scientific reports

OPEN

Developing diagnostic assessment of breast lumpectomy tissues using radiomic and optical signatures

Samuel S. Streeter[1]✉, Brady Hunt[1], Rebecca A. Zuurbier[2,4], Wendy A. Wells[3,4], Keith D. Paulsen[1,4] & Brian W. Pogue[1,4]

High positive margin rates in oncologic breast-conserving surgery are a pressing clinical problem. Volumetric X-ray scanning is emerging as a powerful ex vivo specimen imaging technique for analyzing resection margins, but X-rays lack contrast between non-malignant and malignant fibrous tissues. In this study, combined micro-CT and wide-field optical image radiomics were developed to classify malignancy of breast cancer tissues, demonstrating that X-ray/optical radiomics improve malignancy classification. Ninety-two standardized features were extracted from co-registered micro-CT and optical spatial frequency domain imaging samples extracted from 54 breast tumors exhibiting seven tissue subtypes confirmed by microscopic histological analysis. Multimodal feature sets improved classification performance versus micro-CT alone when adipose samples were included (AUC = 0.88 vs. 0.90; *p*-value = 3.65e−11) and excluded, focusing the classification task on exclusively non-malignant fibrous versus malignant tissues (AUC = 0.78 vs. 0.85; *p*-value = 9.33e−14). Extending the radiomics approach to high-dimensional optical data—termed "optomics" in this study—offers a promising optical image analysis technique for cancer detection. Radiomic feature data and classification source code are publicly available.

Breast cancer is the second-leading cause of cancer death among women in the United States[1]. At the time of diagnosis, the disease is most frequently early-stage, localized cancer, which is predominately treated with surgical resection followed by radiation therapy. The goal of the surgical procedure, termed a lumpectomy or breast-conserving surgery (BCS), is to remove the malignancy with a surrounding layer or margin of non-cancerous tissue. An ideal BCS procedure results in cancer-free tissue on the margins (i.e., negative margins), an outcome known to provide the best prognosis[2–5]. Determining the final margin status of a BCS resection takes a day or more to complete. Depending on clinical context, one or more positive margins confirmed by histopathology may necessitate a re-excision procedure to remove residual disease, given the known association of positive margins with ipsilateral breast tumor recurrence[2,6–9]. Today, ~ 20% of BCS procedures require a follow-up re-excision due to positive margins[10–13]. Several studies report the negative health and financial impacts of BCS re-excision procedures to the patient and healthcare system[14–18]. For these reasons, improving intraoperative margin assessment in BCS is a pressing clinical need. In this study, volumetric imaging by X-rays and surface imaging by optical light were combined in a classification pipeline following conventions described in radiomics. The purpose of the study was to demonstrate how this approach might improve diagnostic assessment of malignancy in breast lumpectomy tissues, and in so doing, provide the foundational work needed to extend the approach to margin assessment in the future.

Standard of care margin assessment techniques include gross tissue inspection by the surgeon[19,20], projection X-ray or radiographic specimen imaging[21], and at some medical centers, frozen section pathology[22,23], imprint cytology[24,25], and post-excision cavity shaving[21]. Gross tissue inspection is useful for detecting palpable invasive cancers but is not as effective at detecting primary lesions intertwined with dense fibroglandular tissue or DCIS, which frequently presents as a cluster of macrocalcifications without a localized mass[26,27]. Projection (i.e., two-dimensional) X-ray imaging coupled with intraoperative reading by a radiologist is a mainstay for analyzing margins, and X-rays offer excellent contrast between adipose and fibrous tissue and sensitivity to microcalcifications[28]. Furthermore, recent advances in volumetric X-ray imaging (i.e., computed tomography

[1]Thayer School of Engineering, Dartmouth College, 14 Engineering Dr., Hanover, NH 03755, USA. [2]Department of Radiology, Dartmouth-Hitchcock Medical Center, 1 Medical Center Dr., Lebanon, NH 03756, USA. [3]Department of Pathology and Laboratory Medicine, Dartmouth-Hitchcock Medical Center, 1 Medical Center Dr., Lebanon, NH 03756, USA. [4]Norris Cotton Cancer Center, Dartmouth-Hitchcock Medical Center, 1 Medical Center Dr., Lebanon, NH 03756, USA. ✉email: Samuel.S.Streeter.TH@Dartmouth.edu

and tomosynthesis) have demonstrated sensing of all six anatomical margins of a specimen in a clinically relevant timeframe with high spatial resolution[29–31]. For this reason, volumetric X-ray imaging is emerging as a powerful ex vivo specimen scanning technology. However, a key limitation of X-ray imaging is its inability to differentiate normal, abnormal benign, and malignant fibrous tissues that may be relevant to diagnosis, thereby posing the risk of frequent false positives[28,29]. Intraoperative frozen section pathology and imprint cytology reduce positive margin rates, but the approaches are not widely adopted, because they are resource-intensive (i.e., requiring pathology staff in the operating room or surgical suite during the procedure) and suffer from slow turnaround times[23,32,33]. Post-excision cavity shaving is also effective at reducing positive margin rates, requiring ≤ 10 min in the operating room with minimal impact to patient cosmesis[34,35]. However, positive margin rates after cavity shaving are variable (6–24%)[34,36–41], suggesting alternative or complementary margin assessment techniques are still needed.

Volumetric X-ray imaging coupled with a second imaging modality could be an attractive solution for rapidly analyzing an entire BCS specimen with increased contrast to fibrous tissues on the margin[32,42]. Pradipta et al. reviewed an array of margin assessment techniques[32], ranging from intraoperative ultrasonography, bioimpedance spectroscopy, and ex vivo magnetic resonance imaging to a host of optical techniques, including optical coherence tomography, ultraviolet-photoacoustic microscopy, and fluorescence probes. Each technique offers advantages and disadvantages. DiCorpo et al. reported that the average surface area of a BCS specimen is ~ 45 cm^2 [30]. A limitation of many proposed techniques is their relatively small field of view and/or long scan time, such that analyzing the entire surface of a specimen would be too time-consuming for clinical translation. Other limitations include: the need for exogenous contrast agents that must preferentially accumulate in target tissues and be proven safe for human use; or lack of evidence demonstrating efficacy when analyzing realistic BCS specimens, which can be amorphous, variable in size, and heterogeneous with a combination of adipose, fibroglandular, and potentially malignant tissues on the margin[42].

In this study, breast tumor tissues were imaged with spatially co-registered micro-computed tomography (micro-CT) scanning and multi-wavelength spatial frequency domain imaging (SFDI). SFDI is a wide-field, noncontact, and rapid optical imaging modality first introduced by Cuccia et al.[43]. The technique involves the projection of one-dimensional, sinusoidal patterns of light (each at a discrete spatial frequency) onto the tissue surface and acquisition of the reflected light. At each spatial frequency, three phase-shifted patterns are projected and imaged. Each set of three phase-shifted images creates a demodulated reflectance map corresponding to the wavelength(s) and spatial frequency of illumination. Additional background information related to the optical imaging modality is available (Supplementary Material Appendix 1), and the interested reader is directed to a recent review of the modality[44].

Previous studies have coupled machine learning models with SFDI data to classify or predict different breast tissues. One study used SFDI-derived optical scattering properties of breast tissue and an explicit mathematical model to predict the epithelial, stromal, and adipose fractions of breast tumor samples[45]. Another study applied high spatial frequency, monochromatic SFDI and a limited set of textural features to classify pairs of breast tumor subtypes using a support vector machine classifier with accuracies ranging from 55 to 95%[46]. Deep learning methods have also determined optical properties from raw SFDI data[47–49]. The use of a limited set of image features may not probe the image data sufficiently to extract the most useful signatures for differentiating tissues. On the other hand, the complexity of deep learning methods hinders model interpretation and thus limits potential for clinical translation. To address these issues, this study used a supervised machine learning pipeline based on a large number of Image Biomarker Standardization Initiative (IBSI)-compliant radiomic features[50]. The pipeline tested the ability of micro-CT alone, SFDI alone, and the combination of micro-CT and SFDI data to classify malignant and non-malignant image samples extracted from wide field-of-view images of breast tumors.

Radiomics involves the quantification of many image features, mining the features to determine diagnostic signatures not readily discerned by visual inspection, and subsequently using the features to build classification models to inform clinical decision making[51,52]. The "radiomics approach" is frequently applied to conventional medical imaging modalities (e.g., CT, magnetic resonance imaging, positron emission tomography). Applying it to optical imaging data is termed "optomics" here, extending the "omic" concept to image features extracted from wide field-of-view optical images. The goals of this study were to: first, determine whether combining X-ray micro-CT and optical SFDI image data improves malignancy classification relative to micro-CT alone based on an "omics" approach; and second, identify the most useful radiomic and optomic features for classifying malignant breast tissues.

Portions of the data collected through this imaging protocol have been analyzed in previous studies[46,49,53,54]. Specifically, SFDI data have demonstrated statistical differences between normal, abnormal benign, and malignant tissue subtypes based on color analysis and diffuse optical properties[53], and separately, using texture analysis of monochromatic (i.e., 490 nm), high spatial frequency (1.37 mm^{-1}) data[46]. SFDI data were also used in a deep learning framework, through which optical properties were approximated directly from raw SFDI data[49]. Monochromatic, high frequency SFDI and micro-CT data were also used in a comparative study to quantify differences in the wide field-of-view optical and micro-CT images[54]. The study found that optical imaging revealed intra-tumoral morphology and malignant-fibrous tissue boundaries that were occult to micro-CT scanning. The present work extends these prior contributions significantly with two key advances: first, by incorporating multi-wavelength, multi-spatial frequency SFDI *and* micro-CT data into the analysis; and second, by evaluating the image data with a supervised machine learning pipeline to quantify binary malignancy classification performance.
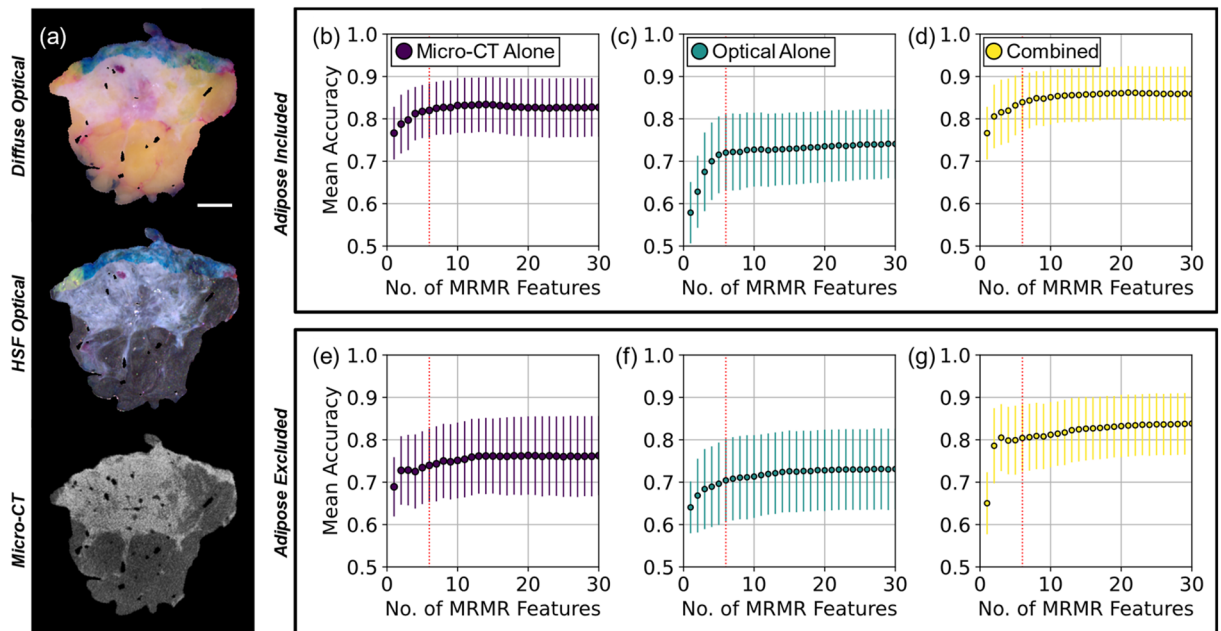
Figure 1. Mean classification accuracy derived from 5×5 mm sub-image samples as a function of the number of optimal MRMR features used in $n = 1000$ Monte Caro cross-validation splits. (**a**) Diffuse optical reflectance, high spatial frequency (HSF, 1.37 mm^{-1}) optical reflectance, and a micro-CT slice of a representative tissue specimen with 1-cm scalebar. Mean accuracies when adipose tissue was included (top row, **b**–**d**) and excluded (bottom row, **e**–**g**). Dashed red vertical lines mark six features, the threshold at which the percent change in mean accuracy dropped below 1% for combined classification in (**d**) and (**g**). Error bars represent ± one standard deviation.

Results

Classification using radiomics and optomics. Adipose was the most frequent tissue type in the breast tumor dataset (Supplementary Material Table S1). Relative to fibroglandular and epithelial components, adipose tissue is straightforward to identify by gross surgical inspection and is readily evident with micro-CT scanning[29,54]. To test the hypothesis that including adipose samples improves performance, the classification pipeline focused on two scenarios: one that included adipose tissue samples and one that excluded them.

Figure 1a shows wide field-of-view optical and micro-CT imagery of a representative tissue specimen. Non-malignant versus malignant classification accuracy is shown for cases when adipose tissue was included (Fig. 1b–d) and excluded (Fig. 1e–g). Accuracies are plotted with respect to the number of optimal features selected by minimum redundancy, maximum relevance (MRMR)[55]. Sub-image samples of tissue ranged in size from 2×2 to 5×5 mm and were extracted from wide field-of-view regions of interest (ROIs) defining distinct tissue subtypes in each specimen. Only results from 5×5 mm sub-image samples are presented here, because this sub-image size provided the best classification performance overall (Supplementary Material Appendix 2 and Fig. S1). A 1% change in average accuracy determined an appropriate minimum number of features to use in each case. Based on the combined radiomic/optomic classification accuracy (Fig. 1d,g), the 1% change in average accuracy threshold required six features, when adipose tissue was both included and excluded. Therefore, subsequent analysis focused on results derived from 5×5 mm sub-image samples and with six radiomic/optomic features. Supplementary Materials Table S2 reports classification performance (i.e., accuracy, recall, precision, receiver operating characteristic (ROC) area under the curve (AUC)) for 5×5 mm sub-image samples and six total features. When adipose tissue was included and using six features, mean accuracies achieved by micro-CT features alone, optical features alone, and combined micro-CT and optical features were 82% (Fig. 1b), 72% (Fig. 1c), and 84% (Fig. 1d), respectively. When adipose tissue was excluded, mean accuracies decreased to 74% (Fig. 1e), 70% (Fig. 1f), and 80% (Fig. 1g) for the same respective feature sets.

Figure 2 compares average ROC curves of optimized classifiers for micro-CT alone, optical alone, and combined feature sets. Classification performance decreased when adipose tissue samples were excluded (compare Fig. 2a with d, Fig. 2b with e, and Fig. 2c with f). Micro-CT data alone performed better than optical data alone when adipose samples were included (AUC = 0.88 vs. 0.78, $p = 3.30e-13$), but the two performed similarly when adipose samples were excluded (AUC = 0.78 vs. 0.75, $p = 0.42$). Classification performance using both modalities was better than using either modality alone (e.g., when adipose tissue was included: micro-CT vs. combined, AUC = 0.88 vs. 0.90, $p = 3.65e-11$; optical vs. combined, AUC = 0.75 vs. 0.90, $p = 3.60e-36$). Improvement in combined feature performance relative to micro-CT alone was more pronounced when adipose samples were excluded (micro-CT vs. combined, AUC = 0.78 vs. 0.85, $p = 9.33e-14$). ROC curves reflect relatively high variance depicted by one standard deviations (shaded regions) and 95% confidence bands (dashed lines), providing a visual depiction of how individual Monte Carlo CV splits performed.
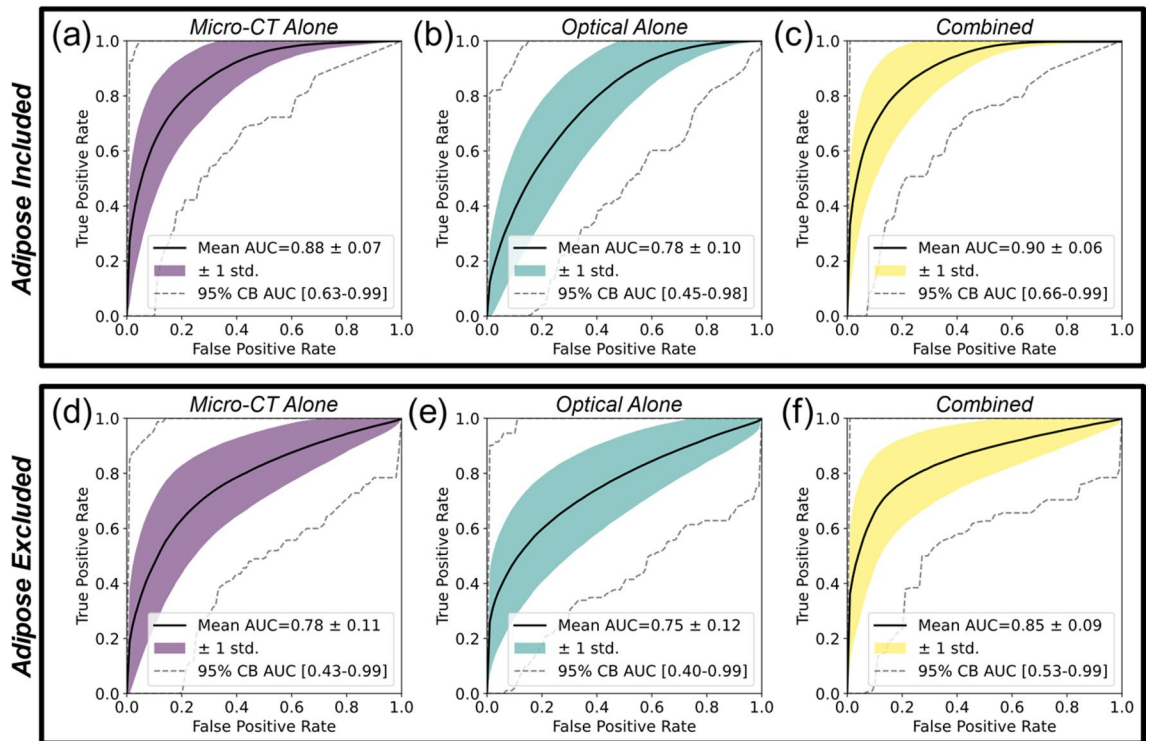
Figure 2. ROC curve analysis based on six radiomic/optomic features derived from 5×5 mm sub-image samples. Adipose tissue was included (top row) and excluded (bottom row). Each shaded region depicts two-dimensional one standard deviation from the mean ROC curve (solid black). AUC 95% confidence bands (CBs) contain 95% of the $n = 1000$ ROC curves that fell closest to the mean curve in each subplot (dashed black).
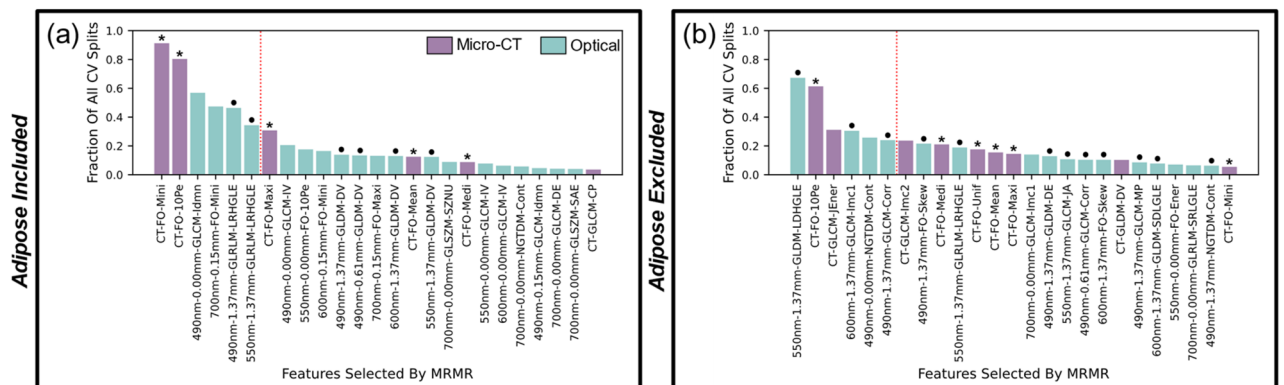

Figure 3. Most frequently selected features by MRMR using combined radiomic/optomic features derived from 5×5 mm sub-image samples when adipose tissue samples were (**a**) included and (**b**) excluded. These subplots relate directly to classification performances shown in Fig. 2c,f, respectively. Vertical axes signify the fraction of all $n = 1000$ Monte Carlo CV splits that MRMR selected each feature. Asterisks (*) signify first-order micro-CT features. Dots (•) signify high spatial frequency (0.61 or 1.37 mm^{-1}) reflectance, SFDI-derived features. Twenty-five features are listed in each subplot, though additional features were selected less frequently. A complete listing of radiomic/optomic feature abbreviations can be found in Supplementary Material Appendix 5.3.

Selected feature importance. The radiomic and optomic features selected by MRMR in each of $n = 1000$ splits were tabulated. Figure 3 shows the distribution of these features for combined data classification and high-lights the fact that both micro-CT and SFDI features were selected for inclusion in the optimal subset of features. This trend was true both when adipose tissue samples were included (Fig. 3a) and excluded (Fig. 3b). (These features correspond to the classification performances shown in Fig. 2c,f, respectively.) Two trends can be inferred from Fig. 3. First, the most important micro-CT radiomic features were first-order histogram statistics (noted by asterisks in the figure), which contain intensity information alone (i.e., no spatial information). This trend was true independent of whether adipose tissue was included. Second, important SFDI optomic features were gener-
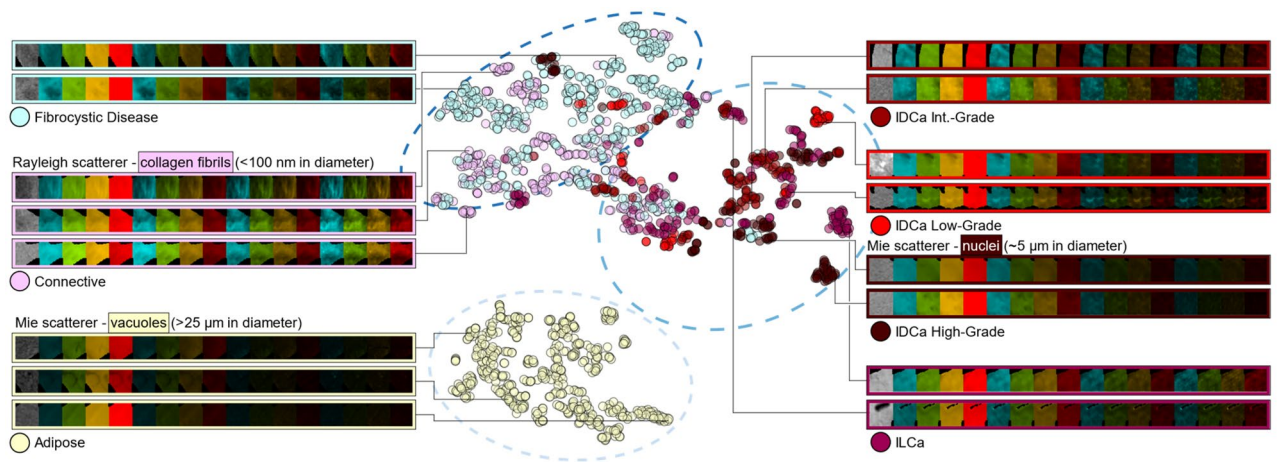
Figure 4. T-distributed stochastic neighbor embedding using six optimal MRMR radiomic/optomic features (i.e., top features shown in Fig. 3b). Dashed ellipses qualitatively delineate three clusters of samples, which can be related to hierarchical clusters depicted in Supplementary Material Fig. S2. Image channel labels show the micro-CT sub-image on the left (grayscale), followed by calibrated reflectance optical channels with spatial frequency increasing from left to right (0.00, 0.15, 0.61, then 1.37 mm^{-1}). Coloring of optical channels illustrates the wavelength of light, increasing from left to right (490, 550, 600, then 700 nm) and repeated for each spatial frequency. IDCa = invasive ductal carcinoma. ILCa = invasive lobular carcinoma.

ally derived from high spatial frequency reflectance (illumination frequencies of 0.61 and 1.37 mm^{-1}; noted by dots in the figure), particularly when adipose tissue was omitted from the classification task. Notably, all but two of the high spatial frequency optomic features in Fig. 3 were derived from second- and higher-order pixel statistics, which depend on the spatial relationships between multiple pixels, and thus, contain textural information.

A t-distributed stochastic neighbor embedding plot, shown in Fig. 4, reduces the dataset to a two-dimensional embedding of only six optimal radiomic/optomic features (using the six most frequently selected features shown in Fig. 3b). Each sub-image sample is color-coded by tissue subtype, and representative samples are labeled with the associated micro-CT and optical image channels. Figure 4 suggests separability of normal (i.e., connective tissue) and abnormal benign (i.e., fibrocystic disease) samples from malignant tissue subtypes is possible when only six radiomic/optomic features are used. Adipose tissue is also clustered effectively under these conditions.

High-dimensional radiomic datasets are frequently visualized using hierarchically clustered heatmaps[51,52,56], and such a visualization is available for the dataset in this study (Supplementary Material Appendix 3 and Fig. S2). Notably, the global structure shown in Fig. S4 is maintained by Fig. 4 using six instead of 1564 radiomic/optomic features. The image channel labels in Fig. 4 exemplify representative tissue subtype image signatures: first, adipose tissue yields relatively low linear attenuation coefficient values in the micro-CT sub-image relative to all other subtypes[54]; second, adipose tissue and high-grade invasive ductal carcinoma are predominantly characterized by Mie scattering vacuoles (> 25 μm in diameter) and nuclei (~ 5 μm in diameter), respectively, explaining the relatively low high spatial frequency optical reflectance for these samples[57]; and third, connective tissue and fibrocystic disease contain collagen fibril structures (< 100 μm in diameter) that are strong Rayleigh-type scatterers, giving rise to an increased backscatter signal in these optical sub-images[57].

Discussion

Analyses indicate overall mean accuracy and AUC achieved were 84% and 0.90 (80% and 0.85 without adipose tissue), respectively, using only six optimal radiomic/optomic features. These results appear to be the first application of radiomics to multi-wavelength, multi-spatial SFDI data and represent a novel approach for validating and extending radiomic feature analysis to optical imaging data. They provide a proof of principle for how radiomic and optomic features can be combined to improve overall classification accuracy. Other important contributions from this study include appropriate image data normalization to linear attenuation coefficient and calibrated reflectance values and use of IBSI-compliant image features in the classification pipeline, considerations that lend to study rigor and reproducibility. Finally, use of defined image features within a supervised machine learning pipeline provides increased model interpretability relative to deep learning methods of tissue classification.

As described in the Introduction, previous studies used portions of the same breast cancer dataset used in this study. The new method revealed in this work relative to previous studies is multimodal radiomic/optomic analysis. The radiomics approach is often deployed with other imaging modalities relevant to the diagnosis and monitoring of breast cancer, mainly magnetic resonance imaging, CT, positron emission tomography, and/or ultrasound. Most studies demonstrate radiomics utility in identifying malignant lesions in the setting of preoperative breast cancer diagnosis, and recently published classification performance AUCs range from 0.57 to 0.98 for this task[58,59]. The work described here suggests that wide field-of-view optical images of breast tumors likely contain useful image features, especially textural information, relevant to diagnosis. The optomics approach provides an alternative to optical property quantification, a process that requires tissue model assumptions and can be computationally intensive to implement (e.g., ~ 1 h for a 2.5 cm × 2.5 cm tissue sample[57]). The study also

demonstrates that the combined radiomics/optomics approach is synergistic for automated diagnostic assessment of breast tissues. Additional findings revealed in this work relative to previous studies include the fact that useful optical features can be extracted from a wide wavelength range (i.e., optimal features span 490–700 nm in Fig. 3a and 490–600 nm in Fig. 3b) and the entire spatial frequency range in the dataset (i.e., optimal features span 0.00–1.37 mm^{-1} in Fig. 3a,b). Notably, the top six features selected by MRMR differed significantly depending on whether adipose tissue was included in the classification task. When adipose was included (Fig. 3a), the top two features were micro-CT first-order histogram statistics. This result is expected given the stark difference in radiodensity between adipose and all other tissues in the breast cancer specimens; purely differences in pixel intensity were sufficient to separate adipose from all other tissues. When adipose was excluded (Fig. 3b), three of the top six features were textural features derived from high spatial frequency optical reflectance, suggesting that this feature type provides value for differentiating tissues that exhibit similar radiodensities. Nevertheless, the second most important feature when adipose tissue was excluded was a micro-CT first-order statistic, reinforcing the claim that micro-CT and optical imaging provide complementary contrast regardless of the presence of adipose tissue.

Importantly, improvements in margin assessment were not directly demonstrated in this study. Rather, freshly resected breast tumor slices were imaged to capture a range of tumor pathologies, and the focus was building differential diagnostic models. It is also necessary to distinguish cancer tissue on the margin that has been thermally denatured or burned by surgical tools during resection, processes known to alter tissue optical properties[60]. However, this limitation was not addressed in this study. Only invasive cancers were analyzed due to a lack of pre-invasive DCIS specimens in the dataset. This limitation is significant, given that relative to other malignant tissue subtypes, DCIS is responsible for one of the largest shares of positive margins leading to re-excision procedures[13]. Future studies should include DCIS samples to demonstrate efficacy in classifying this important pre-invasive subtype. Finally, future work should also involve radiomic/optomic analysis of intact BCS specimens, taking advantage of the volumetric sensing of micro-CT in combination with surface-mapped optical reflectance to evaluate the performance of this approach for sensing cancer-positive tissues at the margin. For example, the radiomic feature quantification package used in this study, PyRadiomics[56], offers a suite of standardized, IBSI-compliant 3D shape and voxel-based radiomic features that could be extracted from sub-volumes of the micro-CT scan. Connected sub-volumes that exhibit malignant/suspicious radiomic signatures that extend to the tissue margin could then be correlated with optomic signatures from the surface tissue to identify potentially involved margins.

This study introduces an optomics paradigm for analyzing high dimensional optical image data and represents a direct and quantitative assessment of the extent to which coupling micro-CT scanning with optical imaging improves classification of malignant breast tumor tissues. Volumetric X-ray imaging is complimentary to surface SFDI, which provides additional sensing of surface tissue subtypes, especially those that are fibrous and exhibit similar radiodensities. These results warrant further research into the combination of specimen X-ray imaging with wide field-of-view, noncontact optical imaging, such as SFDI, for potentially improving intraoperative margin assessment.

Materials and methods

Breast tumor imaging protocol. All methods were carried out in accordance with relevant guidelines and regulations. Specimen imaging was performed at the Dartmouth-Hitchcock Medical Center (DHMC) in Lebanon, New Hampshire. The imaging protocol was approved by the Committee for the Protection of Human Subjects, the Institutional Review Board at Dartmouth College, and all aspects of the study followed the approved protocol. Tissue specimens were procured from patients electing BCS at DHMC who participated in the study under informed consent. Imaging was performed post-operatively during standard of care pathological processing in the specimen grossing laboratory and did not impact tissue processing or diagnostic reporting in any way. One representative, ~ 5-mm thick slice from each BCS tumor was selected by an experienced Pathologists' Assistant for imaging and was de-identified and referenced only by a unique accessioning number. Each imaged slice exhibited a clear cross section of the primary tumor and surrounding tissues, thereby revealing clear regions of several normal, abnormal benign, and malignant breast tissue subtypes. Each slice was firmly positioned between clear acrylic plates, creating a flat tissue surface that mitigated specular reflection and demodulation artifacts in the SFDI data. The top surface of each imaged slice underwent standard of care specimen processing, sectioning, staining with hematoxylin and eosin, and microscopic analysis by a board-certified breast pathologist (WAW). Histologic slides corresponding to the imaged tissue surface underwent whole slide, high resolution digital imaging and were then mosaicked together to confirm wide-field tissue subtype ROIs. These microscopic ROIs were co-registered to the wide field-of-view micro-CT and optical imagery.

Imaging system. Imaging was performed with a customized IVIS SpectrumCT system (PerkinElmer, Hopkinton, MA) containing a cone-beam CT in a "pancake" geometry and retrofitted optical imaging components[61]. Micro-CT scans were acquired with X-ray tube settings of 50 kVp and 1 mA with an exposure time of 100 ms/projection for a total of 720 projections. The reconstructed scan volume was $12 \times 12 \times 3$ cm^3 with 150 μm^3 voxels, and the combined acquisition and reconstruction time was ~ 4 min. These settings were the same as those used in a previous micro-CT BCS specimen study[29]. The SFDI acquisition leveraged the charged coupled device camera native to the IVIS SpectrumCT system (Andor iKon, Andor Technologies Ltd, Belfast, UK) and a digital light projector (CEL5500 Fiber, Digital Light Innovations Inc., Austin, TX) retrofitted in the light-tight, fully shielded imaging cabinet for projecting the structured light patterns. The light source was a supercontinuum laser (SuperK Blue, NKT Photonics, Denmark) with a tunable line filter (SuperK Varia, NKT Photonics, Denmark). SFDI acquisition involved optical wavelengths of 490, 550, 600, and 700 nm, and illumination spatial
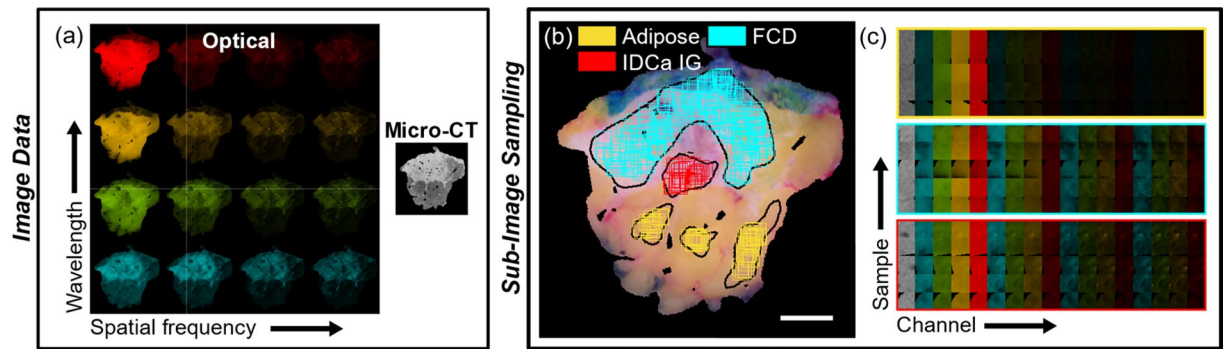
Figure 5. Overview of tissue imaging and sampling protocol. (**a**) Visualization of optical data collection at four wavelengths and four spatial frequencies along with micro-CT scanning of a representative tissue specimen. (**b**) Sub-image sampling from histologically confirmed regions of distinct tissue subtypes, and (**c**) sub-image samples grouped by tissue subtype (color borders) containing 17 channels of image data (16 optical, one micro-CT). FCD = fibrocystic disease. IDCa IG = invasive ductal carcinoma intermediate-grade.

Tissue subtype	Specimen count	ROI count	ROI Size (average ± 1 std mm²)
Healthy and abnormal benign			
Adipose tissue	49	93	43 ± 32
Normal connective tissue	13	20	85 ± 66
Fibrocystic disease	14	22	112 ± 133
Invasive cancer			
IDCa low-grade	10	10	46 ± 33
IDCa intermediate-grade	13	13	59 ± 56
IDCa high-grade	10	10	82 ± 58
Invasive lobular carcinoma	9	9	99 ± 102

Table 1. Breast tumor specimen and ROI totals by tissue subtype in this study. *IDCa* invasive ductal carcinoma.

frequencies of 0.00, 0.15, 0.61, and 1.37 mm⁻¹. Thus, 16 unique wavelength-spatial frequency reflectance images were collected per specimen. SFDI acquisition and reconstruction time was ~ 8 min. Combined with the surface tissue micro-CT slice, every specimen in the dataset had 17 channels of image data (Fig. 5a).

Image data and pre-processing. A total of 70 specimens were imaged under the approved protocol. Sixteen cases failed to meet data analysis eligibility criteria: five were excluded due to inconsistent micro-CT scan settings, six were excluded due to ambiguous histology co-registrations, one was excluded due to a small cross-sectional area (< 2 cm²), and four were omitted, because they presented tissue subtypes that were represented in fewer than three specimens in the dataset. Thus, 54 tumor specimens from 54 BCS patients were analyzed. From these specimens, 177 ROIs were drawn, isolating regions of seven histologically confirmed breast tissue subtypes. Of the 54 specimens, 14 contained only normal and/or abnormal benign tissues, three contained only malignant tissue, and 37 contained both normal or abnormal benign and malignant tissues. Table 1 summarizes the number of breast tumor specimens and the number of ROIs categorized by tissue subtype. The number of sub-images extracted from each ROI was proportional to the size of each ROI, providing an approximate baseline by which all ROIs were equally sampled. Figure 5b illustrates the process of sub-image sampling from the wide field-of-view tissue images. Supplementary Material Table S1 contains totals of sub-image samples categorized by subtype and by sub-image size.

Calibrated reflectance maps were generated at each of 16 wavelength-spatial frequency settings using a previously described normalization process with a reflectance standard[61]. Micro-CT scans were converted to linear attenuation coefficient values based on a 50 kVp X-ray energy. All image data were masked to remove regions with poor tissue coupling to the top acrylic plate, thereby restricting the image analysis to flat, consistent tissue surfaces. Micro-CT scanning and SFDI were completed sequentially, without moving the tissue or acrylic plates between acquisitions, facilitating spatial co-registration between modalities. Tissue subtype ROIs were sampled over a range of square sub-image sizes (2 × 2–5 × 5 mm).

Radiomics package and classification pipeline. The classification pipeline (summarized in Supplementary Material Fig. S3) performed binary stratification of samples as malignant or non-malignant based on radiomic and/or optomic features. Fornacon-Wood et al. recently reviewed several popular radiomic feature quantification software packages, identifying strengths and weaknesses of each package[62]. Here, the pipeline

was implemented in the Python coding language (v3.7.9) using the PyRadiomics package (v3.0.1) for feature quantification, given that it is IBSI-compliant for reproducibility, free of cost, and open-source[56]. The pipeline involved MRMR feature selection, a random forest classifier, and Monte Carlo cross-validation with $n = 1000$ splits partitioned on the patient-level. Additional details related to the classification pipeline are available (Supplementary Material Appendix 4 and Fig. S3).

Features were quantified from only micro-CT data (1 channel), only SFDI data (16 channels), and combined data (17 channels) (Fig. 5c). A total of 92 PyRadiomic features were extracted from each image channel. Thus, up to 1,564 features were quantified from each sub-image sample. Quantified features included a range of first-, second-, and higher order pixel statistics. Supplementary Material Appendix 5 contains additional information about the features quantified, including fixed bin width considerations (Appendix 5.1, Fig. S4) and a complete list of features and associated abbreviations (Appendix 5.2).

Statistical analysis. For every combination of pipeline parameters, ROC curves ($n = 1000$) were generated through Monte Carlo CV. Average ROC curves were derived by interpolating true positive values to a constant range of false positives across all splits. DeLong's test determined whether differences between pairs of average ROC AUC values were statistically significant[63,64]. A p-value ≤ 0.05 was considered significant.

Data availability

Image data in the form of comma-separated value files, classification pipeline source code, and a PyRadiomics parameter file documenting all feature quantification settings are available with this publication in an open-source repository (https://github.com/optmed/radiomics-optomics).

References

1. DeSantis, C. E. *et al.* Breast cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 438–451 (2019).
2. Park, C. C. *et al.* Outcome at 8 years after breast-conserving surgery and radiation therapy for invasive breast cancer: Influence of margin status and systemic therapy on local recurrence. *J. Clin. Oncol.* **18**, 1668–1675 (2000).
3. Cowen, D. *et al.* Local and distant failures after limited surgery with positive margins and radiotherapy for node-negative breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **47**, 305–312 (2000).
4. Fisher, B. *et al.* Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N. Engl. J. Med.* **347**, 1233–1241 (2002).
5. Houssami, N., Macaskill, P., Marinovich, M. L. & Morrow, M. The association of surgical margins and local recurrence in women with early-stage invasive breast cancer treated with breast-conserving therapy: A meta-analysis. *Ann. Surg. Oncol.* **21**, 717–730 (2014).
6. Schnitt, S. J. *et al.* The relationship between microscopic margins of resection and the risk of local recurrence in patients with breast cancer treated with breast-conserving surgery and radiation therapy. *Cancer* **74**, 1746–1751 (1994).
7. Smitt, M. C. *et al.* The importance of the lumpectomy surgical margin status in long-term results of breast conservation. *Cancer* **76**, 259–267 (1995).
8. Smitt, M. C., Nowels, K., Carlson, R. W. & Jeffrey, S. S. Predictors of reexcision findings and recurrence after breast conservation. *Int. J. Radiat. Oncol. Biol. Phys.* **57**, 979–985 (2003).
9. Leong, C. *et al.* Effect of margins on ipsilateral breast tumor recurrence after breast conservation therapy for lymph node-negative breast carcinoma. *Cancer* **100**, 1823–1832 (2004).
10. McCahill, L. E. *et al.* Variability in reexcision following breast conservation surgery. *JAMA* **307**, 467–475 (2012).
11. Landercasper, J., Whitacre, E., Degnim, A. C. & Al-Hamadani, M. Reasons for re-excision after lumpectomy for breast cancer: Insight from the American Society of Breast Surgeons Mastery(SM) database. *Ann. Surg. Oncol.* **21**, 3185–3191 (2014).
12. Kaczmarski, K. *et al.* Surgeon re-excision rates after breast-conserving surgery: A measure of low-value care. *J. Am. Coll. Surg.* **228**, 504-512.e2 (2019).
13. Landercasper, J. *et al.* Factors associated with reoperation in breast-conserving surgery for cancer: A prospective study of American society of breast surgeon members. *Ann. Surg. Oncol.* **26**, 3321–3336 (2019).
14. Moran, M. S. *et al.* Society of Surgical Oncology-American Society for Radiation Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in stages I and II invasive breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* **88**, 553–564 (2014).
15. Wazer, D. E. *et al.* Factors influencing cosmetic outcome and complication risk after conservative surgery and radiotherapy for early-stage breast carcinoma. *J. Clin. Oncol.* **10**, 356–363 (1992).
16. King, T. A. *et al.* Clinical management factors contribute to the decision for contralateral prophylactic mastectomy. *JCO* **29**, 2158–2164 (2011).
17. Heil, J. *et al.* Do reexcisions impair aesthetic outcome in breast conservation surgery? Exploratory analysis of a prospective cohort study. *Ann. Surg. Oncol.* **19**, 541–547 (2012).
18. Abe, S. E. *et al.* Margin re-excision and local recurrence in invasive breast cancer: A cost analysis using a decision tree model. *J. Surg. Oncol.* **112**, 443–448 (2015).
19. Cabioglu, N. *et al.* Role for intraoperative margin assessment in patients undergoing breast-conserving surgery. *Ann. Surg. Oncol.* **14**, 1458–1471 (2007).
20. Nunez, A., Jones, V., Schulz-Costello, K. & Schmolze, D. Accuracy of gross intraoperative margin assessment for breast cancer: Experience since the SSO-ASTRO margin consensus guidelines. *Sci. Rep.* **10**, 17344 (2020).
21. Landercasper, J. *et al.* Toolbox to reduce lumpectomy reoperations and improve cosmetic outcome in breast cancer patients: The American society of breast surgeons consensus conference. *Ann. Surg. Oncol.* **22**, 3174–3183 (2015).
22. Olson, T. P., Harter, J., Muñoz, A., Mahvi, D. M. & Breslin, T. Frozen section analysis for intraoperative margin assessment during breast-conserving surgery results in low rates of re-excision and local recurrence. *Ann. Surg. Oncol.* **14**, 2953–2960 (2007).
23. Garcia, M. T. *et al.* Accuracy of frozen section in intraoperative margin assessment for breast-conserving surgery: A systematic review and meta-analysis. *PLoS ONE* **16**, e0248768 (2021).
24. Bakhshandeh, M., Tutuncuoglu, S. O., Fischer, G. & Masood, S. Use of imprint cytology for assessment of surgical margins in lumpectomy specimens of breast cancer patients. *Diagn. Cytopathol.* **35**, 656–659 (2007).
25. Esbona, K., Li, Z. & Wilke, L. G. Intraoperative imprint cytology and frozen section pathology for margin assessment in breast conservation surgery: A systematic review. *Ann. Surg. Oncol.* **19**, 3236–3245 (2012).

26. Schnitt, S. J. & Connolly, J. L. Processing and evaluation of breast excision specimens: A clinically oriented approach. *Am. J. Clin. Pathol.* **98**, 125–137 (1992).
27. Sahin, A. A. Chapter 34: Surgical Margin Evaluation in Patients Treated with Breast-Conserving Therapy. in *Advanced Therapy of Breast Disease* 341–384 (BC Decker Inc., 2004).
28. Kopans, D. *Breast Imaging* (Lippincott Williams & Wilkins, 2006).
29. McClatchy, D. M. 3rd., Zuurbier, R. A., Wells, W. A., Paulsen, K. D. & Pogue, B. W. Micro-computed tomography enables rapid surgical margin assessment during breast conserving surgery (BCS): Correlation of whole BCS micro-CT readings to final histo-pathology. *Breast Cancer Res. Treat* **172**, 587–595 (2018).
30. DiCorpo, D. *et al.* The role of Micro-CT in imaging breast cancer specimens. *Breast Cancer Res. Treat* https://doi.org/10.1007/s10549-020-05547-z (2020).
31. Kulkarni, S. A. *et al.* High-resolution full-3D specimen imaging for lumpectomy margin assessment in breast cancer. *Ann. Surg. Oncol.* https://doi.org/10.1245/s10434-021-10499-9 (2021).
32. Pradipta, A. R. *et al.* Emerging technologies for real-time intraoperative margin assessment in future breast-conserving surgery. *Adv. Sci. (Weinh)* **7**, 1901519–1901519 (2020).
33. St John, E. R. *et al.* Diagnostic accuracy of intraoperative techniques for margin assessment in breast cancer surgery: A meta-analysis. *Ann. Surg.* **265**, 300–310 (2017).
34. Chagpar, A. B. *et al.* A randomized, controlled trial of cavity shave margins in breast cancer. *N. Engl. J. Med.* **373**, 503–510 (2015).
35. Wang, K., Ren, Y. & He, J. Cavity shaving plus lumpectomy versus lumpectomy alone for patients with breast cancer undergoing breast-conserving surgery: A systematic review and meta-analysis. *PLoS ONE* **12**, e0168705–e0168705 (2017).
36. Kobbermann, A. *et al.* Impact of routine cavity shave margins on breast cancer re-excision rates. *Ann. Surg. Oncol.* **18**, 1349–1355 (2011).
37. Marudanayagam, R. *et al.* Effect of cavity shaving on reoperation rate following breast-conserving surgery. *Breast J.* **14**, 570–573 (2008).
38. Unzeitig, A. *et al.* Influence of surgical technique on mastectomy and reexcision rates in breast-conserving therapy for cancer. *Int. J. Surg. Oncol.* **2012**, 725121 (2012).
39. Janes, S. E. J., Stankhe, M., Singh, S. & Isgar, B. Systematic cavity shaves reduces close margins and re-excision rates in breast conserving surgery. *Breast* **15**, 326–330 (2006).
40. Huston, T. L., Pigalarga, R., Osborne, M. P. & Tousimis, E. The influence of additional surgical margins on the total specimen volume excised and the reoperative rate after breast-conserving surgery. *Am. J. Surg.* **192**, 509–512 (2006).
41. Jones, V. *et al.* Excising additional margins at initial breast-conserving surgery (BCS) reduces the need for re-excision in a predominantly african american population: A report of a randomized prospective study in a public hospital. *Ann. Surg. Oncol.* **23**, 456–464 (2016).
42. Maloney, B. W. *et al.* Review of methods for intraoperative margin detection for breast conserving surgery. *J. Biomed. Opt.* **23**, 1–19 (2018).
43. Cuccia, D. J., Bevilacqua, F., Durkin, A. J. & Tromberg, B. J. Modulated imaging: Quantitative analysis and tomography of turbid media in the spatial-frequency domain. *Opt. Lett.* **30**, 1354–1356 (2005).
44. Angelo, J. P. *et al.* Review of structured light in diffuse optical imaging. *J. Biomed. Opt.* **24**, 1–20 (2018).
45. McClatchy, D. M. *et al.* Light scattering measured with spatial frequency domain imaging can predict stromal versus epithelial proportions in surgically resected breast tissue. *J. Biomed. Opt.* **24**, 1–11 (2018).
46. Streeter, S. S. *et al.* Structured light imaging for breast-conserving surgery, part II: Texture analysis and classification. *J. Biomed. Opt.* **24**, 096003 (2019).
47. Zhao, Y. *et al.* Deep learning model for ultrafast multifrequency optical property extractions for spatial frequency domain imaging. *Opt. Lett.* **43**, 5669–5672 (2018).
48. Chen, M. T., Mahmood, F., Sweer, J. A. & Durr, N. J. GANPOP: Generative adversarial network prediction of optical properties from single snapshot wide-field images. *IEEE Trans. Med. Imaging* **39**, 1988–1999 (2020).
49. Pardo, A. *et al.* Modeling and synthesis of breast cancer optical property signatures with generative models. *IEEE Trans. Med. Imaging* **PP**, (2021).
50. Zwanenburg, A. *et al.* The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* **295**, 328–338 (2020).
51. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, 4006 (2014).
52. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images are more than pictures, they are data. *Radiology* **278**, 563–577 (2016).
53. Maloney, B. W. *et al.* Structured light imaging for breast-conserving surgery, part I: Optical scatter and color analysis. *J. Biomed. Opt.* **24**, 096002 (2019).
54. Streeter, S. S. *et al.* Optical scatter imaging of resected breast tumor structures matches the patterns of micro-computed tomography. *Phys. Med. Biol.* **66**, 115021 (2021).
55. Radovic, M., Ghalwash, M., Filipovic, N. & Obradovic, Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinform.* **18**, 9 (2017).
56. van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
57. McClatchy, D. M. *et al.* Wide-field quantitative imaging of tissue microstructure using sub-diffuse spatial frequency domain imaging. *Optica* **3**, 613–621 (2016).
58. Conti, A., Duggento, A., Indovina, I., Guerrisi, M. & Toschi, N. Radiomics in breast cancer classification and prediction. *Semin. Cancer Biol.* **72**, 238–250 (2021).
59. Sollini, M. *et al.* PET/CT radiomics in breast cancer: Mind the step. *Methods* **188**, 122–132 (2021).
60. Nguyen, J. Q. *et al.* Spatial frequency domain imaging of burn wounds in a preclinical model of graded burn severity. *J. Biomed. Opt.* **18**, 66010 (2013).
61. McClatchy, D. M. *et al.* Calibration and analysis of a multimodal micro-CT and structured light imaging system for the evaluation of excised breast tissue. *Phys. Med. Biol.* **62**, 8983–9000 (2017).
62. Fornacon-Wood, I. *et al.* Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur. Radiol.* **30**, 6241–6250 (2020).
63. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* **44**, 837–845 (1988).
64. Sun, X. & Xu, W. Fast implementation of delong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Process. Lett.* **21**, 1389–1393 (2014).

Acknowledgements

Author contributions

S.S., K.P., and B.P. conceptualized the study. S.S., R.Z., and W.W. performed the research. S.S., B.H., R.Z., and W.W. analyzed the data. S.S. drafted the manuscript. B.H., K.P., and B.P. provided edits. K.P. and B.P. provided supervision. K.P., B.P., and S.S. were involved in funding acquisition. All authors reviewed the manuscript.

Competing interests

B.P., S.S., and K.P. have a patent pending (US Application No.: 17/076,788) related to this study. K.P. is Co-Founder of CairnSurgical, Inc. B.P. is President and Co-Founder of DoseOptics, LLC. Authors in their roles in the medical industry did not in any way impact this study. B.H., R.Z., and W.W. declare no potential conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-01414-z.

Correspondence and requests for materials should be addressed to S.S.S.

Reprints and permissions information is available at www.nature.com/reprints.

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