AI-driven Pathomics in Breast Cancer Metastasis Research: A Postprint

Authors: Fang Xiaoqing, Lü Qingjie, Lü Qingjie

Date: 2025-11-10T00:00:00+00:00

Abstract

Breast cancer constitutes one of the major global diseases. Patient survival rates are substantially reduced when cancer cells spread to adjacent and non-adjacent tissues and organs, causing distant metastasis. Therefore, early diagnosis and prediction of malignant potential are critically important for treatment planning and prognosis in breast cancer patients. Amidst rapid artificial intelligence development, the maturation of pathomics technology has made significant contributions to precision medicine advancement. This article systematically and comprehensively reviews progress in pathomics across three domains: diagnosis of breast cancer metastasis, predictive modeling, and tumor heterogeneity research, by synthesizing relevant literature. Artificial intelligence algorithms can automatically identify key structures such as tumor cells, stromal components, and immune cells in pathological sections, markedly improving the efficiency and accuracy of pathomics feature extraction. Concurrently, multimodal artificial intelligence models integrating diverse information including clinical data and genomic data have further enhanced pathomics performance in breast cancer metastasis diagnosis and prediction, making important contributions to precision medicine. However, due to technical constraints and limitations such as data sharing restrictions from patient privacy protection, this field requires further exploration. This article aims to provide more instructive reference foundations for precision therapy of breast cancer patients and to promote deep application of artificial intelligence and pathomics technology in breast cancer metastasis research and clinical practice.

Full Text

Advances in Integrative Pathomics Enhanced by Artificial Intelligence for Breast Cancer Metastasis Research

FANG Xiaoqing, LYU Qingjie*

Department of Pathology, Shengjing Hospital of China Medical University, Shenyang 110000, China

Corresponding author: LYU Qingjie, Chief physician/Doctoral supervisor; E-mail: lvqjie@163.com

Abstract

Breast cancer remains one of the most prevalent and life-threatening malignancies worldwide. The prognosis of patients markedly worsens once cancer cells metastasize to regional or distant sites. Thus, early detection and accurate assessment of metastatic potential are critical for optimizing treatment strategies and improving clinical outcomes. With the rapid evolution of artificial intelligence (AI), digital pathology has emerged as a powerful tool, significantly advancing the field of precision oncology. This review provides a comprehensive overview of recent advancements in the application of digital pathology in breast cancer metastasis, focusing on three major domains: diagnostic accuracy, predictive modeling, and tumor heterogeneity analysis. AI-driven algorithms enable automated and high-throughput recognition of key histopathological components—such as tumor cells, stromal architecture, and immune infiltrates—substantially enhancing the efficiency and reproducibility of feature extraction. Furthermore, the integration of pathological imaging with multimodal data sources, including clinical parameters and genomic profiles, through advanced AI models has demonstrated improved performance in metastatic risk stratification and outcome prediction. Despite these promising developments, challenges such as technical constraints and limitations in data sharing due to patient privacy concerns continue to hinder broader clinical translation. This review aims to provide a valuable reference for the development of personalized therapeutic approaches and to promote the integration of AI-assisted digital pathology in the management of metastatic breast cancer.

Key words: Breast cancer; Artificial intelligence; Machine learning; Pathomics; Metastasis

Funding: National Natural Science Foundation of China (82072095); Liaoning Provincial Science and Technology Plan Joint Plan (Fund) Project (2023JH2/101800019); Liaoning Province "Xingliao Talent Plan" Medical Master Project (XLYC2412006)

Citation: FANG X Q, LYU Q J. Advances in integrative pathomics enhanced by artificial intelligence for breast cancer metastasis research[J]. Chinese General Practice, 2025. DOI:10.12114/j.issn.1007-9572.2025.0247. [Epub ahead of print]

According to 2022 global cancer statistics, breast cancer (BC) accounted for 2.309 million new cases and over 665,000 deaths. The molecular heterogeneity of metastatic lesions not only increases the complexity of treatment selection but also drives up healthcare costs significantly, necessitating the development

of more precise assessment systems. Histopathology (HP) examination serves as the gold standard for metastasis diagnosis, relying on hematoxylin-eosin (HE) staining morphology and immunohistochemical marker analysis. However, technical bottlenecks such as long detection cycles, high costs, and dynamic loss of tumor markers constrain the precision of metastatic lesion tracing and therapeutic efficacy evaluation. Innovations in digital pathology have created breakthrough opportunities for this field. Whole slide imaging (WSI) technology preserves complete histomorphological details, ushering pathology diagnosis into a new era of quantitative analysis.

Against this backdrop, the integration of pathomics and artificial intelligence (AI) has substantially advanced precision medicine. This technology employs high-throughput feature extraction algorithms to capture quantitative indicators such as morphological topology, cellular spatial distribution, and microenvironmental heterogeneity from multimodal pathological images including HE staining and immunohistochemistry. When combined with AI to parse tumor evolution patterns, it provides a new paradigm for elucidating metastasis mechanisms and enabling personalized therapy. This review systematically examines the latest advances in pathomics technology for BC metastasis research, following the logical framework of traditional pathological assessment → pathomics feature system construction → multi-omics applications, with a focus on AI application strategies in diagnosis, prediction, and tumor heterogeneity studies, aiming to provide theoretical foundations for optimizing BC precision diagnosis and treatment pathways.

1 Literature Search Strategy

We conducted computerized searches of PubMed, CNKI, and Web of Science from database inception to May 2025. Search terms included: "breast cancer," "digital pathology," "computational pathology," "histology," "histopathology," "whole slide image," "metastasis," "pathomics," as well as corresponding Chinese terms. Inclusion criteria comprised literature related to pathomics and BC metastasis diagnosis, prediction, and metastatic lesion heterogeneity. Exclusion criteria included irrelevant topics and unavailable full texts.

2 AI Application Strategies in BC Metastasis Pathomics Research

AI, a branch of computer science, encompasses robotics, speech recognition, image recognition, natural language processing, and expert systems. Machine learning constitutes a crucial component of AI, with primary methods including supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Currently, supervised learning is widely applied in BC metastasis research. Supervised learning data contains target labels, enabling direct feature extraction from pathological images to train optimal models for predicting unknown data. However, this approach requires human expert an-

notation of each pathological image, incurring substantial labor costs. Semi-supervised and unsupervised learning methods can effectively address this limitation. In unsupervised learning algorithms, input data lacks corresponding labels; the method discovers latent structures or features through clustering of unlabeled raw data, optimizing models based on similarity calculations between training samples. Unsupervised learning can also uncover novel metastasis features for tumor heterogeneity quantification and subtype discovery. For scenarios with scarce annotated data, semi-supervised learning combines limited labeled data with large amounts of unlabeled data for training, significantly reducing labor costs while improving model generalization capability.

In the process of building BC metastasis prediction models using machine learning, the pathomics feature system primarily involves the following steps: extraction of regions of interest (ROI) from whole slide images, acquisition of cellular and tissue structures through preprocessing and segmentation, extraction of multidimensional features including morphological, textural, spatial distribution, and deep learning features, followed by feature selection and dimensionality reduction to ultimately construct a unified feature matrix for model training and clinical tasks. This system integrates image analysis with AI to achieve precise quantification of pathological images and support clinical decision-making. The first step in pathomics research is extracting metastasis-related features from BC tissue slice images. Compared with traditional techniques that rely on handcrafted features (i.e., human selection and design of diagnostically useful features), deep learning can automatically learn and extract useful features from large amounts of high-quality data, making it more accurate and efficient for representation learning and predictive model construction.

3 Application of Pathomics in BC Metastasis Diagnosis

Lymph node metastasis and distant metastasis in BC are key factors leading to poor prognosis and increased mortality, representing the most important predictors of overall recurrence and survival. In recent years, with the development of AI algorithms and increased attention to BC, pathomics has been increasingly applied to BC metastasis diagnosis model research. For convenient comparison of model fundamentals, we summarize each model' s publication year, modeling methodology, primary problem addressed, and advantages/disadvantages in Table 1 .

3.1 Classification of Metastases with Unknown Primary Origin

In BC metastasis diagnosis, HP assessment has long relied on HE staining morphology and immunohistochemical marker analysis, yet faces multiple clinical challenges. First, the detection process incurs high economic and time costs (average turnaround time $\geq$ 7 working days). Second, dynamic loss of tumor markers due to heterogeneity (e.g., post-treatment downregulation of estrogen/progesterone receptor expression) may complicate metastatic lesion tracing. Global epidemiological studies show that approximately 2%-5% of metastatic

malignancies have unclear primary sites, with these patients experiencing significantly worse outcomes due to inability to match organ-specific therapies.

Addressing these complex challenges in diagnosing BC metastasis origin, AI technology demonstrates significant potential through multimodal data fusion and deep learning models. CHEN et al. developed a handcrafted-deep learning hybrid model combining manual features (nuclear morphology, texture quantification) with a DenseNet classifier, achieving superior classification performance in identifying primary origins of liver metastases from colon, esophageal, BC, and pancreatic cancers. The highest area under the curve (AUC) reached 0.94 for colorectal metastasis classification, with heatmaps localizing morphological similarities between primary and metastatic regions, providing an interpretable pathway for tumor heterogeneity research. At the whole-slide analysis level, LU et al. developed the Tumor Origin Assessment (TOAD) model, training on over 30,000 HE slides to achieve 95.5% Top-3 accuracy in predicting known primary cases in the training set, validating the universality of deep learning in metastasis tracing. For BC pleural metastasis (incidence 15%-30%) cytology diagnosis challenges, PARK et al. constructed an AI model based on 596 cytology WSIs, achieving 81.13% accuracy in plaque classification tasks, surpassing the average pathologist performance (72.49%) and providing a new approach to improve malignant pleural effusion diagnosis sensitivity (traditional methods only 58%), though external validation is needed. TIAN et al. built the TORCH (Tumor Origin differentiation using cytology and Histology) model based on 57,220 pleural effusion and ascites cytology images, achieving systematic localization of cancers of unknown primary (CUP) with 82.6% first-choice accuracy. The model' s diagnostic efficacy (AUC=0.969) significantly exceeded pathologists (AUC=0.813), and model-assisted therapy extended median patient survival by 10 months, establishing a new paradigm for CUP diagnosis and treatment. These studies demonstrate that AI, by integrating quantitative pathological features with clinical data, is gradually overcoming temporal, spatial, and precision limitations of traditional diagnosis, driving BC metastasis management toward precision medicine.

4 Application of Pathomics in Predicting BC Recurrence and Metastasis

BC recurrence refers to the reappearance of BC after initial treatment and remission. Recurrence can be categorized as local or metastatic: local recurrence indicates spread to nearby breast tissue and lymph nodes, while metastatic recurrence refers to tumor cells migrating to distant organs such as bone, lung, liver, or brain. In recent years, increasing research has focused on constructing BC recurrence and metastasis prognostic models. We summarize these studies in Table 2 .

4.1 Predicting Lymph Node Metastasis

Lymph node metastasis is not merely a local dissemination event but a "sentinel signal" of systemic metastasis. Clinical data show that lymph node-positive patients have a 3.2-fold higher distant metastasis risk within 5 years compared to negative groups (HR=3.2, 95%CI=2.8-3.7), with pathomics features of metastatic lesions (e.g., mitotic index >10/high-power field, necrotic area proportion >30%) enabling further prognostic stratification.

XU et al. utilized WSIs from 1,058 patients to develop the first supervised learning model for predicting SLN metastasis from primary tumor WSIs, achieving good accuracy (0.831). YU et al. proposed a Prototypical Multiple Instance Learning (PMIL) framework that parsed WSIs through weak supervision, achieving an AUC of 0.984 for predicting BC lymph node metastasis with 92% sensitivity for micrometastases (<0.2mm), tripling the efficiency of traditional manual annotation. This technology effectively alleviates the shortage of senior pathologists by focusing on "suspected metastasis hotspots" (e.g., nuclear atypia aggregation areas). PARK et al. employed an innovative multimodal unsupervised learning method to preoperatively predict axillary lymph node metastasis by identifying pathological imaging patterns associated with metastasis status, such as micropapillary growth, infiltration patterns, and necrosis, achieving external validation across 5 cohorts with a maximum AUC of 0.801 (95%CI=0.728-0.873).

4.2 Predicting Distant Metastasis

BC metastatic recurrence is a multi-step dynamic process involving enhanced tumor cell invasion and migration (e.g., epithelial-mesenchymal transition), immune microenvironment remodeling (e.g., T cell exhaustion), and selective barrier breakthrough (e.g., blood-brain barrier penetration). This complexity drives rapid development of multimodal prediction models. YAO et al. integrated pathological image features, clinical staging, and gene mutation profiles (e.g., TP53, PIK3CA) from 198 TCGA patients through weak supervision to construct a deep feature fusion model, achieving an AUC of 0.82 for metastasis risk prediction. YANG et al. extracted HE image texture features (e.g., nuclear-cytoplasmic ratio variation) using CNN and combined them with lymph node metastasis status to build a prognostic model with 67% sensitivity, 83% specificity, and AUC of 0.72, effectively achieving recurrence risk stratification. However, this study only included 127 single-center patients, limiting sample size and requiring cross-center generalization validation. LIU et al. developed a multiscale risk assessment framework by fusing low-level color features (RGB mean/skewness) with wavelet multi-subband texture matrices (WMCM), demonstrating stable performance with AUCs of 0.75 and 0.72 in internal and external validation sets, respectively.

These results demonstrate that multi-omics integration can enhance model biological interpretability but requires overcoming technical limitations of small

samples and single modalities. Studies have shown that tumor-infiltrating lymph node reactions correlate positively with BC prognosis. VERGHESE et al. constructed a supervised deep learning model to automatically quantify germinal center numbers and lymph sinus area on digital WSIs, further predicting metastasis-free survival. Traditional pathomics relies on high-resolution WSIs and professional scanning equipment, which is costly and has limited feature interpretability. BASAAD et al. innovatively introduced a Bidirectional Encoder Representations from Transformers (BERT) model and Graph Neural Network (GNN) to directly parse HP report text (e.g., "invasive ductal carcinoma with vascular invasion") and constructed the BERT-GNN Method for Breast Cancer (BG-MBC) model to predict distant metastasis risk, achieving an AUC of 0.98. This technology improves prediction accuracy by capturing semantically critical information and complex relationships between different medical features, highlighting the advantages of large language models (LLMs).

5 Tumor Heterogeneity Quantification and Subtype Discovery

The complexity of BC metastasis stems from its high heterogeneity. AI systematically parses intra-tumoral molecular and morphological diversity by integrating multi-omics data with pathomics features, driving precise subtype classification. Genomic studies have revealed five classic molecular subtypes of BC (luminal A, luminal B, HER2-enriched, basal-like, and normal-like) and their molecular mechanisms, while also discovering frequent polyclonal coexistence within the same tumor. Notably, molecular subtypes may dynamically evolve between primary and metastatic lesions: basal-like gene signatures in primary lesions have higher predictive value for post-metastasis overall survival, while proliferation-related gene expression in metastases better reflects patient prognosis.

Recent advances in AI-based pathological image analysis provide new tools for quantitative assessment of tumor spatial heterogeneity. YU et al. automatically segmented digital slides, extracted 256$×$256 pixel image patches from foreground at user-specified magnifications, and integrated lncRNA expression profiles, immune cell infiltration scores, and clinical information to construct four immune-metabolic subtypes: immune-active (fatty acid metabolism activation), immune-excluded (amino acid metabolism enrichment), immune-dysfunctional (glucose metabolism enrichment), and immune-desert (folate metabolism enrichment). The deep learning model built on HE slides achieved an AUC of 0.92 for predicting these subtypes, providing a potential stratification tool for immunotherapy. AI models can automatically quantify a range of spatial structural features from HE images, including tumor cell nuclear density distribution, atypia scores, stroma-tumor interface clarity, tumor-infiltrating lymphocyte (TIL) spatial density, glandular arrangement morphology, and necrotic area distribution patterns. These image features can be used not only for subtype discrimination but also for revealing potential prognostic markers. ZHANG

et al. found that samples with higher tumor cell proportions in WSIs were closely associated with increased BC recurrence risk. Further analysis revealed that high-risk groups showed significantly reduced intercellular spacing, indicating higher tumor cell aggregation. This enhanced spatial aggregation correlates significantly with increased tumor invasiveness and higher recurrence rates.

Since single-time-point biopsies cannot comprehensively capture clonal evolution, AI models fail to reflect tumor dynamic spatial heterogeneity. Moreover, cross-platform image feature extraction shows substantial deviation and lacks unified standards, with insufficient clinical validation. Therefore, future research should focus on constructing cross-center, ten-thousand-case-level multimodal BC atlas databases, integrating spatial transcriptomics technology, and developing dynamic AI models with temporal dimensions to achieve the leap from static image representation to spatiotemporal heterogeneity modeling, advancing BC management from precision diagnosis to personalized intervention.

6 Summary and Outlook

Currently, existing machine learning models demonstrate favorable performance in BC metastasis diagnosis and malignant potential prediction, offering advantages such as high diagnostic efficiency, significant reduction in pathologist workload, and narrowing the diagnostic capability gap between experts and non-experts. However, several limitations persist in this field: (1) Study design limitations: Most current models are built on single-center, small-sample retrospective data, generally lacking external validation and raising questions about their clinical generalizability. (2) Unresolved key scientific questions: At the pathological slice analysis level, critical issues such as BC-specific organ metastasis prediction and the impact of molecular subtype differences between metastatic and primary lesions on treatment response remain insufficiently addressed. (3) Data quality constraints on model development: Classic supervised deep learning models rely on high-quality annotated data, yet high-quality labels are scarce in reality. Although studies have attempted to introduce methods like label-cleaning MIL to optimize coarse-grained annotations and apply unsupervised augmentation techniques to improve image signal-to-noise ratios, standardized and generalizable technical systems have yet to be established. (4) Insufficient model interpretability: Most current AI models remain "black boxes," making it difficult to establish explicit associations between imaging features and tumor biological mechanisms, limiting their credible application in clinical decision support systems. (5) The gap from "usable" to "practical" awaits clinical translation loop construction: Although laboratory models based on AI, omics analysis, and multimodal fusion have achieved remarkable results in recent years, their practical clinical application still faces numerous challenges. Systematic planning is needed from three aspects: multi-center retrospective validation, prospective observational studies, and interventional clinical trials, while promoting seamless integration of AI models into actual clinical workflows.

Current prediction models have achieved good predictive performance, but

the key obstacle to clinical translation is the high requirements for software/hardware resources and personnel configuration. To facilitate widespread adoption, models should have lightweight deployment capabilities, supporting operation on conventional hospital servers or edge computing devices. Future research should focus on the following algorithmic innovations and system constructions: First, addressing current bottlenecks of small samples and weak annotations, we should strengthen development of few-shot learning, semi-supervised/self-supervised learning, multiple instance learning, and transfer learning methods to enhance model generalization and robustness under limited data conditions. Second, in multimodal data fusion, we need to explore more efficient mechanisms for multi-source heterogeneous data integration, such as Transformer-based multimodal feature extraction and cross-modal representation alignment techniques, to enhance synergistic reasoning capabilities in the "pathology-clinical data-omics information" multidimensional space. Additionally, improving model interpretability will be key to achieving clinical acceptability and trust. Future efforts should enhance research on biological mechanism-based feature visualization, causal inference, and model uncertainty assessment to increase transparency and traceability of results.

In summary, to achieve a closed-loop path from laboratory research to clinical translation, we must construct a multi-center, large-sample, multimodal collaborative data foundation, develop AI models with broad adaptability and good interpretability, and promote intelligent closed-loop management of BC from precision diagnosis to personalized treatment through a "clinical-pathology-omics" three-dimensional explanatory framework.

References

[1] BRAY F, LAVERSSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2024, 74(3): 229-263. DOI:10.3322/caac.21834.

[2] PARK M, KIM D, KO S, et al. Breast cancer metastasis: mechanisms and therapeutic implications[J]. Int J Mol Sci, 2022, 23(12): 6806. DOI:10.3390/ijms23126806.

[3] DING Q Q, HUO L, PENG Y, et al. Immunohistochemical markers for distinguishing metastatic breast carcinoma from other common malignancies: update and revisit[J]. Semin Diagn Pathol, 2022, 39(5): 313-321. DOI:10.1053/j.semdp.2022.04.002.

[4] HANNA M G, PARWANI A, SIRINTRAPUN S J. Whole slide imaging: technology and applications[J]. Adv Anat Pathol, 2020, 27(4): 251-259. DOI:10.1097/PAP.0000000000000273.

[5] MELO R C N, RAAS M W D, PALAZZI C, et al. Whole slide imaging and its applications to histopathological studies of liver disorders[J]. Front Med, 2020, 6: 310. DOI:10.3389/fmed.2019.00310.

[6] WAQAS A, BUI M M, GLASSY E F, et al. Revolutionizing digital pathology with the power of generative artificial intelligence and foundation models[J]. Lab Investig, 2023, 103(11): 100255. DOI:10.1016/j.labinv.2023.100255.

[7] DEO R C. Machine learning in medicine[J]. Circulation, 2015, 132(20): 1920-1930. DOI:10.1161/circulationaha.115.001593.

[8] PAIXÃO G M M, SANTOS B C, ARAUJO R M, et al. Machine learning in medicine: review and applicability[J]. Arq Bras Cardiol, 2022, 118(1): 95-102. DOI:10.36660/abc.20200596.

[9] BHALLA K, XIAO Q, LUNA J M, et al. Radiologic imaging biomarkers in triple-negative breast cancer: a literature review about the role of artificial intelligence and the way forward[J]. BJR Artif Intell, 2024, 1(1): ubae016. DOI:10.1093/bjrai/ubae016.

[10] CHANG J M, LEUNG J W T, MOY L, et al. Axillary nodal evaluation in breast cancer: state of the art[J]. Radiology, 2020, 295(3): 500-515. DOI:10.1148/radiol.2020192534.

[11] GOWN A M, FULTON R S, KANDALAFT P L. Markers of metastatic carcinoma of breast origin[J]. Histopathology, 2016, 68(1): 86-95. DOI:10.1111/his.12877.

[12] BORCH W R, MONACO S E. Current approach to undifferentiated immunohistochemical stains[J]. Arch Pathol Lab Med, 2023, 147(12): 1364-1373. DOI:10.5858/arpa.2022-0459-RA.

[13] LEE M S, SANOFF H K. Cancer of unknown primary[J]. BMJ, 2020, 371: m4050. DOI:10.1136/bmj.m4050.

[14] CHEN C H, LU C, VISWANATHAN V, et al. Identifying primary tumor site of origin for liver metastases via a combination of handcrafted and deep learning features[J]. J Pathol Clin Res, 2024, 10(1): e344. DOI:10.1002/cjp2.344.

[15] LU M Y, CHEN T Y, WILLIAMSON D F K, et al. AI-based pathology predicts origins for cancers of unknown primary[J]. Nature, 2021, 594(7861): 106-110. DOI:10.1038/s41586-021-03512-4.

[16] CHEN M T, SUN H F, ZHAO Y, et al. Comparison of patterns and prognosis among distant metastatic breast cancer patients by age groups: a SEER population-based analysis[J]. Sci Rep, 2017, 7: 9254. DOI:10.1038/s41598-017-10166-8.

[17] PARK H S, CHONG Y, LEE Y J, et al. Deep learning-based computational cytopathologic diagnosis of metastatic breast carcinoma in pleural fluid[J]. Cells, 2023, 12(14): 1847. DOI:10.3390/cells12141847.

[18] KASSIRIAN S, HINTON S N, CUNINGHAME S, et al. Diagnostic sensitivity of pleural fluid cytology in malignant pleural effusions: systematic review and meta-analysis[J]. Thorax, 2023, 78(1): 32-40. DOI:10.1136/thoraxjnl-2021-217959.

[19] TIAN F, LIU D, WEI N, et al. Prediction of tumor origin in cancers of unknown primary origin with cytology-based deep learning[J]. Nat Med, 2024, 30(5): 1309-1319. DOI:10.1038/s41591-024-02915-w.

[20] ALI S, LI J Q, PEI Y, et al. State-of-the-art challenges and perspectives in multi-organ cancer diagnosis via deep learning-based methods[J]. Cancers, 2021, 13(21): 5546. DOI:10.3390/cancers13215546.

[21] HEGDE S, SHETTY S, RAI S, et al. A survey on machine learning approaches for automatic detection of voice disorders[J]. J Voice, 2019, 33(6): 947.e11-947.e33. DOI:10.1016/j.jvoice.2018.07.014.

[22] ISLAM T, HOQUE M E, ULLAH M, et al. CNN-based deep learning approach for classification of invasive ductal and metastasis types of breast carcinoma[J]. Cancer Med, 2024, 13(16): e70069. DOI:10.1002/cam4.70069.

[23] CHALLA B, TAHIR M, HU Y, et al. Artificial intelligence-aided diagnosis of breast cancer lymph node metastasis on histologic slides in a digital workflow[J]. Mod Pathol, 2023, 36(8): 100216. DOI:10.1016/j.modpat.2023.100216.

[24] WANG R, GU Y, ZHANG T Y, et al. Fast cancer metastasis location based on dual magnification hard example mining network in whole-slide images[J]. Comput Biol Med, 2023, 158: 106880. DOI:10.1016/j.compbiomed.2023.106880.

[25] RETAMERO J A, GULTURK E, BOZKURT A, et al. Artificial intelligence helps pathologists increase diagnostic accuracy and efficiency in the detection of breast cancer lymph node metastases[J]. Am J Surg Pathol, 2024, 48(7): 846-854. DOI:10.1097/PAS.0000000000002248.

[26] KHALILIBOROUJENI S, HE X J, JIA W J, et al. End-to-end metastasis detection of breast cancer from histopathology whole slide images[J]. Comput Med Imaging Graph, 2022, 102: 102136. DOI:10.1016/j.compmedimag.2022.102136.

[27] YU J G, WU Z H, MING Y, et al. Prototypical multiple instance learning for predicting lymph node metastasis of breast cancer from whole-slide pathological images[J]. Med Image Anal, 2023, 85: 102748. DOI:10.1016/j.media.2023.102748.

[28] KIM Y G, SONG I H, CHO S Y, et al. Diagnostic assessment of deep learning algorithms for frozen tissue section analysis in women with breast cancer[J]. Cancer Res Treat, 2023, 55(2): 513-522. DOI:10.4143/crt.2022.055.

[29] CHANG C P, HSU C Y, WANG H S, et al. Detection of metastatic breast carcinoma in sentinel lymph node frozen sections using an artificial intelligence-assisted system[J]. Pathol Res Pract, 2025, 267: 155836. DOI:10.1016/j.prp.2025.155836.

[30] AKRAM M, IQBAL M, DANIYAL M, et al. Awareness and current knowledge of breast cancer[J]. Biol Res, 2017, 50(1): 33. DOI:10.1186/s40659-017-0140-9.

[31] XU F, ZHU C, TANG W Q, et al. Predicting axillary lymph node metastasis in early breast cancer using deep learning on primary tumor biopsy slides[J]. Front Oncol, 2021, 11: 759007. DOI:10.3389/fonc.2021.759007.

[32] PARK D, LEE Y M, EO T, et al. Multimodal AI model for preoperative prediction of axillary lymph node metastasis in breast cancer using whole slide images[J]. NPJ Precis Onc, 2025, 9: 131. DOI:10.1038/s41698-025-00914-9.

[33] YAO Y H, LV Y P, TONG L, et al. ICSDA a multi-modal deep learning model to predict breast cancer recurrence and metastasis risk by integrating pathological, clinical and gene expression data[J]. Brief Bioinform, 2022, 23(6): bbac448. DOI:10.1093/bib/bbac448.

[34] YANG J L, JU J, GUO L, et al. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning[J]. Comput Struct Biotechnol J, 2022, 20: 333-342. DOI:10.1016/j.csbj.2021.12.028.

[35] LIU X Y, YUAN P, LI R L, et al. Predicting breast cancer recurrence and metastasis risk by integrating color and texture features of histopathological images and machine learning technologies[J]. Comput Biol Med, 2022, 146: 105569. DOI:10.1016/j.compbiomed.2022.105569.

[36] LIU F F, HARDIMAN T, WU K L, et al. Systemic immune reaction in axillary lymph nodes adds to tumor-infiltrating lymphocytes in triple-negative breast cancer prognostication[J]. NPJ Breast Cancer, 2021, 7: 86. DOI:10.1038/s41523-021-00292-y.

[37] GRIGORIADIS A, GAZINSKA P, PAI T, et al. Histological scoring of immune and stromal features in breast and axillary lymph nodes is prognostic for distant metastasis in lymph node-positive breast cancers[J]. The Journal Pathology CR, 2018, 4(1): 39-54. DOI:10.1002/cjp2.87.

[38] VERGHESE G, LI M Y, LIU F F, et al. Multiscale deep learning framework captures systemic immune features in lymph nodes predictive of triple negative breast cancer outcome in large-scale studies[J]. J Pathol, 2023, 260(4): 376-389. DOI:10.1002/path.6088.

[39] BASAAD A, BASURRA S, VAKAJ E, et al. A BERT-GNN approach for metastatic breast cancer prediction using histopathology reports[J]. Diagnostics, 2024, 14(13): 1365. DOI:10.3390/diagnostics14131365.

[40] SAMMUT S J, CRISPIN-ORTUZAR M, CHIN S F, et al. Multi-omic machine learning predictor of breast cancer therapy response[J]. Nature, 2022, 601(7894): 623-629. DOI:10.1038/s41586-021-04278-5.

[41] PEROU C M, SØRLIE T, EISEN M B, et al. Molecular portraits of human breast tumours[J]. Nature, 2000, 406(6797): 747-752. DOI:10.1038/35021093.

[42] CEJALVO J M, DE DUEÑAS E M, GALVÁN P, et al. Intrinsic subtypes and gene expression profiles in primary and metastatic breast cancer[J]. Cancer

Res, 2017, 77(9): 2213-2221. DOI:10.1158/0008-5472.CAN-16-2717.

[43] YU Y F, CAI G Y, LIN R C, et al. Multimodal data fusion AI model uncovers tumor microenvironment immunotyping heterogeneity and enhanced risk stratification of breast cancer[J]. MedComm, 2024, 5(12): e70023. DOI:10.1002/mco2.70023.

[44] AHUJA S, ZAHEER S. Advancements in pathology: Digital transformation, precision medicine, and beyond[J]. J Pathol Inform, 2025, 16: 100408. DOI:10.1016/j.jpi.2024.100408.

[45] ZHANG H, YANG F, XU Y, et al. Multimodal integration using a machine learning approach facilitates risk stratification in HR+/HER2- breast cancer[J]. Cell Rep Med, 2025, 6(2): 101924. DOI:10.1016/j.xcrm.2024.101924.

[46] WANG Z Z, SAOUD C, WANGSIRICHAROEN S, et al. Label cleaning multiple instance learning: refining coarse annotations on single whole-slide images[J]. IEEE Trans Med Imaging, 2022, 41(12): 3952-3968. DOI:10.1109/TMI.2022.3202759.

Author Contributions: FANG Xiaoqing and LYU Qingjie conceived and designed the article. FANG Xiaoqing collected and organized literature/materials and drafted the manuscript. LYU Qingjie conducted feasibility analysis, revised the manuscript, was responsible for quality control and final review, and provided overall supervision.

Conflict of Interest: The authors declare no conflict of interest.

ORCID: FANG Xiaoqing https://orcid.org/0009-0007-8746-5087

(Received: June 11, 2025; Revised: August 15, 2025)

(Editor: LI Weixia)

Note: Figure translations are in progress. See original paper for figures.