discrepancies were minor and unlikely to affect patient management, whereas

major discordances were rare and primarily occurred at diagnostic thresholds (e.g., benign versus suspicious/malignant), a known challenge in breast CNB interpretation, irrespective of platform. Even though these incidents may not be statistically significant in a cohort of this size, they should still be carefully considered in routine practice and minimized through focused training, standardized review procedures, and safety precautions, including secondary review for high-impact diagnostic thresholds. Overall, these findings support WSI as a reliable alternative to CLM for routine breast CNB assessment.

The interpretation of biomarkers on WSI showed substantial concordance with CLM, hence affirming the viability of employing digital slides for clinically relevant breast biomarker reporting. ER scoring demonstrated significant consistency across platforms, aligning with its strong interpretative consistency in standard practice. Conversely, PR and Ki-67 exhibited slightly more variability among specific observer pairs, potentially indicating intrinsic biological heterogeneity and subjective thresholding in borderline or weakly positive cases[16]. HER2 interpretation typically demonstrated robust concordance. However, a limited number of cases displayed one-level shifts (e.g., from entirely negative to low-level positivity), indicating that subtle variations in visual perception on digital slides may affect threshold interpretation in ambiguous or faint staining scenarios. These findings underscore the necessity of uniform on-screen calibration and consensus training in the digital interpretation of biomarkers, especially for those with significant interpretative subjectivity.

These findings align with previous validation studies that support the use of WSI for breast pathology primary diagnostics. High concordance between WSI and CLM for breast CNB was reported by the University of Nebraska Medical Center, confirming cross-platform diagnostic equivalency[11]. Comparing high-resolution WSI to traditional microscopy, research from the University of Miami also showed minimal intra-observer variability[12]. Furthermore, a recent multi-observer study supported digital evaluation for prognostic and predictive markers by finding no significant differences between CLM and WSI for the assessment of histologic grade and breast biomarkers in breast CNB samples[23]. This study's concordance patterns support these reports and provide additional data from a tertiary-care facility in Southeast Asia, an institutional setting that is still underrepresented in the literature on digital pathology.

In this study group, observer experience has a greater impact on interpretation efficiency than imaging modalities. In certain biomarker evaluations, WSI interpretation was not slower than CLM, and observers with balanced experience with both platforms showed comparable performance between WSI and CLM. On the other hand, observers who had long relied on traditional microscopy tended to do better with CLM, indicating a learning curve and transition phase when using digital operations. These findings corroborate earlier research showing that exposure and practice enhance digital efficiency[13]. When proper training and quality assurance procedures are in place, digital workflows may offer useful benefits in navigation, standardized viewing, and quick comparison across fields, which can enable uniform evaluation. These benefits go beyond time measurements.

This study has limitations. First, it was conducted at a single tertiary-care institution with a modest cohort size. Although the number of cases met the minimum validation expectations recommended by the CAP recommendation[8], larger multi-center studies are needed to improve generalizability and to evaluate performance across a wider range of practice settings. Second, all cases were scanned using a single WSI platform; therefore, the findings may not be directly transferable to other scanners or viewing software, and potential inter-platform variability warrants further investigation. Third, biomarker interpretation was performed by visual assessment without the use of digital image analysis. Variations in color rendering and monitor calibration may influence the perception of weak staining on digital slides, as reported in prior studies[24,25]. Fourth, interpretation time was not adjusted for lesion complexity. Because breast CNB specimens encompass a broad spectrum of diagnostic difficulty, differences in case mix may have influenced efficiency metrics irrespective of the viewing modality. Future research utilizing case-mix adjustment or predetermined complexity categorization would offer a more reliable evaluation of diagnostic effectiveness. Finally, this validation focused on CNB specimens rather than excision specimens. Additional studies using excision material may be valuable for assessing performance in scenarios requiring more extensive architectural evaluation and histopathologic detail.

Conclusion

This study shows that WSI is practicable,

reliable, and safe for primary breast core needle biopsies in clinical practice. Digital workflows can be used without sacrificing diagnostic accuracy or patient safety, as shown by the high overall concordance and outstanding inter- and intra-observer agreement between digital and glass slide interpretations for morphologic assessment and key biomarkers. This study's findings support the use of WSI in routine breast pathology workflows as an alternative to microscopy. This validation is crucial to the wider adoption of digital pathology, which will enable more efficient, flexible, and technologically sophisticated diagnostic services while preserving the highest accuracy and consistency in patient care.

What is already known about this topic?

- WSI is a practical alternative to conventional microscopy, providing comparable image quality and advantages such as remote access, workflow flexibility, and digital consultation.
- Diagnostic performance depends on scanner quality. High-resolution systems can support primary diagnosis, whereas others may introduce artifacts, requiring validation under standards such as the CAP WSI guideline.
- Prior research shows strong WSI and glass slide concordance for breast CNB. However, Campbell et al. (2014) did not use kappa statistics or B1 to B5 diagnostic categories[11].
- Choi et al. (2024) confirmed WSI accuracy for Nottingham grading and breast biomarkers, though concordance for PR and HER2 was lower, and the study focused solely on carcinoma[23].
- In Thailand, low-cost WSI systems have shown diagnostic accuracy comparable to glass slides for lymph node metastasis, supporting feasibility in resource-limited settings[3].

What does this study add?

- First WSI validation research for breast CNB in Southeast Asia and a resource-limited context shows glass slide accuracy.
- First use of the entire B1 to B5 diagnostic classification for CAP 2021-compliant minor and major discordance assessment.
- WSI consistently has greater IHC staining, especially for HER2, indicating clinically significant discordance and the requirement for scanner-specific calibration.
- Strong intra-observer agreement across modalities confirms that WSI preserves crucial morphological detail.
- WSI is suitable for regular diagnostics and pathology training since interobserver patterns are comparable across expertise levels.

Acknowledgment

The authors express their gratitude for the financial support from the Faculty of Medicine, Thammasat University Hospital, which facilitated this study. The authors express their heartfelt gratitude to the Section of Pathology at Thammasat University Hospital for granting access to biopsy specimens and for technical support during the research procedure.

Authors' contributions

WK: Conceptualization, methodology, investigation, formal analysis, data curation, writing-original draft. WH: Conceptualization, methodology, investigation, formal analysis, writing-review and editing, supervision, project administration. AP: Methodology, investigation, formal analysis, data curation, writing-review and editing, supervision. All authors read and approved the final manuscript.

Data availability statement

The data supporting the findings of this study are not publicly available because they contain institutional pathology materials and potentially identifiable patient-related information. De-identified data may be made available from the corresponding author upon reasonable request, subject to approval by Thammasat University Hospital and the relevant ethics requirements.

Ethics approval and consent to participate

This study was approved by the Human Research Ethics Committee of Thammasat University (Medicine) (COA No. 092/2024). The requirement for informed consent was waived due to the retrospective nature of the study and the use of archived specimens/materials.

Clinical trial registration

Not applicable. This study was not a clinical trial.

Use of artificial intelligence

The authors used ChatGPT-4 only for minor language improvement and grammar editing during manuscript preparation. All output was reviewed and revised by the authors, who take full responsibility for the final content. AI was not used for data analysis, diagnostic interpretation, or scientific conclusion-making.

Conflicts of interest

The authors declare no conflicts of interest related to this study.

References

1. Meric-Bernstam F, Pollock RE. Oncology. In: Brunicardi FC, Andersen DK, Billiar TR, Dunn DL, Hunter JG, Matthews JB, et al., editors. Schwartz's principles of surgery. 10th ed. New York. McGraw-Hill Education; 2015. p. 273-320.
2. Klimberg VS, Hunt KK. Diseases of the breast. In: Townsend CM, Evers BM, Beauchamp RD, Mattox KL, editors. Sabiston's textbook of surgery. 21st ed. St. Louis: Elsevier; 2022. p. 809-55.
3. Luechakiettisak P, Rungkaew P. Breast biopsy: Accuracy of core needle biopsy compared with excisional or incisional biopsy: A prospective study. Thai Journal Surgery 2008;29:6-10.
4. Cadavid-Fernández N, Carretero-Barrio I, Moreno-Moreno E, Rodríguez-Villena A, Palacios J, Pérez-Mies B. The role of core needle biopsy in diagnostic breast pathology. Revista de Senología y Patología Mamaria - Journal of Senology and Breast Disease 2022;35 Suppl 2:S3-12.
5. Dahabreh IJ, Wieland LS, Adam GP, Halladay C, Lau J, Trikalinos TA. AHRQ comparative effectiveness reviews. Core needle and open surgical biopsy for diagnosis of breast lesions: An update to the 2009 report. Rockville, MD: Agency for Healthcare Research and Quality (US); 2014.
6. Fraggetta F, L'Imperio V, Ameisen D, Carvalho R, Leh S, Kiehl TR, et al. Best practice recommendations for the implementation of a digital pathology workflow in the anatomic pathology laboratory by the European Society of Digital and Integrative Pathology (ESDIP). Diagnostics (Basel) 2021;11:2167. doi: 10.3390/diagnostics11112167.
7. Cross S, Furness P, Igali L, Snead D, Treanor D. Best practice recommendation for implementing digital pathology. London: The Royal College of Pathologists; 2018.
8. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Arch Pathol Lab Med 2013;137:1710-22.
9. Mastrosimini MG, Eccher A, Nottegar A, Montin U, Scarpa A, Pantanowitz L, et al. WSI validation studies in breast and gynecological pathology. Pathol Res Pract 2022;240:154191. doi: 10.1016/j.prp.2022.154191.
10. Evans AJ, Brown RW, Bui MM, Chlipala EA, Lacchetti C, Milner DA, et al. Validating whole slide imaging systems for diagnostic purposes in pathology. Arch Pathol Lab Med 2022;146:440-50.
11. Campbell WS, Hinrichs SH, Lele SM, Baker JJ, Lazenby AJ, Talmon GA, et al. Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. Hum Pathol 2014;45:1713-21.
12. Reyes C, Ikpatt OF, Nadji M, Cote RJ. Intra-observer reproducibility of whole slide imaging for the primary diagnosis of breast needle biopsies. J Pathol Inform 2014;5:5. doi: 10.4103/2153-3539.127814.
13. Kantasiripitak C, Laohawetwanit T, Apornvirat S, Niemnapa K. Validation of whole slide imaging for frozen section diagnosis of lymph node metastasis: A retrospective study from a tertiary care hospital in Thailand. Ann Diagn Pathol 2022;60:151987. doi: 10.1016/j.anndiagpath.2022.151987.
14. Ellis IO, Humphreys S, Michell M, Pinder SE, Wells CA, Zakhour HD. Best practice No 179. Guidelines for breast needle core biopsy handling and reporting in breast screening assessment. J Clin Pathol 2004;57:897-902.
15. Fitzgibbons PL, Connolly JL. Template for reporting results of biomarker testing of specimens from patients with carcinoma of the breast. Northfield, Illinois: College of American Pathologists (CAP); March 2023.
16. Salama AM, Hanna MG, Giri D, Kezlarian B, Jean MH, Lin O, et al. Digital validation of breast biomarkers (ER, PR, AR, and HER2) in cytology specimens using three different scanners. Mod Pathol 2022;35:52-9.
17. Allison KH, Hammond MEH, Dowsett M, McKernin SE, Carey LA, Fitzgibbons PL, et al. Estrogen and progesterone receptor testing in breast cancer: ASCO/CAP guideline update. J Clin Oncol 2020;38:1346-66.
18. National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: Breast cancer. Version 5.2025. Plymouth Meeting, PA: NCCN; 2025. Available from: https://www.nccn.org/.
19. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, 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. Arch Pathol Lab Med 2018;142:1364-82.
20. Bardia A, Hu X, Dent R, Yonemori K, Barrios CH, O'Shaughnessy JA, et al. Trastuzumab deruxtecan after endocrine therapy in metastatic breast cancer. N Engl J Med 2024;391:2110-22.
21. Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, et al. Assessment of Ki67 in breast cancer: Updated recommendations from the International Ki67 in Breast Cancer Working Group. J Natl Cancer Inst 2021;113:808-19.
22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
23. Choi JE, Kim KH, Lee Y, Kang DW. Digital validation in breast cancer needle biopsies: Comparison of histological grade and biomarker expression assessment using conventional light microscopy,

whole slide imaging, and digital image analysis. J Pers Med 2024;14:312. doi: 10.3390/jpm14030312.
24. Nassar A, Cohen C, Albitar M, Agersborg SS, Zhou W, Lynch KA, et al. Reading immunohistochemical slides on a computer monitor--a multisite performance study using 180 HER2-stained breast carcinomas. Appl Immunohistochem Mol Morphol 2011;19:212-7.
25. Kondo Y, Iijima T, Noguchi M. Evaluation of immunohistochemical staining using whole-slide imaging for HER2 scoring of breast cancer in comparison with real glass slides. Pathol Int 2012;62:592-9.