npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-026-01441-x

Hybrid supervised deep learning for lung adenocarcinoma diagnosis to optimize surgical strategies

Jianwei Zhao[1,4], Junyan Zhang[2,4], Yihao Wang[3], Xinwen Zhong[3] ✉ & Xiaojiao Guan[1] ✉

Intraoperative frozen section pathological diagnosis of lung adenocarcinoma serves as the gold standard for determining the extent of surgical resection. Due to the dual constraints of intraoperative time limitations and the challenge of manually assessing tumor invasion volume in three-dimensional space, current manual diagnostic approaches and weakly supervised deep learning methods have demonstrated suboptimal diagnostic accuracy. To enhance the accuracy of intraoperative pathological diagnosis of lung adenocarcinoma and provide more precise recommendations for intraoperative resection extent to thoracic surgeons, we have developed a Hybrid-Supervised Framework for Lung Adenocarcinoma (HSFLA). This framework accomplishes the following processes: hybrid-supervised diagnosis of 2D whole slide images (WSIs), automatic annotation of tumor invasive regions, registration of consecutive WSIs, and 3D reconstruction and volume calculation of tumor invasive areas. We evaluated HSFLA on a dataset comprising 1161 WSIs from two centers and three subtypes, achieving an accuracy of 95.6%, which represents a significant improvement over manual review (84.7%) and weakly supervised learning (66.2% ± 3.0%). The consistency of its invasive area automatic annotations with manual pixel annotations was 86.6%. Furthermore, HSFLA's concordance with spatial transcriptomics samples demonstrated its interpretability at the genetic level. Utilizing HSFLA's automatic annotation functionality provided pathologists with a safe and effective diagnostic aid, improving their manual diagnostic accuracy by 22.9% ($n = 3$). We also applied HSFLA in real-world clinical settings for prospective study. Compared to manual diagnosis alone, the "human-machine interaction" diagnostic mode provided more appropriate surgical recommendations for 5 patients (among 70). Overall, HSFLA demonstrates the potential clinical utility of artificial intelligence in supporting intraoperative pathological assessment and surgical decision-making, and may serve as a paradigm for future innovations in AI-assisted clinical workflows.

Lung cancer currently poses the most significant threat to human health among malignant neoplasms[1]. According to the "Global Cancer Statistics 2024 report"[2], lung cancer ranks first in global cancer incidence. With the evolving disease spectrum, lung adenocarcinoma has progressively become the histological type with the highest incidence rate among lung cancers. Lung adenocarcinoma cases typically follow a histopathological evolution process from Atypical Adenomatous Hyperplasia (AAH) to Adenocarcinoma In-Situ (AIS), Minimally Invasive Adenocarcinoma (MIA), and Invasive Adenocarcinoma (IA)[3]. Tumor types are classified as: AIS, where histological sections show tumor cell proliferation confined to the alveolar epithelium without invasion, MIA, with scattered invasive lesion distributed within ≤5 mm in the tissue, and IA, demonstrating extensive invasive lesion[4]. Different surgical strategies are employed for various types. For AIS and MIA, wedge resection, a form of partial

[1]Department of Pathology, Shengjing Hospital, China Medical University, Shenyang, China. [2]Department of Surgical Oncology and General Surgery, First Affiliated Hospital, China Medical University, Shenyang, China. [3]Department of Thoracic Surgery, First Affiliated Hospital, China Medical University, Shenyang, China. [4]These authors contributed equally: Jianwei Zhao, Junyan Zhang.
✉e-mail: xwzhong@cmu.edu.cn; guanxiaojiao@126.com

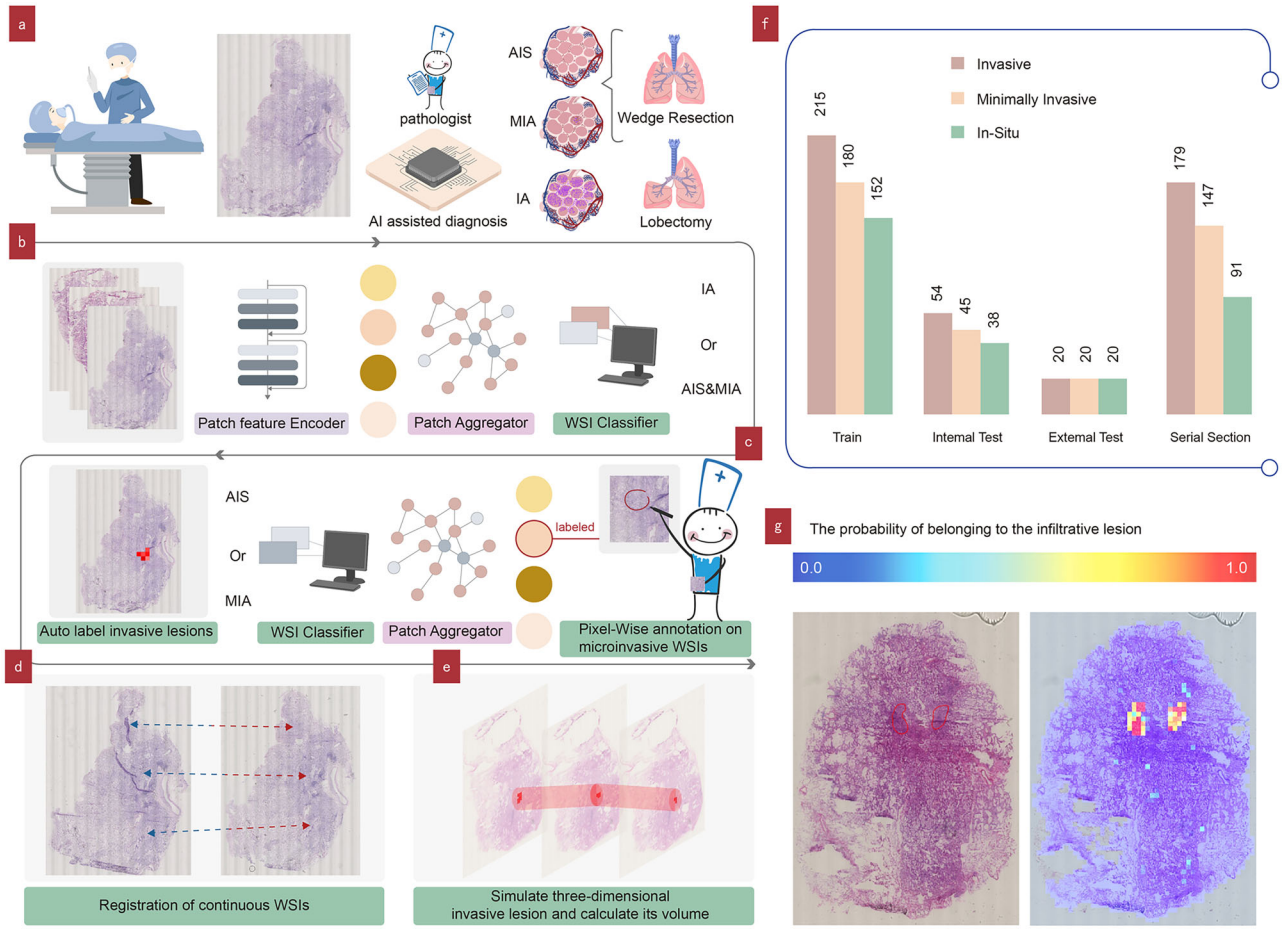

Fig. 1 | Overview of "hybrid-supervised deep learning framework for lung adenocarcinoma". a Intraoperative frozen histopathological diagnosis of lung adenocarcinoma determines different surgical methods. **b** Weakly-supervised step. **c** Supervision of manual annotation. **d** Registration of continuous WSIs. **e** Reconstruction of lesions in three-dimensional space. **f** Description of the dataset. **g** Display of infiltration lesion detection.

resection, is primarily utilized, with a minority of MIA cases requiring additional hilar and mediastinal lymph node dissection. For IA, partial resection involving segmentectomy or lobectomy is necessary, along with dissection of segmental, lobar, and mediastinal lymph nodes, as illustrated in Fig. 1a. Pathologists can confirm histological types through tissue sections, thereby guiding surgical decision-making[5,6]. Consequently, accurate intraoperative frozen section diagnosis and subtyping are crucial for thoracic surgical planning.

However, intraoperative pathological diagnosis during lung adenocarcinoma resection has long been challenged by issues of consistency and reproducibility, constrained by three factors: 1) Intraoperative diagnosis requires pathologists to complete microscopic evaluation within minutes to reduce intraoperative waiting time and minimize the probability of complications[7–9]; 2) Pathologists can only observe two-dimensional section images, making it difficult to assess the true size and spatial distribution of invasive lesions in three-dimensional tissue[10]; 3) MIA lesion sections contain only minute invasive lesion, which are easily overlooked in manual intraoperative diagnosis[11]. These limitations affect the accuracy of manual diagnosis of lung adenocarcinoma, especially for MIA, potentially leading to erroneous thoracic surgical planning decisions in severe cases.

In recent years, the emergence of weakly supervised learning methods, particularly Multiple Instance Learning (MIL)[12–14], has significantly reduced annotation costs through slide-level labels, demonstrating notable advantages in whole slide image (WSI) classification tasks[15]. However, existing methods face fundamental limitations in meeting the requirements of intraoperative lung adenocarcinoma diagnosis: Firstly, when distinguishing between morphologically subtle AIS and MIA, accuracy drops precipitously, with evaluations falling below 65%, making it challenging to differentiate these based on minor differences in single invasive lesion in two-dimensional sections. Secondly, the weakly supervised paradigm lacks precise feature localization capabilities, failing to meet the clinical need for invasive focus measurement[16]. Most critically, 2D image analysis cannot reconstruct true 3D invasion patterns, while current diagnostic criteria explicitly require assessment of the maximum continuous invasion area in 3D space[17]. This deficiency introduces systematic bias when distinguishing between MIA (invasive area length ≤5 mm, while tumor length ≤3 cm) and IA (invasive area length >5 mm)[11]. Imprecise intraoperative pathological diagnosis significantly impacts thoracic surgeons' intraoperative decision-making regarding surgical strategies.

To address the aforementioned challenges, this study proposes a Hybrid-Supervised Framework for Lung Adenocarcinoma (HSFLA). As shown in Fig. 1b–e, the designed HSFLA primarily integrates four collaborative modules: 1. MIL-based preliminary screening for invasive adenocarcinoma: Utilizing an MIL network to identify IA based on its widely distributed tumor characteristics. 2. Hybrid-supervised dual-channel network for further classification: Development of a hybrid-supervised dual-channel network combining 392 meticulously annotated MIA WSIs (all of the collected MIAs, verified by consensus of three senior pathologists) with slide-level labels to enhance subtype classification sensitivity through contrastive learning. 3. Deep feature matching-based slice registration algorithm: Employing the DeeperHistReg[18] architecture to automatically detect anatomical landmarks, enabling sub-pixel level 3D reconstruction. 4. 3D measurement module for invasive lesions: Applying morphological operations to automatically calculate the spatial volume of the maximum

invasive lesion, ensuring strict adherence to the International Association for the Study of Lung Cancer (IASLC) diagnostic criteria[19].

The hybrid-supervised paradigm further provides a universal solution for "limited annotation but complex diagnostic demands" in digital pathology, highlighting the potential of data-driven approaches to complement experience-based pathological diagnosis. Moreover, the preliminary application of HSFLA in real-world clinical environments suggests its feasibility and highlights the potential for broader integration of artificial intelligence technologies in pathological diagnosis.

Results

Results demonstrate that HSFLA exhibits excellent diagnostic performance and practicality. Specifically, HSFLA achieved an accuracy of 95.6% in subtype classification, significantly outperforming manual diagnosis (84.7%) and weakly supervised learning methods (66.2% ± 3.0%). Notably, it attained a 98.15% recall rate for IA identification while improving the classification accuracy of AIS vs MIA to 94.0%. In tests of automatic invasive lesion labeling, HSFLA showed 86.6% concordance with pixel-by-pixel manual annotations. The capability of HSFLA to mark invasive lesion was confirmed to be genetically interpretable in four spatial transcriptomics samples. Consequently, HSFLA's automatic annotation function provides pathologists with a safe and effective diagnostic aid (Fig. 1g). With its assistance, manual diagnostic accuracy improved by 22.9% ($n = 3$, 11.7%, 28.5%, 28.5%). We also conducted a prospective study in a real-world clinical setting. Compared to manual diagnosis alone, the "human-machine interaction" diagnostic mode leaded to more precise surgical resection ranges for 5 patients (among 70). This technology established quantifiable standards for intraoperative subtyping, which may help reduce the risk of insufficient or excessive resection during lung adenocarcinoma surgery. These results not only demonstrate the superior performance of HSFLA in lung adenocarcinoma classification tasks but also highlight its significant potential in reducing diagnostic time and improving patient treatment decisions. This provides robust support for the application of artificial intelligence in clinical pathological diagnosis.

Description of the dataset

In this study, we constructed a diverse pathological image dataset for lung adenocarcinoma. A NanoZoomer 2.0-RS digital slide scanner (Hamamatsu, Japan) was used to scan all glass slides into digital format. The scanner operates at a magnification of 40× (with an additional 10× eyepiece) while maintaining a resolution of 300 dpi, thereby producing images of sufficient high quality for the purposes of this study. From 2022 to 2024, we retrospectively collected 684 WSIs of lung adenocarcinoma from Shengjing Hospital of China Medical University. The dataset comprises three pathological subtypes: IA ($n = 269$), MIA ($n = 225$), and AIS ($n = 190$). The dataset was randomly split into a training set ($n = 547$, 80%) and a test set ($n = 137$, 20%), with the subtype distributions balanced across the two subsets. In the internal dataset, each WSI corresponded to a distinct patient, comprising a total of 684 patients. Therefore, no additional patient-level data splitting was required. To further evaluate the generalizability of the model, an independent external test set was constructed by collecting an additional 60 WSIs (20 images per subtype) from the First Hospital of China Medical University. No explicit stain normalization was applied; instead, data diversity was increased by incorporating staining protocols from two centers with substantial variability. This external test set was specifically designed to assess the model's performance on cross-institutional data. Additionally, from January 2025 to July 2025, we prospectively collected 417 WSIs from 70 lung adenocarcinoma patients (30 with IA, 25 with MIA, and 15 with AIS) at Shengjing Hospital of China Medical University. For each patient, at least five consecutive WSIs were acquired. These serial sections were organized into a three-dimensional pathological image test set to evaluate the model's performance in handling spatially correlated consecutive slices. All cases were consecutively enrolled without additional screening. The clinical baseline characteristics and demographic information of the cases are provided in Supplementary Information 1, 2.

Diagnostic errors in intraoperative frozen section analysis of lung adenocarcinoma are relatively common, primarily due to constraints on examination time and limitations in pathologists' expertise. These misdiagnoses may lead to alterations in surgical approach, potentially resulting in over- or under-resection during the operative procedure. Post-operatively, pathologists conduct a thorough review of the initial intraoperative diagnosis, incorporating additional paraffin-embedded tissue samples and allowing for ample examination time. Thus, during our data collection process, we implemented a comprehensive review and quality control protocol involving three experienced pathologists who re-examined all WSIs. The three pathologists provided a system of mutual oversight, significantly reducing the risk of secondary diagnostic errors. Under the double review, 102 WSIs of misdiagnosis were identified and corrected. To construct a hybrid supervised model, we manually annotated all MIA-type WSIs, delineating all invasive lesion. All WSIs underwent automated tissue detection to remove background regions. Tissue areas were identified using color-based thresholding in the HSV color space to exclude blank or non-tissue regions[20]. Subsequently, each WSI was partitioned into non-overlapping patches of 256×256 pixels at 40× magnification, while Resnet-50 was used as the encoder. During training, all tissue-containing patches were utilized. During inference, patch-level predictions were aggregated using the corresponding model-specific aggregation strategies to generate WSI-level predictions, while patch-level risk scores were retained to visualize invasive lesions.

Diagnosis performance on 2D WSIs

HSFLA was evaluated on the collected dataset (2D images) and demonstrated excellent performance. All algorithms were trained and evaluated on the datasets used in this study. In the first step, the framework achieved 98.71% accuracy in identifying IA. Subsequently, in distinguishing between IA, AIS and MIA, HSFLA achieved a classification ACC of 95.62% and an AUC of 98.71%, as shown in Table 1. Compared to other common weakly supervised learning methods (including MaxMIL[21], MeanMIL[21], MIL-RNN[21], ABMIL[13], CLAM[20], TransMIL[22], DS-MIL[15], IB-MIL[23], CI-MIL[24] and DTFD-MIL[25]), HSFLA exhibited performance improvement. The best-performing weakly supervised learning method, DTFD-MIL, only achieved an ACC of 70.80% and an AUC of 76.86%, indicating their limited ability to differentiate between MIA and IA. Confidence intervals for these metrics were calculated, and statistical evaluations were performed using DeLong's and McNemar's tests (Supplementary Information 7–9).

In addition, a fully supervised learning baseline was included for comparison. This baseline consisted of a ResNet-50 backbone and a fully connected (FC) classifier[26]. The classifier was trained using manual annotations provided by pathologists as ground-truth labels to classify whether each patch corresponded to invasive lesion. During inference, patch-level predictions were aggregated to estimate the extent of invasive lesions, and the AIS, MIA, or IA status was subsequently determined according to established pathological criteria (lesion area). Despite the availability of more precise supervision signals, the performance of the standalone ResNet-50 + FC remained inferior to that of HSFLA, achieving only 90.5% ACC and 92.2% AUC. Due to the occurrence of false positives, some AISs were incorrectly identified as containing invasive lesion, resulting in a recall of only 84.2%. These results underscore the importance of employing hybrid supervision signals, which enable the model to learn subtle features that are difficult to capture using weak supervision alone, while mitigating the impact of patch-level false positives on the final classification.

To assess the model's generalization capability, we conducted validation on an external test set. The results showed that in the task of distinguishing IA from other subtypes, HSFLA maintained a classification recall of 95.0%. In the three-classification task, it achieved a classification accuracy of 91.7% and an AUC of 94.2%. These results not only maintained a high level of performance but also consistently outperformed the compared weakly supervised learning methods, convincingly demonstrating the robustness of the proposed approach. Overall, this study has shown the advantages of the proposed HSFLA in addressing the complex problem of lung adenocarcinoma subtype classification. By effectively integrating limited fine-grained annotation

Table 1 | Diagnosis performance on 2D WSIs

Methods	Internal Test					External Test				
	ACC	AUC	Recall-IA	Recall-MIA	Recall-AIS	ACC	AUC	Recall-IA	Recall-MIA	Recall-AIS
MaxMIL	64.96%	63.80%	90.74%	46.67%	50.00%	56.67%	56.03%	75.00%	45.00%	50.00%
MeanMIL	62.77%	60.09%	83.33%	53.33%	44.74%	53.33%	57.94%	70.00%	40.00%	50.00%
MIL-RNN	67.88%	68.47%	92.59%	55.56%	47.37%	51.67%	55.72%	80.00%	55.00%	20.00%
ABMIL	61.31%	66.29%	94.44%	40.00%	39.47%	55.00%	57.67%	80.00%	60.00%	25.00%
CLAM	68.61%	71.07%	98.15%	48.89%	50.00%	61.67%	62.39%	95.00%	50.00%	40.00%
TransMIL	67.15%	74.43%	96.30%	44.44%	52.63%	63.33%	67.21%	85.00%	30.00%	75.00%
DS-MIL	63.50%	65.45%	92.59%	46.67%	42.11%	61.67%	53.75%	85.00%	0%	100%
IB-MIL	67.15%	68.62%	94.44%	42.22%	57.89%	55.00%	60.94%	90.00%	45.00%	30.00%
CI-MIL	67.88%	70.06%	96.30%	37.78%	63.16%	70.00%	73.81%	90.00%	65.00%	55.00%
DTFD-MIL	70.80%	76.86%	96.30%	53.33%	55.26%	60.00%	69.37%	85.00%	30.00%	65.00%
Resnet50+FC	90.51%	92.17%	94.44%	91.11%	84.21%	86.67%	90.47%	90.00%	**90.00%**	80.00%
HSFLA	**95.62%**	**98.71%**	**98.15%**	**93.33%**	**94.74%**	**91.67%**	**94.18%**	**95.00%**	**90.00%**	**90.00%**

Bold numbers indicate the best performance.

Fig. 2 | Supplementary results of HSFLA Classification Performance. **a**, **b** Diagnosis performance on small training sets. **c** Comparison of Diagnostic Accuracy between Pathologists and HSFLA on 2D WSIs (training set and internal testing set). Confusion matrix of intraoperative pathological diagnoses by pathologists on the internal dataset. The x-axis represents labels after double review, while the y-axis represents diagnoses made by pathologists during surgery. **d** Confusion matrix of intraoperative pathological diagnoses by pathologists on the internal test set. **e** HSFLA's error correction performance on the internal test set, analyzing 21 misdiagnosed WSIs. The x-axis represents labels after double review, while the y-axis represents HSFLA's predictions.

information with a large number of slide-level labels, this method successfully overcame the limitations of traditional weakly supervised methods in identifying subtle morphological differences, providing a more reliable and precise tool for intraoperative pathological diagnosis.

HSFLA maintains high performance on small training sets

In addition to testing on external datasets, we further evaluated the robustness of HSFLA by reducing the volume of training data. We examined the model's classification performance using 10% to 100% of the original training volume (increasing in 10% increments), as illustrated in Fig. 2a, b. The testing data consisted of complete sets encompassing all three types. Notably, even when the training data was reduced to just 10% of the original volume (approximately 60 WSIs), the model still demonstrated impressive performance, achieving 77.37% ACC and 79.45% AUC on the internal test set; 61.67% ACC and 60.82% AUC on the external test set. These performance metrics surpass those of

weakly supervised learning methods trained on the full dataset (as shown in Table 1).

These results showcase the model's exceptional generalization capability when faced with limited training data, which is particularly crucial in the field of medical imaging where large-scale data is often difficult and expensive to obtain. The model's robustness in small-sample scenarios indicates its ability to extract key features effectively from limited annotated information, enhancing learning efficiency and reducing dependence on extensive labeled datasets. In scenarios involving rare cases or emerging diseases where data scarcity is a challenge, HSFLA could potentially provide robust support for rapid and accurate diagnosis.

HSFLA maintains stable performance with limited manual annotations

Given the challenges and high costs associated with acquiring large-scale manually annotated data, recent research trends have increasingly shifted towards weakly supervised learning paradigms. However, in the specific task of lung adenocarcinoma subtyping, weakly supervised learning methods often struggle to identify subtle feature differences between in situ and minimally invasive types, resulting in significantly reduced diagnostic performance. This phenomenon highlights the tension between human resource investment and model diagnostic performance.

This research demonstrates that HSFLA's injection-supervised training method exhibits low dependence on manual annotations. Even when some WSIs lack manual annotations, it does not significantly interfere with the training process. To validate HSFLA's performance stability under limited manual annotations, we randomly masked 50% of the manual annotation data during training. Experimental results show that under these conditions, HSFLA still maintained 89.05% ACC and 88.43% AUC on the internal test set. On the external test set, the model similarly demonstrated excellent generalization capability, achieving 78.33% ACC and 79.15% AUC. These data strongly support HSFLA's ability to meet the high-precision requirements of lung adenocarcinoma subtyping diagnosis, even with limited manual annotations. HSFLA's low dependence on manual annotations significantly reduces the cost of model deployment and application, enabling its effective use in resource-constrained medical environments.

HSFLA demonstrates potential to reduce intraoperative pathological diagnostic errors on 2D WSIs

HSFLA exhibits significant potential in assisting pathologists with intraoperative subtype diagnosis of lung adenocarcinoma. The time constraint for intraoperative pathological review in lung adenocarcinoma surgery is extremely limited, typically requiring pathologists to make a diagnosis within minutes. However, diagnostic accuracy is crucial, as errors can lead to inappropriate surgical approach selection, consequently affecting treatment outcomes. Therefore, we quantified the current accuracy of intraoperative diagnoses and evaluated HSFLA's capability in diagnostic assistance.

Initially, we conducted a detailed analysis of the internal dataset described in Section 2.1. Through double verification, including post-operative review and data enrollment review, we found that the accuracy of intraoperative pathological diagnostic reports was only 85.09% (confusion matrix shown in Fig. 2c, with the x-axis representing labels after double review and the y-axis representing intraoperative diagnostic results). In the internal test set, the intraoperative diagnostic accuracy was 84.67% (Fig. 2d). Subsequently, we introduced HSFLA to analyze 21 WSIs that were misdiagnosed by pathologists in the internal test set. The results demonstrated HSFLA's excellent performance on these challenging samples, achieving a high correction rate of 90.48% (Fig. 2e, with the x-axis representing labels after double review and the y-axis representing HSFLA prediction results). Notably, the prediction execution time was only 27 s on a computer equipped with an Nvidia 3090 GPU.

These results provide evidence of HSFLA's significant advantage in identifying cases that are difficult for human experts to judge. This high-accuracy assistive diagnostic capability not only has the potential to substantially improve the accuracy of intraoperative diagnoses but also offers pathologists a reliable second opinion, especially in time-pressured intraoperative diagnostic settings. The combination of HSFLA's high accuracy and rapid processing time underscores its clinical value, particularly in scenarios where quick and accurate decision-making is critical for patient outcomes.

Detection capability of invasive lesions on 2D WSIs

We further evaluated the proposed HSFLA's capacity to identify and localize invasive lesions in 2D WSIs, which is crucial for accurately differentiating MIA from other subtypes and achieving diagnostic transparency and interpretability. Clinically, invasive lesion distribution within 5 mm and tumor distribution within 3 cm is defined as MIA, necessitating precise detection capabilities from the model. We utilized HSFLA to estimate the probability of each patch belonging to invasive lesion, defined as patch-level risk probability.

Experimental results demonstrated that in the internal test set (containing only minimally invasive WSIs), patches marked as invasive lesion by pathologists had an average risk probability of 86.6%, while non-invasive lesion patches had an average risk probability of merely 3.4%. In the external test set, these values were 81.9% and 5.8%, respectively, further validating the mode's generalization capability. Figure 3 presents selected visualization results, intuitively demonstrating the high concordance between model predictions and pathologists' annotations. These results confirm HSFLA's accuracy in identifying minute invasive lesion and its reliability for clinical application. By accurately localizing invasive lesion, this method provides critical support in differentiating the diagnostically challenging MIA from other subtypes, while simultaneously laying the foundation for subsequent diagnostic applications in 3D pathological images.

Illustrative alignment between HSFLA-identified high-risk regions and spatial transcriptomics patterns

To explore the biological plausibility and interpretability of HSFLA-identified high-risk regions, we conducted an illustrative, qualitative comparison with spatial transcriptomics data derived from a limited number of samples. Spatial transcriptomics data were collected from 4 WSIs in the internal test set, providing gene expression measurements at each spatial location. Gene expression patterns were subsequently analyzed using the Louvain clustering algorithm to characterize the spatial organization of distinct regions, cell populations, and functional areas within the tissue.

Figure 4 presents a comparative visualization of multiple spatial transcriptomics clustering distributions alongside the heat maps generated by HSFLA predictions. In the spatial transcriptomics clustering images, distinct colors demarcate regions with similar expression patterns within the tissue sections, reflecting the spatial heterogeneity of tissue structure and functional compartmentalization. Conversely, the HSFLA prediction heat maps employ a color gradient (with red indicating high risk and blue denoting low risk) to illustrate the model's predictions of invasive lesion within the WSIs. The results demonstrate structural congruence between HSFLA predictions and spatial transcriptomics clustering regions. These findings indicate the algorithm's capacity to capture histological features associated with gene expression patterns. Given the limited sample size, these observations should be interpreted as illustrative rather than quantitative validation. The qualitative spatial correspondence between HSFLA predictions and gene expression clusters suggests biological plausibility and supports the interpretability of the proposed framework.

Utilizing the invasive lesion detection feature to provide diagnostic recommendations for pathologists

To evaluate HSFLA's capability in assisting pathologists with intraoperative diagnosis, we conducted a reader study involving three pathologists with varying levels of experience (working years of 10, 3, and 2, respectively). Initially, the pathologists were tasked with diagnosing the internal test set without prior knowledge of the labels. After one month, they were provided with HSFLA-generated invasive focus detection prompts (as illustrated in

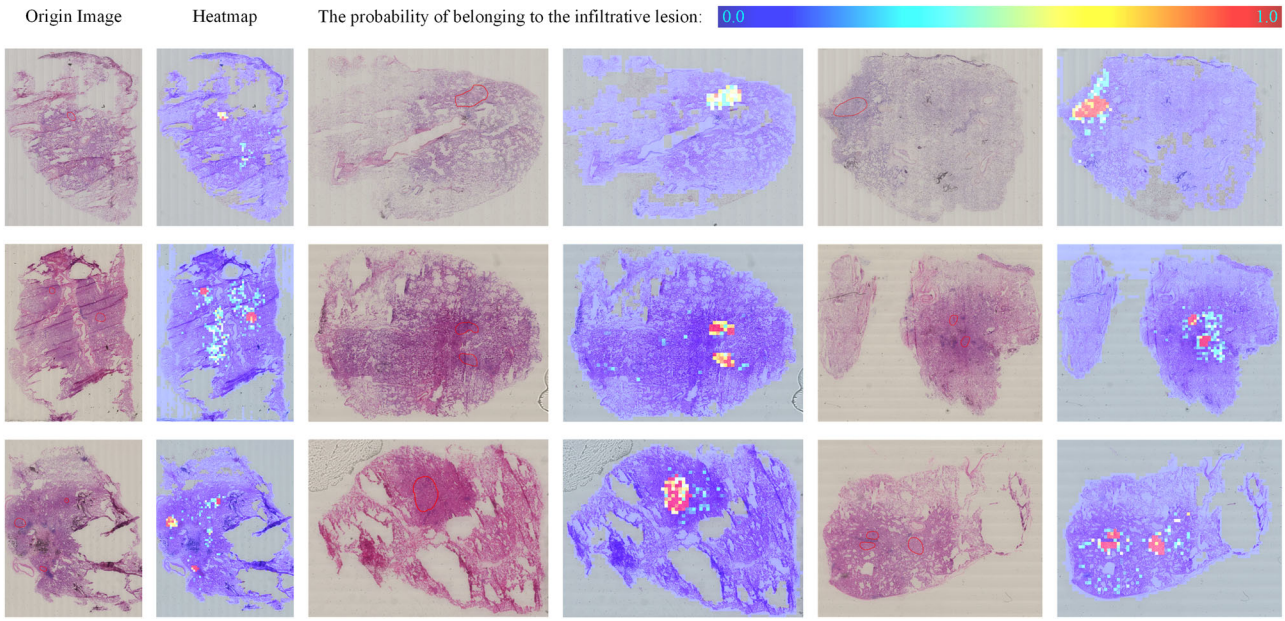

Fig. 3 | Comparison between the detection of infiltration lesions by HSFLA and manual annotation by pathologists.

Fig. 4 | Comparison between the HSFLA's prediction infiltration lesions and the clustering of spatial transcriptomics.

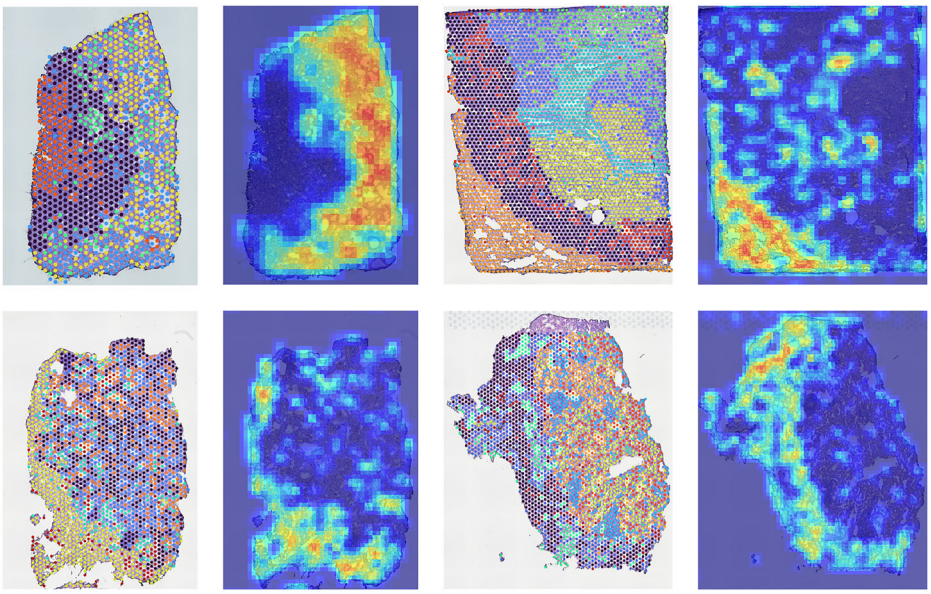

Fig. 3, highlighting potential invasive areas on WSIs) and asked to reassess their diagnoses. Across the two reading sessions, all settings, including the readers and the presentation order of the WSIs, were kept identical. The participating pathologists were informed a priori that the study design involved at least two review sessions separated by an extended washout period. Throughout the entire duration of the study, they remained blinded to the ground-truth diagnoses of the WSIs.

The diagnostic outcomes under different conditions are presented as confusion matrices (Fig. 5a–f). Figure 5a, c, e depicts the initial diagnoses made by the three pathologists, while Fig. 5b, d, f shows their corresponding AI-assisted diagnoses. The x-axis represents the true labels, and the y-axis represents the pathologists' diagnoses. Notably, HSFLA's ROI suggestion feature reduced misdiagnosis rates by 11.7%, 28.5%, and 28.5% for the three pathologists, respectively. These values were obtained by comparing the differences in diagnostic ACC of the pathologists before and after AI-assisted review, with the internal test set ($n = 137$) used as the denominator, representing absolute gains. The substantial improvement demonstrates HSFLA's robust capability in augmenting intraoperative pathological diagnoses, particularly for less experienced pathologists. While Section 2.5 highlighted HSFLA's ability to directly correct pathologists' errors, the current approach of generating suspicious area heatmaps serves only as a suggestion, with humans responsible for confirming the results. This operational model reduces intraoperativepathological diagnostic risks while maintaining excellent clinical practicality and safety.

Furthermore, HSFLA significantly reduced the time required for diagnosis. As shown in Fig. 5g, pathologists initially needed an average of 2.8 min to diagnose a single WSI, with even experienced pathologists requiring an average of 2.15 min per case. HSFLA effectively shortened the diagnostic time for each participating pathologist. Even when accounting for the 27-second AI running time, the overall diagnostic time decreased by an average of 31 s per WSI, underscoring its feasibility for clinical use. This time reduction is particularly crucial in preventing intraoperative complications.

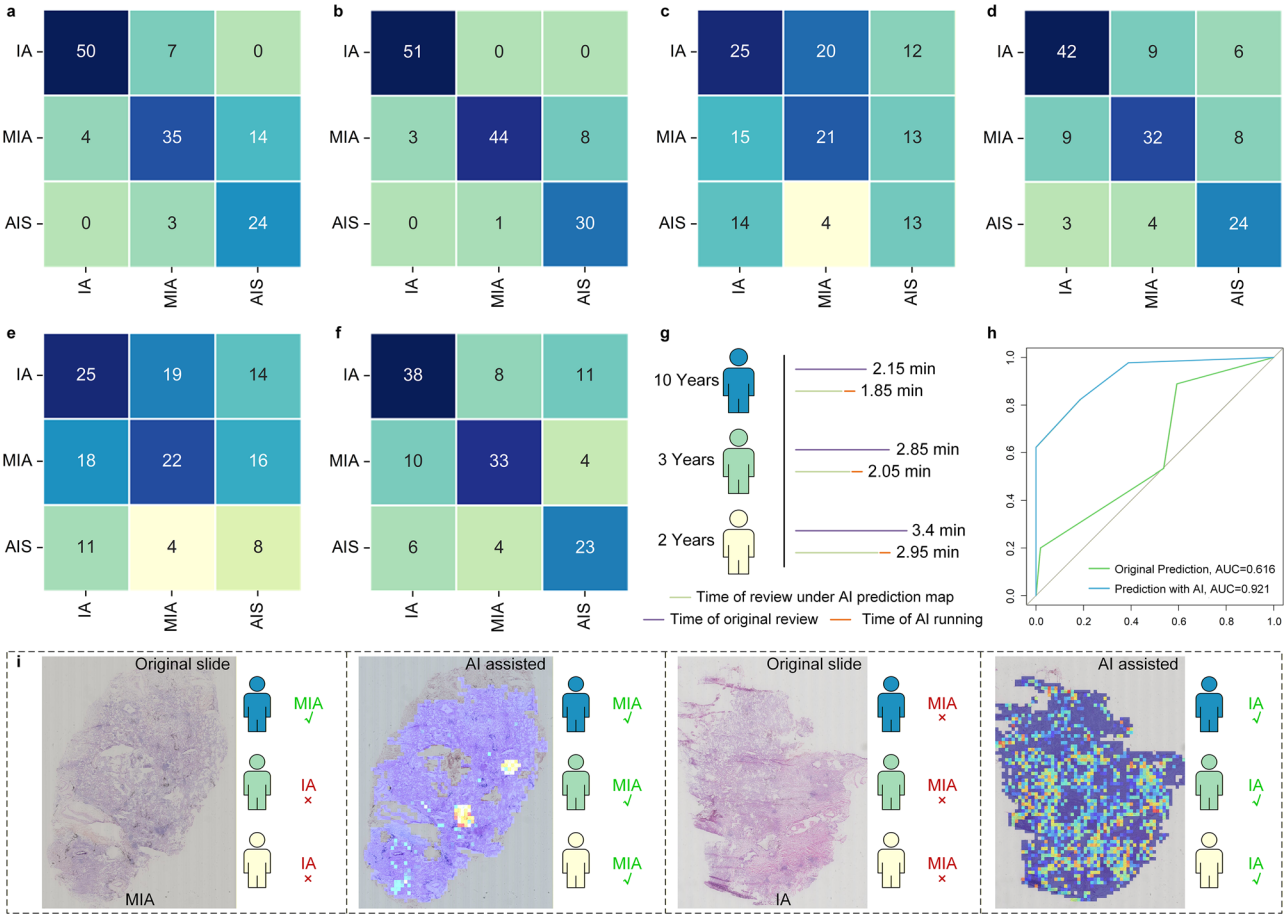

Fig. 5 | Comparison of diagnostic effectiveness among pathologists before and after AI assistance. a–f Confusion matrices depicting the diagnostic performance of three pathologists before and after AI assistance. **a** Initial diagnoses made by a pathologist with 10 years of professional experience. **b** Diagnoses made by a pathologist with 10 years of professional experience with AI assistance. **c** Initial diagnoses made by a pathologist with 3 years of professional experience. **d** Diagnoses made by a pathologist with 3 years of professional experience with AI assistance. **e** Initial diagnoses made by a pathologist with 2 years of professional experience. **f** Diagnoses made by a pathologist with 2 years of professional experience with AI assistance. **g** Comparative analysis of time required by the three pathologists for diagnosis with and without AI assistance. **h** Receiver operating characteristic (ROC) curves comparing the diagnostic performance of the three pathologists before and after AI assistance. **i** Enhanced diagnostic concordance among pathologists for challenging WSIs following the implementation of AI assistance, demonstrating improved inter-observer agreement in complex cases.

Distinguishing between MIA and IA is relatively important in classification, as the corresponding surgical approaches differ significantly, and misdiagnoses can lead to substantial medical risks. This task presents significant challenges, as it requires pathologists to identify invasive lesion and accurately determine their size within a limited timeframe. In this specific classification task, the three pathologists initially achieved accuracy rates of 85.9%, 46.5%, and 47.5%, respectively. With HSFLA's assistance, these rates improved to 96.0%, 74.7%, and 71.7%, demonstrating the effectiveness of HSFLA's support, especially for less experienced pathologists. The three pathologists formed a diagnostic panel, with each correct diagnosis assigned a confidence score of 0.33. After obtaining confidence statistics for the entire test set, we plotted ROC curves for the pathologists' diagnoses before and after AI assistance (Fig. 5h). The analysis revealed that HSFLA improved the pathologists' diagnostic AUC from 0.616 to 0.921. Figure 5i illustrates two challenging WSI cases with initial diagnostic confidence scores of 0.33 and 0, respectively. HSFLA's prompts for different subtypes are visually distinct, with marked differences in the area and distribution of highlighted suspicious regions, providing pathologists with direct and powerful guidance leading to correct diagnoses. These results demonstrate HSFLA's substantial potential in enhancing diagnostic accuracy, reducing diagnostic time, and providing crucial assistance in challenging cases, indicating significant applicability in clinical settings. Notably, the participation of three pathologists limits to fully evaluate the generalizability of this advantage.

Evaluation of consecutive WSI registration

Quantitative evaluation of the WSI registration process is required to ensure diagnostic accuracy in three-dimensional space. The adopted DeeperHistReg[18] registration strategy enhanced anatomically consistent alignment between consecutive sections, supporting accurate three-dimensional reconstruction and invasive lesion volume estimation. Quantitative evaluation of registration performance demonstrated effective alignment, as reflected by increases in dice similarity, mutual information (MI) and structural similarity index. Specifically, dice similarity increased from 0.816 ± 0.044 before registration to 0.857 ± 0.048 after registration, while MI increased from 0.213 ± 0.161 to 0.408 ± 0.183, and structural similarity improved from 0.456 ± 0.066 to 0.551 ± 0.090. These results indicate that reliable anatomical correspondence was achieved in most cases. Detailed quantitative registration results are provided in Supplementary Information 3. However, registration accuracy may be reduced in the presence of severe tissue damage, fragmentation, or pronounced morphological distortion during slide preparation, which is further discussed in the qualitative failure case analysis.

HSFLA assistance to pathologists in 3D space: application in a prospective study

The actual growth pattern of invasive lesion is three-dimensional, rather than the planar representation typically presented in traditional WSIs.

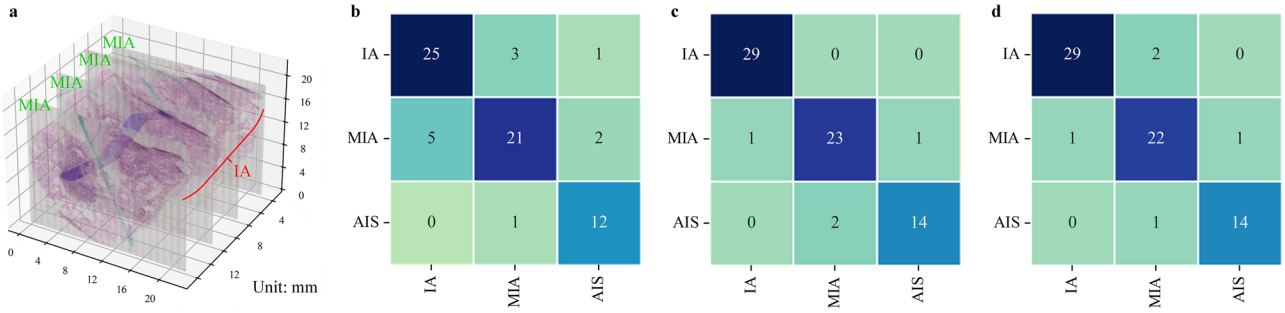

Fig. 6 | HSFLA diagnosis on 3D WSIs. a Schematic representation of HSFLA diagnosis on 3D WSIs. Even when all individual WSIs are predicted as minimally invasive, their combination reveals a larger invasive focus. Consequently, the complete 3D sample is classified as invasive lung adenocarcinoma. **b** Confusion matrix of pathologists' intraoperative pathological diagnoses on 3D WSIs. The x-axis represents labels after double review, while the y-axis indicates diagnoses made by pathologists during surgery. **c** Confusion matrix of HSFLA's classification predictions in three-dimensional space. The x-axis denotes labels after double review, whereas the y-axis shows HSFLA predictions. **d** Confusion matrix of pathologists' final diagnoses with HSFLA assistance. Five patients' surgical methods (among 70) have been corrected due to the intervention of AI assisted diagnosis, demonstrating the potential to improve the efficacy of surgical decision-making.

Table 2 | Diagnosis performance on 3D WSIs

Methods	ACC	Recall-IA	Recall-MIA	Recall-AIS
CLAM	62.86%	76.67%	48.00%	60.00%
TransMIL	65.71%	80.00%	44.00%	73.33%
IB-MIL	57.14%	70.00%	60.00%	26.67%
CI-MIL	58.57%	66.67%	56.00%	46.67%
DTFD-MIL	67.14%	80.00%	52.00%	66.67%
Resnet-50 + FC	88.57%	**96.67%**	88.00%	73.33%
HSFLA	**94.29%**	**96.67%**	**92.00%**	**93.33%**

Bold numbers indicate the best performance.

Consequently, developing an auxiliary diagnostic system in three-dimensional space can provide pathologists with safer and more accurate diagnostic recommendations. HSFLA not only reflects the tumor's true state more comprehensively but also addresses the limitations of two-dimensional analysis. Following ethical approval, we applied HSFLA to real intraoperative pathological diagnosis processes for lung adenocarcinoma. This constituted a prospective study with consecutive data collection. Initially, examined specimens were processed with at least five consecutive sections after staining and fixation. Subsequently, pathologists made their initial diagnoses and recorded them. Finally, these sections were digitally scanned, allowing HSFLA to provide predictive diagnostic results and potential invasive lesion location markers. The pathologist then confirmed the final results and communicated them to the surgical doctor, determining the appropriate surgical approach (local resection or complete lobectomy) for the patient. Post-surgery, a panel of pathologists thoroughly reviewed these samples (including intraoperative frozen section slides and post-operative paraffin-embedded section slides) to obtain definitive diagnoses.

By identifying and registering invasive lesion positions in all continuous slices of a patient, we were able to calculate the volume of complete invasive lesion in three-dimensional space (with 1.5 mm intervals between consecutive slices). According to clinical standards, when the maximum distance range of invasive lesion in space exceeds 5 mm, the patient is diagnosed with IA rather than MIA type. As illustrated in Fig. 6a, even when all individual WSIs are diagnosed as MIA, the combined volume of the invasive focus may still warrant surgical intervention as IA. Consequently, referencing the auxiliary diagnostic results on three-dimensional WSIs offers enhanced clinical safety. In contrast, MIL typically analyzes continuous WSIs through feature aggregation methods when these images are superimposed. This computational approach diverges significantly from the diagnostic workflow outlined in lung adenocarcinoma guidelines. As shown in Table 2, in a three-dimensional image test set comprising 70 patients (417 WSIs in total), HSFLA demonstrated exceptional performance, achieving 94.29% ACC (patient-level). In contrast, the highest accuracy achieved by MIL methods alone was only 67.14%, indicating their near absence of diagnostic capability in three-dimensional space. As ResNet-50 (equipped with a fully connected layer as the classifier) does not explicitly model AIS as a distinct subtype (as achieved by HSFLA), false-positive invasive patches were more likely to be identified in AIS cases. Consequently, its overall accuracy (88.6%) was largely affected by the relatively low recognition rate for AIS (73.33%).

Notably, in actual practice, pathologists face the challenging task of simultaneously reading multiple WSIs from a single patient within a limited timeframe, mentally aligning invasive lesion based on experience to estimate their true volume. Results show that, compared to post-operative comprehensive review outcomes, experts achieved 82.86% diagnostic accuracy in actual intraoperative diagnoses (Fig. 6b). Experts misclassified 16.7% of IA cases as MIA and 12.5% of MIA cases as IA. This could potentially lead to inappropriate surgical approaches for affected patients, presenting significant safety concerns. In contrast, HSFLA exhibited superior performance, attaining a classification accuracy of 94.3% (Fig. 6c), with only one invasive case misclassified as minimally invasive. Ultimately, when pathologists made diagnoses with HSFLA's invasive lesion marking function recommendations, their accuracy substantially improved to 92.9% (Fig. 6d). With HSFLA's intervention, surgical approaches for 5 patients (among 70) were modified to better align with pathological assessment and current surgical guidelines. This prospective study demonstrates the practical efficacy of HSFLA in enhancing diagnostic accuracy and patient care in real-world settings, highlighting its potential for improving surgical decision-making and patient outcomes in lung adenocarcinoma cases. The human-machine interaction between experienced pathologists and HSFLA resulted in more appropriate treatment strategies for patients, representing an advancement in the application of artificial intelligence to clinical tasks.

Discussions

Several limitations of this study should be acknowledged. First, although HSFLA influenced intraoperative diagnostic interpretation and surgical strategy in a subset of patients, the present study was not designed to systematically evaluate downstream clinical outcomes such as postoperative complications, recurrence, or long-term survival. While no adverse events, including recurrence or major complications, have been observed to date in the prospectively enrolled cohort (Supplementary Information 4), the sample size and follow-up duration remain limited. Therefore, conclusions regarding improvements in patient outcomes should be interpreted with caution.

Second, the prospective application of HSFLA was conducted at a single center with experienced thoracic pathologists, which may limit the generalizability of the findings to other institutions or clinical settings.

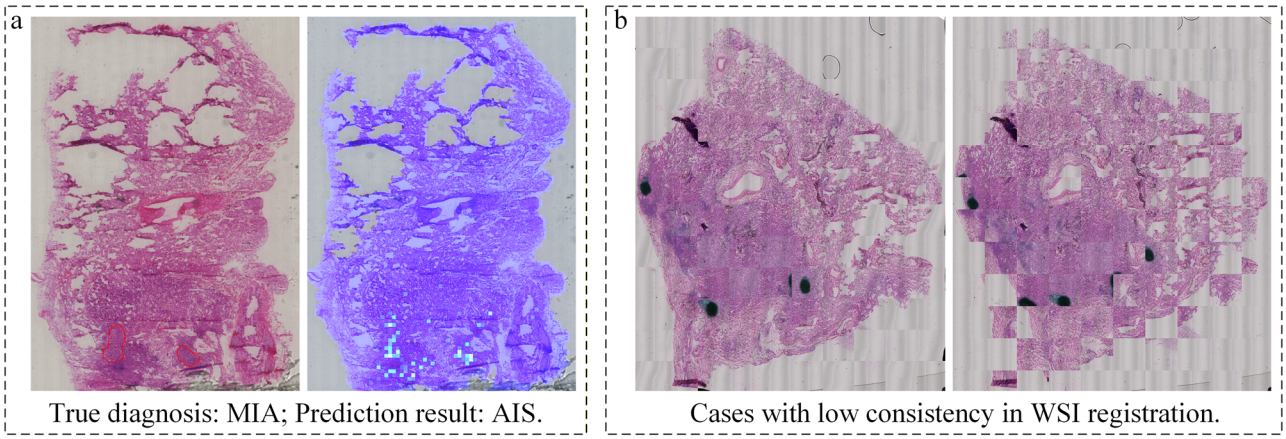

Fig. 7 | Qualitative analysis of failure cases. a Low risk scores assigned by HSFLA to lesions with very small areas, remaining below the predefined threshold for invasive classification (the region circled in red in the left image corresponds to the invasive lesion). **b** Registration distortion occurring in fragmented WSIs. Defining passed registration for 3D spatial diagnosis as a Dice Similarity of >75% for intact tissue boundaries, the failure rate of consecutive WSIs is 2.6%.

Additionally, HSFLA currently functions as a decision-support tool rather than an autonomous diagnostic system, and final clinical decisions remained dependent on human judgment. Future multicenter studies with larger cohorts, standardized surgical endpoints, and long-term follow-up are warranted to determine whether AI-assisted intraoperative pathology can translate into measurable improvements in oncological outcomes and patient prognosis.

Third, a qualitative analysis of cases mis-predicted by HSFLA revealed several representative and challenging pathological scenarios, including extremely small invasive lesion and tissue damage introduced during slide preparation, which remain difficult for both human pathologists and automated methods. Two typical failure modes were identified: (i) false-negative errors arising from unsuccessful recognition of invasive lesions during the classification stage, and (ii) misdiagnoses caused by failed registration of consecutive WSIs, leading to inaccurate estimation of invasive lesion volume in three-dimensional space.

As shown in Fig. 7a, for some lesions with smaller areas, HSFLA assigned higher risk scores to the corresponding regions than to surrounding non-invasive tissue; however, these scores remained below the predefined threshold (>0.5), resulting in false-negative predictions (the region circled in red in the left image corresponds to the invasive lesion). Adaptive or task-aware threshold optimization may represent a potential direction for future improvement. Notably, even relatively low but salient risk cues may still provide supportive information to pathologists and help reduce missed diagnoses under a human–AI collaborative setting. Regarding artifacts specific to frozen sections: the detrimental impact of ice crystals and tissue tearing on patch-level recognition remains manageable. In Fig. 3, significant tissue tearing and staining variability can be observed; however, these factors do not compromise the recognition of ROIs. Furthermore, although ice crystals induce local nuclear distortion and tissue gaps that complicate microscopic morphological assessment, no obvious spurious activations were observed in the heatmaps. Because the training set consists entirely of frozen sections and manual annotation, the model can effectively learn from the pathologists' expertise to disregard the localized morphological blurring caused by ice crystals. Employing encoders pre-trained on multicenter pathological image datasets was proven to mitigating noise interference and improving model performance[27–29]. This advantage is applicable across most frameworks (Supplementary Information 10 for detailed performance gains).

Nevertheless, tissue tearing does indeed reduce the accuracy of the WSI registration process, as illustrated in Fig. 7b. Substantial tissue damage or morphological distortion in some consecutive WSIs led to spatial distortion during three-dimensional reconstruction by existing registration algorithms, thereby affecting invasive volume estimation and final subtype determination. Defining passed registration for 3D spatial diagnosis as a Dice Similarity of >75% for intact tissue boundaries, 2.6% of WSIs are unqualified (Supplementary Information 3). Developing or training more robust, pathology-oriented slide registration algorithms tailored to such data characteristics should be prioritized in future work. Additionally, explicit training aimed at detecting artifacts ought to be integrated.

Overall, this study introduces HSFLA, a hybrid-supervised deep learning framework for lung adenocarcinoma. The framework achieves high-precision lung adenocarcinoma subtype classification on both 2D and 3D pathological images by combining weakly supervised learning with limited fine-grained annotations. Through extensive validation across multiple test sets, HSFLA significantly outperforms existing mainstream weakly supervised deep learning approaches due to its architectural advantages, while also demonstrating strong competitiveness compared to manual review. In terms of automated invasive area annotation, HSFLA maintains high consistency when compared to manual pixel-level annotations and spatial transcriptomics results. Consequently, it provides pathologists with a safe and effective diagnostic aid. This technology adheres to the quantitative standards for intraoperative lung adenocarcinoma subtype classification in pathology and may assist in reducing the risk of under- or over-treatment during surgical decision-making. The diagnostic accuracy demonstrated by HSFLA in real-world clinical settings supports the feasibility and potential value of integrating artificial intelligence technologies into pathological workflows. It illustrates how AI can augment rather than replace human clinical judgment, fostering a collaborative approach that leverages the strengths of both human expertise and machine learning algorithms.

Methods

Weakly supervised MIL network for initial screening of invasive adenocarcinoma

To efficiently classify WSIs into invasive adenocarcinoma versus non-invasive subtypes (microinvasive & in-situ), we formulate the task as a weakly supervised MIL problem. Each WSI is treated as a "bag" of instances (image patches), where only slide-level labels are available during training.

Let $X = \{x_1, x_2, \ldots, x_N\}$ denote a WSI (bag) containing N patches, and $y \in \{0, 1\}$ be its slide-level label (1: invasive, 0: non-invasive). In MIL, the bag label is determined by the presence of at least one positive instance:

$$y = \begin{cases} 1 & if\ \exists f(x_k) = 1 \\ 0 & otherwise \end{cases} \quad (1)$$

Where $f(\cdot)$ is an unknown patch-level classifier. The goal is to learn $f(\cdot)$ while training only on slide-level labels.

Each patch x_i is encoded into a d-dimensional feature vector $h_k \in \mathbb{R}^d$ using a pre-trained ResNet-50 backbone:

$$h_k = ResNet(x_k), h_k \in \mathbb{R}^d \quad (2)$$

To weigh patches by their contribution to the slide-level prediction, we employ an attention mechanism to compute importance scores α_k:

$$\alpha_k = \frac{exp(\omega^T tanh(Vh_k))}{\sum_N exp(\omega^T tanh(Vh_k))} \quad (3)$$

Where $V\mathbb{R}^{L\times d}$ and $w\mathbb{R}^L$ are learnable parameters, and L is the hidden dimension.

The bag embedding Z is a weighted sum of patch features:

$$Z = \sum_N \alpha_i h_i \quad (4)$$

A fully connected layer $fc(\cdot)$ followed by softmax predicts the slide label:

$$p(\mathrm{y} = 1|\mathrm{X}) = fc(\mathrm{W_c z} + \mathrm{b_c}) \quad (5)$$

$$L_{MIL} = -(y \cdot log(p(y = 1|X)) + (1 - y)log(1 - log(p(y = 1|X))) \quad (6)$$

Where $W_c \in \mathbb{R}^{1\times d}$ and $b_c \in \mathbb{R}^{1\times d}$ are learnable weights.

Hybrid-supervised dual-channel network for micro-invasive subtype classification

MIA exhibits limited stromal invasion (<5 mm), requiring precise localization of small invasive lesion amidst predominantly in situ patterns. To address this, we develop a hybrid-supervised dual-channel network that integrates: 1. Patch-level supervision: 180 expert-annotated WSIs with pixel-level masks of micro-invasive lesion (consensus by 3 pathologists). The manually generated pixel-level annotations were converted into patch-level labels, whereby patches (x_k) encompassed by the annotated regions were classified as invasive lesion $\left(y_p^k = 1|x_k\right)$. 2. Slide-level supervision: Standard diagnostic labels for all WSIs. Where the MIL network used a single ResNet-50 backbone, we now deploy a dedicated high-resolution pathway, employ ResNet50 with patch-level supervision:

$$p_{patch}\left(y_p^k = 1|x_k\right) = \sigma\left(W_{patch} \cdot ResNet(x_k) + b_{patch}\right) \quad (7)$$

$$\begin{aligned} L_{patch} = &-\sum_k^N (y_p^k \cdot log(p_{patch}\left(y_p^k = 1|x_k\right)) \\ &+(1 - y_p^k)log(1 - log(p_{patch}\left(y_p^k = 1|x_k\right)) \end{aligned} \quad (8)$$

$$L_{Final} = L_{MIL} + L_{patch} \quad (9)$$

L_{patch} is a composite cross-entropy loss function, by which the network is optimized via gradient backpropagation[30]. Since high-attention patches ($p_{patch}\left(y_p^k = 1|x_k\right) > 0.5$) are likely tumor regions, the model implicitly localizes invasive patterns. At inference, we threshold $p_{patch}\left(y_p^k = 1|x_k\right)$ to extract suspicious regions for pathologist review.

Deep feature-based slice alignment for 3D reconstruction

We utilized the DeeperHistReg[18] framework for the registration of consecutive H&E stained sections. DeeperHistReg is a software framework specifically designed for the registration of WSIs acquired with various staining methods, such as H&E and immunohistochemistry. The registration process using the DeeperHistReg is divided into three steps: preprocessing, initial alignment and nonrigid registration.

During preprocessing, image intensity correction is typically achieved using a linear transformation formula to adjust the color intensity. If the original image intensity is denoted as $I(x, y)$ the transformed intensity becomes $I'(x, y) = a \cdot I(x, y) + b$, where a and b are correction coefficients; x and y are pixel coordinates. This step ensures consistency for subsequent processing.

Initial alignment is performed using affine transformation for linear registration of images. The affine transformation can be expressed as:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (10)$$

where (x, y) and (x', y') are the pixel coordinates before and after transformation, $a_{11}, \ldots, a_{22}$ are elements of the affine transformation matrix, t_x and t_y are translation components. Affine transformation includes scaling, rotation, and translation.

Nonrigid registration employs B-splines or dense displacement fields for more complex image deformation. The B-spline transformation is controlled by control points and basis functions, expressed as:

$$T(x, y) = \sum_i \sum_j B_i(x) \cdot B_j(y) \cdot P_{i,j} \quad (11)$$

where $T(x, y)$ is the transformed coordinate, $B_i(x)$ and $B_j(y)$ are B-spline basis functions, $P_{i,j}$ is the coordinate of the control point. By adjusting the positions of the control points, nonlinear deformation of the image can be achieved.

Dense displacement fields achieve deformation by calculating the displacement of each pixel on the image grid, typically represented as:

$$D(x, y) = (u(x, y), v(x, y)) \quad (12)$$

where $D(x, y)$ is the displacement field; $u(x, y)$ and $v(x, y)$ are the displacements in the horizontal and vertical directions, respectively.

Three-dimensional tumor micro-invasion assessment method

Given m precisely registered consecutive pathological slices $\{X_1, X_2, \ldots, X_m\}$ with an inter-slice spacing of $d = 1.5mm$, each tumor infiltration region $\Omega_i \subset \mathbb{R}^2$ on slice X_i is defined by expert annotation or automated segmentation as a polygonal area with vertex coordinates $\{v_1^{(i)}, v_2^{(i)}, \ldots, v_n^{(i)}\}$. Each 2D infiltration region Ω_i is mapped into 3D space as a planar structure at $z = z_i$, where z_i denotes the position of slice X_i along the stacking axis (unit: mm). For consecutive slices X X_i and X_{i+1}, a truncated conical volume C_i is constructed by linearly interpolating corresponding points between Ω_i and Ω_{i+1}:

$$C_i = \{\beta \cdot (x, y, z_i) + (1 - \beta) \cdot (x', y', z_{i+1}) | (x, y) \in \Omega_i, (x', y') \in \Omega_{i+1}, \beta \in [0, 1]\} \quad (13)$$

This formulation ensures topological continuity between adjacent layers.

The complete 3D tumor infiltration volume V is defined as the union of all interconnected truncated cones:

$$V = \bigcup_{i=1}^{m-1} C_i \quad (14)$$

The maximum infiltration depth $D_{\max}$ along the z-axis is defined as:

$$D_{max} = \underset{1 \le i < j \le m}{max}\left(z_j - z_i\right) \tag{15}$$

Where $z_j - z_i = (j - i) \cdot d$, constrained by the existence of a connected 3D path between Ω_i and Ω_j.

To ensure biological validity, adjacent slice infiltration regions must satisfy overlap criteria: 1. Ω_i and Ω_{i+1} are considered continuous if their orthogonal projections onto a common plane exhibit non-empty intersection. 2. Non-overlapping adjacent slices partition $D_{\max}$ into segment-wise depth measurements. Following International Association for the Study of Lung Cancer (IASLC) guidelines, micro-invasive status is determined by:

$$MIA \iff Vol(V) \le 125mm^3 \; and \; D_{max} \le 5mm \tag{16}$$

Ethics policy explanation

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2024). The study protocol and the use of related data were approved by the Ethics Committees of Shengjing Hospital of China Medical University (internal dataset, approval No. 2025PS1537K) and The First Hospital of China Medical University (external testing set, approval No. AF-SOP-07-1.2-01). Copies of the ethical approval documents have been uploaded (Supplementary Information 5, 6). For the retrospective and prospective patient cohorts: The artificial intelligence system developed in this study is designed solely to assist pathologists in decision-making and does not involve any direct patient intervention. Therefore, the requirement for informed consent regarding AI-assisted intraoperative pathological diagnosis was waived by the Ethics Committee of Shengjing Hospital of China Medical University, deeming the study to present minimal risk to participants. Clinical trial registration was deemed inapplicable as this is a decision-support study rather than an interventional clinical trial. While, informed consent was obtained from the three pathologists participating in the reader study.

Data availability

The complete dataset (include WSIs and manual annotations) and code associated with this study will be made publicly available. Access link: https://github.com/MedFLung/HSFLA.

Received: 30 November 2025; Accepted: 9 April 2026;
Published online: 20 April 2026

References

1. Barta, J. A., Powell, C. A. & Wisnivesky, J. P. Global epidemiology of lung cancer. *Ann. Glob. Health.* **85** (2019).
2. Siegel, R. L., Giaquinto, A. N. & Jemal, A. Cancer statistics, 2024. *CA Cancer J. Clin.* **74**, 12–49 (2024).
3. Travis, W. D. et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.* **6**, 244–285 (2011).
4. Travis, W. D., Brambilla, E., Burke, A. P., Marx, A., & Nicholson, A. G. (Eds.). WHO Classification of Tumours of the Lung, Pleura, Thymus and Heart (4th ed.). (International Agency Research Cancer, 2015).
5. Russell, P. A. et al. Does lung adenocarcinoma subtype predict patient survival? A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. *J. Thorac. Oncol.* **6**, 1496–1504 (2011).
6. Zheng, M. Classification and pathology of lung cancer. *Surg. Oncol. Clin.* **25**, 447–468 (2016).
7. Rosai, J. *Rosai and Ackerman's Surgical Pathology e-book* (Elsevier Health Sciences, 2011).
8. Nakhleh, R. E. Quality in surgical pathology communication and reporting. *Arch. Pathol. Lab. Med.* **135**, 1394–1397 (2011).
9. Azari, F., Kennedy, G. & Singhal, S. Intraoperative detection and assessment of lung nodules. *Surg. Oncol. Clin. North Am.* **29**, 525 (2020).
10. Farahani, N. et al. Three-dimensional imaging and scanning: current and future applications for pathology. *J. Pathol. Inform.* **8**, 36 (2017).
11. Liu, S. et al. Precise diagnosis of intraoperative frozen section is an effective method to guide resection strategy for peripheral small-sized lung adenocarcinoma. *J. Clin. Oncol.* **34**, 307–313 (2016).
12. Carbonneau, M.-A. et al. Multiple instance learning: a survey of problem characteristics and applications. *Pattern Recognit.* **77**, 329–353 (2018).
13. Ilse, M., Tomczak, J. & Welling, M. Attention-based deep multiple instance learning. In *International Conference on Machine Learning* (PMLR, 2018).
14. Quellec, G. et al. Multiple-instance learning for medical image and video analysis. *IEEE Rev. Biomed. Eng.* **10**, 213–234 (2017).
15. Li, B., Li, Y. & Eliceiri, K. W. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2021).
16. Zhou, Z.-H. A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* **5**, 44–53 (2018).
17. Nojima, S. et al. CUBIC pathology: three-dimensional imaging for pathological diagnosis. *Sci. Rep.* **7**, 9269 (2017).
18. Wodzinski, M. et al. RegWSI: Whole slide image registration using combined deep feature-and intensity-based methods: Winner of the ACROBAT 2023 challenge. *Comput. Methods Programs Biomed.* **250**, 108187 (2024).
19. Nicholson, A. G. et al. The 2021 WHO classification of lung tumors: impact of advances since 2015. *J. Thorac. Oncol.* **17**, 362–387 (2022).
20. Lu, M. Y. et al. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat. Biomed. Eng.* **5**, 555–570 (2021).
21. Campanella, G. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).
22. Shao, Z. et al. Transmil: transformer based correlated multiple instance learning for whole slide image classification. *Adv. Neural Inf. Process. Syst.* **34**, 2136–2147 (2021).
23. Lin, T. et al. Interventional bag multi-instance learning on whole-slide pathological images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2023).
24. Lin, W. et al. Boosting multiple instance learning models for whole slide image classification: a model-agnostic framework based on counterfactual inference. In *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI, 2024).
25. Zhang, H. et al. Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2022).
26. Koonce, B. *"ResNet 50." Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*, 63–72 (Apress, 2021).
27. Chen, R. J. et al. Towards a general-purpose foundation model for computational pathology. *Nat. Med.* **30**, 850–862 (2024).
28. Neidlinger, P. et al. Benchmarking foundation models as feature extractors for weakly supervised computational pathology. *Nat. Biomed. Eng.* **9**, 1–11 (2025).
29. Zhang, H. et al. Customized transformer for lymph node metastasis prediction from lung adenocarcinoma histology in a multicentric study. *NPJ Precision Oncol.* **10**, 16 (2025).
30. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444 (2015).

Acknowledgements
Thanks to the First Affiliated Hospital of China Medical University for agreeing this research. This study did not receive funding.

Author contributions
Jianwei Zhao was responsible for writing manuscripts, developing models, and collecting data. Junyan Zhang was responsible for experimental testing and manuscript writing. Yihao Wang was responsible for writing the manuscript and drawing. Xinwen Zhong and Xiaojiao Guan jointly revised the manuscript and supervised the work. Jianwei Zhao and Junyan Zhang have made equal contributions.

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/s41698-026-01441-x.

Correspondence and requests for materials should be addressed to Xinwen Zhong or Xiaojiao Guan.