

<https://doi.org/10.1038/s41746-025-01833-6>

Transformer optimization with meta learning on pathology images for breast cancer lymph node micrometastasis



Jing Huang¹, Jingtao Wang², Junhai Shi¹, Hengli Ni^{3,4}, Shan Xu⁵, Ping Wu⁵, Yuexiang Ren^{5,6}, Lijuan Bian⁴, Chenhan Su¹, Yuxuan Xu¹, Xinyu He¹, Xinjian Chen^{2,7}✉ & Jianming Li^{1,4,8}✉

Lymph node micro-metastasis represents the initial stage of breast cancer spread or metastasis. However, the limited size of these hidden lesions restricts dataset expansion, presenting a significant challenge for manual examination and conventional deep learning techniques. By harnessing the power of meta-learning on limited datasets, we developed a novel network named MetaTrans, equipped with a 34-category dataset (MT-MCD) to effectively pinpoint micro-metastases in lymph nodes from pathological images. MetaTrans demonstrated superior performance on two different multi-center datasets and excelled in the 0-shot task for intraoperative frozen section diagnosis. Beyond breast cancer, MetaTrans efficiently identifies micro-metastases in thyroid and colorectal cancers and can be directly applied to recognize images captured by digital cameras under a microscope. Across all clinical validation scenarios, our method surpasses state-of-the-art baselines, exhibiting robust cross-domain adaptation and task-specific reliability, which highlight its translational potential in diverse pathological settings.

Breast cancer constitutes a predominant health concern, with an annual incidence exceeding 2 million cases. In 95% of countries, breast cancer is the leading cause of female cancer-related mortality¹. Lymph node metastasis is a crucial mechanism in the progression of breast cancer. As the primary site of initial metastasis, axillary sentinel lymph nodes² play key roles in pathological biopsy, serving for subsequent treatment decisions, staging, and prognosis. In some early-stage breast cancer patients, lymph node metastasis is negative. These patients are eligible not only for breast-conserving surgery but also for de-escalated adjuvant therapies, such as hypofractionation, partial breast irradiation, and intraoperative radiotherapy³. However, some lymph nodes diagnosed as negative are not truly negative. While macro-metastases are easily identified, micro-metastases are often overlooked due to the subtle nature of the lesions, especially in more challenging frozen section analyses. Patients who are misclassified in this way may not receive adequate adjuvant therapy or follow-up, increasing their risk of recurrence⁴. Therefore, accurately

identifying micro-metastases that might otherwise be mistaken for negative can help ensure that patients receive more tailored and effective treatment strategies.

Lymph node metastases are typically examined by pathologists through hematoxylin and eosin (H&E) stained histopathology slides. With the development of Whole-slide Images (WSIs), artificial intelligence (AI) diagnostics can greatly assist in this laborious and error-prone process. In recent years, deep learning (DL) methods have shown significant improvements on a wide range of computer vision tasks on natural images, e.g., image classification^{5–7}, object detection^{8,9} and semantic segmentation¹⁰. Similarly, some promising studies have employed DL method for diagnosing lymph node metastasis and achieved good performances^{11–16}, and various computer-aided diagnosis (CAD) systems have been developed to assist pathologists to detect cancer metastasis in WSIs^{17–19}. However, a common challenge is the low accuracy in identifying micro-metastases^{12,20,21}. According to the American joint Committee on Cancer tumor-node-

¹Department of Pathology, Soochow Medical College, Soochow University, Suzhou, China. ²School of Electronics and Information Engineering, Soochow University, Suzhou, China. ³Department of Pathology, Children's Hospital, Soochow University, Suzhou, China. ⁴Department of Pathology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China. ⁵Department of Pathology and Institute of Molecular Pathology, Jiangxi Provincial Key Laboratory for Precision Pathology and Intelligent Diagnosis, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China. ⁶Department of Pathology, The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China. ⁷State Key Laboratory of Radiation Medicine and Protection, Soochow University, Suzhou, China. ⁸The MOE Basic Research and Innovation Center for the Targeted Therapeutics of Solid Tumors, Department of Pathology and Institute of Molecular Pathology, Jiangxi Provincial Key Laboratory for Precision Pathology and Intelligent Diagnosis, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China. ✉e-mail: xjchen@suda.edu.cn; lijming3@sysu.edu.cn

metastasis staging system (8th version), lymph node metastasis of breast cancer is categorized into three groups: isolated tumor cells (ITCs), defined as single tumor cells or small clusters measuring less than 0.2 mm; micro-metastasis, with a size range of 0.2–2.0 mm; and macro-metastasis, exceeding 2.0 mm²². The smaller the metastatic lesion, the more challenging its identification becomes, both for pathologists and algorithms.

Micro-metastasis and ITCs (collectively referred as small metastasis) represent the initial stage of lymph node metastasis and carries significant clinical relevance. Due to the concealed of the lesions, acquiring slides is difficult, resulting in a relatively small dataset. Moreover, owing to the diminutive size, the number of labeled positive patches is considerably smaller than that of macro-metastasis, further limiting dataset expansion. These constraints restrict the performance of big data-driven DL methods in identifying lymph node micro-metastasis. Meta-learning is a focal area of investigation in few-shot learning. Recently, it has received a lot of attention from researchers as a potential solution for challenges such as limited resources, scarcity of labeled samples, and domain shift^{23,24}. Given the limitation of the small metastases' dataset, meta-learning becomes a more suitable method.

Recent studies demonstrate that Vision Transformers (ViT)^{25,26}, leveraging multi-head self-attention mechanisms and long-range dependency modeling, exhibit superior performance over traditional convolutional neural networks (CNNs) in natural image processing. However, in medical imaging analysis, Transformers face two critical challenges: (1) self-attention mechanisms struggle to efficiently capture fine-grained local details (e.g., edge morphology of micro-lesions); (2) their high computational complexity and requiring of large-scale training data limit applicability in clinical settings. To address these limitations, researchers have proposed hybrid architectures that integrate the strengths of CNNs and Transformers. For instance, Gulati et al.²⁷ proposed Conformer, which synergizes parallel convolutional and Transformer blocks to balance local feature retention and global semantic representation. TransUNet²⁸ embeds Transformers Blocks into the UNet encoder, significantly enhancing medical image segmentation performance. Building on these advancements, we propose the MetaTrans network—a novel architecture optimizes Transformer with Meta-learning and CNN to improve the detection of lymph node micro-metastasis in breast cancer across multi-center pathological images, particularly under limited data conditions. MetaTrans is specifically designed for limited-data scenarios. By employing meta-learning strategies, the network effectively adapts to scarce annotated data and achieves high detection accuracy for small metastases, including micro-metastases and ITCs.

Due to the large size of a WSI, it cannot be directly fed into a deep learning model for detection. A common solution is to divide the WSI into numerous smaller patches that fit the model's input size. However, because of hardware limitations during clinical deployment, this approach generates tens of thousands of patches under basic configurations, many of which contain redundant information, resulting in detection times that may reach the hour-level range¹⁵. While high-end hardware and network resources can reduce this time to under five minutes¹³, such configurations incur high costs and are typically limited to research laboratories, making them less feasible for clinical applications²⁹. Another approach is to downscale the WSI, but this leads to the loss of significant information, making it impossible for the model to extract features at the cellular level. To address this challenge, we trained end-to-end a tissue-recognition model for regions of interest at low magnification (Model4x), and a cell-recognition model for high magnification images (Model10x). We propose the Breast Cancer Lymph Node Diagnosis Process in Field of View (FOV) Based on MetaTrans (Fig. 1), inspired by the diagnostic practices of pathologists. They can capture a FOV at 4× magnification using either a microscope or a WSI. The captured FOV is divided into 256 × 256-sized patches (typically around 100), which are then processed by MetaTrans for detection. The model generates a probability distribution for all patches and produces an attention map within 5 s, making it feasible for time-sensitive intraoperative freeze examinations. When this attention map is overlaid onto the original FOV, it

can assist pathologists in diagnosis and guide surgical decisions. Furthermore, pathologists can choose to use a 10× magnified view for the model's detections. This higher magnification offers more detailed information than the 4× view, further improving the reliability of the diagnostic results.

Our contributions include the following: First, we collected and labeled a 34-class WSI dataset specifically for meta-training. A WSI patch-based meta-task construction strategy was proposed to fully utilize the various tissue categories within the WSI, enabling finer tissue identification. Second, we designed the MetaTrans network by combining meta-learning with Transformer blocks to address the challenge of detecting small metastases under limited data conditions. Specifically, the network is trained end-to-end and remains effective even when sample sizes are significantly smaller than those for macro-metastases. Third, we collected and labeled multi-center small metastasis datasets, including private and public data, and conducted cross-dataset and cross-disease experiments. These experiments demonstrated the superiority of our method. Finally, We collected various image types, including whole slide images (WSIs) and digital microscope captures from both paraffin and frozen sections. These experiments highlight the strong potential of our approach for clinical applications.

Results

Cross-validated model performance in BLCN-MiD and camelyon datasets

The paraffin section (FFPE) from BLCN-MiD and Camelyon underwent training using 5-fold cross-validation on training dataset (Supplementary Fig. 1). We conducted a comparative analysis involving several methodologies, including a deep learning-based simple convolutional network (D-Conv-4), lightweight network (squeezenet1_0), classical classification networks (ResNet18, ResNet34, ResNet 50) and transformer architectures (ViT_Small, ViT_Base). Our network also trained via deep learning (dMetaTrans) and meta-learning (iMetaTrans, mMetaTrans, rMetaTrans). Results were reported in Table 1 and Supplementary Tables 1–2, evaluating performance metrics such as ACC, AUC, Precision, Recall, and F1-Score at Model4x and Model10x. Corresponding ROC curves and confusion matrixes were illustrated in Fig. 2a, b and Supplementary Fig. 2. Generally, all networks performed well in Model10x rather than Model4x, indicating greater difficulty at lower magnifications. Notably, our network (dMetaTrans) demonstrated superior performance. Moreover, further enhancements were observed when transitioning to meta-learning, particularly with rMetaTrans achieving the highest performance across all metrics. At 4× magnification, rMetaTrans surpassed the second-ranked network in both BLCN-MiD (by 3.67%, 5.9%, 5.97%, 5.55%, 5.57% in 5 metrics above) and Camelyon (by 6.17%, 5.56%, 5.78%, 6.15%, 6.96%) datasets. Most metrics for rMetaTrans also showed improved scores at Model10x compared to Model4x, although the relative improvement over the second-ranked network was less pronounced at higher magnifications. To visually compare the feature learned by different networks, we utilized split violin plots as shown in Supplementary Fig. 3a. The outputs of network were normalized as positive scores, with the legend representing the patch labels. Patches with a positive score greater than 0.5 were detected to be metastatic, while those with scores less than 0.5 were detected to be non-metastatic. The results indicated that the features obtained by our network exhibit comparable separability to those obtained by other networks. These findings underscore the effectiveness of our approach in optimizing tissue structure feature extraction and recognition, particularly highlighting its efficacy at lower magnifications.

To visualize and interpret the relative importance of each region in the field, we generated attention maps based on the detections of models. The normalized scores were mapped to their corresponding spatial locations in the original field (Fig. 2c). We selected the representative networks with superior performance for comparison. The results showed the trained models are generally capable of delineating the boundary between detected tumor and normal tissue. Notably, the boundary delineated by our network, especially rMetaTrans, aligns more closely with the ground truth annotations compared to other models. Additionally, the attention maps can be

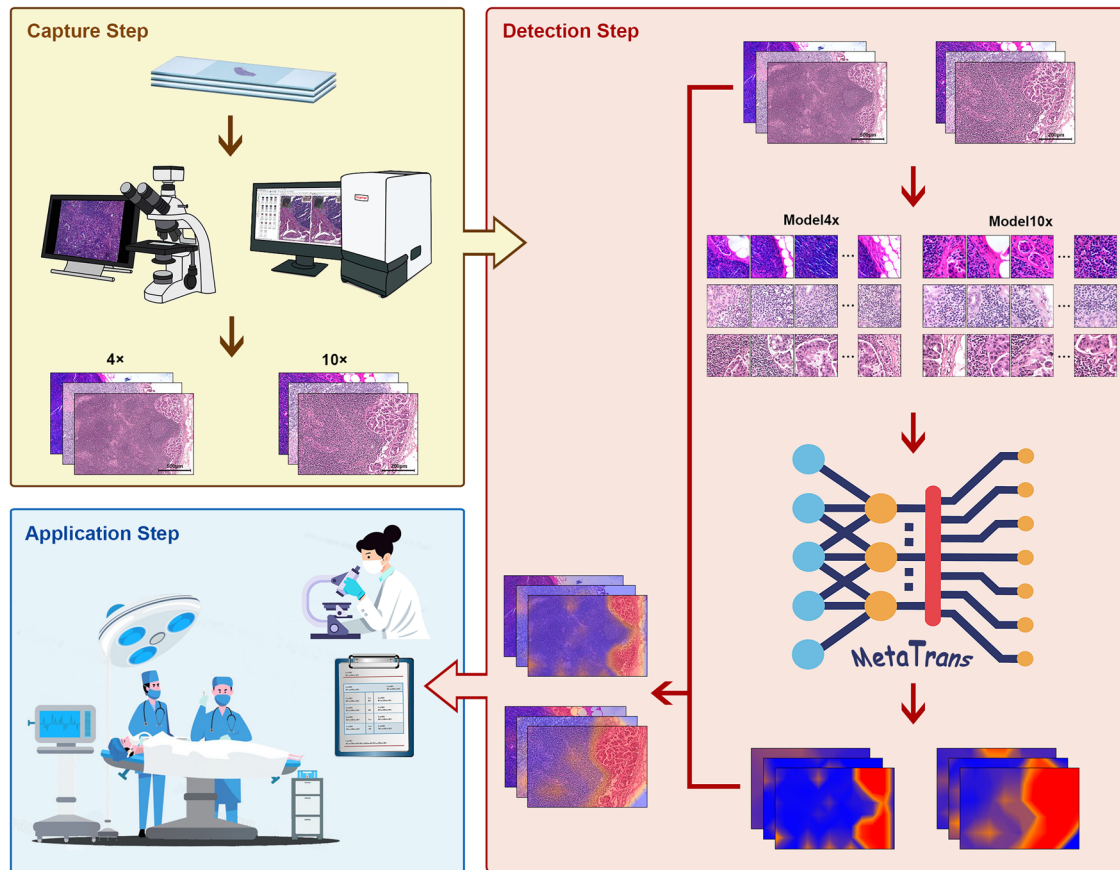


Fig. 1 | Breast cancer lymph node diagnosis process in FOVs based on MetaTrans. Pathologists capture the FOVs using a microscope or WSI as the target. The target is then divided into patches, which subsequently detected by MetaTrans. The model

generates a probability distribution for all patches and draws an attention map, which is overlaid onto the original target. This merged image aids diagnosis and guide surgical procedures. Scale bar, 500 μ m, 200 μ m.

Table 1 | The ACC and AUC of indicated networks on training dataset of FFPE in BCLN-MiD and Camelyon using 5-fold cross-validate

Method	BCLN-MiD				Camelyon			
	Model4x		Model10x		Model4x		Model10x	
	ACC (%)	AUC (%)	ACC (%)	AUC (%)	ACC (%)	AUC (%)	ACC (%)	AUC (%)
dConv_4	77.19 $\pm$ 2.27	79.57 $\pm$ 3.90	78.46 $\pm$ 2.17	81.55 $\pm$ 3.42	77.91 $\pm$ 1.07	80.47 $\pm$ 5.02	82.41 $\pm$ 1.70	89.01 $\pm$ 1.61
dSqueezeNet1_0	78.22 $\pm$ 1.90	49.53 $\pm$ 17.17	81.25 $\pm$ 0.97	71.63 $\pm$ 6.56	77.29 $\pm$ 1.07	31.77 $\pm$ 13.63	85.93 $\pm$ 1.26	54.91 $\pm$ 11.42
dResNet18	78.69 $\pm$ 1.14	82.84 $\pm$ 3.48	83.63 $\pm$ 0.82	88.91 $\pm$ 1.80	78.27 $\pm$ 0.89	83.46 $\pm$ 5.18	85.59 $\pm$ 1.96	92.83 $\pm$ 1.87
dResNet34	79.90 $\pm$ 1.77	82.90 $\pm$ 3.50	83.02 $\pm$ 1.51	88.17 $\pm$ 3.25	78.90 $\pm$ 1.22	84.96 $\pm$ 2.59	85.62 $\pm$ 1.35	93.06 $\pm$ 1.09
dResNet50	79.32 $\pm$ 1.31	84.08 $\pm$ 2.69	83.25 $\pm$ 0.88	90.08 $\pm$ 0.83	78.64 $\pm$ 0.98	83.47 $\pm$ 3.42	87.49 $\pm$ 1.74	94.49 $\pm$ 1.24
dViT_Small	82.29 $\pm$ 1.59	85.55 $\pm$ 3.68	83.64 $\pm$ 0.91	89.82 $\pm$ 2.48	80.64 $\pm$ 1.98	86.40 $\pm$ 3.13	87.86 $\pm$ 1.82	94.97 $\pm$ 1.25
dViT_Base	74.18 $\pm$ 6.10	70.27 $\pm$ 7.76	82.09 $\pm$ 1.26	88.40 $\pm$ 1.81	70.64 $\pm$ 5.29	71.10 $\pm$ 9.98	86.00 $\pm$ 1.96	93.25 $\pm$ 1.52
dMetaTrans (Ours)	82.50 $\pm$ 1.94	87.70 $\pm$ 1.66	83.20 $\pm$ 1.96	88.89 $\pm$ 2.82	83.38 $\pm$ 1.70	88.50 $\pm$ 3.52	86.67 $\pm$ 1.57	93.89 $\pm$ 1.26
iMetaTran (Ours)	84.08 $\pm$ 0.79	90.17 $\pm$ 2.08	84.23 $\pm$ 1.70	90.32 $\pm$ 2.62	85.05 $\pm$ 1.57	91.32 $\pm$ 2.10	88.13 $\pm$ 1.87	94.84 $\pm$ 1.31
mMetaTrans (Ours)	84.75 $\pm$ 1.07	90.68 $\pm$ 1.25	84.52 $\pm$ 1.64	90.77 $\pm$ 3.25	86.05 $\pm$ 0.70	91.55 $\pm$ 1.76	88.43 $\pm$ 2.12	95.22 $\pm$ 1.36
rMetaTrans (Ours)	85.96 $\pm$ 1.33	91.45 $\pm$ 2.11	85.53 $\pm$ 1.35	92.39 $\pm$ 1.35	86.81 $\pm$ 1.50	91.96 $\pm$ 1.89	89.39 $\pm$ 2.09	95.87 $\pm$ 1.33

The bolded values highlight the best-performing result for each metric across the compared methods.

extended to the whole-slide level, where our network consistently generated the most accurate detective attention maps (Supplementary Fig. 3b).

Adaptability to independent test cohorts

The exceptional performance of our method stems from its ability to leverage meta-learning for few-shot adaption, enabling effective identification of concealed lesions. To validate this capability, we directly evaluated

the pre-trained networks above on testing dataset of Camelyon in FFPE. This dataset was further categorized into micro-metastases, ITCs, and negative instances. Additionally, we downloaded 27 macro-metastatic lymph nodes from the Camelyon16 for testing (Fig. 3a). Supplementary Table 3 illustrated the metrics achieved by various networks. Figure 3b–e showed the AUC and corresponding ROC curves in different categories. All networks demonstrated strong performance on macro-metastases, but they

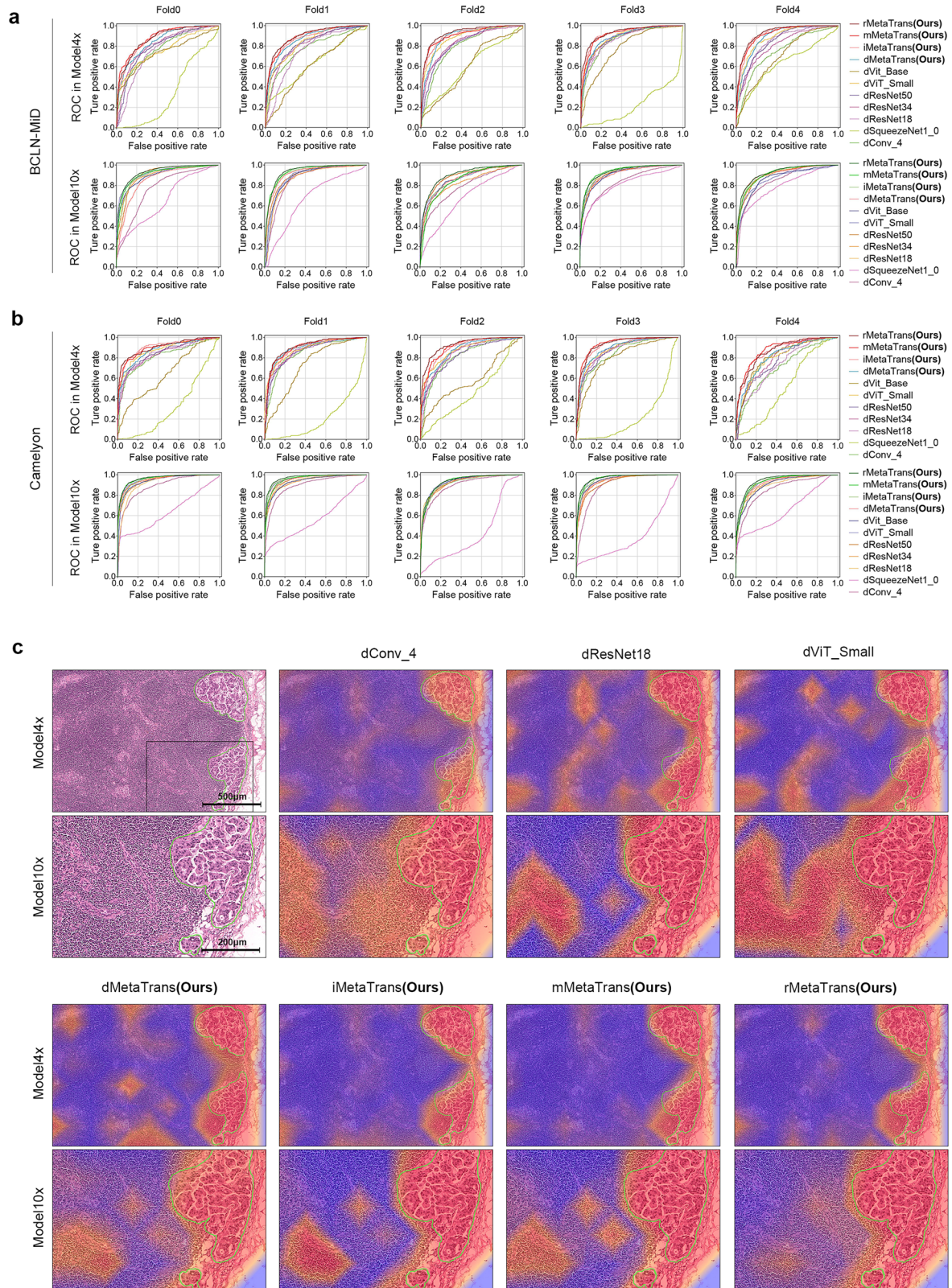


Fig. 2 | Cross-validated model performance and visualization. a, b ROC curves of indicated networks in Model4x and Model10x. The experiments conducted in BCLN-MiD and Camelyon dataset. **c** Visualization of FOV in FFPE of BCLN-MiD.

The FOVs in 4× and 10× magnification were the final evaluation targets. Ground truth metastatic areas were marked with a green border. The color blue to red in attention map represented normalized score of 0–1. Scale bar, 500 μm, 200 μm.

displayed lower AUC on micro-metastases, with many showing even poorer performance on ITCs. The split violin plots also indicated the worst comparable separability in ITCs group (Fig. 3f). Our network maintained high AUC with macro-metastases while showcasing superior performance with

micro-metastases and ITCs. Furthermore, the results highlight that in our network, the AUC of dMetaTrans consistently underperforms compared to iMetaTrans, mMetaTrans, and rMetaTrans. This underscored the superiority of meta-learning approaches in small metastases recognition task.

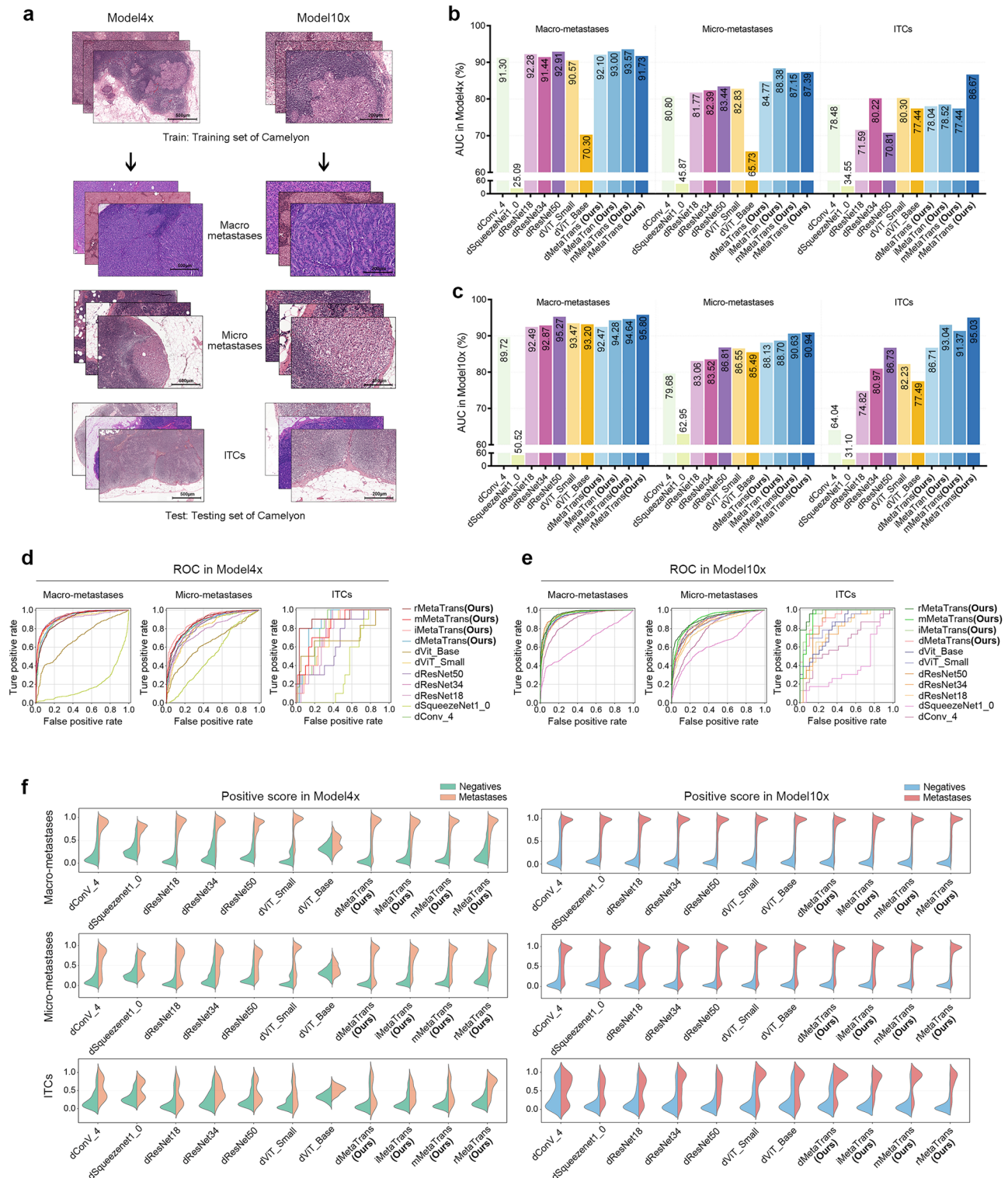


Fig. 3 | Adaptability to independent test cohort in FFPE of Camelyon. **a** Flow chart of experiment. The testing dataset was further categorized into macro-metastatic, micro-metastases, ITCs, and negative instances. The networks were directly evaluated without training again. Scale bar, 500 μ m, 200 μ m. **b**, **c** AUC obtained by

different networks over three categories in Model4x and Model10x. See Supplementary Table 3 for complete results. **d**, **e** ROC curves of indicated networks in Model4x and Model10x. **f** The distribution of positive scores for each patch. The legends represented the labels.

Adaptability for frozen section analysis

To evaluate the adaptability of our method across varying clinical scenarios, we further compared our method and baseline networks on different types of tissue slices. Frozen section (FS) typically requires pathology reports within 30 min to facilitate immediate surgical decision-making. However, in order to prevent tissue shrinkage, thicker FS often suffer from poor clarity

(Fig. 4a). The ice crystals formed by rapid freezing distorted cell morphology, causing swelling and enlarged nuclei. Due to these challenges, experienced pathologists achieve only about 95% accuracy. Networks trained on paraffin sections (FFPE) were directly evaluated on FS test sets without retraining (FFPE $\rightarrow$ FS) (Fig. 4b). The results showed in Fig. 4c, d and Supplementary Table 4. Corresponding ROC curves were illustrated in

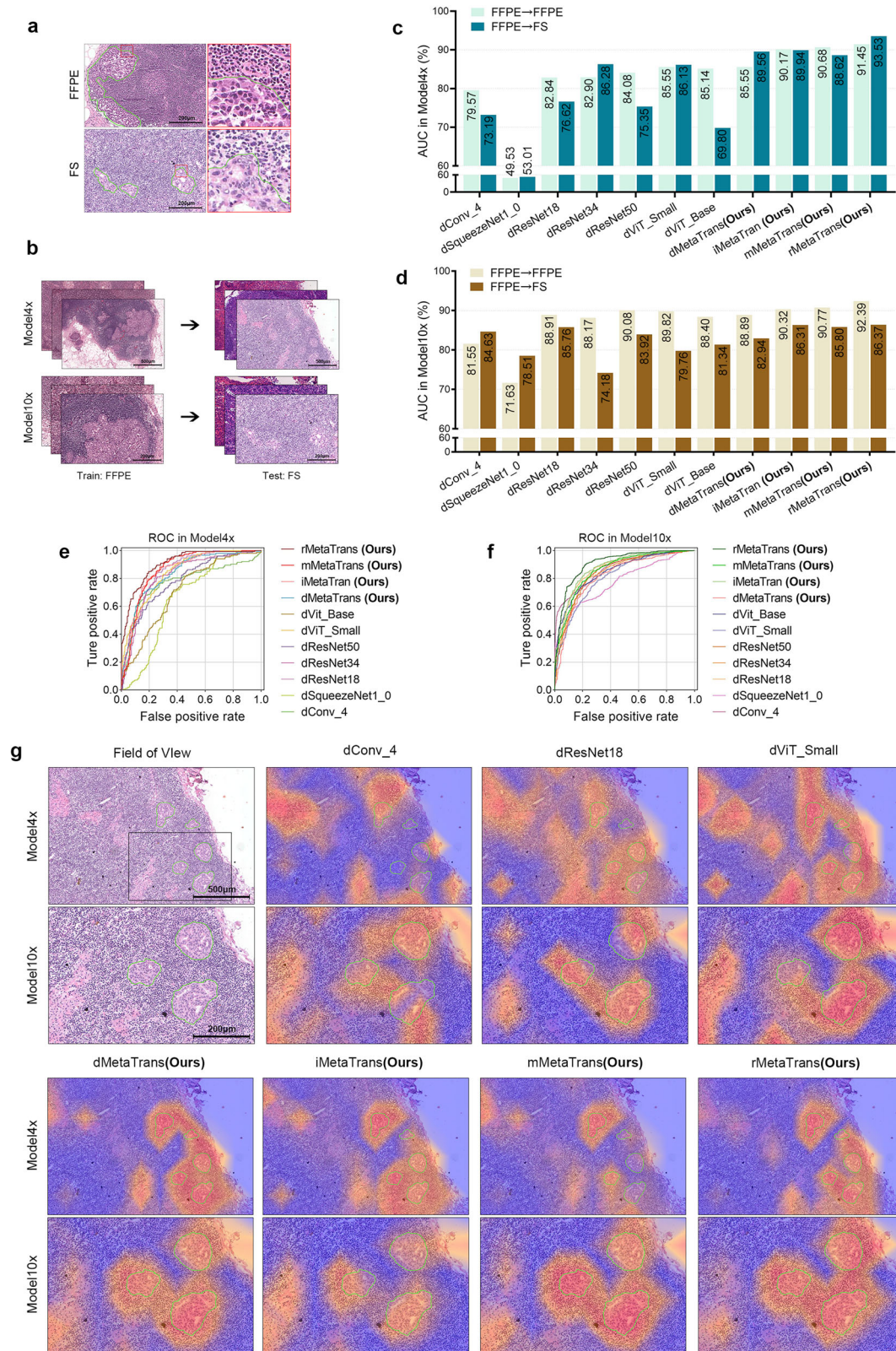


Fig. 4 | Adaptability for frozen section analysis in BCLN-MiD. a Description of differences between FFPE and FS. **b** Flow chart of experiment. The networks were directly evaluated without training again. **c, d** AUC obtained by different networks in Model4x and Model10x. The “FFPE → FFPE” means networks underwent training using 5-fold cross-validation (See Supplementary Table 1 for complete results). The

“FFPE → FS” means the networks trained above were directly evaluated on FS without retraining (See Supplementary Table 4 for complete results). **e, f** ROC curves of indicated networks in Model4x and Model10x in FS. **g** Visualization of Model4x and Model10x detections in FS. Scale bar, 500 μm, 200 μm.

Fig. 4e, f. We found that most networks performed less effectively on FFPE $\rightarrow$ FS compared to FFPE $\rightarrow$ FFPE (The results of BCLN-MiD in Table 1). However, our network maintained the highest scores for metrics when transitioning from FFPE to FS, with minimal performance degradation, particularly rMetaTrans in Model4x: 1.11% in ACC, 4.36% in precision, and 0.05% in F1-score. Additionally, AUC increased by 2.08%, and recall improved by 4.39%. While the performance on Model10x was slightly lower than on Model4x, our network still stood out in most metrics. The attention maps in Fig. 4g demonstrated that rMetaTrans has the smallest false positive area, provided that the metastases are completely identified. This highlights the method's ability to focus on diagnostically relevant features and maintain precision under limited supervision, a key advantage of meta-learning-based approaches.

Adaptability to microscope digital camera imaging

To facilitate the translation of our network to clinical applications, we extra collected 20 FFPE slides and 20 FS slides of small metastatic lymph nodes, with images taken directly by a microscope digital camera. The trained baseline models and MetaTrans models were used to directly evaluate these digital camera images (DCIs) (Fig. 5a). Comprehensive experimental results indicate that MetaTrans, particularly rMetaTrans, consistently achieved superior performance across the majority of evaluation metrics (Supplementary Tables 5–6). The results of MetaTrans models were shown in Fig. 5b, c, with the corresponding ROC curves presented in Fig. 5d, e. Both the Model4x and Model10x performed well on FFPE DCIs, with maximum AUC of 94.41% and 94.28%. The DCIs in FS are slightly worse, but the maximum AUC is still 93.57% and 91.40%. The attention maps in Fig. 5f indicate that the true metastases in DCIs were recognized by all MetaTrans, and the detected area of rMetaTrans was the most consistent with the ground truth annotations in green, whether in FFPE or FS. These findings suggested that rMetaTrans exhibits superior performance in the direct evaluation of DCIs, underscoring its significant potential for clinical applications.

Versatility in detecting thyroid and colorectal cancer

Given the superior performance of our network in detecting small metastases in breast cancer lymph nodes, we extended its evaluation to other tumor types. Colorectal cancer (CRC) is known for its high recurrence and lethality³⁰. Approximately 30–40% of stage II CRC patients initially classified as lymph node-negative experience recurrence due to undetected micro-metastases, which decreases survival rates³¹. Thyroid cancer is the most common head and neck malignancy and often metastasizes early to cervical lymph nodes³². We collected 59 lymph nodes in WSIs with small metastases in thyroid cancer and 14 in CRC. The baseline and MetaTrans models, trained on BCLN-MiD data, were directly applied for evaluation (Fig. 6a). Extensive experimental results confirm that MetaTrans, especially the rMetaTrans consistently outperformed baseline methods across all evaluated metrics (Supplementary Tables 7–8). The results of MetaTrans models are presented in Fig. 6b, c, with corresponding ROC curves shown in Fig. 6d, e. Both Model4x and Model10x performed exceptionally well in detecting CRC metastases, achieving maximum AUC of 97.78% and 99.15%, respectively. For thyroid cancer metastases, Model10x also demonstrated strong performance with an AUC of 98.74%. However, Model4x showed lower performance with an AUC of 86.79%, possibly due to the colloid lakes, a structural feature absent in breast cancer, which may have led to miscalculations by the model. Attention maps (Fig. 6f) revealed that Model4x struggled to define the entirety of thyroid cancer metastases. In contrast, Model10x showed improved performance at 10 $\times$ magnification. In CRC, the metastases detected by MetaTrans closely matched ground truth annotations, particularly with the rMetaTrans model. These findings demonstrate that our network effectively recognizes small metastases in other adenocarcinomas, especially CRC, which has structural similarities to breast cancer.

Ablation experiments on MT-MCD

To systematically evaluate the impact of MT-MCD on cross-task transfer performance, we conducted ablation experiments on the test sets of target

datasets BCLN-MiD and Camelyon. Specifically, we designed three comparison strategies to isolate the effects of MT-MCD and meta-learning. In Method 1 [MT-MCD ($\times$), Meta-Learning ($\times$)], the model is trained directly on target datasets using deep learning (DL) without MT-MCD or meta-learning. In Method 2 [MT-MCD ($\checkmark$), Meta-Learning ($\times$)], a baseline model is first pre-trained on MT-MCD through DL and then fine-tuned on the target datasets, allowing us to assess the effect of MT-MCD alone^{33,34}. Finally, in Method 3 [MT-MCD ($\checkmark$), Meta-Learning ($\checkmark$)], the model undergoes meta-training on MT-MCD using the rMetaTrans strategy to jointly optimize parameters for enhanced cross-task generalization.

All methods selected their optimal models based on validation performance and were evaluated on the test sets to assess cross-domain performance. As shown in Supplementary Table 9, Method 2 outperforms Method1 in 4 out of 5 evaluation metrics at both 4 $\times$ and 10 $\times$ magnifications on the BCLN-MiD dataset. On the Camelyon dataset, Method 2 consistently achieves superior performance across all metrics compared to Method 1. These results highlight that MT-MCD enhances domain adaptability during conventional pre-training by introducing multi-category pathological representations. More notably, although Method 3 shows slightly lower AUC values on the BCLN-MiD dataset relative to Method 1 and Method 2, it demonstrates significant superiority in key performance metrics including Accuracy, Precision, Recall, and F1-Score. On the Camelyon dataset, Method 3 consistently outperforms all methods across the full set of evaluation metrics. These findings suggest that MT-MCD, through its structured meta-task repository, effectively integrates biologically-informed priors into the meta-learning process, and that the diversity of its task construction substantially improves model robustness and adaptability across heterogeneous pathological scenarios.

Discussion

Most AI-based tumor recognition methods rely on large-scale WSI data to identify macro metastases directly^{20,35,36}, which is impractical in small metastases. Although traditional machine learning methods^{37,38} do not require extensive data labeling, their reliance on manually designed features, coupled with limitations in accuracy and model robustness, restricts their clinical applicability. Some semi-supervised or weakly supervised methods^{39–42} alleviate labeling difficulties but still require a sufficient number of samples, which conflicts with the scarcity of small metastasis samples. In contrast, our work utilizes the MetaTrans network with a tailored training strategy that leverages meta-learning for few-shot adaptation, combined with the representational strength of Transformer blocks. This approach allows us to achieve high diagnostic performance, even with limited small metastasis samples. Meta-learning focuses on learning how to adapt to novel tasks without overfitting to previously seen tasks^{43–45}. Typically, it assumes a task distribution and employs a meta-learner to extract shared knowledge across tasks^{46,47}, a framework shown to be effective in few-shot learning settings with minimal supervision^{23,24,48,49}. To support this framework, we adopt a fully supervised learning paradigm using expert-annotated patch-level labels, which is particularly suited for small metastasis detection tasks requiring fine-grained spatial precision. Compared to weakly supervised approaches such as MIL^{12,50}, our design ensures more reliable localization and mitigates the label noise ambiguity often introduced by slide-level annotations. However, considering the gradual increase in available data or the potential inclusion of large volumes of similar pathological images as supplementary data, approaches such as multiple instance learning (MIL), self-supervised learning, or pre-trained medical foundation models may become more suitable choices. In the current study, to addressing the clinical challenge of limited sample resources for detecting small metastases in breast cancer, we are the first to integrate a meta-learning algorithm with a Transformer-based network. Our proposed MetaTrans network and training strategy are specifically designed to accurately detect small metastasis lesions.

MetaTrans demonstrates strong performance even with limited sample sizes due to its few shot learning skills and robust adaptation to diverse pathological patterns. Hence, we constructed the first 34-category

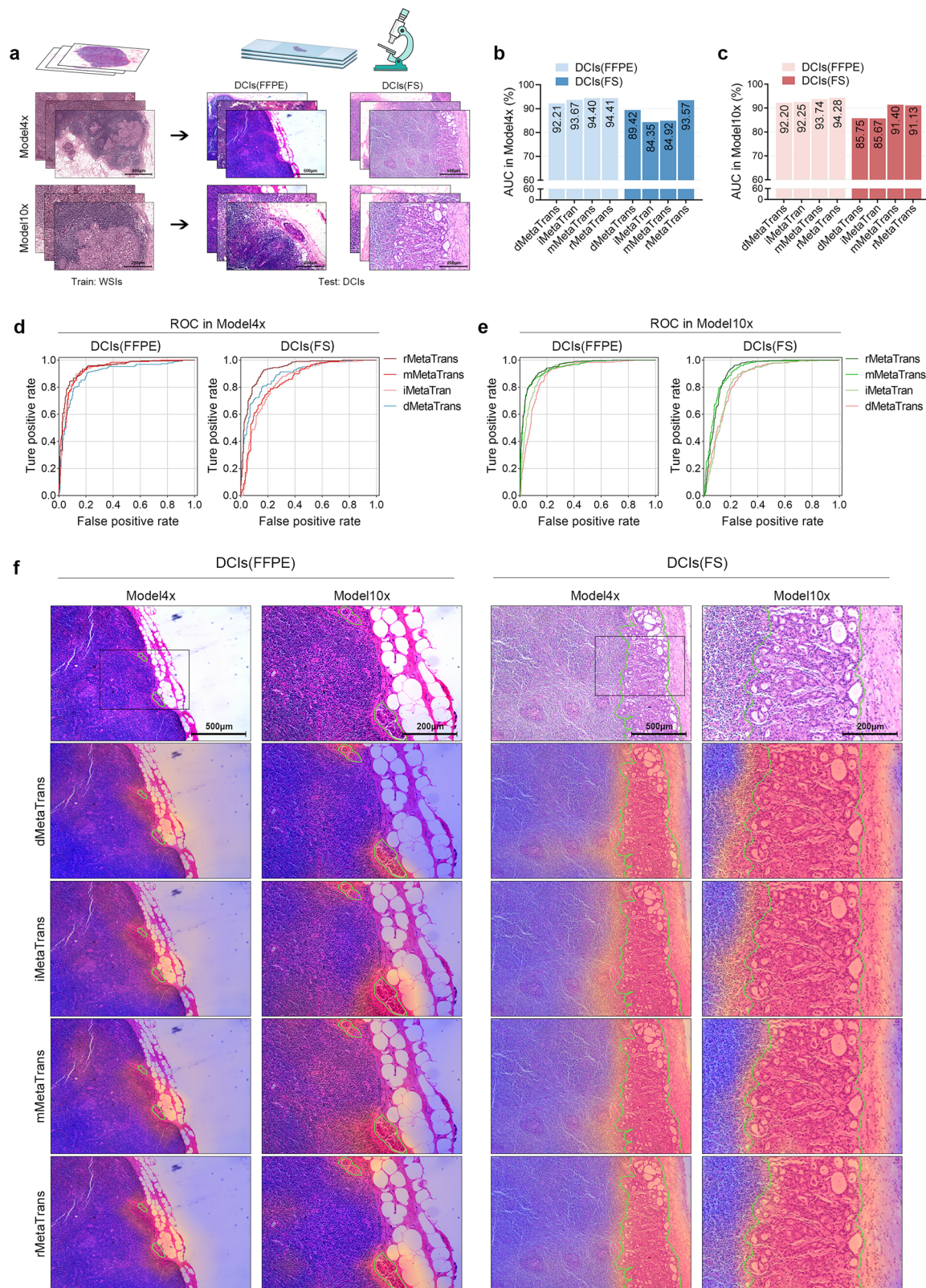


Fig. 5 | Adaptability to microscope digital camera imaging. **a** Flow chart of experiment. DCIs is captured from FFPE and FS. The networks were directly evaluated without training again. **b**, **c** AUC obtained by different networks in Model4x and Model10x. See Supplementary Tables 5-6 for complete results.

d, **e** ROC curves of indicated networks in Model4x and Model10x in DCIs from FFPE and FS. **f** Visualization of Model4x and Model10x detections in DCIs from FFPE and FS. Scale bar, 500 μ m, 200 μ m.

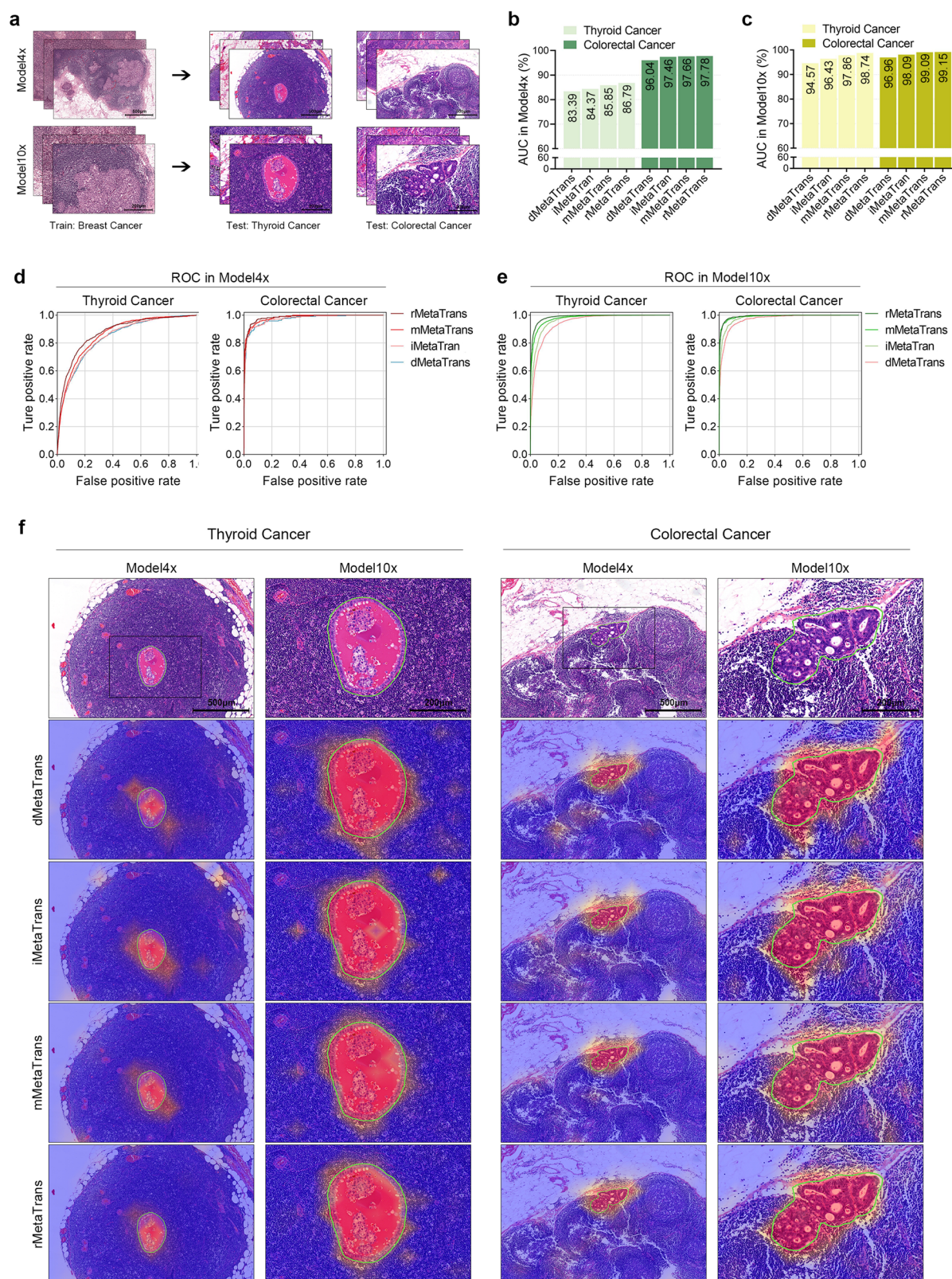


Fig. 6 | Versatility in detecting thyroid and colorectal cancer. **a** Flow chart of experiment. The networks were directly evaluated without training again. **b, c** AUC obtained by different networks in Model4x and Model10x. See Supplementary Tables 7–8 for complete results. **d, e** ROC curves of indicated networks in Model4x

and Model10x in lymph node from thyroid cancer and colorectal cancer. **f** Visualization of Model4x and Model10x detections in lymph node from thyroid cancer and colorectal cancer. Scale bar, 500 μ m, 200 μ m.

pathological dataset specifically for meta-training, named MT-MCD. Both Table 1 and Fig. 2 demonstrate the superiority of the models trained using meta-learning for identifying small metastases in breast cancer lymph nodes. It is noteworthy that our method achieved the highest Recall scores across all experiments, indicating that more metastatic labeled patches were identified. This is particularly significant for tumor metastasis detection, where missed diagnoses—especially of micro-metastases—can lead to serious clinical consequences. In line with this clinical need for high sensitivity, our field-of-view-based policy retains all positive patch-level predictions in the attention map (Fig. 2C). This allows pathologists to focus their review on highlighted suspicious regions rather than scanning the entire slide, thereby improving diagnostic efficiency while minimizing the risk of oversight. Nonetheless, there is still considerable room for expansion of the current MT-MCD. With the more diverse and larger datasets, the enhanced performance of our method will be, subsequently improving the identification of small metastases in breast cancer lymph nodes.

In addition, as shown in Table 1, among CNN-based models, dResNet50 achieved the highest AUC scores on the BCLN-MiD dataset at both FFPE-4× and 10× resolutions (84.08% and 90.08%, respectively), though their accuracy (ACC) was suboptimal. Among Transformer-based models, dViT_Small outperformed dViT_Base significantly at FFPE-4× (AUC: 85.55% vs. 70.27%), while their performance was comparable at 10× resolution (89.82% vs. 88.40%). On the Camelyon dataset, CNN models exhibited resolution-dependent performance: dResNet34 slightly surpassed dResNet50 at 4× (AUC: 84.96% vs. 83.47%), whereas dResNet50 dominated at 10× (94.49% vs. 93.06%). In contrast, the Transformer-based dViT_Small achieved optimal performance across both resolutions. These findings highlight two key insights: (1) model performance does not strictly correlate with network depth or parameter scale but requires architecture customization based on task-specific characteristics (e.g., data scale, image resolution); (2) neither pure CNN nor Transformer architectures can consistently dominate all evaluation metrics across diverse tasks.

Notably, despite the respective strengths of CNN and Transformer architectures—such as the local sensitivity of ResNet and the global modeling capability of ViT—the proposed MetaTrans achieves state-of-the-art performance across all tasks. MetaTrans employs a shallow CNN (Meta Learner) and a pretrained deep Transformer (Transformer Learner), establishing a hierarchical feature extraction and fusion mechanism that enable collaborative representation of fine-grained pathological details (e.g., cell-level nuclear atypia) and global semantic context. Moreover, A key innovation, decoupled meta-learning strategy: During meta-training, only the shallow CNN and classifier are optimized, while the deep Transformer’s ImageNet-pretrained weights remain frozen to preserve their robust general feature representation ability. This approach reduces computational complexity while improves diagnostic accuracy, enabling efficient adaptation to target tasks.

Interestingly, Model4x and Model10x, which were specifically trained to detect small metastases, also performed well in identifying macro metastases. During labeling, we observed that in many of the positive patches from small metastases, the metastatic area occupied a small proportion, increasing the difficulty for the models to recognize them. ITCs presented two significant challenges: the small number of patches and the small metastatic area they occupy, causing most networks to perform poorly on this task. However, the MetaTrans network, particularly rMetaTrans, maintained excellent performance in these conditions. This further demonstrates the superiority of rMetaTrans in handling tasks with small sample sizes.

Frozen sections (FS) are an important tool in intraoperative pathology examination. The 0-shot results on FS highlight the robust cross-domain adaptability of our method, demonstrating its effectiveness even without task-specific fine-tuning. Notably, Model4x outperformed Model10x in these scenarios, likely due to its emphasis on extracting global tissue architecture features, which remains relatively consistent between FFPE and FS. Conversely, Model10x emphasizes cellular morphological features, which are the main distinguishing factors between the two.

The lengthy time required to evaluate an entire WSI is a significant barrier to translating AI for pathological clinical practice. However, reducing computation costs by using network architectures with compressed parameters can significantly decrease model performance. For instance, SqueezeNet1_0 performed poorly in all experiments. Therefore, targeting the FOV as the final evaluation objective ensures high model performance and reduces time consumption while aligning with routine pathology practices. The adaptability of meta-learning enables our network to be applied to digital images captured by microscopes and to the identification of small metastases in lymph nodes for various adenocarcinoma types. The proposed MetaTrans, developed through meta-training on the MT-MCD, demonstrates significant potential for adapting to new clinical diagnostic tasks. It is particularly effective when faced with limited training data, making it highly suited to clinical needs. In summary, we hope our research and methods provide new ways to identify small metastases in routine pathological specimens, thereby advancing computational pathology and improving clinical pathology diagnostic efficiency.

Methods

Dataset

The Camelyon Dataset included Camelyon17 and Camelyon16. These two datasets are public challenge datasets for automated detection and classification of breast cancer metastases. We utilized all lymph nodes which tested micro-metastases, ITCs or negatives from these two public datasets to construct Camelyon dataset. In Camelyon17, 192 lymph nodes (112 negatives, 56 micro-metastases, 24 ITCs) from five medical centers were downloaded. In camelyon16, 320 lymph nodes (238 negative, 82 micro-metastases) from two medical centers were downloaded. We collectively refer to micro-metastases and ITCs as small metastases.

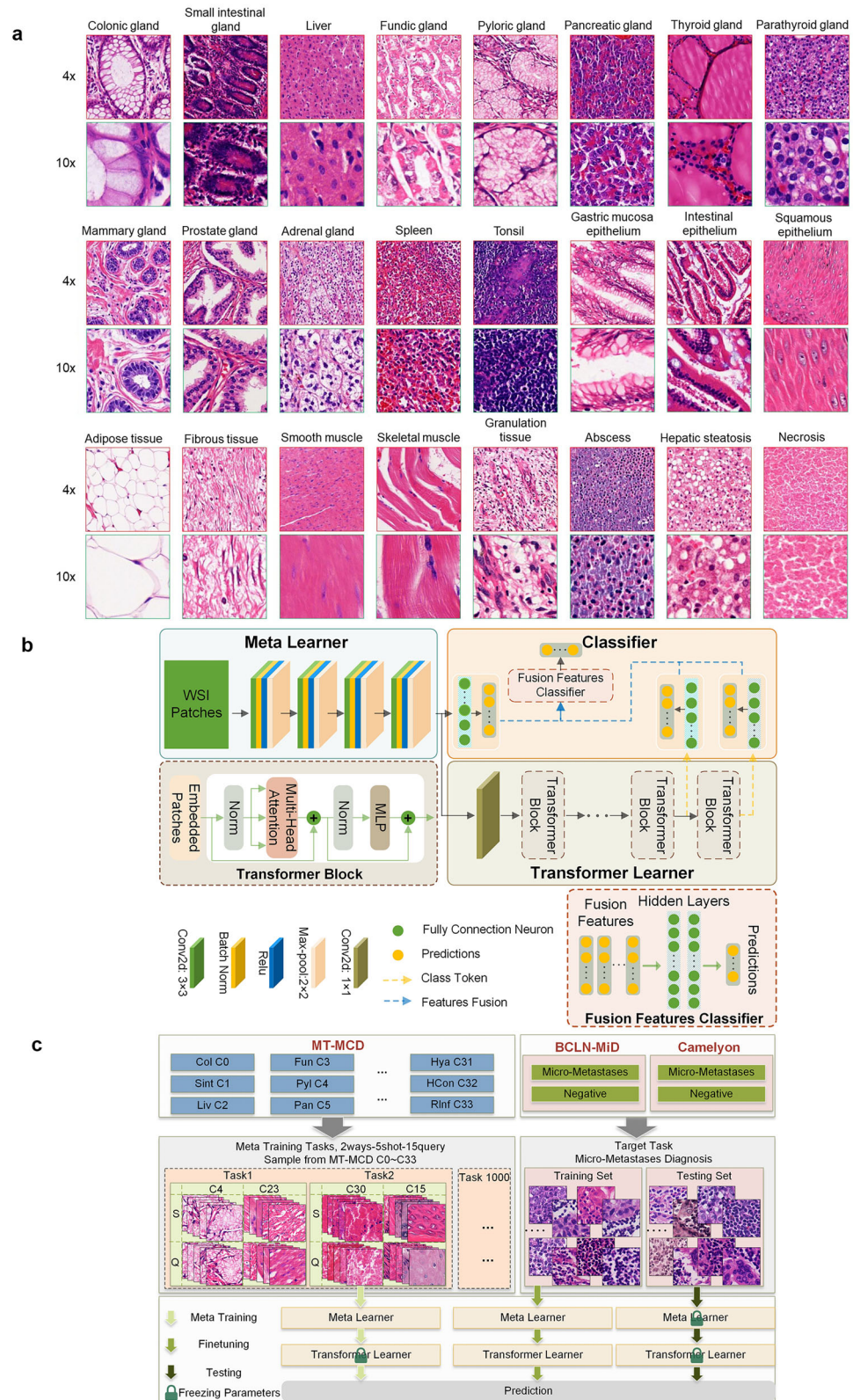
The Meta-Training Multi-Category Dataset (MT-MCD) and Breast Cancer Lymph Node Micro-Metastasis Dataset (BCLN-MiD) were created by ourselves. Each slide was collected from one related patient who was diagnosed at department of pathology. We constructed the MT-MCD in Sun Yat-sen Memorial Hospital. This dataset utilized about 16 different glandular organs in both normal and pathological states, including the liver, stomach, colon, thyroid, breast, pancreas, etc. Each state was represented by 20 collected slides. After scanned into WSIs, they were divided into small patches (256 × 256 pixels) (Fig. 7a). These patches were labeled based on different tissues such as mammary gland, small intestinal gland, colonic gland, thyroid gland, pyloric gland, adipose tissue, smooth muscle, etc. In total, the MT-MCD comprises 34 categories (Supplementary Table 10). The BCLN-MiD collected a total of 275 lymph nodes in paraffin section (FFPE) and 166 lymph nodes in frozen section (FS). Lymph nodes with macro-metastasis were not included in BCLN-MiD. As shown in Supplementary Tables 11–12, the slides were obtained from four hospitals. All lymph nodes were divided into the training set and the test set according to 8:2. Within the training set, 20% of the lymph nodes were sampled to form a validation set for hyperparameter tuning. The test set remained strictly independent of the training process and included data from all centers to prevent data leakage.

The slides were digitized into WSIs at 4× and 10× magnification. Ground truth class labels were extracted from clinical pathology reports, with diagnostic determinations provided by two experienced pathologists in the department. A WSI was cropped into several fields of view in 3072 × 2048 pixels according to the real picture size taken by the OPLENIC microscope digital camera. All patches were resized to 256 × 256 using nearest-neighbor interpolation, with random horizontal flipping (probability = 0.5), random rotation (probability = 0.5) random modifications to brightness, contrast, saturation, and hue as data augmentation. Patches were scaled to [0,1] range and channel-wise normalized (RGB channels normalized independently). All comparative methods adopted the same data pre-processing strategy.

Schematic of Meta-task Construction

Meta-training tasks were constructed using our self-built 34-class WSI dataset (MT-MCD). Each meta-task adopts a 2-way-5-shot-5-query

Fig. 7 | MetaTrans with the designed Meta-training strategy based on multi- category data (MT-MCD). **a** Representative patches in the Meta-Training Multi-Category Dataset (MT-MCD). See Supplementary Table 10 for all MT-MCD information. **b** Network Architecture of MetaTrans. MetaTrans comprises three main components: The Meta Learner for extracting shallow (low-dimensional) features, Transformer Learner for extracting deeper (high-dimensional) features, and a Classifier for classifying the fused features. **c** The scheme of Meta-Tasks generation and Meta-Learning Training Strategy in Our Method. Meta-training tasks were constructed using our self-built 34-class WSI dataset (MT-MCD). Each meta-task adopts a 2-way-5-shot-5-query configuration. The Transformer Learner does not participate in meta-training. Finetuning Stage based on the potential model obtained from meta-training, the model is quickly adjusted using the target task (small metastases) training dataset. Test Stage was the performance evaluation of MetaTrans. The module with “lock” in the dotted box means that parameters of this module keep frozen at this stage.



configuration. For each meta-task τ , two classes (2-way) are randomly selected from the 34 tissue categories to adapt the target task scenario (i.e., distinguishing micro-metastasis from negative cases). Each class includes 5 support-set samples (5-shot, D_{τ}^{train}) and 5 query-set samples (5-query, D_{τ}^{test}) to simulate the micro-metastasis diagnosis under data limitation. Finally, a total of 500 meta-tasks were generated, covering all

combinatorial patterns of the 34 tissue categories. This ensures diverse meta-knowledge extraction and robust adaptability to the target task. The target task (denoted as τ_{target} or τ_{unseen}) is derived from the breast cancer lymph node micro-metastasis dataset (BCLN-MiD), which comprises three subsets: training set D_{target}^{train} , validation set D_{target}^{val} , and test set D_{target}^{test} .

Architecture of MetaTrans designed based on Meta-learning and transformer

Existing meta-learning algorithms often utilize models with simple structures and fewer parameters, such as four-layer convolutional networks^{51–53}. However, these models are inadequate for medical images, which are more complex than natural images, with lesions that are typically small, varied in shape and color, and have blurred boundaries. While transformer models demonstrate superior performance, they are less effective in medical tasks where data volumes are typically smaller⁵⁴. For example, medical images like frozen sections often involve smaller datasets but higher image complexity, making the diagnostic task particularly challenging.

To address these challenges, we developed a new network, MetaTrans, specifically for diagnosing small metastasis, where more complex medical images and smaller dataset (Fig. 7b). MetaTrans consists of three main components: The Meta Learner, Transformer Learner, and a Classifier. When patches of WSI are input into MetaTrans, the Meta Learner first extracts shallow features. These shallow (low-dimensional) features are then passed to the Transformer Learner for deeper (high-dimensional) feature extraction. Finally, the Classifier fuses shallow and deep features to produce more accurate diagnostic results.

Meta Learner: A shallow feature extractor, consists of four 3×3 convolutional layers, along with activation and normalization layers, specifics can be referred to Supplementary Table 13. CNNs are widely recognized for their superior feature extraction capabilities, particularly in image processing. Meta-learning algorithms typically favor simple yet efficient model architectures, with the four-layer convolutional network being a common choice. Consequently, we adopted this straightforward and effective structure as the Meta Learner for extracting shallow feature, to swiftly and accurately extract shallow features from WSI data during meta-training. The Meta Learner, combined with Transformer Learner, is designed to align with the conventional strengths of meta-learning algorithms while addressing the specific demands of complex medical image processing, thereby providing a robust foundation for our model.

Transformer Learner: The deep feature extractor in MetaTrans is composed of six Transformer blocks, each incorporating multi-head self-attention mechanisms, feed-forward neural networks, and residual connections. For the current task's complexity, the deep feature extractor is configured with an embedding dimension of 384 and six attention heads. These six Transformer blocks enable the model to perform multiple layers of deep feature extraction and information interaction while processing input sequences. This design enhances the model's ability to capture the complex structure and semantic information of the input data, leading to more expressive representations and improved adaptability across varied data distributions. In summary, the deep feature extractor effectively performs complex feature extraction and representation learning through multi-layer Transformer blocks, providing robust feature representation for small metastasis diagnosis in breast WSI.

Fusion Features Classifier: The purpose of the fusion features classifier is to effectively integrate features extracted from both shallow and deep layers and utilize the fused features for final detection. To achieve this, we implemented three design strategies: feature selection, feature fusion, and classification. In feature selection part, the feature fusion module receives inputs from both shallow and deep feature extractors. Specifically, we selected the final layer output of the shallow feature extractor, and the outputs from each of the last three Transformer blocks of the deep feature extractor for feature fusion. In feature fusion part, the feature fusion module first applies a linear transformation to unify the dimensions of features from different scales, and then concatenates them to form the fused features. This step facilitates the effective integration of feature information from different depths, enhancing the network's capacity to capture complex features. In classification part, the fused features are subsequently fed into a neural network with two hidden layers for further processing. These hidden layers allow the model to learn higher-order feature representations, ultimately leading to the final detection output.

Training strategy of MetaTrans based on Meta-learning

Meta-learning focuses on developing strategies that allow models to quickly adapt to new tasks by leveraging prior experiences, much like human intelligence⁵⁵. This study, leveraging the meta-tasks constructed from the WSI 34-category dataset MT-MCD, applies a meta-learning algorithm to acquire prior knowledge from these meta-tasks, facilitating rapid adaptation to the target task of small metastasis diagnosis. The meta-learning algorithm enables reliable performance in small metastasis diagnosis tasks by effectively adapting to limited training data and capturing task-relevant features.

An approach is model-agnostic meta-learning, which offers wide applicability and flexibility while enhancing data efficiency and enabling adaptation across diverse domains^{51,53,56}. Three typical meta-learning algorithms (MAML⁵¹, iMAML^{52,56}, and Reptile⁵³) are integrated in our approach (Algorithm 1). Meta-learning consists of two phases: meta-train and meta-test. A meta-training example is a task τ sampled from some distribution of tasks $p(\tau)$. τ is called episode including a training split $\tau^{(tr)}$ (support set $\mathcal{D}_{\tau}^{train}$) to optimize the base-learner and a test split $\tau^{(te)}$ (query set $\mathcal{D}_{\tau}^{test}$) to optimize the meta-learner. In particular, meta-training aims to learn from a number of episodes τ sampled from $p(\tau)$. An unseen task τ_{unseen} in meta test will start from that experience of the meta-learner and adapt the base-learner. The final evaluation is done by testing a set of unseen datapoints $\tau_{unseen}^{te}(\mathcal{D}_{target}^{test})$.

Algorithm 1: Meta-Learning Algorithms

Input: $p(\tau)$: Distribution over tasks
 θ : Meta-learner's initialization meta parameters

Output: θ' : Updated meta-learner's weights

```

1: while not done do
2: # meta training
3: Sample batch of tasks  $\{\tau_i\}_{i=1}^B \sim p(\tau)$ 
4: for all  $\tau_i$  do
5: Update  $\theta$  for specific meta training tasks  $\tau_i$ :
    $\phi_{\tau_i} \leftarrow \text{Optimize}(L_{\tau_i}(\theta, \mathcal{D}_{\tau_i}^{train}))$ 
6: Update meta-learner:
    $\theta' \leftarrow \text{Optimize}(L_{\tau}(\phi_{\tau}, \mathcal{D}_{\tau_i}^{test}))$ 
7: end for
8: Evaluate on all tasks:
    $\theta' \leftarrow \text{Optimize}(\sum_{i=1}^B L_{\tau_i}(\phi_{\tau_i}, \mathcal{D}_{\tau_i}^{test}))$ 
9: # Meta test
10: for each  $\tau_{unseen}$  do
11: for k iterations do
12: Update meta parameters  $\theta'$  for  $\tau_{unseen}$ :
    $\theta_{\tau_{unseen}} \leftarrow \text{Optimize}(\theta', \mathcal{D}_{\tau_{unseen}}^{train})$ 
13: end for
14: Evaluate  $Loss(\theta_{\tau_{unseen}}, \mathcal{D}_{\tau_{unseen}}^{test})$ 
15: end for
16: end while
17: return  $\theta'$ 

```

Meta-Training phase aims to learn a meta-learner from multiple episodes. In each episode, meta-training has a two-stage optimization. Stage-1 is called base-learning, where the cross-entropy loss is used to optimize the parameters of the base-learner. Stage-2 (Meta-learning) contains feed-forward test on episode test datapoints. In base-learning stage, given an episode $\tau \in p(\tau)$, we optimize the initialization parameters θ of Meta Learner through cross-entropy loss on support set $\mathcal{D}_{\tau}^{train}$:

$$\phi_{\tau} = \theta - \alpha \nabla_{\theta} \mathcal{L}_{ce}(\theta, \mathcal{D}_{\tau}^{train}) \quad (1)$$

where α denotes the learning rate, $\mathcal{L}_{ce}$ represents the cross-entropy loss function. The objective of base learning is to minimize the training loss on current batch meta tasks, ensuring optimal performance of MetaTrans on the current batch tasks. In Meta-learning stage, we calculate test loss on

query set $\mathcal{D}_{\tau}^{\text{test}}$ and update parameters θ :

$$\theta \leftarrow \theta - \beta \nabla_{\phi} \mathcal{L}_{\text{ce}}(\phi_{\tau}, \mathcal{D}_{\tau}^{\text{test}}) \quad (2)$$

where β denotes the meta-learning rate. The meta-learning objective aims to minimize test losses across all tasks, thereby extracting shared knowledge from meta tasks distribution $p(\tau)$. Through training on all meta-tasks, the final optimization can be formulated as follows:

$$\theta' = \arg \min_{\phi} \sum_{\tau \in p(\tau)} \mathcal{L}_{\text{ce}}(\phi_{\tau}, \mathcal{D}_{\tau}^{\text{test}}) \quad (3)$$

The number of θ updates correspond to the number of meta-task batches, with each update performed on a newly sampled set of meta-tasks to ensure the model can extract universal knowledge from diverse tasks.

The meta-training phase aims to generate a set of potential initialization parameters that facilitate efficient adaptation to new target tasks, followed by fine-tuning using limited training data from the target task. In model initialization stage, the Meta Learner is initialized with meta parameters θ' , obtained through meta-training to enable efficient adaptation to downstream tasks. The Transformer Learner is initialized with ImageNet pre-trained weights. Initialized parameters of MetaTrans denoted as $\theta_{\text{potential}}$. In finetuning stage, $\theta_{\text{potential}}$ is further optimized using cross-entropy loss on the limited training data of the target task:

$$\theta_{\text{final}} = \theta_{\text{potential}} - \gamma \nabla_{\theta} \mathcal{L}_{\text{CE}}(\theta_{\text{potential}}, \mathcal{D}_{\text{target}}^{\text{train}}) \quad (4)$$

where γ denotes the fine-tuning learning rate, $\mathcal{D}_{\text{target}}^{\text{train}}$ represents the limited labeled training set of the target task. In model evaluation stage, the fine-tuned model achieving the highest validation accuracy ($\mathcal{D}_{\text{target}}^{\text{val}}$) was retained for final testing ($\mathcal{D}_{\text{target}}^{\text{test}}$):

In this study, we construct test tasks (unseen target tasks) using BCLN-MiD and Camelyon. The meta-learner, trained on 34-class synthetic data, is then quickly adapted to these unseen target tasks. Given the limitations of meta-learning algorithms in optimizing complex models, we designed a separable network architecture named MetaTrans. During the meta-training phase, the meta multiple tasks constructed from the 34-class data MT-MCD are only used to optimize a simple network consisting of a four-layer convolutional Meta-learner and classifier, while the deep, complex network based on the Transformer Learner is excluded from parameter optimization (Fig. 7c). A significant advantage of the Transformer is its pre-trained parameters derived from the large-scale ImageNet dataset, which offer superior performance compared to traditional CNNs like ResNet⁷ and VGG⁶. Consequently, MetaTrans' Transformer Learner directly utilize these pre-trained ImageNet parameters during meta-training. This stage establishes a reliable and potential model for our target task, small metastasis diagnosis.

During the fine-tuning stage, we quickly adjust the potential model using the target task's training data, since this potential model is generated by MetaTrans through meta-training, its strong task-adaptability and the reliable pre-trained parameters from ImageNet enable MetaTrans to produce a dependable end-to-end model for small metastasis diagnosis, even when the target task's training samples are limited.

Experimental settings

The hyperparameter configurations in this study were primarily determined based on domain literature references, pilot exploratory experiments, and validation set performance evaluation, with considerations for task-specific characteristics (e.g., small-sample data scale) and hardware constraints (GeForce RTX A6000, 48 G). The key settings are listed in Supplementary Table 14. The Meta Inner/Outer LR/Loop and optimizer configurations were adapted from classical meta-learning implementations^{51–53,56,57,58}, with adjustments for convergence stability based on pilot experiments. iMetaTrans/mMetaTrans/rMetaTrans used 25/5/5 inner-loop iterations, respectively, to balance computational efficiency and parameter update

sufficiency. The 2-way setting aligns with the target task (micro-metastasis/negative). The 5-shot 5-query and 500 meta tasks settings maintain multi-task diversity within GPU memory constraints. Meta-training employed SGD(*momentum* = 0.9)⁵⁹ Adam ($\beta_1 = 0.9$, $\beta_2 = 0.999$)⁶⁰ for parameter optimization. Finetuning/Deep Learning optimizer (Adam, $\beta_1 = 0.9$, $\beta_2 = 0.999$), initial learning rate (0.001), and batch size (16) followed standard configurations in medical image analysis^{60,61}, with validation set monitoring to prevent overfitting.

Evaluation metrics

Accuracy, Precision, F1-Score, Recall and the area under the receiver operating characteristic curve (AUC) were used to fully and fairly evaluate the performance, where TN, TP, FN, and FP represent true negative, true positive, false negative, and false positive, respectively. These metrics are defined as:

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ \text{Precision} &= \frac{TP}{TP+FP} \\ \text{F1score} &= \frac{2TP}{2TP+FP+FN} \end{aligned} \quad (5)$$

Data availability

The Camelyon datasets used in this study is publicly available (<https://camelyon16.grand-challenge.org/> and <https://camelyon17.grand-challenge.org/>). However, the MT-MCD and BCLN-MiD datasets, which were constructed by the authors in collaboration with medical institutions, will not be shared due to hospital data privacy policies. Extended data derived from out analyses are available from the corresponding author upon reasonable request.

Code availability

Source code available at <https://github.com/wangjingtao93/metatrans>.

Received: 20 September 2024; Accepted: 23 June 2025;

Published online: 09 July 2025

References

- Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **71**, 209–249 (2021).
- Huo, Y. et al. Thematic trends and knowledge structure map of sentinel lymph node biopsy for breast cancer: a bibliometric analysis from 2010 to 2019. *Transl. Cancer Res.* **11**, 3092–3107, <https://doi.org/10.21037/tcr-21-2841> (2022).
- de Jonge, C. et al. Breast conserving surgery with intraoperative electron beam radiation therapy for low-risk breast cancer: Five-year follow-up of 306 patients. *Int. J. Cancer* **155**, 1237–1247, <https://doi.org/10.1002/ijc.35033> (2024).
- de Boer, M. et al. Micrometastases or Isolated Tumor Cells and the Outcome of Breast Cancer. *N. Engl. J. Med.* **361**, 653–663, <https://doi.org/10.1056/NEJMoa0904832> (2009).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Commun.ACM.* **60**, 84–90 (2017).
- Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition.* 770–778 (2016).
- Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition.* 580–587 (2014).

9. Girshick, R. Fast r-cnn. *In Proc. IEEE International Conference on Computer Vision*. 1440–1448 (2015).
10. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 3431–3440 (2015).
11. Wang, X. et al. Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. *Nat. Commun.* **12**, 1637. <https://doi.org/10.1038/s41467-021-21674-7> (2021).
12. Yu, J. G. et al. Prototypical multiple instance learning for predicting lymph node metastasis of breast cancer from whole-slide pathological images. *Med. Image Anal.* **85**, 102748. <https://doi.org/10.1016/j.media.2023.102748> (2023).
13. Lu, M. Y. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570. <https://doi.org/10.1038/s41551-020-00682-w> (2021).
14. Hou, L. et al. Patch-based convolutional neural network for whole slide tissue image classification. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*. 2424–2433 (2016).
15. Li, Y. & Ping, W. Cancer metastasis detection with neural conditional random field. *arXiv preprint arXiv:1806.07064* (2018).
16. Zhu, W., Xiang, X., Tran, T. D., Hager, G. D. & Xie, X. Adversarial deep structured nets for mass segmentation from mammograms. *In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI)*. 847–850 (2018).
17. Elmore, J. G. et al. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama-J. Am. Med. Assoc.* **313**, 1122–1122. <https://doi.org/10.1001/jama.2015.1405> (2015).
18. Brem, R. F., Rapelyea, J. A., Zisman, G., Hoffmeister, J. W. & DeSimio, M. P. Evaluation of breast cancer with a computer-aided detection system by mammographic appearance and histopathology. *Cancer* **104**, 931–935. <https://doi.org/10.1002/cncr.21255> (2005).
19. Lin, H. et al. Fast scanet: fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection. *IEEE Trans. Med. Imaging* **38**, 1948–1958 (2019).
20. Bejnordi, B. E. et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama-J. Am. Med. Assoc.* **318**, 2199–2210. <https://doi.org/10.1001/jama.2017.14585> (2017).
21. Bándi, P. et al. From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 Challenge. *IEEE Trans. Med. Imaging* **38**, 550–560. <https://doi.org/10.1109/Tmi.2018.2867350> (2019).
22. Giuliano, A. E. et al. Breast cancer—major changes in the american joint committee on cancer eighth edition cancer staging manual. *CA Cancer J. Clin.* **67**, 290–303. <https://doi.org/10.3322/caac.21393> (2017).
23. Rusu, A. A. et al. Meta-learning with latent embedding optimization. *In International Conference on Learning Representations (ICLR)*. 1–17 (2019).
24. Liu, Y. et al. Learning to propagate labels: Transductive propagation network for few-shot learning. *In International Conference on Learning Representations (ICLR)*. 1–14 (2019).
25. Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. *In International Conference on Learning Representations (ICLR)*. 1–21 (2021).
26. Vaswani, A. Attention is all you need. *In Adv. Neural Inform. Process. Syst.* 6000–6010 (2017).
27. Gulati, A. et al. Conformer: Convolution-augmented transformer for speech recognition. *arXiv preprint arXiv:2005.08100* (2020).
28. Guo, S., Sheng, S., Lai, Z. & Chen, S. Trans-U: Transformer enhanced U-Net for medical image segmentation. *In International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*. 628–631 (2022).
29. Wang, S., Yang, D. M., Rong, R., Zhan, X. & Xiao, G. Pathology image analysis using segmentation deep learning algorithms. *Am. J. Pathol.* **189**, 1686–1698. <https://doi.org/10.1016/j.ajpath.2019.05.007> (2019).
30. Siegel, R. L., Wagle, N. S., Cercek, A., Smith, R. A. & Jemal, A. Colorectal cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 233–254. <https://doi.org/10.3322/caac.21772> (2023).
31. Yamamoto, H. Micrometastasis in lymph nodes of colorectal cancer. *Ann. Gastroenterol. Surg.* **6**, 466–473. <https://doi.org/10.1002/ags3.12576> (2022).
32. Li, M. M., Dal Maso, L. & Vaccarella, S. Global trends in thyroid cancer incidence and the impact of overdiagnosis. *Lancet Diabetes Endocrinol.* **8**, 468–470 (2020).
33. Chao, R. et al. Semi-supervised pairwise transfer learning based on multi-source domain adaptation: a case study on EEG-based emotion recognition. *Knowl. Based Syst.* **305**, 112669. <https://doi.org/10.1016/j.knosys.2024.112669> (2024).
34. Xuetong, W., Jonathan, H. M., Uwe, A. & Jingge, Z. Online transfer learning: negative transfer and effect of prior knowledge. *In IEEE International Symposium on Information Theory (ISIT)*, 1540–1545. <https://doi.org/10.1109/isit45174.2021.9517974> (2021).
35. Bayramoglu, N., Kannala, J. & Heikkilä, J. Deep learning for magnification independent breast cancer histopathology image classification. *In 2016 23rd International conference on pattern recognition (ICPR)*. 2440–2445 (2016).
36. Spanhol, F. A., Oliveira, L. S., Petitjean, C. & Heutte, L. Breast cancer histopathological image classification using convolutional neural networks. *In 2016 International Joint Conference on Neural Networks (IJCNN)*. 2560–2567 (2016).
37. Aksebzeci, B. H. & Kayaalti, Ö. Computer-aided classification of breast cancer histopathological images. *In 2017 Medical Technologies National Congress (TIPTEKNO)*. 1–4 (2017).
38. Luo, W., Wang, L. & Sun, J. Feature selection for cancer classification based on support vector machine. *In 2009 WRI Global Congress on Intelligent Systems*. **4**, 422–426 (2009).
39. Zemmal, N., Azizi, N., Dey, N. & Sellami, M. Adaptive semi supervised support vector machine semi supervised learning with features cooperation for breast cancer classification. *J. Med. Imaging Health Inform.* **6**, 53–62 (2016).
40. Sun, W., Tseng, T.-L. B., Zhang, J. & Qian, W. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. *Comput. Med. Imaging Graph.* **57**, 4–9 (2017).
41. Li, J.-B., Yu, Y., Yang, Z.-M. & Tang, L.-L. Breast tissue image classification based on semi-supervised locality discriminant projection with kernels. *J. Med. Syst.* **36**, 2779–2786 (2012).
42. Huang, Q., Zhang, F. & Li, X. Few-shot decision tree for diagnosis of ultrasound breast tumor using BI-RADS features. *Multimed. Tools Appl.* **77**, 29905–29918 (2018).
43. Snell, J., Swersky, K. & Zemel, R. Prototypical networks for few-shot learning. *In Adv. Neural Inform. Process. Syst.* 4080–4090 (2017).
44. Sung, F., Yang, Y., Li, Z., Tao, X. & Hospedales, T. M. Learning to compare: relation network for few-shot learning. *In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1199–1208 (2018).
45. Oreshkin, B. N., Lacoste, A. & Rodriguez, P. TADAM: Task dependent adaptive metric for improved few-shot learning. *In Adv. Neural Inform. Process. Syst.* **31**, 719–729 (2018).
46. Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K. & Wierstra, D. Matching networks for one shot learning. *In Adv. Neural Inform. Process. Syst.* **29**, 3637–3645 (2016).
47. Larochelle, S. Optimization as a model for few-shot learning. *In International Conference on Learning Representations (ICLR)*. 1–11 (2017).
48. Lee, Y. & Choi, S. Gradient-based meta-learning tree with learned layerwise metric and subspace. *In International Conference on Machine Learning*. 2927–2936 (2018).

49. Lee, K., Maji, S., Ravichandran, A. & Soatto, S. Meta-learning with differentiable convex optimization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10657–10665 (2019).
50. Chen, Y. et al. dMIL-Transformer: multiple instance learning via integrating morphological and spatial information for lymph node metastasis classification. *IEEE J. Biomed. Health Inf.* **27**, 4433–4443, <https://doi.org/10.1109/JBHI.2023.3285275> (2023).
51. Finn, C., Abbeel, P. & Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*. 1126–1135 (2017).
52. Khadka, R. et al. Meta-learning with implicit gradients in a few-shot setting for medical image segmentation. *Comput. Biol. Med.* **143**, 105227 (2022).
53. Nichol, A., Achiam, J. & Schulman, J. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999v3* (2018).
54. Hu, W. et al. A state-of-the-art survey of artificial neural networks for Whole-slide Image analysis: from popular Convolutional Neural Networks to potential visual transformers. *Comput. Biol. Med.* **161**, 107034. <https://doi.org/10.1016/j.compbiomed.2023.107034> (2023).
55. Schmidhuber, J. Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook (diploma thesis). www.bibsonomy.org/bibtex/2a96f7c3d42103ab94b13bade5d869f0/brazovayeye (1987).
56. Rajeswaran, A., Finn, C., Kakade, S. M. & Levine, S. Meta-learning with implicit gradients. In *Adv. Neural Inform. Process. Syst.* **32**, 113–124 (2019).
57. Nichol, A. & Schulman, J. Reptile: a scalable metalearning algorithm. *arXiv preprint arXiv:1803.02999v1* (2018).
58. Lu, L., Cui, X., Tan, Z. & Wu, Y. MedOptNet: Meta-learning framework for few-shot medical image classification. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **21**, 725–736 (2023).
59. Keskar, N. S. & Socher, R. Improving generalization performance by switching from adam to sgd. *arXiv preprint arXiv:1712.07628* (2017).
60. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
61. Malik, H., Naeem, A., Sadeghi-Niaraki, A., Naqvi, R. A. & Lee, S.-W. Multi-classification deep learning models for detection of ulcerative colitis, polyps, and dyed-lifted polyps using wireless capsule endoscopy images. *Complex Intell. Syst.* **10**, 2477–2497, <https://doi.org/10.1007/s40747-023-01271-5> (2024).

Acknowledgements

This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China (No. 82203273, 82303584), the Natural Science Foundation of Jiangsu Province, China (No. BK20220495), the Major Project of the Jiangxi Provincial Natural Science Foundation (No. 20242BAB27004), the Key Support Project of the National Natural Science Foundation of China Regional Innovation and Development Joint Fund

(U24A201290), the Key Program of the National Natural Science Foundation of China (No. 82130080), the National Key R&D Program of China (No. 2018YFA0701700), the National Natural Science Foundation of China (No. U20A20170) and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 21KJB320018).

Author contributions

Study concept and design: J.H., J.W., X.C. and J.L.; acquisition of patient specimens: H.N., S.X., Y.R., and L.B.; Label of data: J.H., H.N., S.X., Y.R., L.B., J.S., C.S., Y.X. and X.H.; analysis and interpretation of data: J.H., J.W., and J.S.; drafting of the manuscript: J.H. and J.W.; critical revision of the manuscript: X.C. and J.L.; supervised the research: X.C. and J.L.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-025-01833-6>.

Correspondence and requests for materials should be addressed to Xinjian Chen or Jianming Li.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2025