# Weakly supervised deep learning for human breast cancer-related optical biomarker discovery in label-free virtual histopathology

Jindou Shi[1,2], Haohua Tu[1,2], Jaena Park[2,3], Marina Marjanovic[2,3,4], Anna M. Higham[5], Natasha N. Luckey[5], Kimberly A. Cradock[5], Z. George Liu[5], and Stephen A. Boppart[1,2,3,4,6]*

1 Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

2 Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

3 Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

4 Carle Illinois College of Medicine, University of Illinois at Urbana-Champaign, Urbana, IL, USA

5 Carle Foundation Hospital, Urbana, IL, USA

6 Cancer Center at Illinois, University of Illinois at Urbana-Champaign, Urbana, IL, USA

*Corresponding Author

boppart@illinois.edu

## Abstract

With the latest advancements in optical bioimaging, rich structural and functional information is generated from biological samples, which calls for powerful computational tools to identify patterns and uncover relationships between optical characteristics and various biological conditions. Here we present a weakly supervised deep learning framework for human breast cancer-related optical signature discovery based on virtual histopathology enabled by simultaneous label-free autofluorescence multiharmonic microscopy. This framework consists of (1) cancer diagnosis by leveraging inexact and incomplete supervision, and (2) cancer-related optical signature identification via the integration of multiple model interpretation techniques. The framework has achieved an average area under the curve (AUC) of 0.975 on the cancer diagnosis task. In addition to well-known cancer biomarkers, non-obvious cancer-related patterns were revealed by the framework, including NAD(P)H-rich extracellular vesicles observed in normal appearing breast cancer tissue, which facilitate new insights into the tumor microenvironment and field cancerization. This framework can be further extended to diverse optical signature discovery tasks and a variety of imaging modalities.

## Introduction

Advancements in optical bioimaging have revealed rich structural and functional information from biological samples based on intrinsic structural, molecular, and metabolic contrasts[1-5]. However, the rich information from the multimodal bioimages may easily exceed the capabilities of human visual inspection. Extracting information and detecting patterns using computational tools could therefore enhance perceptions and inspire new scientific discoveries.

Machine Learning (ML), which excels at extracting patterns and insights from large volume of data, has demonstrated superhuman performance in a significant number of tasks[6-9]. However, generating accurate and fine-grained manual annotations for ML model training can be time-consuming, costly, and sometimes unfeasible[10]. To alleviate the dependency on the strong supervision, numerous weakly supervised learning approaches have been proposed to exploit inexact and incomplete supervision[11-13]. Among those approaches, Multiple Instance Learning (MIL), which leverages inexact supervision, has gained popularity in many application fields, including content-based image retrieval[14,15], computer-aided diagnosis[16,17], and object localization[18]. In MIL, instances are grouped into bags, and only bag-level labels are available for training. Based on the objective of application, MIL models can be trained to generate bag-level predictions as well as to inform discriminative instances.

Recent studies have demonstrated substantial success achieved by MIL in computational pathology (CPATH)[19-21], where whole slide images (WSI) are defined as bags and the extracted patch images are treated as instances. Cancer diagnosis is one of the major topics in CPATH. The fine-grained local annotations are often hard to obtain in many CPATH cancer diagnosis tasks, while the labels for WSIs are available more readily[10]. It was reported that MIL-based decision support systems have achieved clinical-grade performance for cancer diagnosis when being trained on large-size datasets without cellular-level annotations[22]. Additionally, MIL has also been applied to a wide variety of pathological tasks, including tumor segmentation[23], cancer prognosis[24], and tumor origin prediction[25].

The majority of MIL studies in CPATH were based on hematoxylin-eosin (H&E) histopathology, which serves as the gold standard for cancer diagnosis. However, the H&E sample preparation process requires tissue fixation, sectioning, and staining, which may unavoidably cause tissue distortions and feature disappearance. Therefore, a label-free virtual histopathology technology is fulfilled by nonlinear optical techniques which enable the observation of the unperturbed biochemical microenvironment[3,26]. In this study, we leverage MIL for human breast cancer-related optical signature identification based on label-free virtual histopathology, where only ambiguous whole-image-level annotations are available. The virtual histopathology slides were generated using simultaneous label-free autofluorescence-multiharmonic (SLAM) microscopy which enables simultaneous and efficient acquisition of two- and three-photon-excited autofluorescence (2PF and 3PF, respectively) as well as second and third harmonic generation (SHG and THG, respectively) with strict spatial and temporal co-registration (Fig. 1a,b)[3,27]. Compared to conventional histopathology, the label-free virtual histopathology offers an unperturbed and spatially-resolved way to study living or freshly resected tissues without the need of tissue fixation and staining[27,28]. In addition to structural information, the label-free virtual histopathology provides the molecular information of a sample in its native state, demonstrating its great potential for the exploration of novel molecular biomarkers in early stages of cancer, as well as for new fundamental investigations into carcinogenesis.

Unlike many ML-based cancer diagnosis applications which focus primarily on tissue regions with well-established cancer-associated morphologies, this study aims to explore unconventional structural and functional optical signatures for which the correlations with breast cancer might not be fully investigated. In addition to primary tumors, we generated the SLAM virtual histology slide images of tumor microenvironment (TME) and peritumoral regions to reveal potential evidence of field cancerization [29,30]. Studies have shown various interactions between tumor and TME linked with tumor angiogenesis and peripheral immune tolerance, which may affect the growth and evolution of cancerous cells[31]. The replacement of large areas of normal cell population by cancer-primed cell population is often referred to field cancerization that may not involve clinically detectable morphological changes[32]. Due to the unperturbed nature and the rich morphological and molecular information, SLAM may unveil unconventional cancer-related optical signatures, which are mainly undetectable using conventional histology techniques.

Constrained by the current knowledge of the unconventional functional and structural information generated by SLAM, the annotations for the whole images (bag-level labels) can be ambiguous. Namely, tissue images from tumor surrounding regions without conventional cancer biomarkers can hardly be assigned certain labels (cancer or normal). This leads to a unique learning task where ML models need to be trained with inexact and incomplete supervision at the same time. Notably, unlike the assumptions of conventional approaches that deal with incomplete supervision (e.g., Semi-Supervised Learning (SSL))[33,34], the positive instances in unlabeled slides may come from separate inherent distributions as those from labeled slides (Supplementary Fig. 1). In other words, the cancer-related optical signatures from uncertain regions (e.g., tumor surrounding areas) may differ from those inside the well-known cancer areas (i.e., primary tumor). Therefore, with the aim of unveiling potential cancer-related patterns from uncertain examples, a ML framework that exploits inexact and incomplete supervision is needed for the optical signature identification task. Learning from inexact or incomplete supervision has been studied separately in MIL and SSL, while the combination of those two types of supervision with the separate inherent distribution assumption has rarely been explored.

Here, we propose Mix-and-Match Multiple Instance Learning (MM-MIL) as a weakly supervised deep learning framework for human breast cancer-related optical signature discovery based on ambiguous and coarse-grained annotations. This framework involves a MIL model for global (bag-level) and local (instance-level) cancer diagnosis along with the optical signature identification enabled by the integration of model interpretation techniques (Fig. 2). The target of this framework is to reveal cancer-related patterns in both the primary tumor regions as well as tumor surrounding regions, where potential optical markers for field cancerization and carcinogenesis may exist. In this study, the cancer-related patterns are defined as optical characteristics exhibited only in tissues from cancer subjects while being unobservable in tissues from normal subjects with no medical history of cancer.

We generated 724 SLAM images (virtual histology slides, “slides” for brevity) from 48 subjects, where “cancer”, “normal”, and “uncertain” labels were provided on the whole-slide level. Each SLAM slide is a mosaic image, with a medium field-of-view (FOV) of 750 µm × 750 µm. SLAM slides that exhibit well-known cancer biomarkers (e.g., tumor cells, tumor-associated collagen fibers) were annotated as “cancer”. “Uncertain” slides, which do not show well-known cancer signatures, were mainly from areas outside the primary tumor (several millimeters to centimeters away from the tumor boundary) and slides from normal subjects were labeled as “normal”. Aiming at leveraging the inexact and incomplete supervision, positive bags in MIL were redefined as groups of truly-positive examples (“cancer” slides)

and potentially-positive examples ("uncertain" slides), so that explicit labels ("positive") could be assigned to bags based on the standard assumption of MIL[35]. Slides were then tiled into patch images ("tiles" for brevity) as instances in MIL. By iterative selection and learning of discriminative instances from positive and negative bags, MM-MIL could make whole-slide level cancer diagnosis, as well as inform cancer-related regions in predicted positive slides. To gain insights into the model prediction, several model interpretation methods were adopted to inspire the discovery of cancer-related optical signatures. The latent feature space of the model was visualized in two dimensions for modal introspection. Based on the unique property of multi-modal images, we integrated channel occlusion with channel-wise attribution maps, aiming to measure the contribution of each SLAM channel for cancer prediction, and for informing salient structures on the pixel level in each channel.

We show that the MM-MIL framework is capable of extracting a variety of breast cancer-related optical signatures from the tumor microenvironment and peri-tumoral fields. In addition to well-established cancer biomarkers, MM-MIL revealed non-obvious cancer-related signatures that appear predominantly in tumor surrounding areas while being unobservable in breast tissues from normal subjects. Those optical signatures may inspire the discovery of new cancer markers inside and outside of the tumor microenvironment and promote research on carcinogenesis, tumor progression, and field cancerization.

## Results

### 1. Global prediction evaluation

Based on the custom-designed Mix-and-Match bagging policy, MM-MIL can be trained and evaluated on the SLAM slides with either certain or uncertain labels. To investigate cancer-related patterns at different scales, MM-MIL models were trained on tiles at three magnification levels (i.e., high-magnification (128 µm × 128 µm), medium-magnification (256 µm × 256 µm), and low-magnification (512 µm × 512 µm). All tile images were resized to the same sizes, with high-magnification level tiles involving the most detailed local information, and low-magnification level tiles having the largest FOV. The dataset was partitioned into a training-validation set (70% of slides), and a held-out test set (about 30% of slides) using subject-level partition. MIL models were trained on the training-validation set using 5-fold Monte Carlo cross-validation. For each cross-validated fold, the trained model was evaluated on the held-out test set.

When being tested on bags, among which positive bags were groups of "cancer" and "uncertain" slides, MM-MIL achieved average AUCs of 0.975, 0.971, and 0.969 on high-, medium-, and low-magnification settings respectively (Fig. 3a). In addition, MM-MIL models were evaluated on slides with only certain labels ("cancer" or "normal"). For models trained on high-magnification level, an average AUC of 0.939 was achieved when "uncertain" slides were removed from positive bags. While for medium- and low- magnification settings, the slide-level prediction performance decreased with average AUCs to be 0.810 and 0.816 respectively (Fig. 3b).

The differences in AUCs between bag- and slide-level predictions indicate that the "uncertain" slides in positive bags have considerable contribution to the positive class (cancer) prediction. To be noted, during the training phase of MM-MIL, only bag-level labels were used for model optimization, while slide-level labels were hidden to encourage models to "look" for cancer-related optical signatures in both "cancer" and "uncertain" slides. For both bag-level and slide-level global prediction tasks, we observed that MM-MIL models trained on the high magnification level achieved the best performance.

### 2. Local prediction evaluation

To quantitatively evaluate the local (tile-level) predictions, known tumor areas in "cancer" slides from the test set were annotated by human experts based on the appearance of known cancer signatures (Fig. 3e). The ratio ($R_{coverage}$) of tiles that were predicted positive (cancer probability >0.5) among the annotated-positive tiles was calculated and reported in Fig. 3c,d. For models trained on the high-magnification mode, 0.823±0.094 (mean ± s.d.) of annotated-positive tiles were predicted as positive, while the models achieved 0.629±0.148 and 0.794±0.022 of $R_{coverage}$ for medium- and low-magnification modes. Considering that areas outside human-annotated tumor regions may also exhibit non-obvious cancer-related patterns, other commonly used evaluation metrics (e.g., Intersection over Union (IoU)) are not suitable for local prediction evaluation in this study.

In addition, $R_{positive}$, which represents the ratio of predicted positive tiles in each group, was calculated for "cancer", "uncertain", and "normal" slides individually (Fig. 3c,d). For "normal" slides from healthy subjects, whose tissue samples were not expected to show evidence of cancer, the $R_{positive}$

ratios are 0.049±0.033, 0.043±0.039, and 0.036±0.030 under the high-, medium-, and low-magnification settings, respectively. While for the "cancer" slides, the $R_{positive}$ ratios at different magnification levels are 0.249±0.062, 0.231±0.050, and 0.231±0.132, respectively. Interestingly, a good number of tile images in "uncertain" slides were predicted as positive, with $R_{positive}$ to be 0.178±0.036, 0.203±0.059, and 0.259±0.103 for the corresponding magnification levels, indicating that "uncertain" slides may show evidence of cancer according to the MM-MIL models. However, from the perspective of human experts, the positive tiles in "uncertain" slides do not exhibit well-known cancer-associated patterns based on current knowledge, thus requiring further analysis and interpretation.

Overall, MM-MIL models trained on the high-magnification level achieved the highest AUC scores for the global prediction tasks, as well as highest $R_{coverage}$ for the local prediction tasks. In the following sections, we will focus on the high magnification level which has the most detailed local information about the tissue sample.

### 3. Visualization of latent feature space for model introspection

To gain insight into model predictions, the latent feature representations of tiles were visualized in two dimensions using t-distributed stochastic neighbor embedding (t-SNE) (Fig. 4,5). Tiles with similar feature representations according to the model were grouped close to each other. Noticeably, a large proportion (around 70%) of positive tiles from "uncertain" slides formed separate clusters with positive tiles from "cancer" slides in the latent feature space (Fig. 4c,d), indicating that "uncertain" slides may exhibit cancer-related patterns that were undetected in "cancer" slides. Such cancer-related patterns were not observed in "normal" slides (Fig. 4b), which implies that, based on SLAM microscopy, unique cancer-related patterns may be revealed from tumor-adjacent and peri-tumoral regions.

Through the visualization of corresponding tile images in the t-SNE plot (Fig 4e), the normal patterns presented in negative tiles (cancer probability ≤ 0.5) include collagen (green, from second harmonic generation - SHG) and adipocytes (cyan, from three-photon fluorescence - 3PF). While in positive tiles, significant optical heterogeneity is observed, implicating the existence of various cancer-related patterns. This is further demonstrated in Fig. 5, where the latent feature space of positive tiles was visualized. The apparent discrepancy of cancer-related patterns from "cancer" and "uncertain" slides can also be observed in Fig. 5c, where tiles from both "cancer" and "uncertain" slides appear on the left half of the t-SNE plot, while positive tiles on the right mainly are from "uncertain" slides. Tile images that correlate to dots in the t-SNE plot (Fig. 5d) further show the diversity of optical characteristics of cancer-related tiles revealed by MM-MIL. Those cancer-related patterns were further investigated and discussed in the following sections.

From these observations, we can infer that MM-MIL is able to identify cancer-related patterns from both "uncertain" and "cancer" slides to differentiate positive bags from the negative ones during the training process. This property makes MM-MIL useful for optical signature discovery in uncertain areas, where cancer-related patterns may not exist in "cancer" slides.

## 4. Channel-wise pixel-level model interpretation

To further investigate the cancer-related patterns revealed by MM-MIL, two model interpretation techniques (i.e., channel occlusion and attribution map) were leveraged, which was aimed to identify the informative SLAM channels for cancer prediction and the salient structures in cancer-related tiles.

In this study, an occlusion-based method was utilized to inform the importance of each channel for cancer prediction. This is achieved by measuring the change of cancer prediction score when the targeted channel is replaced by uninformative signals. It was observed that the highly informative channels (the ones inducing significant change in prediction score after occlusion) differ among positive tiles (Fig. 6), which indicates that the importance of each channel for cancer prediction may vary within positive tiles. Among them, for the positive tiles in "uncertain" slides (Fig. 6g-l), the 3PF channel has predominant contribution to the predictions, while 2PF, THG, and SHG channels have high contribution for cancer tile predictions in "cancer" slides (Fig. 6a-f). Although the channel occlusion method provides indications of the channel contribution to cancer prediction, it is not used to compare the importance of channels since the correlation between channels is not considered. To gain insights into model predictions on the pixel level, channel-wise attribution maps were generated using gradient-based saliency map techniques, which inform the prominent structures in each channel for cancer prediction (shown in Fig. 6).

## 5. Cancer-related optical signature discovery via model interpretation

Based on the forementioned localization predictions and model interpretation methods, several cancer-related patterns were revealed by MM-MIL, including well-known cancer biomarkers and non-obvious cancer-related patterns.

### 5.1 Well-known cancer biomarkers

Dense clusters of tumor cells are visible in the positive tiles from "cancer" slides (Fig. 5d, rows (d) and (e)). Those breast cancer tumor cells are reported to have high THG signal intensity (magenta colored) in SLAM images[36], which match with our findings that the THG channel has a significant contribution to cancer prediction as shown in Fig. 6d-f.

Additionally, collagen fibers (visible in SHG channel) are highlighted in positive tiles (Fig. 6d-f). In breast cancer, stromal collagen fibers participate in the migration of metastatic tumor cells which may promote tumorigenesis[37]. Based on the morphology and their interactions with tumor cells, collagen fiber patterns in breast cancer can be categorized into different types of tumor-associated collagen signatures (TACS)[38,39]. Using the attribution maps and channel occlusion results, collagen fibers that co-occur with tumor cells were determined as prominent structures for cancer diagnosis by MM-MIL (Fig. 6d-f). This specific pattern matches with the characteristics of TACS-6, which is defined by the disordered alignment of collagen fibers that enables multidirectional tumor cell migration[8]. According to pathological reports, such collagen fiber patterns appeared in SLAM images from subjects who were diagnosed with invasive ductal carcinoma. This confirms the previous observation that TACS-6 appears at the invasive stage of tumor development[8]. Constrained by the FOV of tiles at all magnification levels (i.e., 128 µm × 128 µm for high, 256 µm × 256 µm for medium, and 512 µm × 512 µm for low-

magnification level), other large-scale TACS signatures (mainly defined on the millimeter scale) could not be observed in MM-MIL predictions.

### 5.2 Non-obvious cancer-related patterns

The non-obvious cancer-related patterns, which mainly come from "uncertain" slides, are visualized in Fig. 5d(f)-(j). Among those patterns, we found that the cyan point-like objects in the 3PF channel were repeatedly highlighted by attribution maps (Fig. 6g-l). Occluding the 3PF channel leads to a significant drop in the prediction score (Fig. 6g-l). Such pattern can also be observed as clusters among positive tiles in the latent feature space (Fig. 4i).

Previous studies showed that these cyan dots are NAD(P)H-rich extracellular vesicles (EVs), which appear as diffraction-limited punctuated pixels in the SLAM 3PF channel[40-42]. EVs play an important role in intercellular communication between cancer cells and the tumor microenvironment[41]. They contribute to cancer growth and metastasis with multiple functionalities, including the suppression of immune response, the recruitment of stromal cells, and the determination of organotrophic metastasis[43,44]. The high concentration of NAD(P)H has also been reported in cancer cells[45,46], which is considered to be linked to the antioxidant defense mechanism[47] and Warburg effect[48]. According to some studies, EVs from breast cancer cells have significantly higher NAD(P)H concentration compared with EVs from nontumorigenic cells[39]. It is also reported that EVs outside the visible tumor area (as far as 5 cm away from the tumor boundary) show even higher NAD(P)H concentration compared to EVs within the tumor, demonstrating the far-reaching impact of EVs in carcinogenesis[49,50]. Those reported results support the findings by MM-MIL that NAD(P)H-rich EVs (or cyan dots in the 3PF channel) are associated with breast cancer, and that such pattern can be more easily observed in peri-tumoral regions (mainly represented by "uncertain" slides) than tumor areas (shown in the "cancer" slides).

In addition to forementioned patterns, we also observed some suspicious cancer-related patterns that are difficult to attribute to known biological processes or mechanisms, including the FAD-rich punctuated patterns surrounding adipocytes (Fig. 6a-c). Further studies are needed to verify the correlation between these unveiled optical signatures and human breast cancer.

## 6. Comparison between different bagging policies

To substantiate the claim that the Mix-and-Match bagging policy is capable of learning from inexact and incomplete supervision, experiments were conducted to compare its performance with different bagging strategies (i.e., slide-level and subject-level). For the slide-level bagging policy, similar to the majority of MIL studies in CPATH, each slide was treated as a bag. Slides from cancer subjects were labeled as positive. While for subject-level bagging policy, slides from individual subjects were grouped into a bag with labels being determined by the subject info (positive for cancer subjects, and negative for normal subjects). It is important to emphasize that due to the lack of obvious cancer biomarkers, some tissue from a cancer subject might have been imaged only in tumor surrounding regions.

When being trained under the same conditions and tested on the same held-out test set with certain slide labels, the MIL models achieved an average AUC of 0.939, 0.808, and 0.781 for Mix-and-Match, slide-level, and subject-level bagging strategies respectively (Supplementary Fig. 3). In addition, the large variance in AUC was observed for slide-level and subject-level bagging strategies, with the standard derivation to be 0.020, 0.211, and 0.257 for the three bagging strategies respectively.

Considering the large number of “uncertain” slides in the dataset, slide-level and patient-level bagging policies cannot guarantee the accuracy of bag labels, which leads to the poor testing results. In contrast, the Mix-and-Match bagging policy tackle the challenge by grouping certain and uncertain examples to new bags, which ensures the accuracy of bag labels and enables the robust training of MIL models.

## Discussion

In this study, we developed a weakly supervised deep learning framework (MM-MIL) for the discovery of human breast cancer-related optical signatures based on label-free virtual histopathology (SLAM microscopy) which contains rich structural and functional information in tissue samples. The task of this optical signature discovery has three main characteristics: (1) inexact supervision, (2) incomplete supervision, and (3) positive instances in labeled and unlabeled data that may come from separate inherent distributions. MM-MIL combines the inexact and incomplete supervision into one MIL task via the Mix-and-Match bagging policy, which ensures the completeness of bag-level labels. The established framework detected cancer-related patterns in both primary tumor areas and surrounding peri-tumoral regions. Based on model interpretation, a variety of human breast cancer-related patterns were revealed from SLAM virtual histopathology slides, including well-known cancer signatures as well as non-obvious patterns. Among them, the presence of NAD(P)H-rich EVs support our previous suggestion that breast-cancer-associated EVs could have extensive impact on carcinogenesis, which would be associated with field cancerization[40,42]. The optical signatures found by MM-MIL inspire new hypotheses for cancer biomarker discovery and translational clinical applications, and merit further investigation.

MM-MIL is capable of learning cancer-related patterns from both labeled and unlabeled data. One potential concern with the Mix-and-Match bagging policy is that models may take shortcuts by making positive predictions only on “cancer” slides, which may still accomplish the MIL task, but be less informative for the optical signature discovery task. However, it is observed that MM-MIL tracked cancer-related patterns in both “cancer” and “uncertain” slides (Fig. 3e, Fig. 5). This can be explained by the fact that during the early stage of the training process, the selection of discriminative tiles for model weights updates is mainly random. Since slide-level labels are hidden during training, it is unlikely for the model to pick tiles only from “cancer” slides under the circumstances that “uncertain” slides also contain cancer-related patterns. In general, the goal of MM-MIL is to reveal patterns that are not observable from “normal” slides, regardless of the slide-level labels (i.e., “cancer” or “uncertain” slides).

Another unique feature of our framework is the integration of multiple model interpretation techniques, which assist us to verify model reliability and inspire the discovery of new cancer-related optical signatures. Given the uniqueness of the multi-modal images (i.e., each channel of SLAM provides distinct structural and functional information about the tissue, and all the channels are captured simultaneously, thereby yielding both spatial and temporal co-registration), knowing the contribution of each channel and the corresponding structures is beneficial for the understanding of model behavior. Thus, two channel-wise model interpretation methods (i.e., channel occlusion and channel-wise attribution maps) were leveraged. Using these techniques, several well-known cancer biomarkers as well as non-obvious cancer-related patterns were revealed in both the primary tumor areas and the surrounding peri-tumoral regions, which align with the before-mentioned assumption of separate distributions of positive instances. These observations further substantiate the claim that MM-MIL is capable of extracting unique cancer-related patterns from both “cancer” and “uncertain” slides. Such property distinguishes MM-MIL from SSL-based approaches[51-53] which assume that both labeled and unlabeled data are generated from the same inherent distribution.

Additionally, MM-MIL is also applicable for inexact or incomplete supervision scenario separately. When labels are coarse-grained but complete, MM-MIL behaves similarly as conventional MIL frameworks since individual whole slide images are treated as one bag. When fine-grained but incomplete labels are available (e.g., whole slide images with annotated tumor regions and uncertain areas), MM-MIL can look for unique positive patterns in labeled and unlabeled regions by grouping tiles in truly-positive regions with others in uncertain areas. Similarly, the distinction between MM-MIL and SSL-based methods still holds given the difference in the distribution assumption of unlabeled data.

The MM-MIL framework has some limitations. In this study, MM-MIL focuses on local patterns that exhibit direct association with cancer. The bag-level predictions are generated using the non-trainable Max-Pooling function, which avoid potential overfitting of bag-level classification when the number of MIL bags is limited. However, such method often misses the correlation between tiles (e.g., the co-occurrence of several normal-looking patterns may be linked with cancer). To enable the learning of instance-wise correlations, trainable bag aggregation methods based on instance prediction scores[54] or feature representations[55] can be embedded to the MM-MIL framework. In addition, the cancer-related patterns learned by MM-MIL are constrained by the size of the FOV of the tile images. In this study, we focus on patterns at the scale of 128 µm, 256 µm, and 512 µm. To learn larger-scale patterns (e.g., large-scale tumor-associated collagen signatures), larger tile FOVs would be recommended. Further improvements can be made to enable optical signatures discovery for multi-class scenarios (e.g., cancer prognosis) with enhanced capability in extracting correlations of patterns across regions and scales. In addition, further validation of our framework is needed to verify its generalizability on data from different microscopy systems.

Our weakly supervised deep learning framework offers a practical solution to gain insights into multiplexed optical bioimages. With the advancements in optical imaging techniques, many crucial biological questions can be answered based on the rich structural and functional information revealed from samples. It is widely believed that Machine Learning is proficient at extracting knowledge from large volume of data. Nevertheless, strong supervision is often difficult to obtain for optical signature discovery tasks. This study demonstrates the capability of the proposed framework in recognizing patterns and extracting correlations between optical features with human breast cancer based on inexact and incomplete supervision, which could be extended to various types of biological conditions and imaging modalities.

# Methods

## Dataset description

- **SLAM virtual histopathology dataset**

SLAM virtual histopathology slides were collected from fresh human breast tissue specimens using a custom-built benchtop SLAM microscope[27], and under a protocol approved by the Institutional Review Boards at Carle Foundation Hospital and the University of Illinois at Urbana-Champaign. Tissue from 48 human subjects was used in this study (Fig. 1c), among which 36 subjects were diagnosed with ductal carcinoma *in situ* (DCIS) or invasive breast cancer (invasive ductal carcinoma, invasive lobular carcinoma, lobular carcinoma *in situ*, invasive papillary carcinoma) by board-certified pathologists. Normal tissue samples were collected from 12 subjects undergoing breast reduction surgery, with no reported history of cancer. Information on subject demographics can be found in Supplementary Table 1. With a special focus on potential optical signatures for field cancerization, both tumor and tumor adjacent regions in cancer tissue samples were imaged using the SLAM microscope. For images collected in tumor adjacent regions, the distances between the imaging site and the tumor boundary (annotated by pathologists) ranged from several millimeters to several centimeters. SLAM slide images were collected from 4 channels (i.e., 2PF, 3PF, SHG, THG), each of which provides unique structural and/or metabolic information of the tissue samples. The medium size of the SLAM slides was 1500 × 1500 pixels (750 µm × 750 µm).

- **Dataset partition**

The dataset was randomly split into a training-validation set (about 70% of slides) and a test set (about 30% of slides) using subject-level partition such that slides from one subject can only exist in one of these sets. During the training process, models were trained and validated using 5-fold Monte Carlo cross-validation. The performance evaluation of trained models was performed on the held-out test set. Since different number of slides was available from each subject (Supplementary Table 1), the number of slides and subjects in each set may vary among cross-validated folds. Thus, the average number of subjects and slides in each set were reported in Fig. 1c.

- **Slide annotation**

Slide-level annotations were generated by human experts based on pathological reports, optical characteristics in SLAM, and the corresponding hematoxylin-eosin (H&E) histological slides (Supplementary Fig. 2). Each annotator assigned labels ("cancer", "normal" or "uncertain") to all SLAM slides independently (Fig. 2a). "Cancer" slides mainly came from tumor regions with well-known cancer-associated signatures (e.g., tumor cells, tumor-associated collagen fibers). SLAM slides from cancer subject were labeled as "uncertain" if they were imaged outside the primary tumor area and do not show known cancer biomarkers. SLAM slides from normal subjects (with no history of cancer) were labeled as "normal". The final slide-level labels were generated by integrating annotations from every annotator. A SLAM slide was labeled as "cancer" only if all annotators labeled it as "cancer", otherwise it was labeled as an "uncertain" slide. Among the 724 SLAM slides, 89 slides (12.3%) were labeled as "cancer", 435 slides (60.1%) were labeled as "uncertain", and 200 slides (27.6%) were labeled as

"normal". In addition, certain tumor regions in the test "cancer" slides were annotated on the pixel-level by human experts. Those pixel-level tumor annotations were used for the local prediction evaluation during model testing.

**Dataset preprocessing**

Each channel of SLAM slides was normalized using Z-score standardization, with the mean and variance calculated for each individual channel. SLAM slides were augmented by flips and rotations: horizontal flip, vertical flip, 90-degree rotation, 180-degree rotation, 270-degree rotation.

With the goal of learning features at different scales, slides were cropped into tile images at high-, medium-, and low-magnification levels. For the high-magnification level, slides were cropped to 256 × 256-pixel tiles (128 µm × 128 µm) with no overlap between tiles. Medium- and low-magnification levels tiles were generated by tiling slides into 512 × 512-pixel images (256 µm × 256 µm), and 1024 × 1024-pixel images (512 µm × 512 µm), respectively, and downscaling to 256 × 256-pixel tiles was performed using nearest-neighbor interpolation. These tiles cover a larger FOV, but lose details compared to high-magnification tiles. Overlap ratios of 50% and 80% were used for medium- and low-magnification levels, respectively. For each magnification level, the total numbers of tiles in the training, validation, and test sets are reported in Supplementary Table 2.

**MM-MIL for optical signature discovery**

To take advantage of the inexact and incomplete supervision, we utilized the Mix-and-Match Multiple Instance Learning (MM-MIL) framework, which is capable of learning useful information for cancer diagnosis based on ambiguous coarse-grained annotations. The MM-MIL framework consisted of two main stages: (1) global and local cancer prediction, and (2) optical signature identification. In the first stage, the training procedure involved the MM-MIL bag formation, cancer-related tile localization, and bag-level prediction. For the second stage, three model interpretation methods were leveraged to verify model reliability and to inspire the discovery of cancer-related optical signatures.

- **Mix-and-Match bagging policy**

To ensure the completeness of bag-level labels, a custom-designed bagging policy (Fig. 2c) was adopted, which groups SLAM slides into bags with certain labels (positive or negative). For a particular set (i.e., training, validation, or test), the subsets of slides with "cancer", "normal", and "uncertain" labels can be denoted as $S_+$, $S_-$, and $S_\pm$, respectively. A positive bag $B_+^{(i)} = \{X_1, \dots, X_{N_i}\}$, is defined as a collection of SLAM slides that are randomly selected from $\{S_+, S_\pm\}$, which satisfies

$$\exists X_j \in B_+^{(i)} \ni X_j \in S_+.$$

Namely, at least one slide in $B_+$ is labeled as "cancer". Thus $B_+$ can be assigned an explicit label (positive) based on the standard assumption of MIL[35]. Since "normal" slides came from normal subjects whose tissue samples were not expected to show evidence of cancer, each of them can form an individual bag with a certain negative label:

$$B_{-}^{(i)} = \{X_j\}, X_j \in S_{-}.$$

To maximize the number of bags, every $B_+$ contained only one "cancer" slide in this study. While "uncertain" slides were randomly distributed to positive bags. A description of bags is shown in Supplementary Table 2. With the aim of enforcing the model to look for cancer-related patterns from both "cancer" and "uncertain" slides, only bag-level labels (i.e., "positive" or "negative") were used for training, while slide-level labels were hidden.

- **Cancer-related tile classification**

Once grouped into bags, each SLAM slide $X_j$ was sliced into $m_j$ tiles $\{x_k: k = 1, 2, \ldots, m_j\}$ (known as instances). To identify cancer-related tiles, a DNN-based model $f_{ins}$ maps $x_k$ to its cancer prediction $\hat{y}_k$. This model was trained in a two-step manner[22,56], which involved an inference step and a training step. Each training epoch started with the inference step, where the model generated prediction scores $\hat{Y}_{train}$ for all the tiles in every training bag. The training sample pool $T_{train} = \{T_{top}, T_{bottom}\}$ was then constructed by combining the top $K_t$ instances (denoted as $T_{top}$) with the highest prediction scores in each bag and the bottom $K_b$ instances (denoted as $T_{bottom}$) with the lowest prediction scores:

$$T_{top} = \{x \in B^{(i)}: |\{x' \in B^{(i)}: f_{ins}(x) < f_{ins}(x')\}| < K_t, i = 1,2, \ldots, N\},$$

$$T_{bottom} = \{x \in B^{(i)}: |\{x' \in B^{(i)}: f_{ins}(x) > f_{ins}(x')\}| < K_b, i = 1,2, \ldots, N\},$$

where $K_t$ and $K_b$ are user-defined parameters, and $N$ is the number of bags. The top $K_t$ instances shared the labels with their bags, while all bottom $K_b$ instances were labeled as negative samples. The instance classifier was then trained on the training sample pool $T$ for one epoch. The training procedure is further demonstrated in Algorithm 1 using the aforementioned notations. In this study, we chose a ResNet34[57] as the backbone of the model. The model was initialized using weights pretrained on ImageNet[58] except for the first convolution layer and the last fully connected layer, of which the input and output dimensions were customized for the task, were randomly initialized.

Algorithm 1: Optimization of the instance-level classifier.

**input:** $B_{train}$, $Y_{train}$, $B_{val}$, $Y_{val}$, $f_{ins}$, $K_t$, $K_b$, $N_{epoch}$
**output:** trained model ($f_{ins}$), validation results ($records$)
$epoch$ = 1
$records$ = {}
**while** $epoch \leq N_{epoch}$ **do**
    /* inference step */
    $\hat{Y}_{train}$= model_inference ($B_{train}$, $f_{ins}$)
    $T_{top}$ = select_top_instances($B_{train}$, $Y_{train}$, $\hat{Y}_{train}$, $K_t$)
    $T_{bottom}$ = select_bottom_instances($B_{train}$, $\hat{Y}_{train}$, $K_b$)
    $T_{train}$ = $\{T_{top}, T_{bottom}\}$
    /* training step */
    $f_{ins}$, $\mathcal{L}_{train}$ = train_model($f_{ins}$, $T_{train}$)
    /* validation */
    $\hat{Y}_{val}$ = model_inference($B_{val}$, $f_{ins}$)
    $metrics$ = prediction_evaluation($Y_{val}$, $\hat{Y}_{val}$)
    $records$ = $\{records, (epoch, \mathcal{L}_{train}, metrics)\}$

```
    /* early stopping */
    flag = determine_early_stopping(records)
    if flag = true then
        return f_ins, records
    end if
    epoch = epoch + 1
end while
return f_ins, records
```

- **Bag aggregation method**

Bag-level predictions were obtained by integrating the prediction scores of corresponding instances. In this study, we chose a nonparametric method $f_{bag}$ (i.e., max-pooling operation) for bag-level aggregation. That is, if at least one instance in a particular bag $B^{(i)}$ is positive (probability>0.5), the whole bag is predicted as positive:

$$f_{bag}\left(B^{(i)}\right) = \begin{cases} 1, & if\ \exists x_k^{(i)} \in B^{(i)}: f_{ins}\left(x_k^{(i)}\right) = 1; \\ 0, & otherwise. \end{cases}$$

Compared to trainable bag aggregation methods[22,54,55], the max-pooling operation is more sensitive to spurious positive instances, thus imposes higher requirements on the reliability of instance-level classifiers.

- **Visualization of latent feature space**

To visualize the latent feature space of the model, we randomly sampled 1,500 tile images from the test set, in addition to all predicted positive tiles (probability>0.5) by a ResNet34 model that was trained on the high-magnification level. The latent feature representations (512-element vectors) of those tiles were extracted from the last layer of the feature extractor (i.e., adaptive average pooling layer of the ResNet34 model). The feature space was visualized in two dimensions using t-distributed stochastic neighbor embedding (t-SNE)[59]. Tiles (converted to RGB images) corresponding to the points in the t-SNE plot were sampled and visualized for model introspection.

- **Occlusion-based channel contribution measurement**

To investigate the relevance of each individual SLAM channel for cancer prediction, we occluded each channel of the positive tiles and measured the decrease of prediction score, which can be seen as an indicator of channel importance regarding to the prediction task. The more the prediction score dropped, the more relevant the channel would be to cancer prediction. To avoid the potential out-of-distribution (OOD) test examples, the channel occlusion was achieved by replacing the targeted channel with Gaussian noise that imitates uninformative signals in that channel. The mean and variance of the Gaussian noise were measured based on manually annotated background areas in 460 predicted positive tiles. When tested on trained MM-MIL models, the predicted cancer probabilities of the Gaussian noise input were close to 0 ($7.92\times10^{-7}\pm3.41\times10^{-8}$), compared to 0.72±0.04 when using an all-zero input (mainly due to OOD). Channel importance measurements were displayed as bar charts in Fig. 6, with bar values representing the decrease of prediction score when the channel was occluded.

- **Attribution maps**

To gain insights into the salient structures on the pixel-level, channel-wise attribution maps were generated using Integrated Gradients (IG)[60]. The baseline input of IG was Gaussian noise (same as the informative signal used in channel occlusion). To reduce the visual noise in the attribution maps, SmoothGrad (SG)[61] was implemented, with the number of randomly generated examples set to be 5, and the standard deviation of SG Gaussian noise to be 0.02. The composite attribution maps were generated by aggregating channel-wise attribution maps (Fig. 6). Each channel of the attribution maps is displayed separately in Supplementary Fig. 4.

**Training details**

Each model was trained for at least 50 epochs and at most 200 epochs. Model performance was evaluated on the validation set every 5 epochs, where bag-level balanced accuracy was recorded. Early stopping would be triggered when the validation accuracy did not increase for 30 consecutive epochs. The model with highest validation accuracy was saved for final performance evaluation on the test set. The objective function of the instance-level classifier is the cross-entropy loss:

$$\mathcal{L}(y_k, \hat{y}_k) = -w_1[y_k \log(\hat{y}_k)] - w_0[(1 - y_k) \log(\hat{y}_k)],$$

where $w_0, w_1$ are negative and positive class weights, respectively, which are set to deal with the unbalanced class frequency within the training sample pool. The optimization of model parameters was achieved by stochastic gradient descent using the Adam optimizer, with a learning rate of $1.5 \times 10^{-4}$, and a weight decay of $1 \times 10^{-5}$. A detailed description of hyperparameters used in the experiments can be found in Supplementary Table 3.

**Model evaluation**

1. Global prediction evaluation

The bag-level predictions, which were obtained by aggregating tile-level predictions, were evaluated as a binary classification task. For each magnification level, 5 models were trained using the same configurations individually. The receiver operating characteristic (ROC) curves and the associated area under the curve (AUC) were reported in Fig. 3a. In addition, we assessed the model performance based on slides with certain labels (i.e., “cancer” and “normal”). The results are reported in Fig. 3b.

2. Local prediction evaluation

The coverage ratio $R_{coverage}$ was used to measure the coverage of the model’s cancer predictions on known tumor regions. The coverage ratio was calculated by

$$R_{coverage} = \frac{m_{PP}}{m_{TP}} \times 100\%,$$

where $m_{TP}$ is the number of tumor tiles, of which at least 50% of the tile area is annotated as tumor region., and $m_{PP}$ is the number of predicted positive tiles among the $m_{TP}$ tumor tiles. In addition, the positive tile ratios in each group of slides (i.e., “cancer”, “uncertain”, and “normal”) were calculated by

$$R_{positive} = \frac{m_{positive}}{m_{total}} \times 100\%,$$

where the total number of tiles from a particular group is denoted as $m_{total}$, and $m_{positive}$ is the number of positive tiles in that group. The local prediction evaluation results are reported in Fig. 3c,d.

**Hardware and software**

All data processing and model development were conducted on a workstation computer, equipped with an Intel Xeon W-2195 central processing unit (CPU), four Nvidia RTX 8000 graphics processing units (GPUs), 256 gigabytes of memory, and 6 terabytes of solid-state-drive (SSD) local storage. The workstation operates on the Ubuntu system (version 18.04). During the model optimization, each batch (i.e., 600 tiles per batch) required 0.365 s for the inference step, and 0.694 s for the training step with four GPUs. At the prediction time, MM-MIL generates diagnoses at an average speed of 23.39 slides/s in high-magnification mode, and 26.17 slides/s in low-magnification mode on 2000×2000-pixel SLAM slides. To be noted, the diagnosis process includes tile-level prediction and slide-level integration.

MM-MIL was implemented in Python 3 based on the Pytorch framework (version 1.4.0). SLAM slides, which were stored in TIFF format, were accessed via ImageIO (version 2.9.0). Pillow (version 7.2.0) and OpenCV-Python (version 4.4.0) were used for image processing. Scikit-learn (version 0.23.2) was used for dimensionality reduction and the calculation of evaluation metrics. Functions in Captum (version 0.3.1) were utilized to generate saliency maps for model interpretation. Plots were generated using Matplotlib (version 3.2.2) and Seaborn (version 0.11.0). Other Python libraries including Numpy (version 1.19.1), Pandas (version 1.1.2), and SciPy (version 1.5.2) were used to assist data analysis.

**Ethics statement**

This study was conducted under a protocol approved by the Institutional Review Boards at the University of Illinois at Urbana-Champaign and Carle Foundation Hospital, Urbana, IL. All pathology reports, SLAM virtual histopathology slides, and H&E slides were de-identified before analysis and did not contain protected information.

**Data availability**

The SLAM human breast cancer datasets that support the findings of this study are available from the corresponding author (S.A.B.) upon reasonable request and through collaborative investigations.

**Code availability**

The source code of the MM-MIL framework and training configuration files are available at https://github.com/Biophotonics-COMI/MM-MIL. The tool for latent feature space visualization used in this study, including a demo pipeline implemented in Jupyter Notebook[62], can be downloaded from https://github.com/Biophotonics-COMI/Visual-LF-Space.

**Acknowledgements**

The authors thank all the human subjects who consented to having their tissue be part of our imaging studies, as well as all the staff at the Carle Research Office for assisting in the coordination and consenting of subjects. We thank Sixian You, Eric J. Chaney, Minh Doan, Chih-Yang Hsu, and Kamyar Farhangi, Anirudh Choudhary for their technical support. This work was supported in part by grants from NIH (R01CA2131479 and R01CA241618). J.S. was additionally supported by GlaxoSmithKline. We thank all the team members from the Biophotonics Imaging Laboratory for their suggestions and support. Additional information can be found at: http://biophotonics.illinois.edu.

**Author contributions**

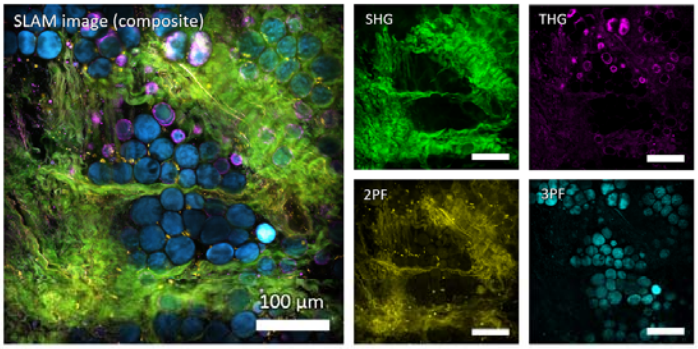
J.S., H.T., and S.A.B. conceived the project and designed experiments; J.S. wrote the code and performed the analysis of the data; A.H., K.C., and N.L. collected the surgical specimens; Z.L. provided histopathological interpretation; J.P. contributed to the raw image data collection and organization; J.S., M.M, and S.A.B. wrote the manuscript with assistance and feedback of all the other co-authors. S.A.B. obtained funding.

**Competing interests**

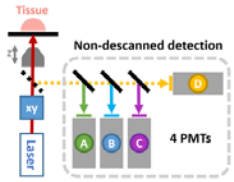
S.A.B. and H.T. are named inventors on patents filed by the University of Illinois at Urbana-Champaign related to the laser source technology used in this study. S.A.B. and H.T. are co-founders and hold equity interest in LiveBx, LLC, Champaign, IL, which is commercializing the laser source technology used in this study. All other authors declare that they have no competing financial interests.

## Figures

a




b




| ID | Detection Modality |
|---|---|
| A | Second Harmonic Generation (SHG) |
| B | 3 Photon Autofluorescence (3PF) |
| C | Third Harmonic Generation (THG) |
| D | 2 Photon Autofluorescence (2PF) |

| Optical marker | THG | 3PF | SHG | 2PF |
|---|---|---|---|---|
| Intracellular target | cytoplasm, mitochondria | NADH | | FAD |
| Extracellular target | interface, tube, vesicle | lipid | collagen, myosin | elastin |

c

| Dataset | Total subjects | Cancer subjects | SLAM slides | "Cancer" slides | "Uncertain" slides | "Normal" slides | Size (Mpixels) |
|---|---|---|---|---|---|---|---|
| Training | 23 | 17 | 428 | 59 | 237 | 132 | 1,077 |
| Validation | 7 | 5 | 92 | 7 | 61 | 24 | 203 |
| Test | 18 | 14 | 204 | 23 | 137 | 44 | 484 |
| **Total** | **48** | **36** | **724** | **89** | **435** | **200** | **1,764** |

**Fig. 1 | Description of SLAM imaging system and the human breast cancer dataset.** The human breast cancer dataset used in this study is generated by a SLAM microscopy imaging system. **a,** A SLAM virtual histopathology slide image (slides for short) from a normal subject. SLAM images were obtained simultaneously from four channels: SHG, THG, 2PF, 3PF, each of which provides unique structural and functional information about the tissue sample. **b,** An overview of SLAM system. (upper left) schematic of SLAM microscope. (upper right table) SLAM consists of four detection modalities that target various intracellular and extracellular optical markers (bottom table). **c,** A description of the human breast cancer dataset. The dataset consists of 724 SLAM slides coming from 48 subjects, 36 of which are cancer subjects. The dataset was randomly split to training, validation, and test sets using subject-level partition. The medium size of the slide images is 1500×1500 pixels, which cover 750 µm×750 µm area of the tissue. Slides were tiled into patch images at different magnification levels. The proposed MM-MIL framework aims at making cancer predictions on both tile- and slide-level while only whole-slide-level incomplete annotations are available.

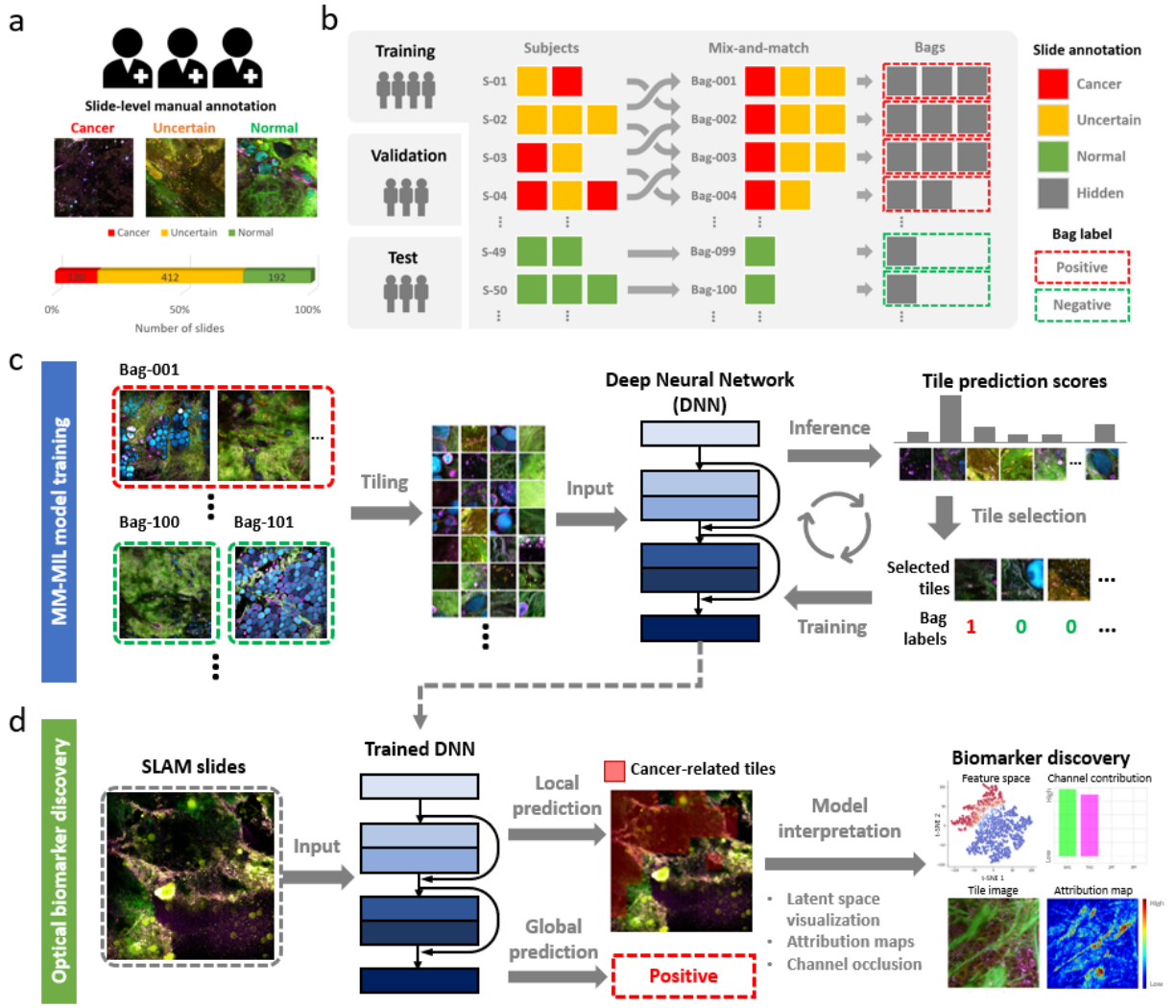
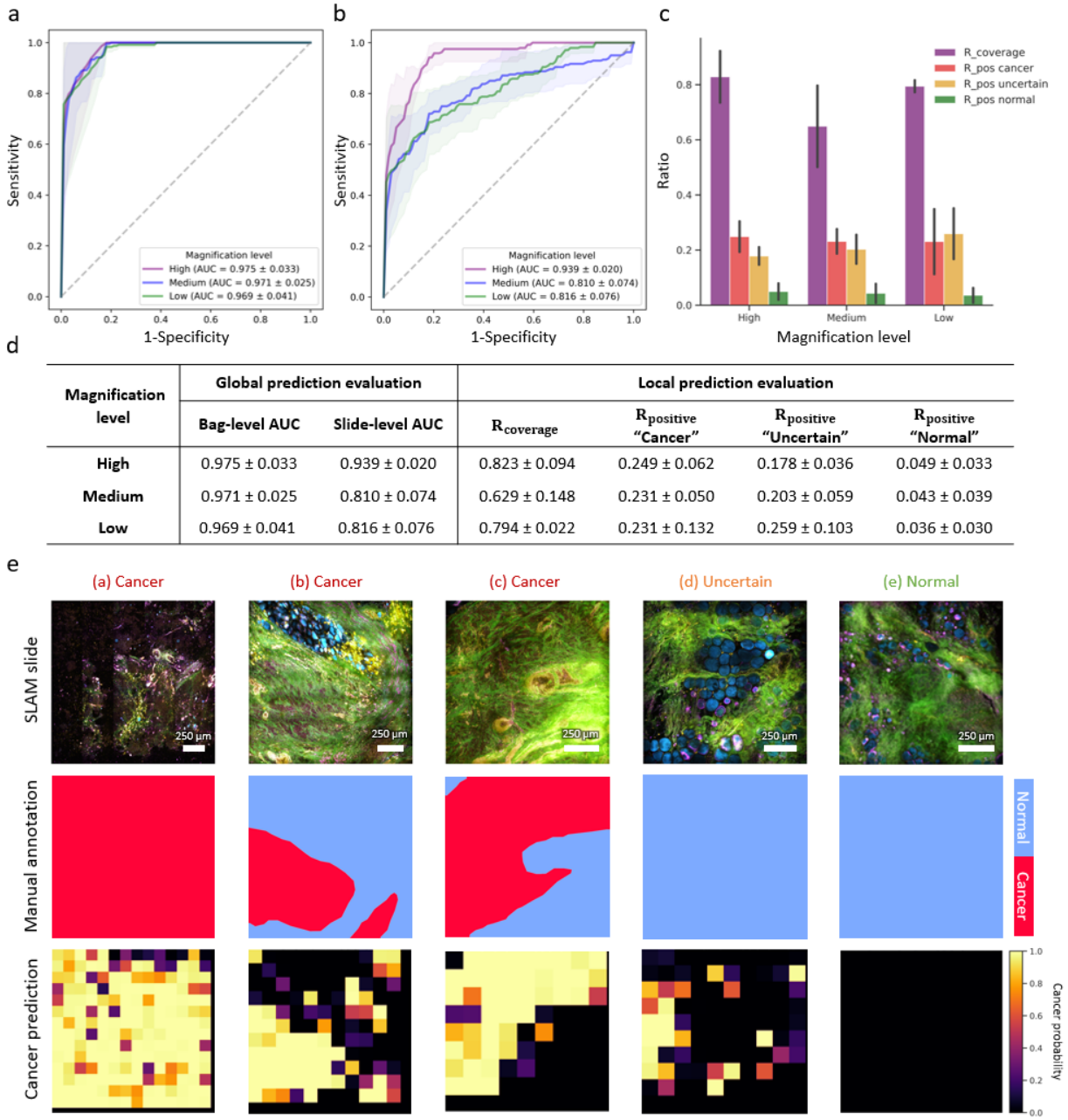



**Fig. 2 |** Overview of the explainable weakly supervised deep learning framework (MM-MIL) in this study. **a,** SLAM virtual histopathology slide images (slides for short) were assigned labels ("cancer", "uncertain", or "normal") by human experts based on the pathological reports, histopathology slides, and SLAM optical characteristics. Among all SLAM slide images, 57% (412 slides) were acquired from outside the tumor, in areas without obvious cancer-associated signatures, and were labeled as "uncertain" slides by human experts. **b,** The Mix-and-Match bagging policy randomly groups "cancer" slides with "uncertain" slides to form positive bags, so that at least one slide in the bag is "cancer". This allows MM-MIL models to extract cancer-related patterns from both types of slides. "Normal" slides, which came from healthy subjects with no history of cancer, were treated as individual negative bags. All slide-level labels were hidden once bags were created, while only bag-level labels ("positive" or "negative") were used for model training. **c,** During the training procedure, MM-MIL models were optimized by iteratively selecting and learning discriminative tile images. **d,** Once the DNN-based model is trained, it produces both local (tile-level) and global (bag/slide-level) cancer predictions. Based on these predictions, multiple model interpretation methods are utilized for the discovery of cancer-related optical signatures.




| Magnification level | Global prediction evaluation | | Local prediction evaluation | | | |
|---|---|---|---|---|---|---|
| | Bag-level AUC | Slide-level AUC | $R_{coverage}$ | $R_{positive}$ "Cancer" | $R_{positive}$ "Uncertain" | $R_{positive}$ "Normal" |
| **High** | 0.975 ± 0.033 | 0.939 ± 0.020 | 0.823 ± 0.094 | 0.249 ± 0.062 | 0.178 ± 0.036 | 0.049 ± 0.033 |
| **Medium** | 0.971 ± 0.025 | 0.810 ± 0.074 | 0.629 ± 0.148 | 0.231 ± 0.050 | 0.203 ± 0.059 | 0.043 ± 0.039 |
| **Low** | 0.969 ± 0.041 | 0.816 ± 0.076 | 0.794 ± 0.022 | 0.231 ± 0.132 | 0.259 ± 0.103 | 0.036 ± 0.030 |

**Fig 3. | Cancer global and local prediction analysis. a,** Receiver operating characteristic (ROC) curves of MM-MIL models evaluated on all test slides based on the Mix-and-Match bagging policy, where positive bags are groups of "cancer" and "uncertain" slides. For each magnification level, five models were trained and evaluated. **b,** ROC curves of MM-MIL models evaluated on slides with certain labels (i.e., "cancer" and "normal"). **c,** The local prediction metrics are reported in a bar chart where error bars indicate the standard deviations. (denoted as "R_coverage" in the plot) measures the coverage ratio of the model's cancer predictions on known tumor regions annotated by pathologists; (denoted as "R_pos" in the plot) is the ratio of predicted positive tiles in all the slides from each group (i.e., "cancer", "uncertain", and "normal" groups). **d,** Global and local evaluation metrics (mean ± s.d.) are recorded in a table. **e,** Cancer localization predictions of slides from "cancer" group (a)-(c),

“uncertain” group (d), and “normal” group (e) produced by a MM-MIL model trained on the high-magnification setting. For each slide in the test set, the pixel-level cancer manual annotation was generated by human experts (second row). The cancer probabilities of each tile in the slide are visualized as heatmaps in the third row.

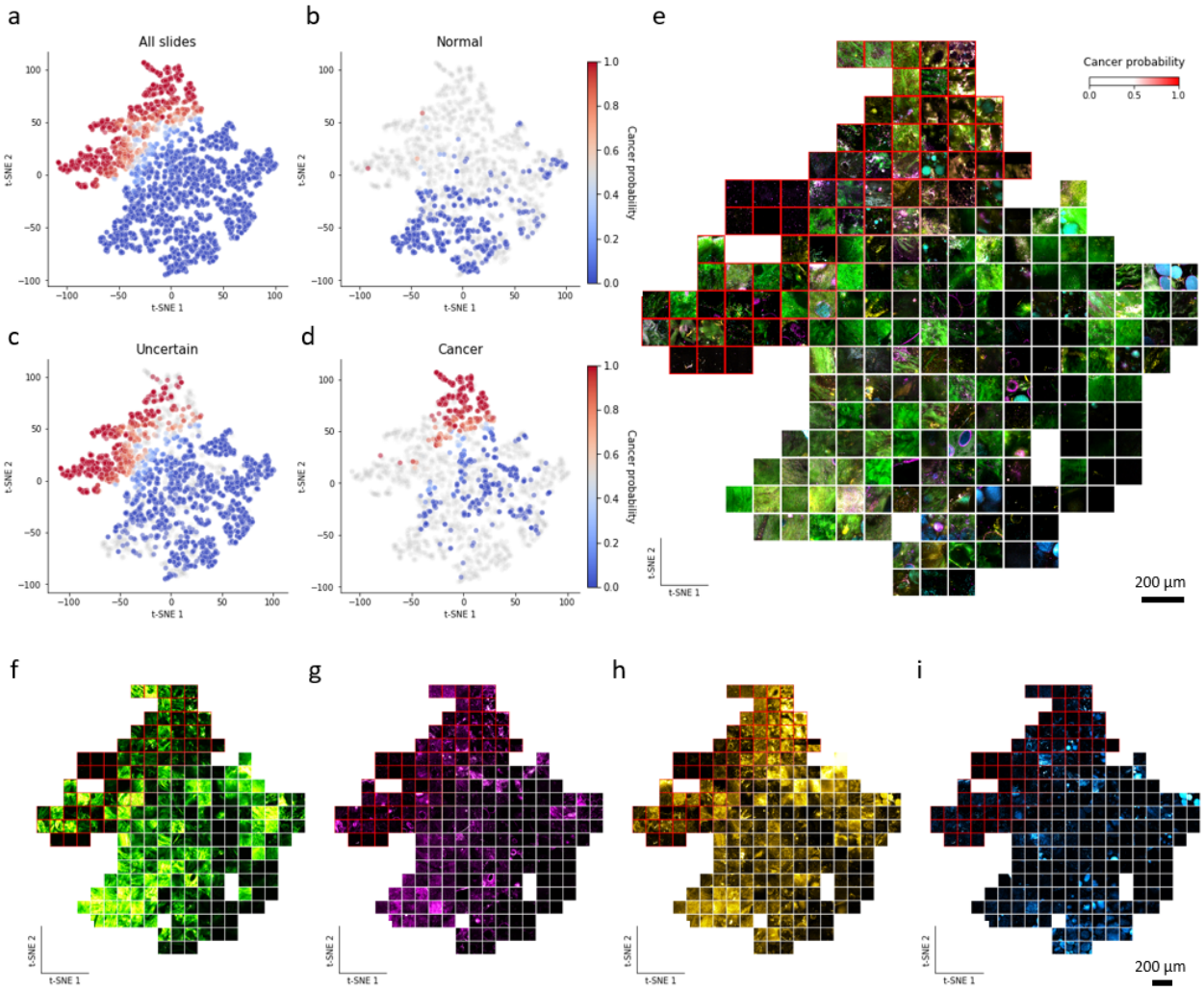



**Fig. 4 |** t-SNE visualization of latent feature space. **a,** t-SNE plot of all slides in the test dataset. Each dot in the plot represents one SLAM tile image. Dots are colored based on the cancer probabilities of the corresponding tiles predicted by the MM-MIL model. **b-d,** t-SNE plots of tiles from "normal", "uncertain", and "cancer" SLAM slides, are visualized separately. Cancer-related patterns in "uncertain" SLAM slides form separate clusters with patterns existing in "cancer" slides, while such patterns are not observed in "normal" slides. **e,** SLAM tiles corresponding to the dots in the t-SNE plot were sampled and visualized in the same locations. Significant optical heterogeneity is observed from predicted positive tile, which indicates that MM-MIL model learned different kinds of cancer-related patterns during training. **f-i,** Each channel of the tile images (f: SHG, g: THG, h: 2PF, i: 3PF) is visualized separately in the t-SNE plots.

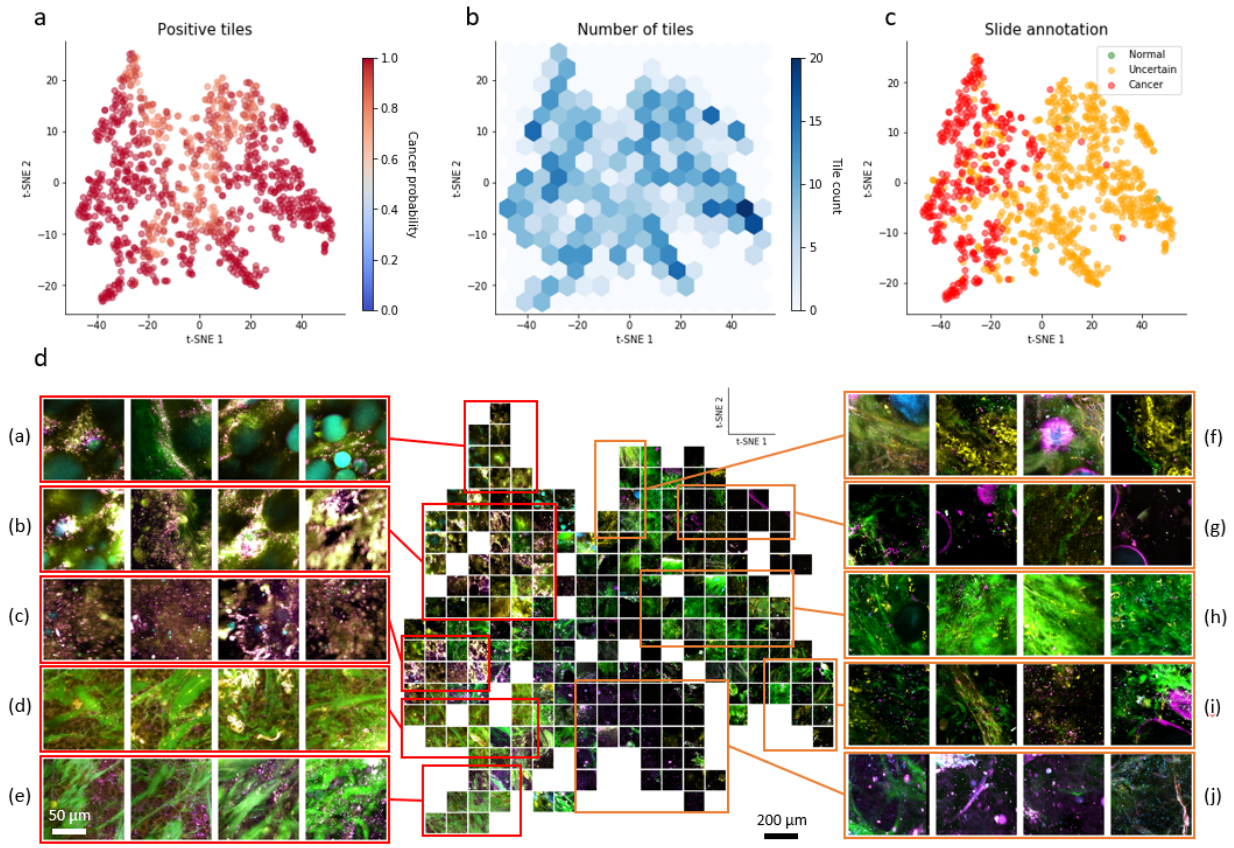



**Fig. 5 |** Visualization of latent feature representations of predicted positive tiles. **a,** t-SNE plot of positive tiles predicted by the model (cancer probability > 0.5). **b,** A hexagonal heatmap showing the number of tiles in each region of the t-SNE plot. **c,** Cancer-related patterns from "cancer" and "uncertain" SLAM slides exhibit both similarities (left part, with tiles from both "cancer" and "uncertain" slides) and differences (right part, with mainly tiles from "uncertain" slides). **d,** Positive SLAM tiles corresponding to the dots in the t-SNE plot were sampled and visualized in the same locations. Positive tiles in the red bounding boxes ((a)-(e)) mainly came from "cancer" slides, while positive tiles inside orange bounding boxes ((f)-(j)) came predominantly from "uncertain" slides.

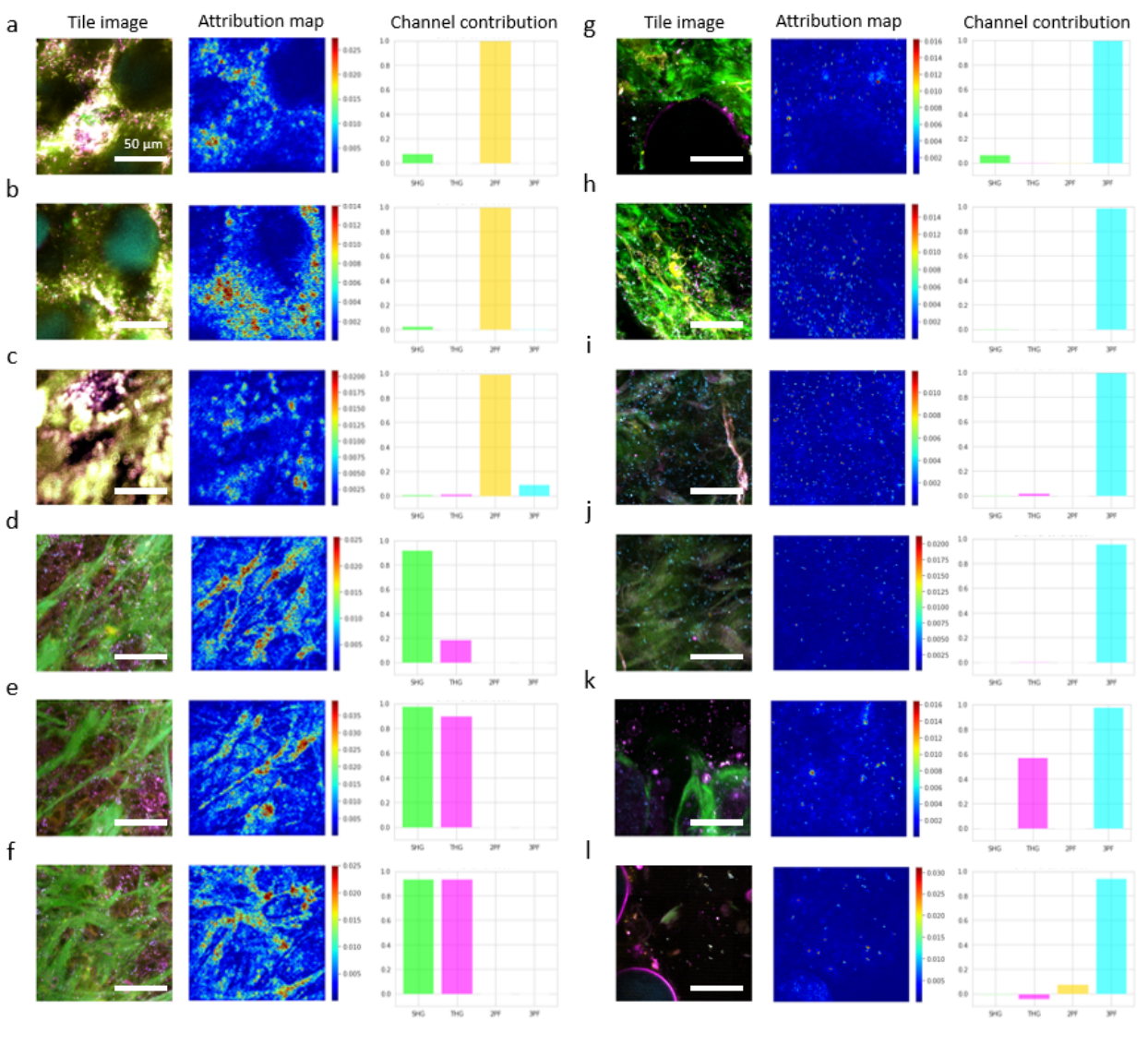
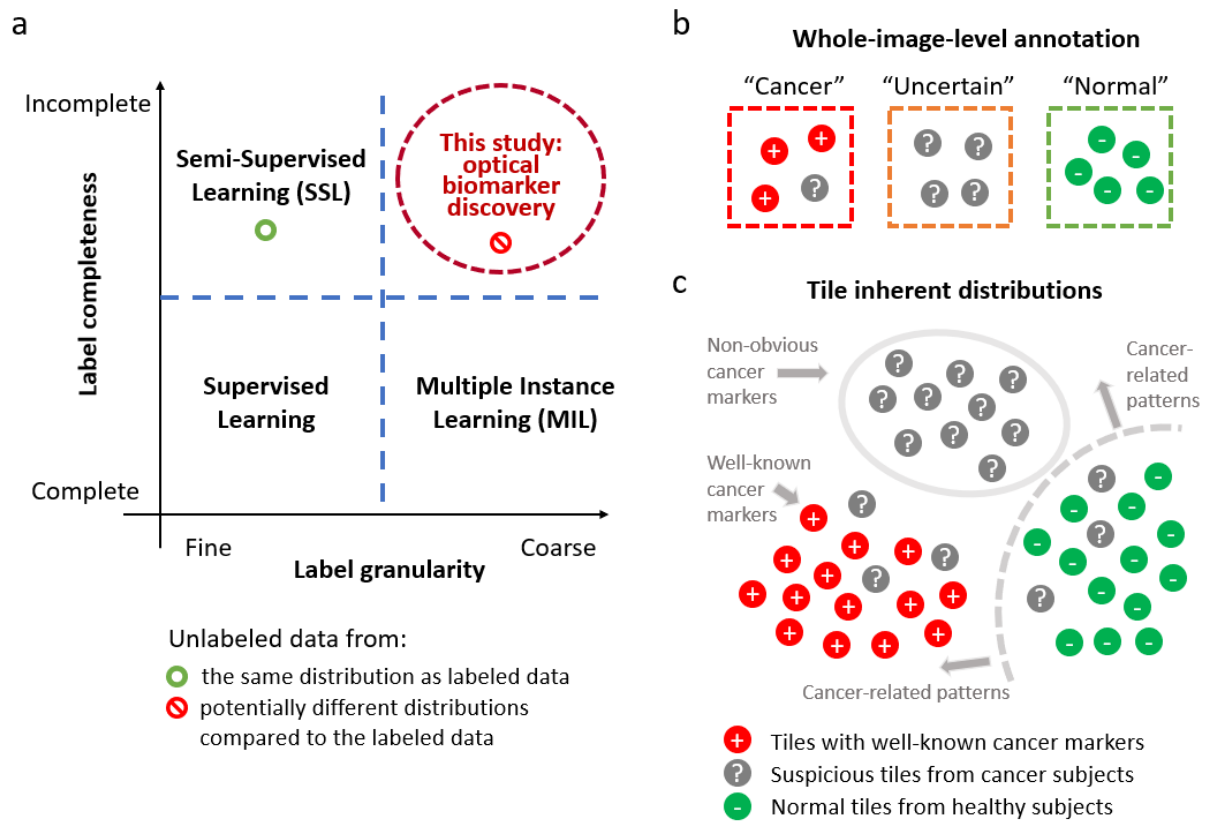
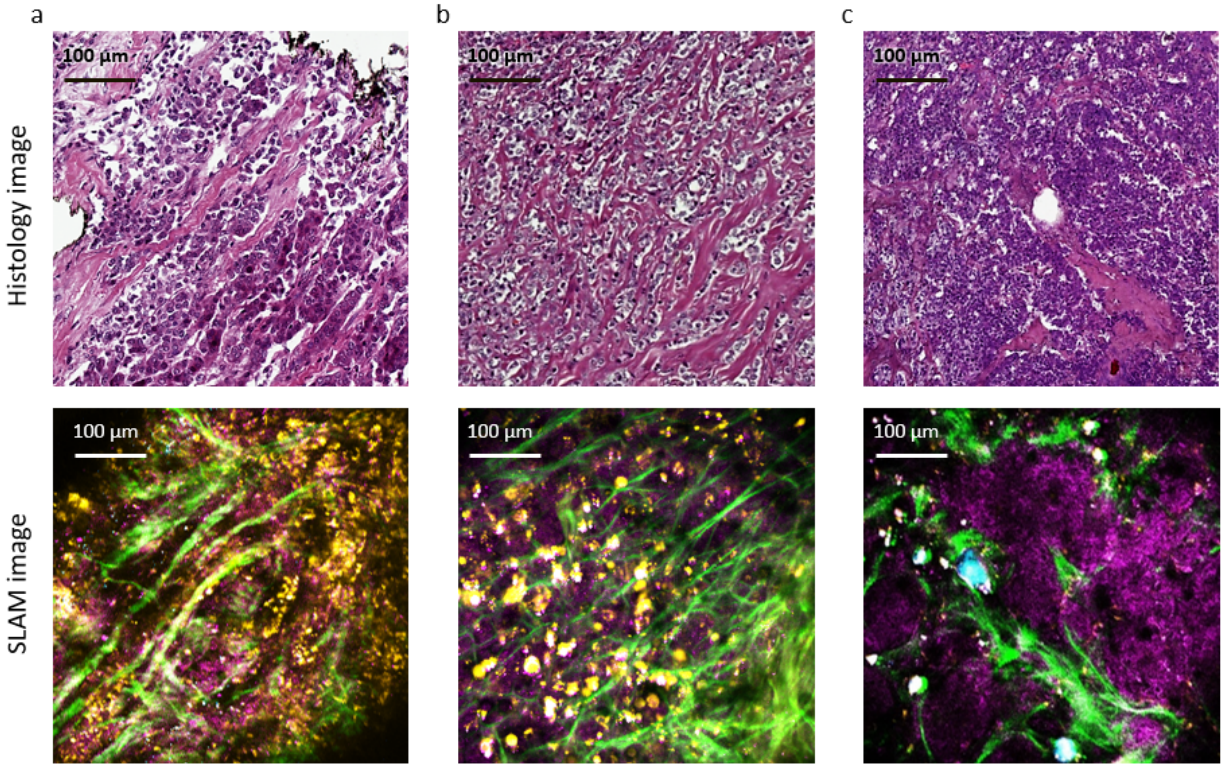



**Fig. 6 |** Informative channels and patterns for cancer prediction revealed via model interpretation. Channel contribution plots show the prediction score drop (y axis) when each SLAM channel is occluded, which can be seen as indicator of the importance of each channel. Attribution maps, generated using Integrated Gradients and SmoothGrad, highlight the salient structures in tile images for cancer prediction. **a-f,** The highly informative patterns found by MM-MIL match with well-known cancer-related signatures, including tumor cells in 2PF channel (**a-c**), tumor cells with high THG intensity (**d-f**), and tumor-associated collagen fibers in the SHG channel (**d-f**). **g-l**, MM-MIL unveiled non-obvious cancer-related patterns from "uncertain" slides. The NAD(P)H-rich extracellular vesicles (cyan dots in the 3PF channel) were determined as a highly informative feature for cancer prediction. All scale bars: 50 µm.

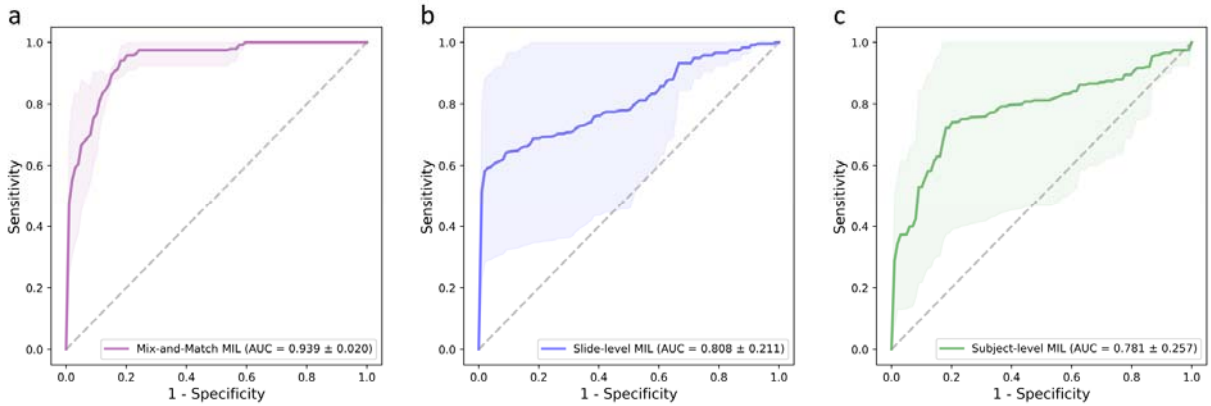
## Supplementary Figures



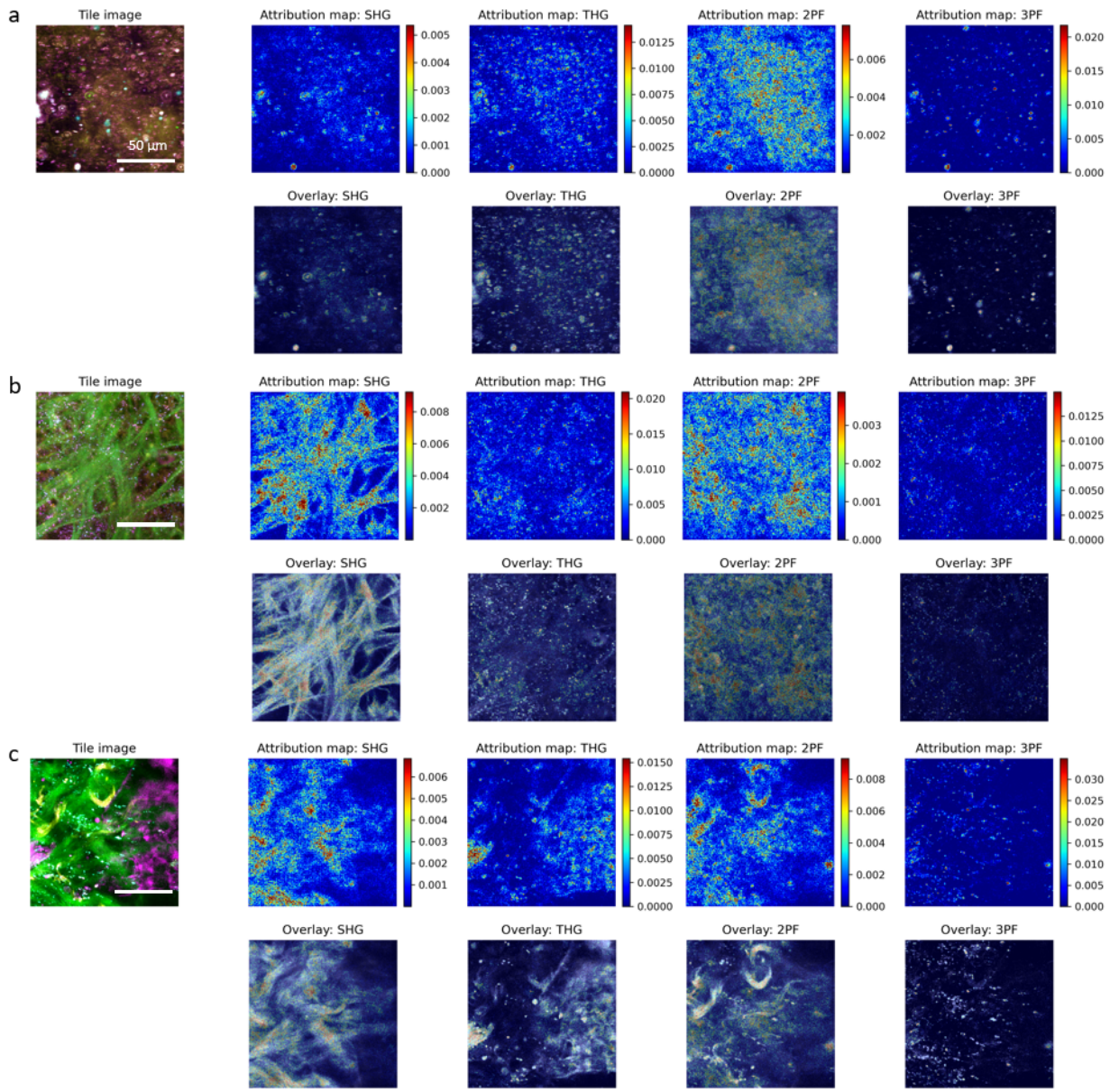

**Supplementary Fig. 1 |** The task of optical signature discovery in the study. **a,** This study focuses on the identification of human breast cancer-related optical signatures by leveraging inexact and incomplete supervision. **b,** Whole-image-level annotations were generated for SLAM virtual histology slide images. **c,** A diagram showing the potential inherent distributions of SLAM tile images. We define the cancer-related patterns as the optical characteristics that appear only in tissues from cancer subjects, while being unobservable in normal tissues from healthy subjects. The potentially new optical biomarkers (non-obvious cancer-related optical signatures) may emerge from different distributions compared to well-known cancer markers.




**Supplementary Fig. 2 |** SLAM virtual histopathology images and the corresponding histology of different types of breast cancer. **a,** Invasive ductal carcinoma (IDC). **b,** Invasive lobular carcinoma (ILC). **c,** Invasive micro-papillary carcinoma (ImPC).




**Supplementary Fig. 3 |** ROC curves of MIL models with different bagging policies evaluated on slides with certain labels ("cancer" and "normal"). Models were trained and evaluated using 5-fold Monte Carlo cross-validation. The mean and standard deviation of AUC are shown in the plots (mean ± s.d.). **a,** MIL models trained on bags generated via the Mix-and-Match bagging policy. **b,** Each SLAM slide image was treated as an individual bag. Slides from cancer subjects were treated as positive bags, otherwise they were treated as negative bags. **c,** Slides from each subject were grouped into a bag (positive for cancer subjects, and negative for normal subjects).




**Supplementary Fig. 4 |** Channel-wise attribution maps highlight salient structures in each channel of the SLAM tile images. **a**, A positive tile image from the primary tumor region. **b**, A positive tile image with tumor-associated collagen fibers. **c**, A positive tile image from an "uncertain" slide. All scale bars: 50 µm.

## Supplementary Tables

**Supplementary Table 1.** Demographic information on human subjects, pathological diagnoses, and SLAM imaging sights. This study included 36 breast cancer subjects (S-001 to S-036) and 12 healthy cancer-free subjects (S-037 to S-048). IDC, invasive ductal carcinoma; DCIS, ductal carcinoma *in situ*; ILC, invasive lobular carcinoma; LCIS, lobular carcinoma *in situ*; IPC, invasive papillary carcinoma. For subject S-001 and S-002, who had IDC before the neoadjuvant therapy prior to surgery, primary tumor was not found in their tissue samples. Their SLAM slides were labeled as "uncertain".

| Subject ID | Age | Pathologic Stage | Type of surgery | Type of cancer | # SLAM slides |
|---|---|---|---|---|---|
| **S-001** | 48 | 0 | Bilateral breast mastectomy | No primary tumor found | 36 |
| **S-002** | N/A | 0 | Bilateral breast mastectomy | No primary tumor found; fibroadenoma | 20 |
| **S-003** | 51 | 0 | Left breast lumpectomy | DCIS | 16 |
| **S-004** | 82 | I | Bilateral breast mastectomy | IPC, DCIS | 28 |
| **S-005** | 65 | I | Right breast mastectomy | IDC, DCIS | 24 |
| **S-006** | 76 | I | Right breast lumpectomy | IDC, DCIS | 24 |
| **S-007** | 47 | I | Bilateral breast mastectomy | IDC, LCIS, DCIS | 8 |
| **S-008** | 68 | I | Right breast lumpectomy | IDC, DCIS | 8 |
| **S-009** | 52 | I | Right breast mastectomy | IDC, LCIS, DCIS | 16 |
| **S-010** | 38 | I | Right breast mastectomy | IDC, DCIS | 4 |
| **S-011** | 71 | I | Right breast mastectomy | IDC, DCIS | 4 |
| **S-012** | 57 | I | Right breast lumpectomy | IDC, DCIS | 8 |
| **S-013** | 64 | I | Right breast lumpectomy | IDC | 16 |
| **S-014** | 52 | I | Left breast mastectomy | IDC, DCIS | 4 |
| **S-015** | 61 | I | Left breast lumpectomy | IDC, LCIS | 12 |
| **S-016** | 35 | I | Right breast mastectomy | IDC, DCIS | 8 |
| **S-017** | 53 | I | Right breast mastectomy | IDC, DCIS | 12 |
| **S-018** | 76 | II | Right breast mastectomy | IPC, DCIS | 16 |
| **S-019** | 56 | II | Bilateral breast mastectomy | IDC, DCIS | 24 |
| **S-020** | 71 | II | Bilateral breast mastectomy | IDC | 20 |
| **S-021** | 82 | II | Right breast mastectomy | IDC, DCIS | 32 |
| **S-022** | 81 | II | Right breast mastectomy | IDC, DCIS | 12 |
| **S-023** | 89 | II | Left breast mastectomy | ILC, IDC, DCIS | 4 |
| **S-024** | 73 | II | Right breast lumpectomy | IDC, DCIS | 4 |
| **S-025** | 71 | II | Right breast lumpectomy | IDC, DCIS | 20 |
| **S-026** | 53 | II | Right breast lumpectomy | IDC, DCIS | 8 |
| **S-027** | 59 | III | Left breast mastectomy | IDC | 4 |
| **S-028** | 46 | III | Right breast mastectomy | IDC, DCIS | 28 |
| **S-029** | 77 | III | Left breast mastectomy | ILC, LCIS, DCIS | 16 |
| **S-030** | 64 | III | Left breast mastectomy | IDC, DCIS | 16 |
| **S-031** | 63 | N/A | N/A | N/A | 8 |

| | | | | | |
|---|---|---|---|---|---|
| **S-032** | 60 | N/A | Left breast mastectomy | ILC, DCIS | 20 |
| **S-033** | 57 | N/A | Prophylactic mastectomy | LCIS, lobular hyperplasia | 12 |
| **S-034** | 43 | Benign | Right breast lumpectomy | Benign | 8 |
| **S-035** | 56 | Benign | Right breast lumpectomy | LCIS | 8 |
| **S-036** | 62 | Benign | Left breast lumpectomy | Benign | 16 |
| **S-037** | 19 | Normal | Breast reduction surgery | — | 8 |
| **S-038** | 70 | Normal | Breast reduction surgery | — | 28 |
| **S-039** | 40 | Normal | Breast reduction surgery | — | 36 |
| **S-040** | 56 | Normal | Breast reduction surgery | — | 28 |
| **S-041** | 50 | Normal | Breast reduction surgery | — | 24 |
| **S-042** | 41 | Normal | Breast reduction surgery | — | 8 |
| **S-043** | 45 | Normal | Breast reduction surgery | — | 20 |
| **S-044** | N/A | Normal | N/A | — | 16 |
| **S-045** | 36 | Normal | Prophylactic bilateral mastectomy | — | 8 |
| **S-046** | 18 | Normal | Breast reduction surgery | — | 8 |
| **S-047** | 22 | Normal | Breast reduction surgery | — | 8 |
| **S-048** | 18 | Normal | Bilateral mastectomy | — | 8 |

**Supplementary Table 2.** Details of bags and tiles under three magnification settings. Considering that the number of SLAM slides varies among subjects, the number of bags and instances in each set may change in each fold of cross validation. Here the average number of bags and instances among cross-validation folds are reported.

| **Tile FOV: 256×256 pixels; overlap ratio: 0%** | | | | |
|---|---|---|---|---|
| **Set** | # Bags | # Positive bags | # Instances | # Positive instances |
| **Training** | 1,098 | 354 | 98,490 | 72,300 |
| **Validation** | 186 | 42 | 18,582 | 12,960 |
| **Test** | 456 | 192 | 44,316 | 32,424 |
| **Tile FOV: 512×512 pixels; overlap ratio: 50%** | | | | |
| **Set** | # Bags | # Positive bags | # Instances | # Positive instances |
| **Training** | 1,098 | 354 | 85,588 | 62,196 |
| **Validation** | 186 | 42 | 15,582 | 10,805 |
| **Test** | 456 | 192 | 37,578 | 27,300 |
| **Tile FOV: 1024×1024 pixels; overlap ratio: 80%** | | | | |
| **Set** | # Bags | # Positive bags | # Instances | # Positive instances |
| **Training** | 1,098 | 354 | 95,136 | 70,043 |
| **Validation** | 186 | 42 | 15,978 | 10,944 |
| **Test** | 456 | 192 | 40,626 | 29,226 |

**Supplementary Table 3.** List of parameters used in our experiments across three magnification levels.

| Parameters | Magnification levels | | |
|---|---|---|---|
| | **High** | **Medium** | **Low** |
| **Tile FOV (pixels)** | 256×256 | 512×512 | 1024×1024 |
| **Tile size (pixels)** | 256×256 | 256×256 | 256×256 |
| **Tile overlap ratio** | 0% | 50% | 80% |
| **Epoch** | 200 | 200 | 200 |
| **Batch size** | 600 | 600 | 600 |
| **Learning rate** | $1.5\times10^{-4}$ | $1.5\times10^{-4}$ | $1.5\times10^{-4}$ |
| **Optimizer** | Adam $\beta_1, \beta_2 = (0.9, 0.999)$ weight_decay $= 1\times10^{-5}$ | Adam $\beta_1, \beta_2 = (0.9, 0.999)$ weight_decay $= 1\times10^{-5}$ | Adam $\beta_1, \beta_2 = (0.9, 0.999)$ weight_decay $= 1\times10^{-5}$ |
| **Number of top-K instances ($K_t$)** | 15 | 15 | 30 |
| **Number of bottom-K instances ($K_b$)** | 15 | 15 | 0 |
| **Backbone of model $f_{ins}$** | ResNet34 | ResNet34 | ResNet34 |
| **Positive class weight ($w_1$)** | 0.66 | 0.65 | 0.66 |
| **Negative class weight ($w_0$)** | 0.34 | 0.35 | 0.34 |
| **Model Selection** | Highest validation accuracy | Highest validation accuracy | Highest validation accuracy |
| **Validation frequency** | Every 5 epochs | Every 5 epochs | Every 5 epochs |

## References


1. Débarre, D., *et al.* Imaging lipid bodies in cells and tissues using third-harmonic generation microscopy. *Nature Methods* **3**, 47-53 (2006).
2. Tu, H., *et al.* Stain-free histopathology by programmable supercontinuum pulses. *Nature Photonics* **10**, 534-540 (2016).
3. You, S., *et al.* Intravital imaging by simultaneous label-free autofluorescence-multiharmonic microscopy. *Nature Communications* **9**, 1-9 (2018).
4. Zipfel, W.R., *et al.* Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation. *Proceedings of the National Academy of Sciences* **100**, 7075-7080 (2003).
5. You, S., *et al.* Raman spectroscopic analysis reveals abnormal fatty acid composition in tumor micro-and macroenvironments in human breast and rat mammary cancer. *Scientific Reports* **6**, 1-10 (2016).
6. Grigorescu, S., Trasnea, B., Cocias, T. & Macesanu, G. A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics* **37**, 362-386 (2020).
7. Miotto, R., Wang, F., Wang, S., Jiang, X. & Dudley, J.T. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics* **19**, 1236-1246 (2018).
8. Nassif, A.B., Shahin, I., Attili, I., Azzeh, M. & Shaalan, K. Speech recognition using deep neural networks: A systematic review. *IEEE Access* **7**, 19143-19165 (2019).
9. Young, T., Hazarika, D., Poria, S. & Cambria, E. Recent trends in deep learning based natural language processing. IEEE *Computational Intelligence Magazine* **13**, 55-75 (2018).
10. Cheplygina, V., de Bruijne, M. & Pluim, J.P. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis* **54**, 280-296 (2019).
11. Amores, J. Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence* **201**, 81-105 (2013).
12. Settles, B. Active Learning Literature Survey. (2009).
13. Zhu, X.J. Semi-supervised Learning Literature Survey. (2005).
14. Vijayanarasimhan, S. & Grauman, K. Keywords to visual categories: Multiple-instance learning forweakly supervised object categorization. in *2008 IEEE Conference on Computer Vision and Pattern Recognition* 1-8 (IEEE, 2008).
15. Song, X., Jiao, L., Yang, S., Zhang, X. & Shang, F. Sparse coding and classifier ensemble based multi-instance learning for image categorization. *Signal Processing* **93**, 1-11 (2013).
16. Tong, T., *et al.* Multiple instance learning for classification of dementia in brain MRI. *Medical Image Analysis* **18**, 808-818 (2014).
17. Xu, Y., *et al.* Weakly supervised histopathology cancer image segmentation and classification. *Medical Image Analysis* **18**, 591-604 (2014).
18. Karpathy, A. & Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3128-3137 (2015).
19. Campanella, G., Silva, V.W.K. & Fuchs, T.J. Terabyte-scale deep multiple instance learning for classification and localization in pathology. *arXiv preprint arXiv:1805.06983* (2018).
20. Li, J., *et al.* A multi-resolution model for histopathology image classification and localization with multiple instance learning. *Computers in Biology and Medicine* **131**, 104253 (2021).
21. Lu, M.Y., *et al.* Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering* **5**, 555-570 (2021).

22. Campanella, G., *et al.* Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine* **25**, 1301-1309 (2019).
23. Lerousseau, M., *et al.* Weakly supervised multiple instance learning histopathological tumor segmentation. in *International Conference on Medical Image Computing and Computer-Assisted Intervention* 470-479 (Springer, 2020).
24. Yao, J., Zhu, X., Jonnagaddala, J., Hawkins, N. & Huang, J. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Medical Image Analysis* **65**, 101789 (2020).
25. Lu, M.Y., *et al.* AI-based pathology predicts origins for cancers of unknown primary. *Nature* **594**, 106-110 (2021).
26. Boppart, S.A., You, S., Li, L., Chen, J. & Tu, H. Simultaneous label-free autofluorescence-multiharmonic microscopy and beyond. *APL Photonics* **4**, 100901 (2019).
27. You, S., *et al.* Slide-free virtual histochemistry (Part I): development via nonlinear optics. *Biomedical Optics Express* **9**, 5240-5252 (2018).
28. Orringer, D.A., *et al.* Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nature Biomedical Engineering* **1**, 1-13 (2017).
29. Willenbrink, T.J., *et al.* Field cancerization: definition, epidemiology, risk factors, and outcomes. *Journal of the American Academy of Dermatology* **83**, 709-717 (2020).
30. Arneth, B. Tumor microenvironment. *Medicina* **56**, 15 (2019).
31. Mbeunkui, F. & Johann, D.J. Cancer and the tumor microenvironment: a review of an essential relationship. *Cancer Chemotherapy and Pharmacology* **63**, 571-582 (2009).
32. Curtius, K., Wright, N.A. & Graham, T.A. An evolutionary perspective on field cancerization. *Nature Reviews Cancer* **18**, 19-32 (2018).
33. Van Engelen, J.E. & Hoos, H.H. A survey on semi-supervised learning. *Machine Learning* **109**, 373-440 (2020).
34. Zhu, X. & Goldberg, A.B. Introduction to semi-supervised learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* **3**, 1-130 (2009).
35. Foulds, J. & Frank, E. A review of multi-instance learning assumptions. *The Knowledge Engineering Review* **25**, 1-25 (2010).
36. You, S., *et al.* Slide-free virtual histochemistry (Part II): detection of field cancerization. *Biomedical Optics Express* **9**, 5253-5268 (2018).
37. Wang, W., *et al.* Single cell behavior in metastatic primary mammary tumors correlated with gene expression patterns revealed by molecular profiling. *Cancer research* **62**, 6278-6288 (2002).
38. Provenzano, P.P., *et al.* Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Medicine* **4**, 1-15 (2006).
39. Xi, G., *et al.* Large-scale tumor-associated collagen signatures identify high-risk breast cancer patients. *Theranostics* **11**, 3229 (2021).
40. You, S., *et al.* Label-free visualization and characterization of extracellular vesicles in breast cancer. *Proceedings of the National Academy of Sciences* **116**, 24012-24018 (2019).
41. Tu, H., *et al.* Concurrence of extracellular vesicle enrichment and metabolic switch visualized label-free in the tumor microenvironment. *Science Advances* **3**, e1600675 (2017).
42. Sun, Y., *et al.* Intraoperative visualization of the tumor microenvironment and quantification of extracellular vesicles by label-free nonlinear imaging. *Science Advances* **4**, eaau5603 (2018).
43. Becker, A., *et al.* Extracellular vesicles in cancer: cell-to-cell mediators of metastasis. *Cancer Cell* **30**, 836-848 (2016).

44. Muralidharan-Chari, V., Clancy, J.W., Sedgwick, A. & D'Souza-Schorey, C. Microvesicles: mediators of extracellular communication during cancer progression. *Journal of Cell Science* **123**, 1603-1611 (2010).
45. Skala, M.C.*, et al.* In vivo multiphoton microscopy of NADH and FAD redox states, fluorescence lifetimes, and cellular morphology in precancerous epithelia. *Proceedings of the National Academy of Sciences* **104**, 19494-19499 (2007).
46. Georgakoudi, I. & Quinn, K.P. Optical imaging using endogenous contrast to assess metabolic state. *Annual Review of Biomedical Engineering* **14**, 351-367 (2012).
47. Panieri, E. & Santoro, M. ROS homeostasis and metabolism: a dangerous liason in cancer cells. *Cell Death & Disease* **7**, e2253-e2253 (2016).
48. Vander Heiden, M.G., Cantley, L.C. & Thompson, C.B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029-1033 (2009).
49. Willms, E., Cabañas, C., Mäger, I., Wood, M.J. & Vader, P. Extracellular vesicle heterogeneity: subpopulations, isolation techniques, and diverse functions in cancer progression. *Frontiers in Immunology* **9**, 738 (2018).
50. Costa-Silva, B.*, et al.* Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nature Cell Biology* **17**, 816-826 (2015).
51. Sansone, E., De Natale, F.G. & Zhou, Z.-H. Efficient training for positive unlabeled learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**, 2584-2598 (2018).
52. Du Plessis, M.C., Niu, G. & Sugiyama, M. Analysis of learning from positive and unlabeled data. *Advances in Neural Information Processing Systems* **27**(2014).
53. Miller, D.J. & Uyar, H. A mixture of experts classifier with learning based on both labelled and unlabelled data. *Advances in Neural Information Processing Systems* **9**(1996).
54. Hou, L.*, et al.* Patch-based convolutional neural network for whole slide tissue image classification. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2424-2433 (2016).
55. Ilse, M., Tomczak, J. & Welling, M. Attention-based deep multiple instance learning. in *International Conference on Machine Learning* 2127-2136 (PMLR, 2018).
56. Andrews, S., Tsochantaridis, I. & Hofmann, T. Support vector machines for multiple-instance learning. *Advances in Neural Information Processing Systems* **15**(2002).
57. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 770-778 (2016).
58. Deng, J.*, et al.* Imagenet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248-255 (Ieee, 2009).
59. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* **9**(2008).
60. Sundararajan, M., Taly, A. & Yan, Q. Axiomatic attribution for deep networks. in *International Conference on Machine Learning* 3319-3328 (PMLR, 2017).
61. Smilkov, D., Thorat, N., Kim, B., Viégas, F. & Wattenberg, M. Smoothgrad: removing noise by adding noise. *arXiv preprint arXiv:1706.03825* (2017).
62. Kluyver, T.*, et al. Jupyter Notebooks-a publishing format for reproducible computational workflows*, (2016).