Real-time diagnosis and visualization of tumor margins in excised breast specimens using fluorescence lifetime imaging and machine learning

JAKOB UNGER,[1,5,6] **CHRISTOPH HEBISCH,**[1] **JENNIFER E. PHIPPS,**[1] **JOÃO L. LAGARTO,**[1] **HANNA KIM,**[2] **MORGAN A. DARROW,**[3] **RICHARD J. BOLD,**[4] **AND LAURA MARCU**[1,5,7]

[1]*Department of Biomedical Engineering, University of California Davis, California, CA 95616, USA*
[2]*Department of Otolaryngology, University of California Davis, California, CA 95817, USA*
[3]*Department of Pathology and Laboratory Medicine, University of California Davis, California, CA 95817, USA*
[4]*Department of Surgery, University of California Davis, California, CA 95817, USA*
[5]*Corresponding authors.*
[6]*jaunger@ucdavis.edu*
[7]*lmarcu@ucdavis.edu*

Abstract: Tumor-free surgical margins are critical in breast-conserving surgery. In up to 38% of the cases, however, patients undergo a second surgery since malignant cells are found at the margins of the excised resection specimen. Thus, advanced imaging tools are needed to ensure clear margins at the time of surgery. The objective of this study was to evaluate a random forest classifier that makes use of parameters derived from point-scanning label-free fluorescence lifetime imaging (FLIm) measurements of breast specimens as a means to diagnose tumor at the resection margins and to enable an intuitive visualization of a probabilistic classifier on tissue specimen. FLIm data from fresh lumpectomy and mastectomy specimens from 18 patients were used in this study. The supervised training was based on a previously developed registration technique between autofluorescence imaging data and cross-sectional histology slides. A pathologist's histology annotations provide the ground truth to distinguish between adipose, fibrous, and tumor tissue. Current results demonstrate the ability of this approach to classify the tumor with 89% sensitivity and 93% specificity and to rapidly ($\sim$ 20 frames per second) overlay the probabilistic classifier overlaid on excised breast specimens using an intuitive color scheme. Furthermore, we show an iterative imaging refinement that allows surgeons to switch between rapid scans with a customized, low spatial resolution to quickly cover the specimen and slower scans with enhanced resolution (400 μm per point measurement) in suspicious regions where more details are required. In summary, this technique provides high diagnostic prediction accuracy, rapid acquisition, adaptive resolution, nondestructive probing, and facile interpretation of images, thus holding potential for clinical breast imaging based on label-free FLIm.

1. Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death among females worldwide [1]. Breast-conserving therapy tends to be the preferred surgical procedure after an early breast cancer diagnosis. The surgeon attempts to excise the entire tumor volume including a surrounding layer (margins) of healthy tissue, to minimize the risk of local recurrence. A major limiting factor for complete surgical resection is the physician's ability to identify the complex tumor margins. On one end, excessive removal of normal tissue to ensure complete cancer removal can compromise the cosmetic outcome and impair functionality. On

the other end, insufficient removal of normal tissue leads to potentially incomplete removal of cancer. Patients are recommended to undergo a second operation to ensure that margins are negative for malignancy, with the consequences of delayed subsequent therapy initiation (i.e., radiation or chemotherapy) and excess healthcare resource utilization. Furthermore, numerous studies have demonstrated that re-excision is associated with a higher risk of recurrence [2,3].

Postoperative analysis of histopathological sections is common practice for surgical margin assessment. Preparation of the specimen involves fixing, sectioning, and staining with hematoxylin and eosin (H&E). Interpretation and treatment planning is finally based on a microscopic assessment of the specimen. If the tumor is extending to the surface (positive margin), re-excision is required. The surgeon correlates the excised specimen to find critical locations where tumor cells are close to the surface and whether margin involvement is focal or extensive. However, the correlation between the excised specimen can be challenging and imprecise. Furthermore, standard histopathological assessment is time-consuming, causes stress for the patient and potential risks of surgical infection, limiting the potential for rapid intraoperative consultation.

In this respect, there is an unmet need for enhanced intrasurgical imaging techniques to ensure clear margins right away at the operating table. New technologies, such as intraoperative cytologic (touch-prep cytology) or pathologic (frozen-section) analysis, can address some of the weaknesses of conventional histopathologic analysis. However, they are either time-consuming, require special training, or have intrinsic variability in subjective interpretation. Several studies indicated that optical techniques have the potential to overcome these limitations due to non-destructive tissue probing. Moreover, they provide information about physiological, chemical, and morphological changes associated with cancer. Recent studies address breast cancer diagnostics using Raman spectroscopy [4,5], optical coherence tomography [6,7], photoacoustic tomography [8,9], fluorescence lifetime imaging [10,11], diffuse reflectance spectroscopy [12,13] and micro-elastography [14].

Although some of these techniques achieved high accuracy, none have been widely adopted into regular clinical practice. All methods have their pitfalls, such as reduced sensitivity, lack of acquisition speed to quickly cover a larger tissue area, requirement of expert skills, and destructive characteristics [15]. Another important need for acceptance by the wider clinical community is real-time capability combined with an intuitive visualization of conclusive diagnostic information. A recent approach combined near-infrared fluorescence with augmented real-time imaging and navigation to assess breast tumor margins in real-time [16]. NIR fluorescence image-guided surgery demonstrated great potential for rapid intraoperative visualization of tumors. However, this near-infrared fluorescence technique requires injection of a contrast agent, which exposes patients to risks of allergic reactions, and the timing of the surgery need to account for contrast agent delivery time and uptake.

Label-free fluorescence lifetime imaging (FLIm) allows for rapid data acquisition and processing of tissue diagnostic data derived from breast tumor specimens. Recent studies have demonstrated that FLIm images can be acquired either during surgery [17–19] or on excised specimens [10] to characterize biochemical features and associated pathologies. In particular, FLIm has been demonstrated to identify breast tumor regions [10,20], glioma tumors [21], oropharyngeal cancer [22] and atherosclerotic lesions [18]. In a recent study, the ability to identify breast tumors was demonstrated for small regions correlated visually with histology slides [10]. However, an ad hoc real-time visualization of diagnostic information was not provided in this earlier study. Furthermore, detection ability was demonstrated for small regions correlated visually with histology slides. Validation of classification algorithms and imaging technology demands a precise match between the optical measurements and the histopathology, which is considered as the gold standard for the evaluation of surgical margins. In this paper, we pursue the next steps towards a practical implementation focusing on: (1) Demonstrating real-time tissue diagnosis based on parameters derived from fluorescence decay, and (2) Intuitive visualization of tissue

type. Due to possible inter-patient variability of fluorescence signatures we transform a Random Forest classifier output into a probability distribution over classes and finally into a simple color scheme representing the different classes tumor, adipose, and fibrous tissue. The visualization is based on probabilistic classification outputs encoding the type of tissue and the certainty of the probabilistic output. Thus, regions that are identified with insufficient certainty are labeled as such and can be considered by the surgeon. In addition, (3) we show an iterative imaging refinement that allows a surgeon to switch between rapid scans with a low spatial resolution and slower scans with enhanced resolution in suspicious regions where more details are required. The supervised learning scheme and its evaluation were based on a well characterized set of samples (N=18 patients) leveraging on a previously developed registration technique between autofluorescence imaging data and cross-sectional histology slides [23].

2.   Materials and methods

2.1.   Breast specimens

Eighteen tissue specimens were obtained from eighteen patients who underwent lumpectomy (n=6) or mastectomy surgeries (n=12) at the University of California Davis Health System (UCDHS) and imaged within an hour of resection. All patients provided informed consent. For each patient, one piece of tissue with a diameter of 15 to 30 mm that was assumed to contain a tumor was studied. Prior to imaging, a 405nm continuous-wave (CW) laser diode was used to generate 4 to 8 clearly observable landmarks (fiducial points) on the tissue block in order to enable an accurate registration between the video images and the corresponding histology slides [23]. Histology confirmed 14 specimens with the diagnosis of invasive cancer and 4 specimens with DCIS. Note that the specimens used for a recent image registration study [23] have been also used in this study.

2.2.   Instrumentation and imaging setup

A prototype time-domain multispectral time-resolved fluorescence spectroscopy (ms-TRFS) system [24] with an integrated aiming beam (Fig. 1(a)) was used in the study. The ms-TRFS system consisted of a fluorescence excitation source, a wavelength-selection module, and a fast-response fluorescence detector. A pulsed 355 nm laser (Teem Photonics, France) was exciting the fluorescence through a fiber (Pulse duration: 650 ps, Energy per pulse: 1.2 µJ, Repetition rate of the laser: 120 Hz). Autofluorescence was spectrally resolved into four spectral bands (channels) using dichroic mirrors and bandpass filters: 390/40 nm (Channel 1), 470/28 nm (Channel 2), 542/28 nm (Channel 3), and 629/53 nm (Channel 4). Specifically, the fluorescence emission in these four bands is adapted to resolve collagen, NADH, FAD, and porphyrins. Each channel outputs the autofluorescence to an optical delay fiber with an increasing length from channel 1 to channel 4. This organization enabled temporal multiplexing of the spectral channels so that the decay waveforms arrive sequentially at distinct time points at the detector (single microchannel plate photomultiplier tube, MCP-PMT, R3809U-50, Hamamatsu, 45 ps FWHM). Subsequently, the signal is amplified (RF amplifier, AM-1607-3000, 3-GHz bandwidth, Miteq) and digitized (PXIe-5185, National Instruments, 12.5-GS/s sampling rate).

Digitizing and deconvolution (see Sect. 2.6) as well as image processing, classification, and visualization tasks are performed on separate computers communicating via TCP/IP protocol. Digitizing and deconvolution was performed on an Intel Celeron dual-core T3100 CPU, 3 GB RAM using LabVIEW (National Instruments). Image processing, classification and visualization were implemented in C++ and OpenCV running on an Intel Core i7-3632QM CPU (4 kernels) equipped with 16 GB of RAM. The whole system was assembled on a cart (Fig. 1(c)) for mobility on demand. Matlab (Mathworks, Inc.) was used for registering autofluorescence imaging data and cross-sectional histology slides [23].
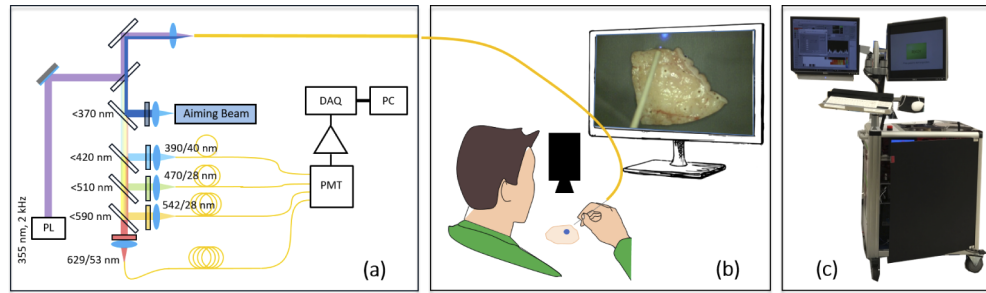
Fig. 1. (a) Schematic of the ms-TRFS instrumentation used for imaging purposes. A single fiber is used for excitation and autofluorescence collection. PL: Pulsed Laser, DAQ: Data Acquisition, PMT: Photomultiplier. (b) Imaging setup. A hand-guided scan was performed for each specimen. An aiming beam is integrated into the optical path and serves as a marker to overlay fluorescence data on the video where the measurement was taken. (c) FLIm system and computers assembled on a cart equipped with two screens that was used to image the breast specimens.

2.3. Imaging protocol

For all specimens, the fiber probe was hand-guided during imaging (Fig. 1(b)). During the scan, the distance from the sample to the probe was kept at a few millimeters. In case that the fiber tip moved too far from the sample, the computer recognized a diminished signal amplitude triggering an acoustic alarm to remind the imaging person to get closer to the sample. If the fiber tip accidentally touched the sample, a reference measurement was performed after the scan. If deviations were recognized, the fiber tip was cleaned, and the scan was repeated. Scans were also repeated if the sample was accidentally moved. The scanning time per sample ranged from 2 to 6 minutes to cover the entire specimen. This corresponds to a scanning time per area of approximately 0.4 - 0.5 s/mm^2. After imaging, specimens were placed in formalin and processed routinely for histologic analysis.

2.4. Aiming beam principle

An external camera (Point Gray Chameleon3 1.3 MP Color USB3 vision with Fujinon HF9HA-1B 2/3"9 mm lens) captured the specimen during the scanning procedure. A 445 nm laser diode (TECBL-50G-440-USB, World Star Tech, Canada) was integrated into the optical path (Fig. 1(a)) delivering to the measured area via the same fiber-optic probe. The incident power of the aiming beam is approximately 0.35 mW. The blue aiming beam served as an optical marker on the sample to highlight the point where the current measurement is carried out (Fig. 1(b)). It was tracked in the video image by transforming the image into the HSV color space and thresholding the hue and saturation channels [20]. Information obtained from the fluorescence measurements was successively overlaid on the marker positions during the scan creating an artificial color overlay on the sample.

2.5. Histological preparation and registration

After imaging the specimens, histology sections were cut in parallel to the imaging plane at 4 μm thickness using a microtome. The first continuous large slice that covers the whole area of the specimen was then stained with hematoxylin and eosin (H&E) and scanned with an Aperio Digital Pathology Slide Scanner (Leica Biosystems). The pathologist assessed the pathology slides and delineated regions of fibrous tissue, normal ducts and lobules, fat, invasive and ductal carcinoma in situ using Aperio ImageScope (Leica Biosystems). The pathologist's annotations (the delineations and tissue labels) were automatically exported and further processed with a

custom made registration tool registering the fluorescence data with the histology annotations. This allowed assigning the fluorescence signatures with histological findings to build up a sophisticated training set for the classifier and validate its performance.

The registration tool relies on a previously presented method to register data acquired with a point-scanning spectroscopic imaging technique from fresh surgical tissue specimen blocks with corresponding histological sections [23]. The lasermarks, generated with a 405-nm CW laser diode, served as fiducial markers (diameter ≈ 100 *μm*) and were visible in both the camera image and the histology slide. The registration pipeline was built as a two-stage process. First, a rough alignment was achieved from a rigid registration by minimizing the distances between corresponding laser marks in the histology and camera image. Second, a piecewise shape matching was used to match the outer shape of the specimen in the camera image and the histology section and to refine the initial rigid registration [23].

2.6. Deconvolution and parameter extraction

Mathematically, the measured autofluorescence response (y) from tissue to an excitation laser pulse can be modeled as convolution of the fluorescence impulse response function (fIRF, h) with the instrument impulse response function (iIRF, I) stemming from delay components and modal dispersion. Thus,

$$y(k) = \sum_{i=0}^{k} I(k-i))h(i) + \epsilon_k, \tag{1}$$

where $t_i = i\Delta t$ with $i = 1 \ldots N - 1$ are discrete time points for N uniform sampling intervals and ϵ_k is an additive white noise component. In order to estimate h, a constrained Laguerre model was used [25],

$$h(k) = \sum_{i=0}^{L-1} c_l b_l(k; \alpha), \tag{2}$$

where $b_l(k; \alpha)$ is an ordered set of Laguerre basis functions with the maximum order L and $0 \leq \alpha \leq 1$ with constraints that $h(t)$ is strictly convex, positive and monotonically decreasing for $0 \leq t < \infty$. In this study, we used $L = 12$ and $\alpha = 0.8$.

From the decay profile, we extract average lifetime τ_{avg} and intensities I_{avg}, each for the four spectral channels,

$$\tau_{avg} = \frac{\Delta t \sum_{k=0}^{N-1} k h(k)}{\sum_{k=0}^{N-1} h(k)}, \tag{3}$$

and

$$I_{avg} = \sum_{k=0}^{N-1} h(k). \tag{4}$$

In order to make the intensity parameter insensitive to extrinsic factors, the intensity ratios were used

$$I_{ratio}^{ch} = \frac{I_{avg}^{ch}}{\sum_{k=1}^{4} I_{avg}^{k}} \tag{5}$$

where $ch = \{1 \ldots 4\}$ specifies the spectral channel. The set of parameters spanning the feature space serving as input for the classifier had 56 dimensions. It consisted of 4 channels, each characterized by 12 Laguerre coefficients, 4 lifetime and 4 intensity parameters.

2.7. Classifier setup and training

The training pipeline is illustrated in Fig. 2. The semiautomatic histology registration procedure was used to map the histology slide and the pathologist's annotations on the video image serving

as ground truth for the training. Due to potential registration errors [23], the delineated regions of tumor, fibrous, and adipose tissue were shrank down by 0.5 *mm* using morphological erosion, and the laser markers are excluded from the training set. The numbers of pixels before and after morphological erosion are given in Table 1. If the markers are evenly distributed, the shrinkage ensures that possible registration errors do not exceed 1 *mm* (in accordance with 0.5 *mm* erosion) and therefore, do not impact the quality of training data [23]. A random forest classifier was trained with the 56 dimensional feature vector obtained from the morphological filter results.
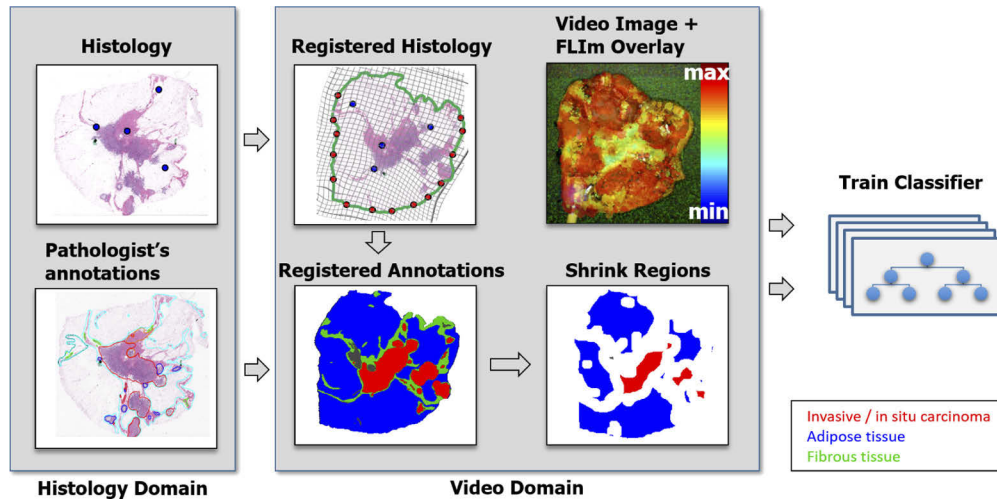

Fig. 2. The supervised training pipeline involves registration of cross-sectional histology with the video image using a hybrid registration method [23]. Pathologist tracings from the histology are mapped to the video domain. In order to account for possible registration errors, regions are narrowed by 0.5 mm. Fluorescence parameters from the resulting regions are fed into a random forest classifier.

Random forest [26] is an ensemble of classification or regression trees using a random subset of features at each candidate split in the learning process that is induced from bootstrap samples of the training data. Predictions are derived from averaging or majority voting of all individual regression trees. Therefore, Random Forests are intrinsically suited for multi-class problems. The combination of less correlated trees combined with randomized node optimization and bagging has demonstrated a reduced variance and sensitivity to overfitting. The maximum number and depth of trees were set to 100 and 10, respectively, in order to limit the size and complexity of the trees. In each split, $\sqrt{N}$ features were used where N denotes the total number of features. The classes were balanced according to their sample size.

For training and validation, the leave-one-subject-out strategy was pursued. This involved sequentially leaving data from a single patient out of the training set, then testing the classification accuracy on that single specimen that was left out for training. This procedure was repeated for all specimens. For each FLIm point measurement, the output of the random forest was compared against the registered histology mappings. Due to imbalanced data (tumor, fibrous and adipose tissue), the Matthews correlation coefficient (MCC) [27] and Receiver operating characteristic (ROC) curves were evaluated for each point measurement in order to assess the predictive power of the classifier.

2.8. Visualization and overlay refinement

The output of the random forest classifier was transformed into a simple color scheme representing the probability that the interrogated point scanned with the fiber optic was identified as adipose,

Table 1. Number of pixels of tissue types obtained from registered histology in video domain

Case no.	after morphological erosion			before morphological erosion		
	Tumor	Fibrous	Adipose	Tumor	Fibrous	Adipose
1	15866	621	2127	21715	3975	8023
2	32604	0	0	37836	697	614
3	16759	7486	1863	22057	16205	5578
4	1161	1406	4864	3120	9610	12444
5	3045	26	19436	6862	6192	32247
6	1085	79	12207	2285	4181	23233
7	3817	2469	5790	5429	13632	12905
8	2215	506	4969	3402	1262	7923
9	1805	2915	1454	3624	7175	3490
10	10996	231	1843	13273	1056	5104
11	18434	7	1172	20172	436	2332
12	0	5520	12570	0	10726	18342
13	11272	214	10323	13149	2480	14621
14	3356	841	11049	5574	3178	14080
15	11440	1810	9844	14710	5698	13234
16	5079	2587	11277	6156	6232	15411
17	6068	1300	8438	8321	5421	11796
18	0	1216	29564	0	6678	40977

fibrous, or tumor tissue. Unlike multilayer perceptrons and other variations of artificial neural networks, random forests do not inherently provide posterior class probability estimates. Instead, the outputs can be transformed into a probability distribution over classes by averaging the unweighted class votes by the trees of the forest, where each tree votes for a single class [28]. This is in contrast to Support Vector Machines, where outputs have to be transformed into a probability distribution over classes, e.g. using Platt scaling [29]. The visualization scheme encodes tissue types in different colors: tumor in red, fibrous in green and adipose in blue. The interrogated position was colored according to

$$C_{output} = 255\{p_{tumor}\delta_{tumor}, p_{fibrous}\delta_{fibrous}, p_{adipose}\delta_{adipose}\}, \tag{6}$$

where p_i denote the posterior class probabilities of tumor, fibrous and adipose tissue types and $\delta_i = 1$, if the majority of trees voted for class i and $\delta_i = 0$ in any other case. Thus, the output color C_{output} encodes the type of tissue in color and the certainty of the probabilistic output in saturation (see Fig. 3). If the output is close to black $(0, 0, 0)$, the feature vector is close to decision borders within the feature space. The visible aiming beam, delivered through the optical probe, enables creating the color overlay scheme as described in Eq. (6) on the video in real-time. However, inhomogeneous tissue properties and different wavelengths of the excitation light and the aiming beam result in a difference between the aiming beam size and the area over which fluorescence is measured [20,30]. In a recent study [20], our group described a strategy where the size of the element is linearly scaled with the scanning speed providing an optimal coverage but at the cost of reduced spatial resolution. The size of region that is colored per point measurement according to the FLIm measurement constitutes a trade-off between resolution and scanning speed. If the size is chosen too large (Fig. 3(a)), the sample can be covered quickly but the overlay becomes blurred and delineation might become imprecise. Contrarily, if the size is chosen small, scanning takes a longer time (Fig. 3(b)).
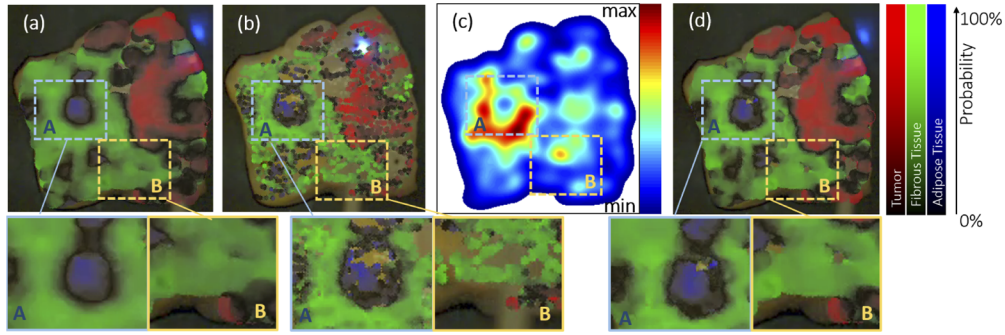
Fig. 3. Augmented classification overlay with fixed element size $l_0 = 6$ (a) and $l_0 = 1$ (b). If the d_p is too large, the overlay gets blurred and imprecise. Contrarily, a small diameter d_p will lead to poor coverage of the sample. The adaptive refinement (d) is based on the sample accumulation f_h (Eq. (7)) quantifying the local sampling density shown in (c). It acquires a high level of detail for densely sampled regions (such as area A) and maintain a good coverage in sparsely sampled regions (such as area B).

Here, we pursue a different strategy aiming for a local refinement of resolution at areas of interest. The diameter d_p of the circular element is adjusted according to the local sampling density. The scan starts with $d_p = l_{max}d_f$, where d_f is the fiber core diameter (400 μm) and l_{max} the maximum scaling factor. The choice of l_{max} thus defines the initial size of the element. If the fiber probes an area of interest for a longer time period, d_p is reduced automatically to increase spatial resolution locally. Accumulating the sampling points

$$f_h(\mathbf{x}) = \sum_{i=0}^{N} K(\mathbf{x} - \mathbf{x_i}) \tag{7}$$

approximates the number of sampling points in the local neighborhood of the current probing position $\mathbf{x}$ where $\mathbf{x_i}$ are sampling points that have been already scanned with the probe and

$$K = \text{rect}\left(\frac{\mathbf{x}}{h_0}\right). \tag{8}$$

The diameter

$$d_p(\mathbf{x}) = (l_{max} - l_0)\, d_f \tag{9}$$

thus depends on the local sampling density where

$$l_0 = \min\left\{\left\lfloor\left|\frac{f_h(\mathbf{x})}{\Delta_0}\right|\right\rfloor, l_{max} + 1\right\} \tag{10}$$

is the local refinement level. The sampling step size Δ_0 specifies the number of samples in the local environment that lead to the next refinement level. This refinement strategy enables an initial rough scan of the sample. For inhomogeneous regions, the color scheme produces a black output indicating a mixed tissue decomposition. Re-scanning the ambiguous region leads to a local refinement and increase of the local spatial resolution. This principle sample is illustrated in Fig. 3. Based on the local sampling point density $K(\mathbf{x})$, shown in (c), the refinement (d) provides a complete coverage of the sample but also exhibits local details at regions with a high sampling density.

3. Results

3.1. Visualization

Figure 4 depicts the scanning process for two specimens, including the color overlay for different time stamps, the histology slide with delineated regions, and the histology projection registered to the video domain. The final overlay exhibits a high level of agreement with the registered histology findings. Note that the fiducial markers (marked with a white circle in the histology projection) stem from a 405 nm CW laser diode to generate clearly observable landmarks on the tissue block to enable registration with the video image [23]. As previously reported, the autofluorescence signatures in cauterized or burnt tissues can be considerably altered [31–33]. Thus, the regions of the laser markers used as fiducials for registration purposes (white circles in Fig. 4(c) and (e)) were excluded from the training and validation set. The color scheme immediately reveals the tissue composition is confirmed with the pathologist's registered histology annotations, the 'gold standard' for the evaluation of surgical margins. The color scheme exhibits a lower probability at the border between two different tissue types. This is due to a mixed tissue composition within the area over which fluorescence is measured pushing the feature vector towards the decision boundaries. The local refinement level was chosen to start with a maximum scaling factor of $l_{max} = 6$ which corresponds to a diameter of $d_p \approx 4.4$ *mm* (Eq. (9)), thus covering 36 times the area at the finest level ($l_0 = 1$). The system was running with approximately 20 frames per second where the bottleneck was clearly on the image processing side. Beam tracking and
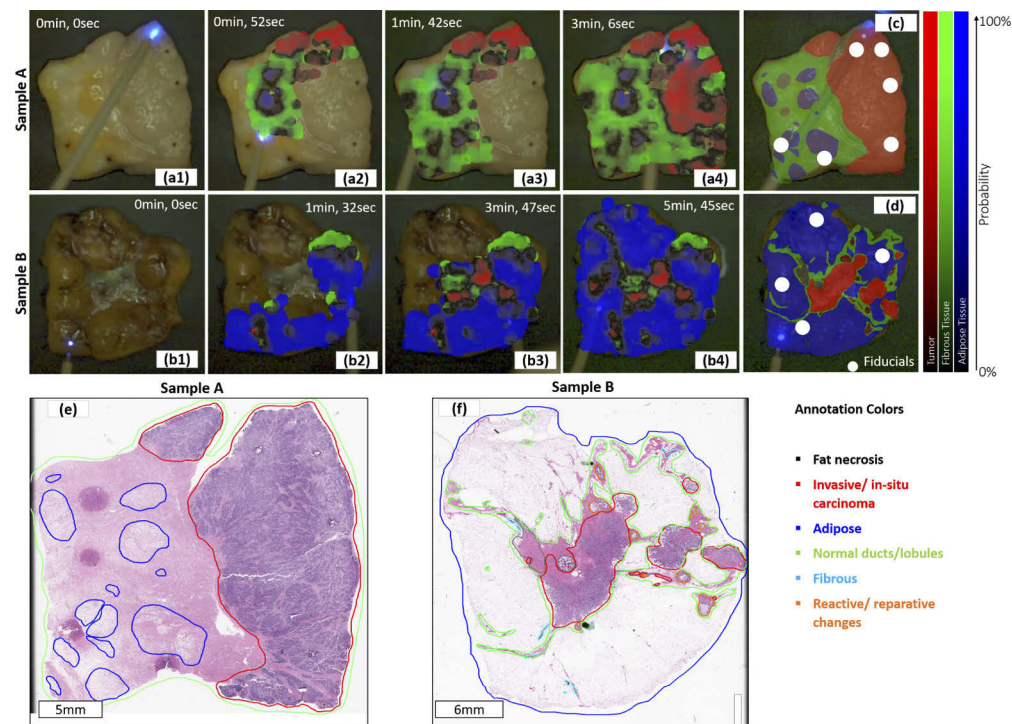

Fig. 4. Two examples (invasive carcinoma) of the augmented real-time overlay for different scanning times (sample A: a1-a4 and sample B: b1-b4, see supplemental videos Visualization 1 and Visualization 2), corresponding H&E histology slides with the pathologist's annotations (sample A: e and sample B: f) and the corresponding registered annotations, mapped to the video domain (sample A: c and sample B: d) and providing the ground truth obtained from histology overlaid onto the video image.

image preprocessing were most demanding and took ≈70% of the time, while classification and visualization only made up 9% and 7%. The rest was related to TCP/IP communication and image acquisition.

3.2. Classification accuracy

For the three-class problem tumor vs. adipose vs. fibrous based on the Random Forest classifier applied to 18 samples, the Matthews correlation coefficient was 0.86 for identifying tumor, 0.81 for adipose, and 0.70 for fibrous tissue.

Figure 5 depicts the receiver operating characteristic (ROC) curve illustrating the classification accuracy for the 2-class problem (tumor vs. no tumor). The area under the ROC curve was 0.96 and the Matthews correlation coefficient was 0.86 corroborating a well-fitted classification model. The ROC curve shows the performance as a trade-off between sensitivity and specificity. Assuming the same costs for true negative and false positive predictions, the classifier yields 88.78% sensitivity and 93.14% specificity. A case-wise analysis of the classifier performance is given in Fig. 6. Samples 2, 12, and 18 were homogeneous tissue samples (tumor only / no tumor only) that were classified correctly, thus being excluded from the ROC curve analysis. A high consistency can be seen across the samples providing stable classifier performances. However, a performance loss was observed for one case (case 7). The ROC curve clearly stands out showing a shift towards an increased false positive rate. The augmented overlay and the corresponding
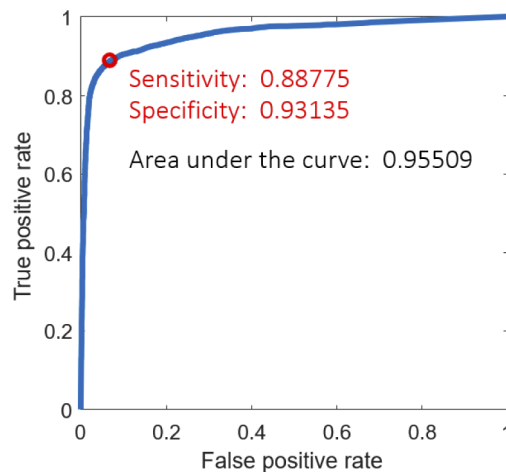

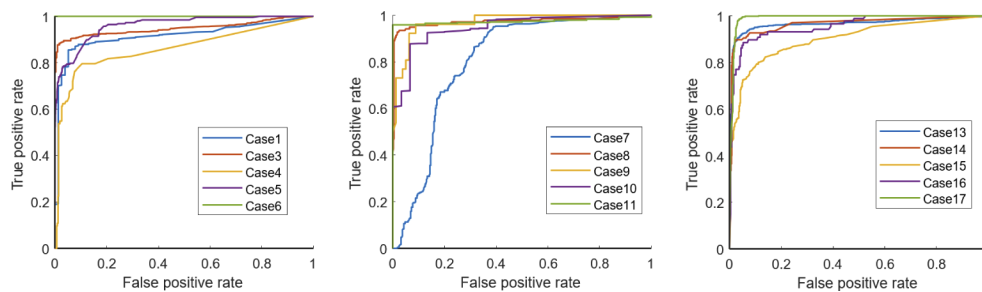**Fig. 5.** ROC curve for the two class problem tumor vs. no tumor for all specimens.

Fig. 6. ROC curves for each individual case. Three cases where histology revealed tumor /no tumor only were excluded from the analysis.

histology projection of case 7 are depicted in Fig. 7. On the left side of the specimen, fibrous tissue is erroneously classified as a tumor, reaching an a posteriori probability of up to 70%.
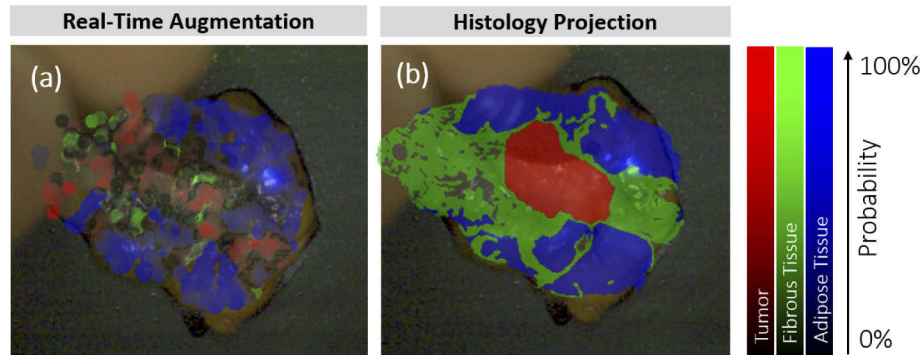

Fig. 7. Augmented overlay and histologic ground truth of case 7 (invasive carcinoma) showing a considerable deterioration of prediction accuracy in Fig. 6. The overlay exhibits an erroneous classification of tumor

4. Discussion

Fluorescence lifetime imaging (FLIm) is a label-free modality for in situ characterization of biochemical alterations in tissue. Here we demonstrated the potential of FLIm to distinguish in real-time between adipose, fibrous, and cancerous regions in breast specimens from women undergoing lumpectomies and mastectomies, and to augment the diagnostic information on the surgical specimen. Specifically, we developed and tested a supervised classification and visualization approach that allows for automated display of diagnostic information and intuitive visualization of such information with adaptive spatial resolution.

Cytology and frozen sections are among the most common techniques to delineate breast tumor margins but they require significant time and cost. In order to overcome these crucial limitations, a variety of optical imaging techniques have been combined with machine learning algorithms. Among them are deep neural networks applied on optical coherence tomography (OCT) images [34], a modified version of deep convolutional U-Net architecture for hyperspectral imaging data [35] and ultrasound combined with a random forest classifier [36] reporting sensitivity; specificity of 91.7%; 96.3%, 80-98%; 93-99% and 75.3%; 82.0%. Although a direct comparison of these modalities remains difficult due to varying datasets and metrics, ultrasound will probably not fulfill requirements to solve the margin status problem [15]. OCT has been applied ex vivo to image excised breast specimens [7] and invivo scanning the cavity after resection [6] with high overall accuracy. However, OCT still has limited ability to distinguish between cancerous and fibrous breast tissue due to potentially similar structural features of these tissue types [7]. Hyperspectral imaging has also demonstrated high accuracy for margin assessment and a fast scanning speed, but its transition to in vivo imaging requires a complicated instrument due to complex imaging geometry of a surgical cavity [37] and possible specular reflections of the liquid on the tissue surface [38]. Diffuse reflectance spectroscopy has shown promising results for discriminating healthy breast tissue from tumor tissue [39,40]. A recent study demonstrates a combination of diffuse reflectance spectroscopy with a Support Vector Machine for in vivo detction of breast cancer [40]. The authors report an excellent accuracy and Matthews Correlation Coefficient of 0.93 and 0.87, respectively. Drawbacks are a relatively slow acquisition speed and influence of ambient light.

FLIm allows for rapid and label-free imaging of tissue diagnostic data derived from breast tumor specimens and has already been demonstrated for invivo use [17–19]. FLIm-based discrimination

between breast tissue types is based on endogenous fluorescence of the fluorophores: collagen fibers, NADH, and FAD. Consistent with our recent findings [10], we observed longer lifetimes in spectral channels 2, 3 and 4 for adipose tissue. For fibrous tissue, shorter lifetimes were seen but longer than carcinoma. A recent study demonstrated the feasibility of FLIm to distinguish between adipose, fibrous, and cancerous regions. In this study, we developed a pipeline that combines FLIm with machine learning and advanced visualization technique, which enables immediate feedback on the interrogated location along with an intuitive visualization of the classifier output. The visualization uses a simple color scheme for displaying tissue type (output of the classification results) as well as the probability of prediction. A coarse scan without refinement provides a good overview. A low probabilistic output (encoded in black paintings) indicates either inhomogeneous tissue composition or a general uncertainty of the classifier. Rescanning the suspicious region will increase the local resolution to uncover details that cannot be concluded from the initial coarse scan. The adaptive resolution principle used in this study has a clear advantage of rapid data collection on the regions of interest from merely the initial scan. Thus, the scanning time, a crucial factor in breast-conserving therapy [15], can be considerably reduced as breast tissue predominantly consists of fat.

The results of this study have demonstrated that spectroscopic FLIm features derived from a large set of point measurements (1,000 - 10,000 point measurements per sample) can delineate the cancerous regions in breast specimens with high accuracy (89% sensitivity and 93% specificity). The classifier performed very well on all samples, except one. The classification accuracy can be influenced by multiple factors. For example, tissue sections were cut in parallel to the imaging plane so that the section used for histology staining might be a few micrometer off the imaged plane. To investigate the impact of a potential offset, in two cases, multiple 4 μm sections were cut within the 300 μm imaged volume. As no substantial differences were seen, the remaining study has been performed with a single section. The other possible cause is a low inter-patient variability due to an insufficient number of patients for training. In order to prevent and limit such drawbacks, future studies will require multiple sections for histology from one specimen and a more extensive database from a larger number of patients. A larger cohort will also be necessary to investigate, whether FLIm can distinguish between invasive cancer and DCIS.

Due to the manipulation of mechanically flexible specimens during the standard histopathological protocol, the difference in shape between the pathology slide and the acquired imaging data is one of the major challenges for any training for machine learning techniques [37]. Unlike our previous study [10], here we register the histology slide with the fluorescence measurements using a method [23] that accounts for possible tissue deformations and loss resulting from histological preparation. This approach allowed for better registration of FLIm maps with histopathological section and pathologist annotations. Using a sophisticated registration that accounts for tissue deformations leads to a higher number and quality of labeled samples. Especially including critical FLIm data acquired close to the margins leads to more accurate classification of FLIm signatures. Although the registration is providing reliable training set, small focal tumors with a diameter <1 mm are not evaluated in the statistics as registration accuracy cannot be guaranteed to provide a correct labeling of the measurement.

While this study focused on excised specimens, the long-term goal is to establish FLIm as an intraoperative tool scanning the surgical bed with a fiber optic. Recent studies from our group demonstrated the applicability of FLIm for in vivo applications [17,22,41]. The transition from ex vivo to in vivo settings will involve a few adaptations of the procedure. To account for the volumetric information that is necessary to facilitate accurate surgical resection, a stereo camera setup during the scan can be used to collect depth information [42]. Thus, the adaptive refinement will require re-triangulation for each measurement to build up a color 3d profile of the scanned region. Moreover, parameters Δ_0 and l_{max} will be redefined for in vivo scans as they define a trade-off between resolution and scanning speed. Real-time segmentation of the aiming beam

will also be more challenging as it has to deal with a variety of geometries, specular reflections, and illumination conditions. To account for these scenarios, we consider a more robust tracking procedure [42] where the aiming beam is pulsed with 50Hz to easily identify the beam position in a subtraction of subsequent images with aiming beam on and off. Finally, we also expect altered fluorescence signatures in in vivo applications due to cauterized tissue and altered metabolism making supervised training of an in vivo classifier a highly challenging task. Standard classifiers are not able effectively account for changes in data distributions between training and test phases. Generative Adversarial Networks (GANs) have been used to adapt synthetic input data from a simulator to a set of sparsely labeled real data (Domain Adaptation) using a neural network [43,44]. Domain Adaptation methods could adapt well-labeled ex vivo training data to the in vivo domain. Due to the simple color-coding scheme and real-time capability, such method holds great potential to guide surgeons intraoperatively when scanning the surgical bed to see if additional tissue needs to be excised.

5. Conclusions

This study demonstrates the potential application of FLIm for margin assessment of resected breast lumpectomy specimens when combined with machine learning and real-time visualization. The presented method here holds many desirable features including high prediction accuracy, rapid acquisition speed that allows for rapid large area scan, spatial refinement capability for suspicious regions, and nondestructive probing. This method is easy to use as it provides an intuitive visualization of tissue characteristics and demonstrates high tumor classification sensitivity (89%) and specificity (93%). Current method provides a path for future applications in vivo to identifying tumor infiltrations in the surgical bed. This future application could potentially provide the surgeon with additional information that reflects the histopathology of the tissue at the resection margin.

Funding

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," Ca-Cancer J. Clin. **65**(2), 87–108 (2015).
2. A. H. Mandpe, A. Mikulec, R. K. Jackler, L. H. Pitts, and C. D. Yingling, "Comparison of response amplitude versus stimulation threshold in predicting early postoperative facial nerve function after acoustic neuroma resection," Am. J. Otol. **19**(1), 112–117 (1998).
3. I. Gage, S. J. Schnitt, A. J. Nixon, B. Silver, A. Recht, S. L. Troyan, T. Eberlein, S. M. Love, R. Gelman, J. R. Harris, and J. L. Connolly, "Pathologic margin involvement and the risk of recurrence in patients treated with breast-conserving therapy," Cancer **78**(9), 1921–1928 (1996).
4. M. D. Keller, E. Vargis, A. Mahadevan-Jansen, N. de Matos Granja, R. H. Wilson, M. Mycek, and M. C. Kelley, "Development of a spatially offset raman spectroscopy probe for breast tumor surgical margin evaluation," J. Biomed. Opt. **16**(7), 077006 (2011).
5. G. Thomas, T.-Q. Nguyen, I. J. Pence, B. Caldwell, M. E. O'Connor, J. Giltnane, M. E. Sanders, A. Grau, I. Meszoely, M. Hooks, M. C. Kelley, and A. Mahadevan-Jansen, "Evaluating feasibility of an automated 3-dimensional scanner using raman spectroscopy for intraoperative breast margin assessment," Sci. Rep. **7**(1), 13548 (2017).
6. S. J. Erickson-Bhatt, R. M. Nolan, N. D. Shemonski, S. G. Adie, J. Putney, D. Darga, D. T. McCormick, A. J. Cittadine, A. M. Zysk, M. Marjanovic, E. J. Chaney, G. L. Monroy, F. A. South, K. Cradock, Z. G. Liu, M. Sundaram, P. S. Ray, and S. Boppart, "Real-time imaging of the resection bed using a handheld probe to reduce incidence of microscopic positive margins in cancer surgery," Cancer Res. **75**(18), 3706–3712 (2015).
7. F. T. Nguyen, A. M. Zysk, E. J. Chaney, J. G. Kotynek, U. J. Oliphant, F. J. Bellafiore, K. M. Rowland, P. A. Johnson, and S. A. Boppart, "Intraoperative evaluation of breast tumor margins with optical coherence tomography," Cancer Res. **69**(22), 8790–8796 (2009).

8. D. Piras, W. Xia, W. Steenbergen, T. G. van Leeuwen, and S. Manohar, "Photoacoustic imaging of the breast using the twente photoacoustic mammoscope: present status and future perspectives," IEEE J. Sel. Top. Quantum Electron. **16**(4), 730–739 (2010).
9. R. Li, P. Wang, L. Lan, F. P. Lloyd, C. J. Goergen, S. Chen, and J. Cheng, "Assessing breast tumor margin by multispectral photoacoustic tomography," Biomed. Opt. Express **6**(4), 1273–1281 (2015).
10. J. E. Phipps, D. Gorpas, J. Unger, M. Darrow, R. J. Bold, and L. Marcu, "Automated detection of breast cancer in resected specimens with fluorescence lifetime," Phys. Med. Biol. **63**(1), 015003 (2017).
11. V. Sharma, S. Shivalingaiah, Y. Peng, D. Euhus, Z. Gryczynski, and H. Liu, "Auto-fluorescence lifetime and light reflectance spectroscopy for breast cancer diagnosis: potential tools for intraoperative margin detection," Biomed. Opt. Express **3**(8), 1825–1840 (2012).
12. M. D. Keller, S. K. Majumder, M. C. Kelley, I. M. Meszoely, F. I. Boulos, G. M. Olivares, and A. Mahadevan-Jansen, "Autofluorescence and diffuse reflectance spectroscopy and spectral imaging for breast surgical margin analysis," Lasers Surg. Med. **42**(1), 15–23 (2010).
13. B. S. Nichols, C. E. Schindler, J. Q. Brown, L. G. Wilke, C. S. Mulvey, M. S. Krieger, J. Gallagher, J. Geradts, R. A. Greenup, J. A. Von Windheim, and N. Ramanujam, "A quantitative diffuse reflectance imaging (qdri) system for comprehensive surveillance of the morphological landscape in breast tumor margins," PLoS One **10**(6), e0127525–25 (2015).
14. K. M. Kennedy, L. Chin, R. A. McLaughlin, B. Latham, C. M. Saunders, D. D. Sampson, and B. F. Kennedy, "Quantitative micro-elastography: imaging of tissue elasticity using compression optical coherence elastography," Sci. Rep. **5**(1), 15538 (2015).
15. B. Maloney, D. McClatchy, B. Pogue, K. Paulsen, W. Wells, and R. Barth, "Review of methods for intraoperative margin detection for breast conserving surgery," J. Biomed. Opt. **23**(10), 1–19 (2018).
16. B. Mondal, S. Gao, N. Zhu, G. Sudlow, K. Liang, A. Som, W. Akers, R. Fields, J. Margenthaler, R. Liang, V. Gruev, and S. Achilefu, "Binocular goggle augmented imaging and navigation system provides real-time fluorescence image guidance for tumor resection and sentinel lymph node mapping," Sci. Rep. **5**(1), 12117 (2015).
17. D. Gorpas, J. Phipps, J. Bec, D. Ma, S. Dochow, D. Yankelevich, J. Sorger, J. Popp, A. Bewley, R. Gandour-Edwards, L. Marcu, and D. G. Farwell, "Autofluorescence lifetime augmented reality as a means for real-time robotic surgery guidance in human patients," Sci. Rep. **9**(1), 1187 (2019).
18. J. Bec, J. E. Phipps, D. Gorpas, D. Ma, H. Fatakdawala, K. B. Margulies, J. A. Southard, and L. Marcu, "In vivo label-free structural and biochemical imaging of coronary arteries using an integrated ultrasound and multispectral fluorescence lifetime catheter system," Sci. Rep. **7**(1), 8960 (2017).
19. P. V. Butte, A. N. Mamelak, M. Nuno, S. I. Bannykh, K. L. Black, and L. Marcu, "Fluorescence lifetime spectroscopy for guided therapy of brain tumors," NeuroImage **54**, S125–S135 (2011).
20. D. Gorpas, D. Ma, J. Bec, D. R. Yankelevich, and L. Marcu, "Real-time visualization of tissue surface biochemical features derived from fluorescence lifetime measurements," IEEE Trans. Med. Imag. **35**(8), 1802–1811 (2016).
21. A. Alfonso-Garcia, J. Bec, S. Sridharan Weaver, B. Hartl, J. Unger, M. Bobinski, M. Lechpammer, F. Girgis, J. Boggan, and L. Marcu, "Real-time augmented reality for delineation of surgical margins during neurosurgery using autofluorescence lifetime contrast," J. Biophotonics **13**(1), e201900108 (2020).
22. B. W. Weyers, M. Marsden, T. Sun, J. Bec, A. F. Bewley, R. F. Gandour-Edwards, M. G. Moore, D. G. Farwell, and L. Marcu, "Fluorescence lifetime imaging for intraoperative cancer delineation in transoral robotic surgery," Trans. Biophotonics **1**(1-2), e201900017 (2019).
23. J. Unger, T. Sun, Y. Chen, J. E. Phipps, R. J. Bold, M. A. Darrow, K. Ma, and L. Marcu, "Method for accurate registration of tissue autofluorescence imaging data with corresponding histology: a means for enhanced tumor margin assessment," J. Biomed. Opt. **23**(1), 1–11 (2018).
24. D. R. Yankelevich, D. Ma, J. Liu, Y. Sun, Y. Sun, J. Bec, D. S. Elson, and L. Marcu, "Design and evaluation of a device for fast multispectral time-resolved fluorescence spectroscopy and imaging," Rev. Sci. Instrum. **85**(3), 034303 (2014).
25. J. Liu, Y. Sun, J. Qi, and L. Marcu, "A novel method for fast and robust estimation of fluorescence decay dynamics using constrained least-squares deconvolution with laguerre expansion," Phys. Med. Biol. **57**(4), 843–865 (2012).
26. L. Breiman, "Random forests," Mach. Learn. **45**(1), 5–32 (2001).
27. B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," Biochim. Biophys. Acta **405**(2), 442–451 (1975).
28. H. Boström, "Calibrating random forests," in *2008 Seventh International Conference on Machine Learning and Applications*, (2008), pp. 121–126.
29. J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in large margin classifiers*, (MIT Press, 1999), pp. 61–74.
30. D. Ma, J. Bec, D. Gorpas, D. R. Yankelevich, and L. Marcu, "Technique for real-time tissue characterization based on scanning multispectral fluorescence lifetime spectroscopy (ms-trfs)," Biomed. Opt. Express **6**(3), 987–1002 (2015).
31. J. L. Lagarto, J. E. Phipps, L. Faller, D. Ma, J. Unger, J. Bec, S. Griffey, J. Sorger, D. G. Farwell, and L. Marcu, "Electrocautery effects on fluorescence lifetime measurements: An in vivo study in the oral cavity," J. Photochem. Photobiol., B **185**, 90–99 (2018).

32. M.-G. Lin, T.-L. Yang, C.-T. Chiang, H.-C. Kao, J.-N. Lee, W. Lo, S.-H. Jee, Y.-F. Chen, C.-Y. Dong, and S.-J. Lin, "Evaluation of dermal thermal damage by multiphoton autofluorescence and second-harmonic-generation microscopy," J. Biomed. Opt. **11**(6), 064006 (2006).
33. M. Kaiser, A. Yafi, M. Cinat, B. Choi, and A. Durkin, "Noninvasive assessment of burn wound severity using optical technology: a review of current and future modalities," Burns **37**(3), 377–386 (2011).
34. A. R. Triki, M. B. Blaschko, Y. M. Jung, S. Song, H. J. Han, S. I. Kim, and C. Joo, "Intraoperative margin assessment of human breast tissue in optical coherence tomography images using deep neural networks," Comput. Med. Imag. Grap. **69**, 21–32 (2018).
35. E. Kho, B. Dashtbozorg, L. L. de Boer, K. K. V. de Vijver, H. J. C. M. Sterenborg, and T. J. M. Ruers, "Broadband hyperspectral imaging for breast tumor detection using spectral and spatial information," Biomed. Opt. Express **10**(9), 4496–4515 (2019).
36. J. Shan, S. K. Alam, B. Garra, Y. Zhang, and T. Ahmed, "Computer-aided diagnosis for breast ultrasound using computerized bi-rads features and machine learning methods," Ultrasound Med. Biol. **42**(4), 980–988 (2016).
37. S. A. Boppart, J. Q. Brown, C. S. Farah, E. Kho, L. Marcu, C. M. Saunders, and H. J. C. M. Sterenborg, "Label-free optical imaging technologies for rapid translation and use during intraoperative surgical and tumor margin assessment," J. Biomed. Opt. **23**(2), 1–10 (2017).
38. G. Lu, D. Wang, X. Qin, L. Halig, S. Muller, H. Zhang, A. Chen, B. W. Pogue, Z. G. Chen, and B. Fei, "Framework for hyperspectral image processing and quantification for cancer detection during animal tumor surgery," J. Biomed. Opt. **20**(12), 126012 (2015).
39. L. L. de Boer, B. G. Molenkamp, T. M. Bydlon, B. H. W. Hendriks, J. Wesseling, H. J. C. M. Sterenborg, and T. J. M. Ruers, "Fat/water ratios measured with diffuse reflectance spectroscopy to detect breast tumor boundaries," Breast Cancer Res. Treat. **152**(3), 509–518 (2015).
40. L. de Boer, T. Bydlon, F. van Duijnhoven, M.-J. T. F. D. Vranken-Peeters, C. E. Loo, G. A. O. Winter-Warnars, J. Sanders, H. J. C. M. Sterenborg, B. H. W. Hendriks, and T. J. M. Ruers, "Towards the use of diffuse reflectance spectroscopy for real-time in vivo detection of breast cancer during surgery," J. Transl. Med. **16**(1), 367 (2018).
41. J. E. Phipps, J. Unger, R. Gandour-Edwards, M. G. Moore, A. Beweley, G. Farwell, and L. Marcu, "Head and neck cancer evaluation via transoral robotic surgery with augmented fluorescence lifetime imaging," in *Biophotonics Congress: Biomedical Optics Congress 2018*, (Optical Society of America, 2018), p. CTu2B.3.
42. J. Unger, J. Lagarto, J. Phipps, D. Ma, J. Bec, J. Sorger, G. Farwell, R. Bold, and L. Marcu, "Three-dimensional online surface reconstruction of augmented fluorescence lifetime maps using photometric stereo," in *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XV*, vol. 10054 International Society for Optics and Photonics (SPIE, 2017), p. 65.
43. A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017).
44. S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," Mach. Learn. **79**(1-2), 151–175 (2010).