Science Bulletin 69 (2024) 1748–1756

Contents lists available at ScienceDirect

Science Bulletin

journal homepage: www.elsevier.com/locate/scib

Article

Potential rapid intraoperative cancer diagnosis using dynamic full-field optical coherence tomography and deep learning: A prospective cohort study in breast cancer patients

Shuwei Zhang [a,1], Bin Yang [b,c,1], Houpu Yang [a,1], Jin Zhao [a], Yuanyuan Zhang [d], Yuanxu Gao [e], Olivia Monteiro [e], Kang Zhang [e,f,*], Bo Liu [g,*], Shu Wang [a,*]

[a] Breast Center, Peking University People's Hospital, Beijing 100044, China
[b] China ESG Institute, Capital University of Economics and Business, Beijing 100070, China
[c] Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[d] Department of Pathology, Peking University People's Hospital, Beijing 100044, China
[e] Center for Biomedicine and Innovations, Faculty of Medicine, Macau University of Science and Technology, Macao 999078, China
[f] College of Future Technology, Peking University, Beijing 100091, China
[g] School of Mathematical and Computational Sciences, Massey University, Auckland 0745, New Zealand

ARTICLE INFO

Article history:
Received 27 October 2023
Received in revised form 18 March 2024
Accepted 19 March 2024
Available online 29 March 2024

Keywords:
Cancer diagnosis
Breast neoplasms
Dynamic full-field optical coherence tomography
Deep learning
Image classification

ABSTRACT

An intraoperative diagnosis is critical for precise cancer surgery. However, traditional intraoperative assessments based on hematoxylin and eosin (H&E) histology, such as frozen section, are time-, resource-, and labor-intensive, and involve specimen-consuming concerns. Here, we report a near-real-time automated cancer diagnosis workflow for breast cancer that combines dynamic full-field optical coherence tomography (D-FFOCT), a label-free optical imaging method, and deep learning for bedside tumor diagnosis during surgery. To classify the benign and malignant breast tissues, we conducted a prospective cohort trial. In the modeling group (n = 182), D-FFOCT images were captured from April 26 to June 20, 2018, encompassing 48 benign lesions, 114 invasive ductal carcinoma (IDC), 10 invasive lobular carcinoma, 4 ductal carcinoma in situ (DCIS), and 6 rare tumors. Deep learning model was built up and fine-tuned in 10,357 D-FFOCT patches. Subsequently, from June 22 to August 17, 2018, independent tests (n = 42) were conducted on 10 benign lesions, 29 IDC, 1 DCIS, and 2 rare tumors. The model yielded excellent performance, with an accuracy of 97.62%, sensitivity of 96.88% and specificity of 100%; only one IDC was misclassified. Meanwhile, the acquisition of the D-FFOCT images was non-destructive and did not require any tissue preparation or staining procedures. In the simulated intraoperative margin evaluation procedure, the time required for our novel workflow (approximately 3 min) was significantly shorter than that required for traditional procedures (approximately 30 min). These findings indicate that the combination of D-FFOCT and deep learning algorithms can streamline intraoperative cancer diagnosis independently of traditional pathology laboratory procedures.

© 2024 Science China Press. Published by Elsevier B.V. and Science China Press. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Approximately 19.3 million people are diagnosed with cancer annually worldwide [1], and surgery remains the most effective treatment option offering the potential for cure. Intraoperative diagnosis is essential for immediate decision-making in the operating room. For example, on-table margin assessment during breast-conserving surgery (BCS) significantly reduces the need for reoperation [2]. Currently, the most popular intraoperative diagnostic method for cancer is frozen section analysis based on hematoxylin and eosin (H&E) histology. However, frozen section involves tissue transport, intricate specimen preparations and manual interpretation [3], with each step creating a potential barrier to effective and timely surgical management.

Dynamic full-field optical coherence tomography (D-FFOCT) is an innovative optical technology that allows light interference imaging to capture intracellular metabolic features [4]. It can also create internal contrast and provide high-resolution visualization of live cells to establish a "virtual histology". D-FFOCT has garnered

* Corresponding authors.
E-mail addresses: kang.zhang@gmail.com (K. Zhang), b.liu@massey.ac.nz (B. Liu), shuwang@pkuph.edu.cn (S. Wang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.scib.2024.03.061
2095-9273/© 2024 Science China Press. Published by Elsevier B.V. and Science China Press.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

recognition for its breakthrough in the imaging of retinal organoids [5], intracellular dynamics, and cytoskeletal variations in HeLa cells [6]. In our previous study [7], D-FFOCT was used to distinguish normal, benign, and cancerous breast tissues. The results demonstrated that D-FFOCT could reveal microscopic tissue architectural features without requiring tissue preparation or destruction, making it a promising alternative to intraoperative histological analysis.

However, the interpretation of either D-FFOCT images or H&E histological images requires trained clinicians and is labor-intensive. Currently, there is a common shortage of pathology services worldwide [8,9]; even in some developed countries, including the United States and Germany, there is a 30% gap in the available supply of pathologist services relative to the estimated demand. Thus, it is not conducive to the availability of effective intraoperative cancer diagnosis for patients [10,11]. Preliminary data have shown that artificial intelligence (AI) can achieve human-level accuracy for medical image classification in various diagnostic fields such as radiology [12], ophthalmology [13] and pathology [14]. Herein, we aimed to develop a streamlined approach for intraoperative cancer diagnosis by combining the advantages of AI and D-FFOCT imaging and integrating it into the workflow of the operating room.

In the current study, we established an intraoperative diagnostic workflow to examine breast samples obtained from breast surgery and to classify tumors on D-FFOCT images using the Swin Transformer [15], a state-of-the-art computer vision deep-learning model. We assessed the accuracy of benign and malignant tumor classification and simulated the evaluation of BCS margins. The significance of this study lies in its potential to offer an accurate, rapid and nondestructive alternative to current intraoperative practices.

2. Materials and methods

2.1. Participants

All study procedures were approved by the Ethics Committee of Peking University People's Hospital (2016PHB210-001). Patients scheduled to undergo breast surgery (lumpectomy for benign lesions or BCS/mastectomy for breast cancer) were eligible. All patients provided written informed consent for the scientific use of their specimens and data. The enrollment of 129 patients commenced on April 26, 2018 and ended on August 17, 2018.

2.2. Dataset and labeling

In this study, the patient enrollment timeline (with June 20, 2018 as the dividing point) led to the creation of two sets: A training/validation (modeling) set and an independent testing set. Breast specimens were obtained from resected breasts. Tissue visualization using D-FFOCT was performed using a commercial D-FFOCT imager (Light-CT; LLTech, Paris, France), following the procedures outlined in our previous report [7]. Because D-FFOCT captures *en face* images under a flat piece of glass, we defined a large image as a slide of the D-FFOCT. The ratio of the slides in the training/validation set to those in the testing set was approximately 4:1. We employed a 5-fold cross-validation approach to eliminate potential sample imbalances by dividing the training/validation set into five segments (Fig. S1a online). This allowed sequential training on four folds and fine-tuning on the remaining fold.

In cases involving multifocal lesions, one slide was obtained for each target lesion. Taking benign tissues from mastectomy resection specimens of patients with breast cancer was allowed. These situations resulted in the collection of multiple slides per patient. In the training/validation set, in situations where tumors measured $\geq$ 2 cm, we partitioned the specimen into 2–4 areas for imaging, creating multiple slides in the same case. A total of 182 D-FFOCT slides were captured from 95 patients, spanning the period from April 26 to June 20, 2018. In the testing set, one slide was captured for each lesion, and 42 D-FFOCT slides were captured from 34 patients, covering the period from June 22 to August 17, 2018. After D-FFOCT imaging, all undamaged specimens without any destruction were transported to the pathology department for routine formalin-fixed, paraffin-embedded histopathological analysis.

The labeling process for D-FFOCT images consisted of two parts: slide-level and patch-level labeling. Slide-level labeling refers to the labeling of the entire slide, whereas patch-level labeling refers to the labeling of the patches measuring 1150 $\times$ 1150 pixels generated from the original slides. Tumors were manually segmented in the D-FFOCT images using the QuPath software [16]. Pathologists with expertise in breast histology provided labels for each D-FFOCT slide based on the corresponding pathological diagnosis of the specimen and features observed on the D-FFOCT slide. Patch labels were assigned based on the presence of malignant pixels within the patches.

2.3. Network architecture

To select an efficient neural network architecture, we compared the average area under the curve (AUC) in five validation folds measured by receiver operating characteristic (ROC) analysis and average training time between two commonly used convolutional neural network architectures and a tiny version of the Swin Transformer (Swin-T) architecture. The Swin-T model showed a relatively higher accuracy with the least training time (Table S1 online) and was therefore selected as our default backbone. As shown in Fig. S1b (online), Swin-T features a hierarchical transformer structure that utilizes window self-attention (WSA) and pixel-shifted WSA (SWSA). It comprises four stages, with each stage composed of repeated units, progressively reducing the resolution of the input feature maps while expanding the receptive field. Initially, Swin-T divides the input feature map into multiple fixed-size small windows using a patch partition layer, thus restricting the computation of self-attention within these small local windows and significantly reducing computational complexity. Subsequently, the partitioned patches are converted into a series of feature vectors through a linear layer. As the network layer deepens, Swin-T reduces the resolution of feature mapping by merging adjacent feature blocks while simultaneously increasing the dimensionality (number of feature channels) of each block. This hierarchical approach generates feature representations of different scales at various levels, effectively capturing the multi-scale information of images. Next, Swin-T uses shifted window partitioning in successive blocks, allowing pixels or feature points across window boundaries to interact in subsequent layers, thereby enhancing the model's ability to capture global information across local regions. Finally, the output feature vector is sent to a global average pooling, layer normalization, and full connection layer to predict the final classification result. The model was pretrained on image classification datasets (ImageNet-1 K and ImageNet-22 K), and then the learned weights were transferred to fine-tune the D-FFOCT classification task. The optimizer was "AdamW" with the following parameters: learning rate = 1×10^{-6}, weight decay = 5×10^{-2}, and momentum = 0.9. The cross-entropy function was selected as the loss function.

All software code for this study was written in Python using the PyTorch module (1.12.1). Model training and validation were performed on a Linux workstation with 32 GB RAM, 4 TB disk space, an Intel CPU, and a GeForce GTX3090 GPU (Nvidia, California, USA).

2.4. Model fine-tuning and patch-slide mapping

The training loss, validation loss and patch-level accuracy for each fold in the five-fold cross-validation are shown in Fig. S2 (online). During the testing phase, each patch in the testing set was processed individually using the model obtained after the iteration of each fold. The average of the five output values was considered the final output for each patch during testing.

Given that slide-level labels are histology-correlated and represent the ground truth, it is essential to align the patch prediction results with the slide levels to accurately determine the prediction performance. To capture semantically rich patch-level feature representations and map them to the slide level, we selected the top and bottom 10 probability patches from each D-FFOCT slide. Subsequently, we extracted features based on statistical measures such as mean, maximum, minimum, sum, median, standard variance, variance, skewness, and kurtosis from all patches, the top 10 patches, and the bottom 10 patches. This resulted in the extraction of a 25-dimensional feature vector from each slide. The normalized 25-dimensional feature vectors were then inputted into a support vector machine classifier to distinguish between malignancy and benignancy in the D-FFOCT slide.

2.5. Model interpretation and visualization

The class activation map (CAM) [17] and t-SNE [18] use feature visualization to explore the working mechanism and judgment basis of deep learning networks and are widely used in the field of deep learning interpretability. In this study, we built activation maps using Score-CAM [19] and Guided Score-CAM, and also used t-SNE to visually analyze the performance of the proposed Swin-T model. The heatmap of the Score-CAM visualization algorithm was the result of downsampling 32 times from the last norm layer of the deep-learning model. Therefore, bicubic interpolation was used to scale the image to its original size.

2.6. Simulated breast conserving surgery margin evaluation

In the training/validation and testing processes, the size of the D-FFOCT slides was generally much larger for experimental purposes. To illustrate the potential of D-FFOCT for margin status assessment, we selected four margins along the specimen edges from the periphery of each slide in the testing set. Each margin consisted of two patches. A total of 336 patches and 168 margins were obtained. The margins were labeled as negative or positive by an experienced pathologist based on the corresponding histopathological diagnosis and features observed on the D-FFOCT slide. The prediction outcomes of the selected patches were subsequently aggregated to the margin level using a machine learning model with the same approach as the initial testing dataset.

2.7. Statistical analysis

Using pathological diagnoses and clinician-designated labels as the gold standard, the primary endpoint was the overall binary diagnostic accuracy of the deep-learning model at the slide and patch levels. The sensitivity and specificity were also calculated. The AUC of the ROC curve at the patch level was assessed. Diagnostic performance was also evaluated using the simulated margin dataset. Fisher's exact test was used to compare the diagnostic accuracy in different subgroups. All the hypotheses tested were 2-sided, and a P value of less than 0.05 was considered significant.

3. Results

3.1. Participants and the dataset

A total of 224 D-FFOCT slides of fresh breast tissues from 129 patients were included in this study, and 13,497 patches were generated. The median number of patches contained in one D-FFOCT slide was 60 (4–251). In total, 182 slides consisting of 10,357 patches were included in the training/validation set and 42 slides consisting of 3140 patches were included in the testing set (Fig. S1a online). Malignant patches accounted for 52.76% (7121/13,497) of all patches. The distribution of malignant patches in both sets was well balanced.

The baseline characteristics of the slides and patches are presented in Table 1. The numbers of slides per fold are shown in Fig. S1a (online). In addition to the common breast cancer histological types, there were eight slides from rare breast tumor types in our dataset, including three mucinous carcinomas, one tubular carcinoma, three papillary carcinomas and one metaplastic carcinoma. Among these, one mucinous carcinoma and one papillary carcinoma were included in the testing set. In terms of molecular subtypes, our dataset consisted of four common molecular subtypes of breast cancer, and their distribution was balanced between the training/validation and the testing sets.

3.2. Model performance on the validation set and independent testing set

The diagnostic results of the cross-validation experiment are presented in Table 2. We found a mean AUC of 0.900 for the binary prediction of patches (Fig. 1a). Fold 5 yielded the best performance,

Table 1
Characteristics of the included D-FFOCT slides.

Characteristics	Total number	Train and validation set	Testing set
Slides included	224	182	42
Breast cancer slides	166	134	32
Histological type	–	–	–
IDC	143	114	29
ILC	10	10	0
DCIS	5	4	1
Rare tumors	8	6	2
Molecular subtypes of invasive breast cancer slides	–	–	–
Luminal A	45	33	12
Luminal B	79	69	10
HER-2 enriched	13	7	6
TNBC	19	18	1
Uncertain	5	3	2
Benign lesions or normal breast slides	58	48	10
Patches included	13,497	10,357	3140
Patches including breast cancer	7121	5214	1907
Benign patches	6376	5143	1233

DCIS: ductal carcinoma *in situ*; IDC: invasive ductal carcinoma; ILC: invasive lobular carcinoma; TNBC: triple negative breast cancer.

Table 2
Prediction performance at patch level among different folds.

Validation set	No. of patches	Accuracy	Sensitivity	Specificity
Fold 1	1352	0.855	0.849	0.860
Fold 2	2322	0.782	0.876	0.689
Fold 3	3014	0.785	0.762	0.804
Fold 4	2635	0.833	0.799	0.882
Fold 5	1034	0.867	0.852	0.880

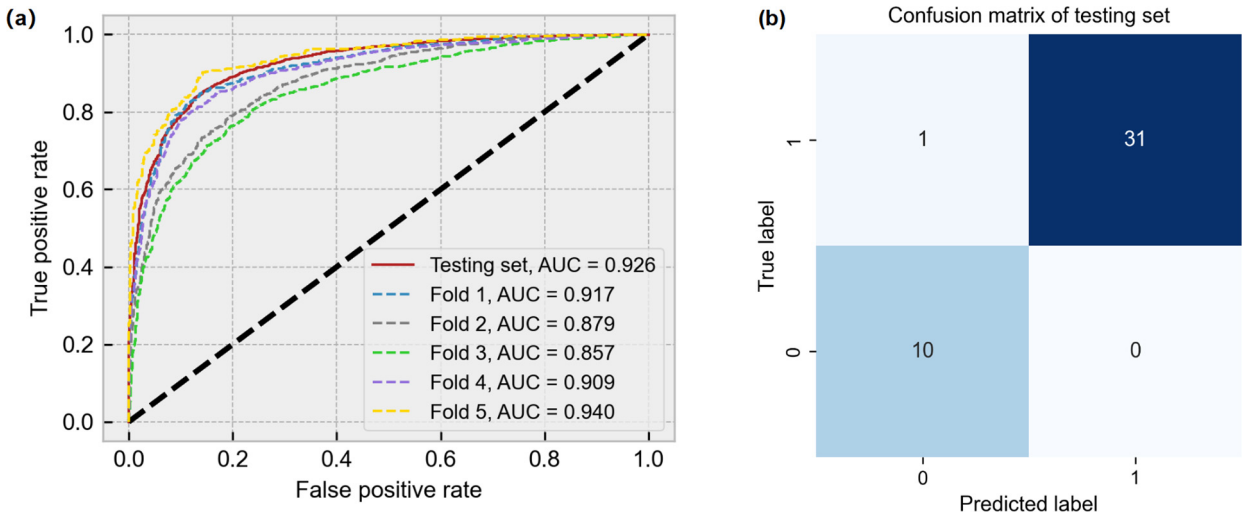


Fig. 1. The performance of the deep learning model on the validation set and testing set. (a) Receiver operating characteristic curves between malignant and non-malignant breast patches in different folds of the validation sets and testing set. (b) Diagnostic results of the deep learning model at slide level in the testing set. AUC: area under the receiver operating characteristic curve.

with an accuracy of 0.867 and AUC of 0.940 (95% confidence interval (CI): 0.921–0.958).

To verify the performance of our deep learning model, we deployed five models on an independent testing set and used the average output as the final prediction result. In the testing set including 3140 patches, the diagnostic performance of our model at the patch level was relatively good for determining the nature of breast tissue, with an AUC of 0.926 (95% CI: 0.907–0.943) (Fig. 1a) and a sensitivity/specificity of 0.861 (1641/1907) / 0.846 (1043/1233).

The results of the independent testing set at the slide level, which contained 42 slides, are shown in Fig. 1b. The deep learning model correctly classified 97.62% (41/42) of the D-FFOCT slides in the testing set with a sensitivity and specificity of 96.88% and 100%, respectively. Specifically, as shown in the confusion matrix, the model misdiagnosed one case as a false negative. The misclassified case was a slide of invasive ductal carcinoma (IDC). Consequently, we calculated the prediction accuracy for different tumor types as follows: 96.55% (28/29) for IDC and 100% for ductal carcinoma in situ (DCIS), mucinous carcinoma, and papillary carcinoma (Fig. 2a). No statistically significant differences in accuracy were observed for the four molecular subtypes of breast cancer (Luminal A, Luminal B, triple negative, and HER-2 enriched). For the misclassified slide, malignancy was observed in a small upper-right corner area (Fig. S3a online) by experienced pathologists, comprising 8 out of 114 patches. The overall diagnostic accuracy for these 114 patches was 89.47% (102/114), with the deep learning model accurately identifying four of the eight malignant patches (Fig. S3b online). For the 106 benign patches, the diagnostic specificity was 92.45% (98/106).

We used three visualization algorithms to investigate the interpretability of our deep learning model and to help doctors locate suspected tumor areas. Fig. 2b–d shows typical Score-CAM heatmaps of D-FFOCT images of IDC (Fig. 2b) and two other rare types of breast cancer (mucinous carcinoma (Fig. 2c) and papillary carcinoma (Fig. 2d)). We observed that despite the variations in tissue structures among different tumor types, distinctive patterns of highly metabolically active cells and deformed collagen distribution could be identified on D-FFOCT slides, similar to the features in the pathological slides. The deep learning model also prioritizes regions where these active cells (cancer cells) tend to cluster, which is consistent with the corresponding pathological slides. This demonstrates that our deep learning model learned clinically relevant features for evaluating tumor diagnosis, irrespective of the tumor type.

Furthermore, Fig. S4 (online) shows the visual results of other typical patches using the activation maximization images generated through Score-CAM and Guided Score-CAM. For the malignant patches (Fig. S4a–c online), the regions with high scores (red and yellow regions) were mainly clustered cancer cells, and their contribution to the classifier gradually decreased with increasing distance from the center of the cancer cell. This demonstrates that the neural networks made malignant classification decisions based mainly on the morphology of cancer cells, which is consistent with expert experience. For the benign patches shown in Fig. 4d–f (online), the neural networks made classification decisions based on the collagen fibers (yellow arrows) and adipocytes (red arrows) rather than the breast duct (black arrows) and breast lobules (white arrows). Fig. S5 (online) shows the *t*-SNE plot of the testing set. Swin-T effectively clustered malignant and non-malignant patches in the two-dimensional feature space of *t*-SNE, indicating that our deep learning network learned the diagnostic characteristics.

3.3. Simulated intraoperative margin evaluation during cancer surgery

Our simulated margin evaluation yielded an accuracy of 95.24%, with 160 out of the 168 margins correctly diagnosed. Fig. 3 shows a typical case of margin assessment within a malignant slide. The predictive results can be used to discern the orientation of the "residual tumor". The imaging procedure for each set of four-margin evaluations required approximately 2 min.

3.4. Novel workflow of utilizing D-FFOCT and the deep learning model

Based on the above results, we propose an intraoperative cancer diagnosis workflow that integrates D-FFOCT with a deep learning model (Fig. 4). D-FFOCT provides high-resolution imaging at a near-histological level to visualize tissue. We trained and validated the model using 10,357 D-FFOCT images and proved the efficacy of the algorithm on an independent testing dataset of 3140 images. The accuracy at the slide level was 97.62%. In our preliminary results of the simulated intraoperative margin diagnosis during surgery, we achieved an initial accuracy rate of 95.24%, with an

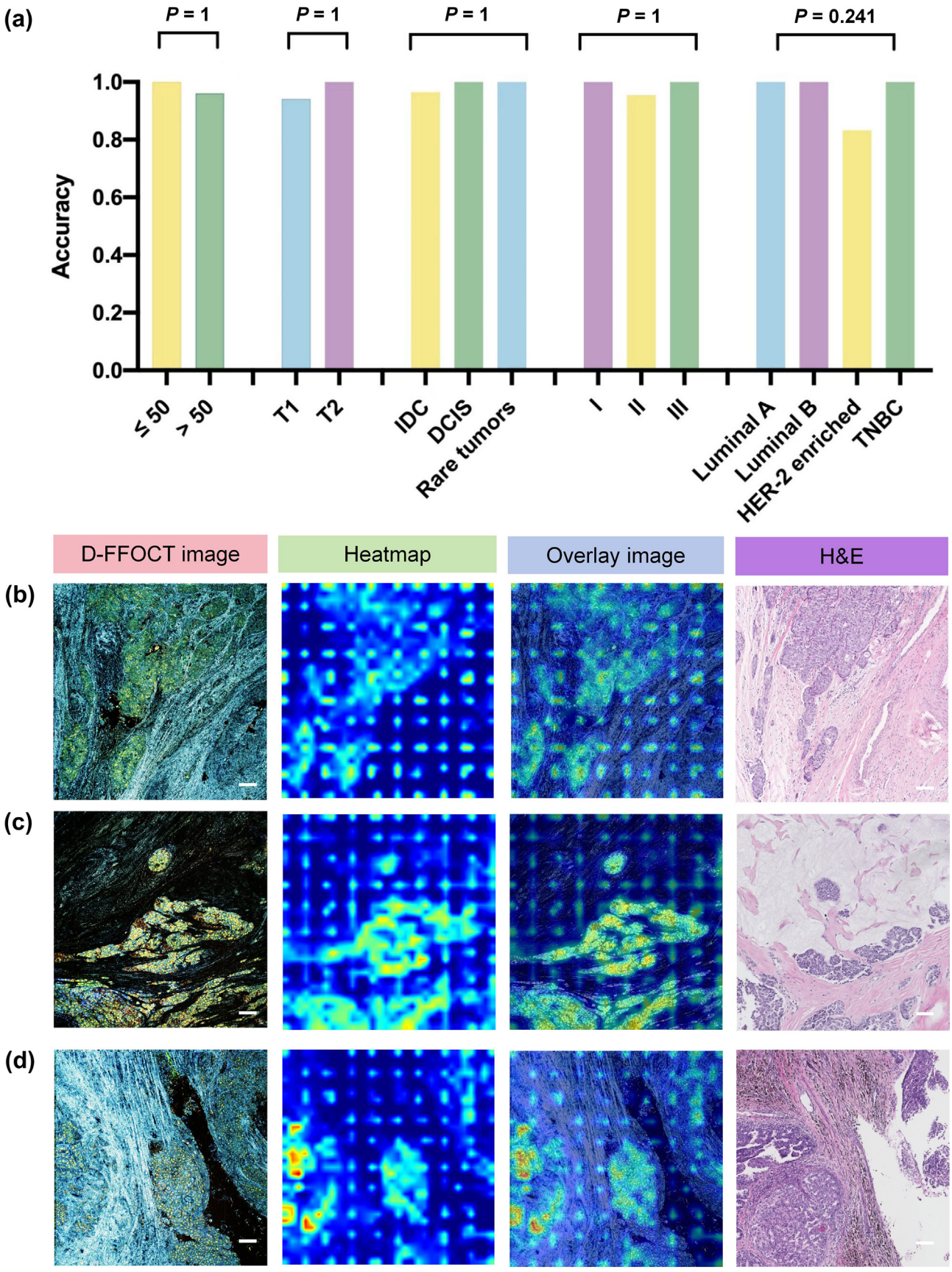


Fig. 2. D-FFOCT reveals typical histopathological features of different types of breast tumors, and different subgroups exhibit high accuracy. (a) Subgroup accuracy for different ages, tumor sizes, histological types, histological grades and breast molecular subtypes. Accuracy among different subgroups was statistically insignificant (Fisher's exact test). (b–d) show D-FFOCT images, Score-CAM heatmaps, overlay images and corresponding H&E images of IDC (b), mucinous carcinoma (c), and papillary carcinoma (d), respectively. Scale bar = 100 μm. DCIS: ductal carcinoma *in situ*; IDC: invasive ductal carcinoma; TNBC: triple negative breast cancer.

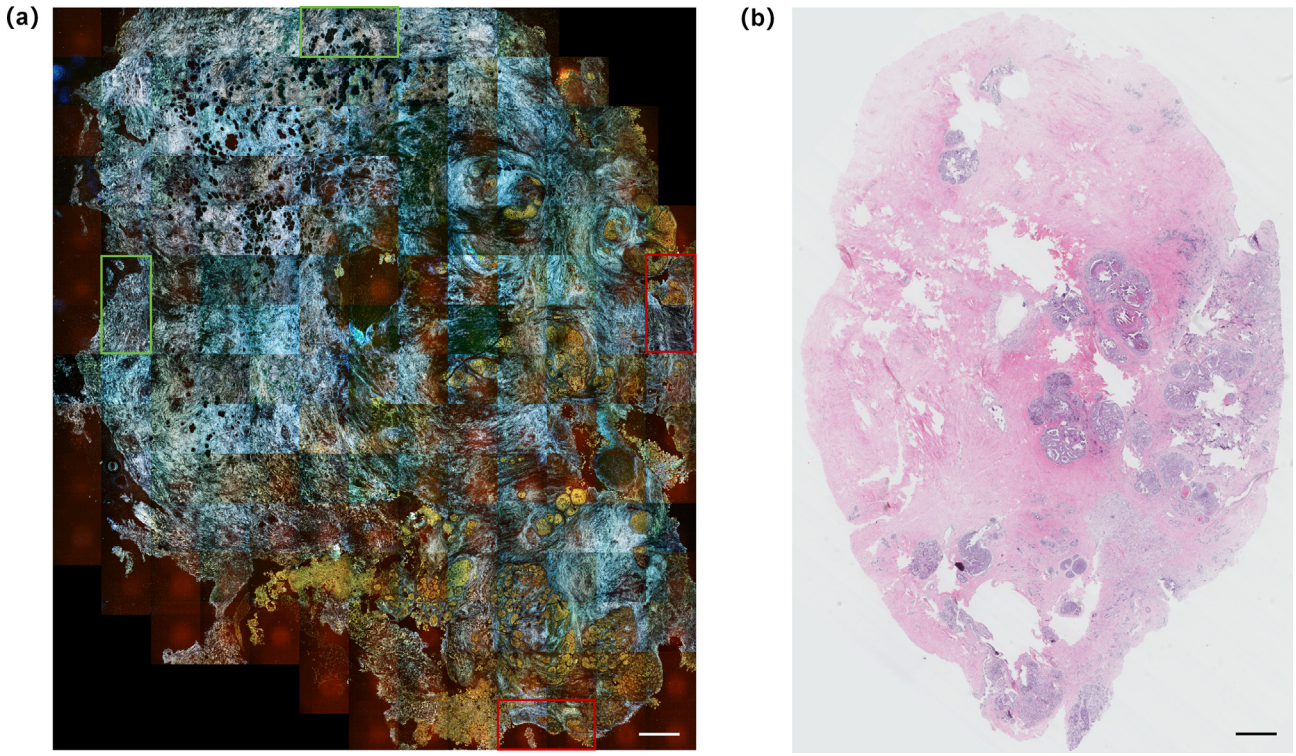


Fig. 3. Simulated margin assessment of a malignant slide. In the D-FFOCT slide (a), four margins (eight patches) indicated by rectangles were chosen for analysis. The red rectangles signify positive margins, while the green rectangles are negative margins. All four margins were correctly diagnosed and were consistent with the cancer area suggested by the H&E histology image (b). The lower-right corner of the D-FFOCT slide was filled with malignant yellow cells, corresponding to H&E histology. Scale bar = 1 mm.

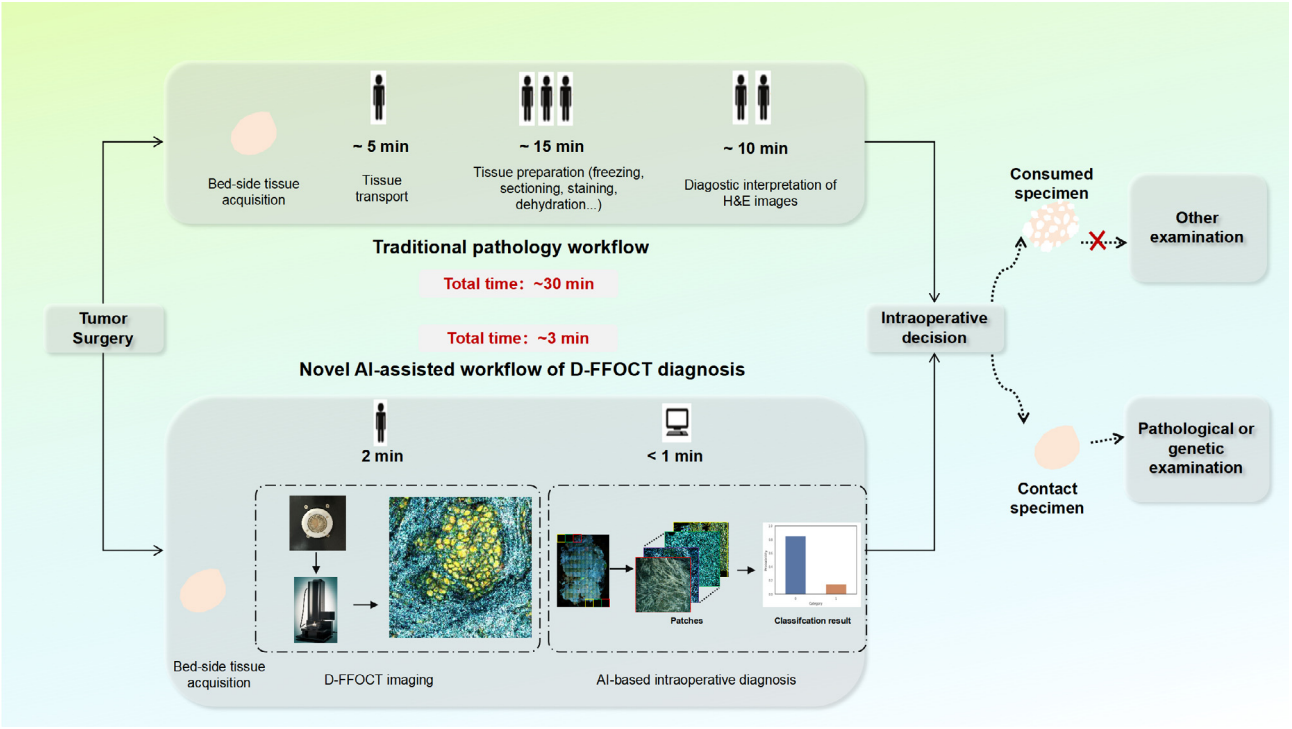


Fig. 4. Intraoperative diagnosis workflows for both conventional H&E-based histology and D-FFOCT plus deep learning are shown in parallel. Our pipeline could provide a diagnosis in approximately 3 min. Moreover, compared with traditional pathology, our novel workflow substantially decreased the labor requirements for the surgical procedure while preserving the entire tissue sample for postoperative examinations.

imaging time of 2 min and total processing time of approximately 3 min, decreasing the time to diagnosis by a factor of 10 compared to conventional intraoperative histology. Fig. 4 illustrates the labor required for the two diagnostic processes. For traditional frozen section, the involvement of specimen transportation, preparations, and image interpretation is very labor-intensive. In contrast, our workflow is remarkably labor-cost-effective. Furthermore, no tissue destruction was observed during optical imaging and analysis. In conclusion, our workflow provides a transparent means for delivering a rapid and accurate intraoperative diagnosis when pathological resources and surgical tissues are limited.

4. Discussion and conclusion

In this paper, we reported a near-real-time, non-specimen-destructive, and automated workflow for breast cancer diagnosis. On an independent testing set, we achieved an accuracy of 97.62%. Furthermore, when simulating margin diagnosis, the entire process required only 3 min, which is significantly shorter than that of traditional procedures. These findings indicate that the combination of D-FFOCT imaging with a deep learning algorithm has the potential to simplify intraoperative cancer diagnosis, thus reducing reliance on traditional pathology laboratory procedures.

In an era in which molecular pathology plays a pivotal role in guiding tumor diagnosis, prognosis, and therapeutic decisions [20], the preservation of specimen integrity is of paramount importance, particularly for small cancers and tiny biopsy samples. Traditional intraoperative histological analysis involves substantial specimen destruction during tissue trimming and slide preparation, potentially compromising the accuracy of tumor diagnosis, staging, or subsequent genetic tests. Therefore, nondestructive imaging techniques present a substantial advantage in modern and future medical practice.

In recent years, various digital optical techniques have been explored for intraoperative guidance without tissue destruction, including optical coherence elastography [21], which is a method for assessing tissue elasticity and structure by measuring the reflection or scattering of light; Raman Spectroscopy [22], which measures the vibrational frequencies of molecules in tissues that can be excited by a laser; and multiphoton microscopy [23], etc. However, none of these techniques have been widely adopted for clinical use due to time-consuming scans and limited spatial resolution.

Full-field optical coherence tomography (FFOCT) provides a high axial and transverse resolution of approximately 1 μm [24]. This resolution is close to that of traditional pathological assessments using microscopy. Thus FFOCT has been extensively studied in various medical fields [24–27]. D-FFOCT, developed from FFOCT, allows for the capture of interferogram videos in which cellular contours are defined and metabolic indices are quantifiable [4,28]. Despite the remarkable attributes that make it a suitable technique for cancer diagnosis, D-FFOCT has limited clinical application. Three years ago, our team pioneered the exploration of its diagnostic potential for breast cancer [7]. Currently, we are taking a step further by integrating AI analysis to enhance diagnostic accuracy and speed.

The rise of AI offers potential solutions by aiding image interpretation and reducing time costs. The Food and Drug Administration has approved more than 500 AI- and machine-learning-enabled medical devices, particularly in radiology [29]. AI holds great promise in medical diagnostics owing to its consistent performance [30]. The application of AI in standard pathology is an expanding field of research with successful examples in primary cancer detection [31,32], cancer subtyping [33], and lymph node diagnosis [34]. In this study, we used Swin-T, a robust computer vision backbone for medical image classification [35,36], to diagnose D-FFOCT. The model demonstrated excellent performance on the testing set, indicating its ability to learn relevant histomorphological, cytological, and nuclear features for image classification, including those traditionally used by pathologists [37]. Despite being developed and validated primarily for breast tissue, deep-learning-based image analysis can potentially transfer our research to various other tumor types, as the features detected in our model appear to be conserved characteristics in oncology diagnosis [38].

In our study, we included 13,497 patches from 224 slides and confirmed the feasibility of using AI-augmented D-FFOCT for breast tissue classification. Our study yielded excellent performance, with an accuracy of 97.62%. Importantly, this high level of accuracy was consistently maintained across various tumor types, including IDC, DCIS, and rare tumors. Furthermore, the accuracy of our method remained consistent across different molecular types, including Luminal, triple negative, and HER-2 enriched breast subtypes. In the margin simulation process in which two patches from one margin were used as the predictive set, the accuracy remained consistently high at 95.24%. However, it is worth noting that the assessment of rare tumor types was based on a limited sample pool. Further investigation and validation are required to ensure the reliability of our method for these less common tumor types. Additionally, the margin issue should be thoroughly investigated in prospective clinical trials to confirm its real-world applicability and effectiveness.

A similar study was reported by Scholler et al. [39], where the training/validation set contained 47 patients (37 for training and 10 for validation), and they manually selected regions of interest (ROIs) to build the model. The model achieved an accuracy of 94% for self-verification, with two false positives and one false negative. The manual ROI selection strategy, representing 10%–20% of all specimens, made it challenging to interpret the model, and the extremely small sample size limited model robustness. Although no independent tests were conducted, their preliminary attempt to combine D-FFOCT and AI showed potential in the field of cancer diagnosis.

Our D-FFOCT testing set yielded no false-positive results; however, one case of IDC was misdiagnosed as a false-negative result. Malignancy was observed in a small upper-right corner area by experienced pathologists, comprising 8 out of 114 patches. The overall diagnostic accuracy for these 114 patches was 89.47% (102/114), with the deep learning model accurately identifying four out of the eight malignant patches. Notably, the misclassified slide had an extremely low proportion of positive patches compared with other malignant slides. However, current machine learning algorithm mapping results from the patch to slide level resembles a black box. We do not know whether the non-tumor portions in the malignant slides affect the final diagnosis. Addressing this issue will involve incorporating more samples with minimal tumor content and employing interpretable models to aggregate results at the patch level.

In addition to the demonstrated capabilities, our study can be further improved in several aspects. First, more advanced deep-learning models may help to further enhance predictive performance; for example, multi-modal information fusion may improve the ability to identify challenging samples, and weakly-supervised learning [40] can help reduce the labeling workload. In addition, owing to the sample size limitations of rare tumors or uncommon molecular subtypes, we did not differentiate tumor types in the D-FFOCT images. Expanding the number of these specimens would aid in refining networks to classify more tumor subtypes.

In conclusion, we developed a novel workflow using deep learning algorithms for intraoperative tumor diagnosis based on D-

FFOCT images. This approach reduces the workload of clinicians and offers a fast, reproducible, nondestructive, and accurate tissue assessment. Similar approaches that integrate D-FFOCT and deep learning may find application in detecting residual tumors and providing microscopic information during surgery.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by the Capital's Funds for Health Improvement and Research (CHF 2020-2Z-40812), Beijing Natural Science Foundation (7242281), Beijing Municipal Science and Technology Project (Z201100005520081), the National Key Research and Development Program of China (2016YFC0901300), the National Natural Science Foundation of China (62076015), Macau Science and Technology Development Fund, Macao, China (0070/2020/A2, 0003/2021/AKP) and Macao Young Scholars Program (AM2023024).

Author contributions

Shu Wang conceived of the study; Shu Wang, Kang Zhang and Houpu Yang designed the study; Shuwei Zhang, Jin Zhao, Houpu Yang, Yuanxu Gao and Olivia Monteiro performed the research and collected the data; Yuanyuan Zhang, Shuwei Zhang and Houpu Yang outlined and annotated the region of malignancy in D-FFOCT images; Bin Yang, Shuwei Zhang and Bo Liu designed the deep learning method and ran the experiments; Shuwei Zhang, Bin Yang and Houpu Yang wrote the paper; Shu Wang, Bo Liu and Kang Zhang did the conceptualization, writing-review and editing; All authors read and approved the final version of the article.

Appendix A. Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.scib.2024.03.061.

Data availability

All data including the imaging data supporting the findings of this study are available on requests for non-commercial and academic purposes from the corresponding author Shu Wang (shuwang@pkuph.edu.cn).

References

[1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
[2] Racz JM, Glasgow AE, Keeney GL, et al. Intraoperative pathologic margin analysis and re-excision to minimize reoperation for patients undergoing breast-conserving surgery. Ann Surg Oncol 2020;27:5303–11.
[3] Olson TP, Harter J, Muñoz A, et al. Frozen section analysis for intraoperative margin assessment during breast-conserving surgery results in low rates of re-excision and local recurrence. Ann Surg Oncol 2007;14:2953–60.
[4] Apelian C, Harms F, Thouvenin O, et al. Dynamic full field optical coherence tomography: Subcellular metabolic contrast revealed in tissues by interferometric signals temporal analysis. Biomed Opt Express 2016;7:1511–24.
[5] Scholler J, Groux K, Goureau O, et al. Dynamic full-field optical coherence tomography: 3D live-imaging of retinal organoids. Light Sci Appl 2020;9:140.
[6] Park S, Nguyen T, Benoit E, et al. Quantitative evaluation of the dynamic activity of HeLa cells in different viability states using dynamic full-field optical coherence microscopy. Biomed Opt Express 2021;12:6431–41.
[7] Yang H, Zhang S, Liu P, et al. Use of high-resolution full-field optical coherence tomography and dynamic cell imaging for rapid intraoperative diagnosis during breast cancer surgery. Cancer 2020;126:3847–56.
[8] Naritoku WY, Furlong MA, Knollman-Ritschel B, et al. Enhancing the pipeline of pathologists in the united states. Acad Pathol 2021;8:23742895211041725.
[9] Guarner J, Fleming K. Cancer in low- and middle-income countries: where to start? is it pathology? Am J Clin Pathol 2014;142:435–6.
[10] Robboy SJ, Weintraub S, Horvath AE, et al. Pathologist workforce in the United States: I. development of a predictive model to examine factors influencing supply. Arch Pathol Lab Med 2013;137:1723–32.
[11] Märkl B, Füzesi L, Huss R, et al. Number of pathologists in Germany: Comparison with European countries, USA, and Canada. Virchows Arch 2021;478:335–41.
[12] Duong MT, Rauschecker AM, Rudie JD, et al. Artificial intelligence for precision education in radiology. Br J Radiol 2019;92:20190389.
[13] Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 2019;103:167–75.
[14] Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol 2019;20:e253–61.
[15] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proc IEEE/CVF Int Conf Comput Vis 2021:10012–22.
[16] Bankhead P, Loughrey MB, Fernández JA, et al. Qupath: Open source software for digital pathology image analysis. Sci Rep 2017;7:16878.
[17] Selvaraju RR, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proc IEEE Int Conf Comput Vis 2017:618–26.
[18] Van der Maaten L, Hinton G. Visualizing data using *t*-SNE. J Mach Learn Res 2008;9:2579–605.
[19] Wang H, Wang Z, Du M, et al. Score-cam: Score-weighted visual explanations for convolutional neural networks. Proc IEEE/CVF Conf Comput Vis Pattern Recognit Workshops (CVPRW) 2020:24–25s.
[20] Malone ER, Oliva M, Sabatini PJB, et al. Molecular profiling for precision cancer therapies. Genome Med 2020;12:1–19.
[21] Kennedy BF, McLaughlin RA, Kennedy KM, et al. Investigation of optical coherence microelastography as a method to visualize cancers in human breast tissue. Cancer Res 2015;75:3236–45.
[22] Haka AS, Volynskaya Z, Gardecki JA, et al. *In vivo* margin assessment during partial mastectomy breast surgery using raman spectroscopy. Cancer Res 2006;66:3317–22.
[23] Perrin L, Bayarmagnai B, Gligorijevic B. Frontiers in intravital multiphoton microscopy of cancer. Cancer Rep 2020;3:e1192.
[24] Assayag O, Antoine M, Sigal-Zafrani B, et al. Large field, high resolution full-field optical coherence tomography: A pre-clinical study of human breast tissue and cancer assessment. Technol Cancer Res Treat 2014;13:455–68.
[25] Grieve K, Thouvenin O, Sengupta A, et al. Appearance of the retina with full-field optical coherence tomography. Invest Ophthalmol Vis Sci 2016;57:OCT96-104.
[26] Dalimier E, Salomon D. Full-field optical coherence tomography: A new technology for 3D high-resolution skin imaging. Dermatol 2012;224:84–92.
[27] Nandy S, Sanders M, Zhu Q. Classification and analysis of human ovarian tissue using full field optical coherence tomography. Biomed Opt Express 2016;7:5182–7.
[28] Groux K, Verschueren A, Nanteau C, et al. Dynamic full-field optical coherence tomography allows live imaging of retinal pigment epithelium stress model. Commun Biol 2022;5:575.
[29] US Food and Drug Administration. Artificial intelligence and machine learning (AI/ML)-enabled medical devices. October 5, 2022, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.
[30] Kelly BS, Judge C, Bollard SM, et al. Correction to: Radiology artificial intelligence: a systematic review and evaluation of methods (raise). Eur Radiol 2022;32:7998–8007.
[31] Wong ANN, He Z, Leung KL, et al. Current developments of artificial intelligence in digital pathology and its future clinical applications in gastrointestinal cancers. Cancers 2022;14:3780.
[32] Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: A blinded clinical validation and deployment study. Lancet Digit Health 2020;2:e407–16.
[33] Couture HD, Williams LA, Geradts J, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. npj Breast Cancer 2018;4:30.
[34] Wang T, Yan D, Liu Z, et al. Diagnosis of cervical lymph node metastasis with thyroid carcinoma by deep learning application to CT images. Front Oncol 2023;13:1099104.
[35] Cai H, Feng X, Yin R, et al. MIST: Multiple instance learning network based on swin transformer for whole slide image classification of colorectal adenomas. J Pathol 2023;259:125–35.
[36] Tummala S, Kim J, Kadry S. BreaST-net: multi-class classification of breast cancer from histopathological images using ensemble of swin transformers. Mathematics 2022;10:4109.
[37] Winchester DP, Cox JD. Standards for diagnosis and management of invasive breast carcinoma. CA Cancer J Clin 1998;48:83–107.
[38] Noorbakhsh J, Farahmand S, Foroughi Pour A, et al. Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. Nat Commun 2020;11:6367.
[39] Scholler J, Mandache D, Mathieu MC, et al. Automatic diagnosis and classification of breast surgical samples with dynamic full-field OCT and machine learning. J Med Imaging 2023;10:034504.

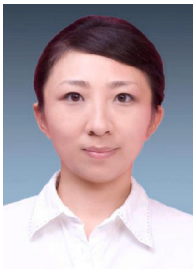
[40] Lu MY, Williamson DFK, Chen TY, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 2021;5:555–70.

Shuwei Zhang received her Bachelor degree in 2018 from Peking University. She is currently a M.D. candidate at Peking University People's Hospital. Her current research focuses on the translational research for breast cancer diagnosis.

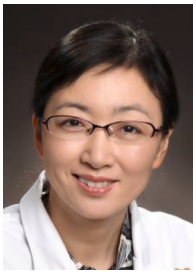
Bin Yang received the M.S. degree in Software Engineering from Beijing University of Technology in 2020. He is currently engaged in research at the China ESG Institute of the Capital University of Economics and Business and is pursuing his Ph.D. degree in Software Engineering at Beijing University of Technology. His research interest includes deep learning, computer vision and medical image analysis.

Houpu Yang received his M.D. degree from Peking University in 2009. Currently, he works as a chief physician at Peking University People's Hospital. His research is centered around comprehensive treatment, plastic surgery and diagnostic imaging of breast cancer.

Kang Zhang, M.D., PhD, is a chair professor and vice dean of the Faculty of Medicine, Macau University of Science. He obtained his M.D. degree from Harvard University and Massachusetts Institute of Technology, and Ph.D. degree in Genetics from Harvard University. He is the first Chinese ophthalmology resident at Johns Hopkins University Wilma Eye Center. His research is centered in genetics, epigenetics, stem cells, nanoengineering and 3D printing, clinical trials, and artificial intelligence.



Bo Liu received the B.S. degree from Beijing Institute of Technology and the M.S. and Ph.D. degrees from Tsinghua University in 2003 and 2008, respectively. She has worked at various institutions including NEC Laboratory China, The University of Chicago, Argonne National Laboratory, Beijing University of Technology, and Massey University. Her current research interests include big data, data mining, machine learning, cloud computing, scientific workflow, Semantic Web and ontology reasoning.



Shu Wang, M.D., Ph.D., serves as the director of the Breast Center at Peking University People's Hospital. She is a member of the Expert Committee on Breast Cancer of the Chinese Society of Clinical Oncology (CSCO-BC) and the Chinese Breast Cancer Society (CBCS). She is dedicated to advancing precision medicine and translational research in breast cancer. Her focus spans AI assisted diagnosis, organoid and biomarker guided systemic treatment, and precision surgery.