

# Identification of different types of tumors based on photoacoustic spectral analysis: preclinical feasibility studies on skin tumors

Mengjiao Zhang<sup>a,†</sup>, Long Wen<sup>b,†</sup>, Chu Zhou<sup>b</sup>, Jing Pan<sup>a</sup>, Shiyong Wu<sup>a</sup>,  
Peiru Wang<sup>b</sup>, Haonan Zhang<sup>a,b</sup>, Panpan Chen<sup>a</sup>, Qi Chen<sup>b</sup>, Xiuli Wang<sup>b,\*</sup>  
and Qian Cheng<sup>a,c,d,\*</sup>

<sup>a</sup>Tongji University, Institute of Acoustics, School of Physics Science and Engineering, Shanghai, China  
<sup>b</sup>Tongji University, Institute of Photomedicine, Shanghai Skin Disease Hospital, School of Medicine, Shanghai, China

<sup>c</sup>National Key Laboratory of Autonomous Intelligent Unmanned Systems, Shanghai, China

<sup>d</sup>Frontiers Science Center for Intelligent Autonomous Systems, Ministry of Education, China

**ABSTRACT.** **Significance:** Collagen and lipid are important components of tumor microenvironments (TME) and participates in tumor development and invasion. It has been reported that collagen and lipid can be used as a hallmark to diagnosis and differentiate tumors.

**Aim:** We aim to introduce photoacoustic spectral analysis (PASA) method that can provide both the content and structure distribution of endogenous chromophores in biological tissues to characterize the tumor-related features for identifying different types of tumors.

**Approach:** *Ex vivo* human tissues with suspected squamous cell carcinoma (SCC), suspected basal cell carcinoma (BCC), and normal tissue were used in this study. The relative lipid and collagen contents in the TME were assessed based on the PASA parameters and compared with histology. Support vector machine (SVM), one of the simplest machine learning tools, was applied for automatic skin cancer type detection.

**Results:** The PASA results showed that the lipid and collagen levels of the tumors were significantly lower than those of the normal tissue, and there was a statistical difference between SCC and BCC ( $p < 0.05$ ), consistent with the histopathological results. The SVM-based categorization achieved diagnostic accuracies of 91.7% (normal), 93.3% (SCC), and 91.7% (BCC).

**Conclusions:** We verified the potential use of collagen and lipid in the TME as biomarkers of tumor diversity and achieved accurate tumor classification based on the collagen and lipid content using PASA. The proposed method provides a new way to diagnose tumors.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.28.6.065004](https://doi.org/10.1117/1.JBO.28.6.065004)]

**Keywords:** photoacoustic spectral analysis; tumor-related biomarkers; endogenous chromophores; support vector machine

Paper 230082GR received Mar. 28, 2023; revised May 19, 2023; accepted May 22, 2023; published Jun. 13, 2023.

\*Address all correspondence to Qian Cheng, [q.cheng@tongji.edu.cn](mailto:q.cheng@tongji.edu.cn); Xiuli Wang, [wangxiuli\\_1400023@tongji.edu.cn](mailto:wangxiuli_1400023@tongji.edu.cn)

†These authors contributed equally to this paper.

## 1 Introduction

In recent years, an increasing number of studies have shown that tumors are closely related to tumor microenvironments (TMEs).<sup>1-3</sup> The TME is a complex system mainly composed of blood vessels, collagen, lipid, etc.<sup>3-5</sup> Lipid can provide energy for tumors and collagen mainly constitutes the scaffold of the tumor.<sup>6,7</sup> Collagen and lipid content can affect the proliferation and differentiation of tumor cells.<sup>8-11</sup> Alterations in the content of lipid and collagen are hallmarks of many diseases, including breast cancer,<sup>12</sup> prostate cancer,<sup>13,14</sup> renal cell carcinoma,<sup>15</sup> skin cancers,<sup>16</sup> and others. Therefore, the collagen and lipid content in the TME can be used for diagnosis and differentiation of tumors.

Currently, several methods are available for detecting the lipid and collagen of tumors. Second-harmonic generation microscopy, which relies on the nonlinear interaction of a laser with non-centrosymmetric molecules, has been used to image fibrillar collagen within tissues.<sup>17</sup> It can provide submicron resolution but has limited detection depth and does not provide information about lipid. Magnetic resonance imaging using chemoselective fat-suppression pulse sequences enables lipid detection. However, it fails to provide effective contrast when there is less fat within the tissue.<sup>18</sup> In recent years, Raman spectroscopy has been successfully used to determine the molecular composition and structure based on the inelastic scattering of different molecules to light.<sup>19</sup> However, the penetration depth of Raman spectroscopy is limited, and only superficial skin information can be obtained.<sup>19,20</sup> To overcome these issues, new diagnostic methods with high sensitivity that can detect both collagen and lipid with sufficient penetration depths are required.

Photoacoustics (PA) is a novel non-invasive detection technique that combines high optical absorption contrast with the high penetration depth of ultrasound.<sup>21</sup> PA is a physical process of "light in and sound out."<sup>22</sup> A pulsed laser is used to irradiate biological tissues wherein the energy is wavelength-selectively absorbed by endogenous chromophores within the tissues, generating ultrasonic waves (PA signals) through thermoelastic expansion. Because ultrasonic waves scatter much less than optical waves, PA technology has a greater detection depth than optical detection technology, which shows great promise for clinical applications.<sup>23,24</sup> The endogenous chromophores, such as hemoglobin, collagen, lipid, and water, have different optical absorption spectra in the visible and infrared bands.<sup>25</sup> By irradiating biological tissues with pulsed lasers of different wavelengths, PA can provide rich endogenous chromophores information about tumors.<sup>26-29</sup> Lei et al.<sup>30</sup> investigated the feasibility of assessing collagen contents to detect fibrosis in Crohn's disease using PA imaging. Wilson et al.<sup>31</sup> implemented multiparametric spectroscopic PA imaging to assess the lipid content, oxygen saturation, and total hemoglobin to identify the development of four types of breast cancer. Conventional PA imaging is mainly based on the amplitude of the envelope of time-domain PA signals to quantify the endogenous chromophore concentrations in biological tissues,<sup>32,33</sup> ignoring the frequency and phase information of PA signals associated with the absorbers. Besides, the envelope of the time-domain PA signal is subjected to the effects of noise and transducer response. It is quite difficult to achieve reliable results in quantifying the size and concentration of absorber with sizes smaller than the system resolution. Considering the different sizes of the absorbers, the ultrasonic spectrum shows significant advantages. PA spectral analysis (PASA) method can remove the low-frequency system noise and high-frequency measurement noise to provide objective results and repeatable measurements.<sup>34,35</sup> Further, frequency analysis has proved feasible in detecting absorbers with sizes smaller than the system resolution.<sup>36,37</sup> Recently, PASA, which analyzes the frequency domain power distribution of PA signals, has demonstrated the ability to assess the content and corresponding microstructure of endogenous chromophore in biological tissues simultaneously.<sup>38-41</sup> Moreover, the spectral parameters extracted from the PA spectrum, e.g., slope, power-weighted mean frequency, can be used to characterize the microstructures of endogenous optical absorbers.<sup>38,42-44</sup> Xu et al.<sup>39</sup> implemented the PASA to assess the changes of lipid for fatty liver identification. They further quantified the Gleason score of prostate cancer based on the tissue microscopic architecture using PASA.<sup>45</sup> Our group combined the PASA with machine learning to better mine the data information and achieved a high diagnostic accuracy of prostate cancer,<sup>46-49</sup> osteoporosis,<sup>50</sup> and breast cancer.<sup>51,52</sup> PASA has shown considerable potential in evaluating the endogenous chromophore in biological tissues for tumor diagnosis.

Therefore, in this study, we investigated the feasibility of non-invasive PASA for characterizing the tumor-related features of lipid and collagen content in the TME to identify different types of tumors. We took skin cancers as the research objects, for which an invasive biopsy is the gold standard for diagnosis. *Ex vivo* experiments were conducted using three types of human skin tissue: normal, squamous cell carcinoma (SCC), and basal cell carcinoma (BCC). The content of lipid and collagen in skin tissue was calculated semi-quantitatively by PASA at different wavelengths. With the help of machine learning classification methods, tumors were successfully identified and tumor types were automatically classified based on quantified PASA parameters.

## 2 Materials and Methods

### 2.1 Ethics Statement

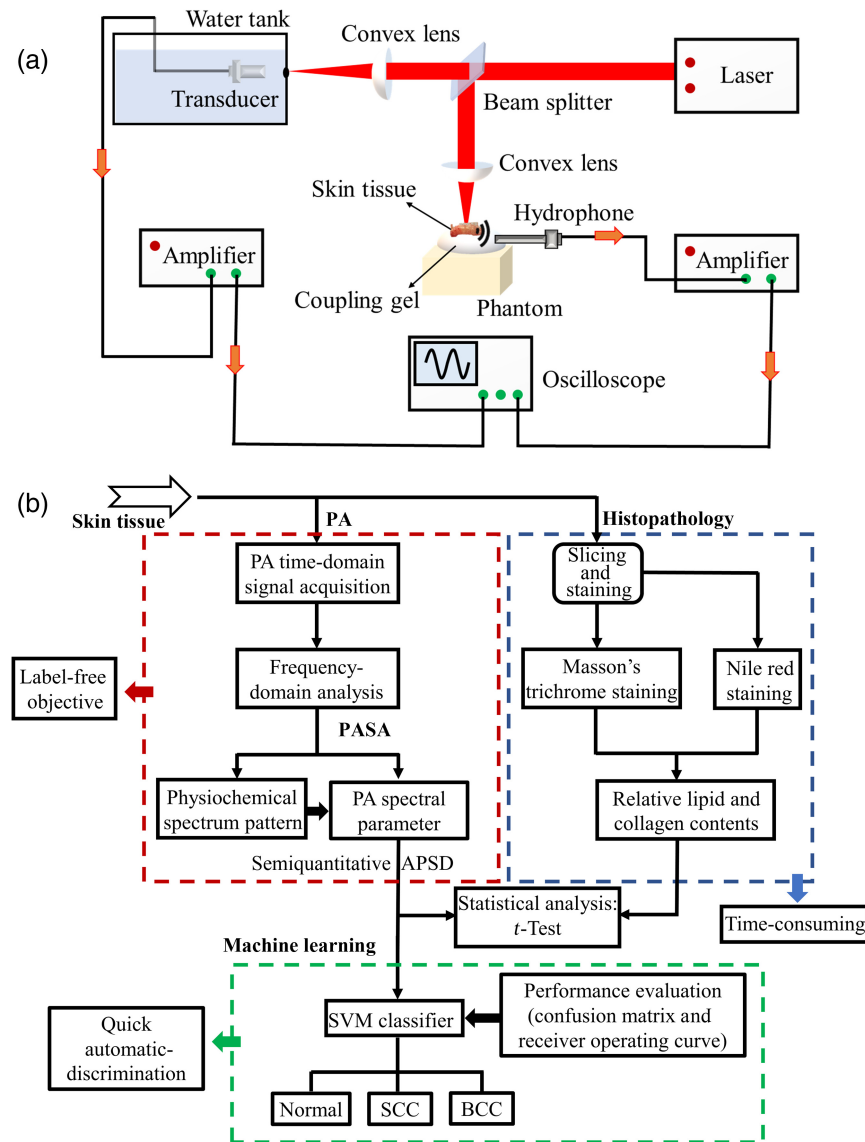
The study protocol was approved by the Ethical Committee of the Shanghai Skin Disease Hospital and was performed in accordance with the tenets of the Declaration of Helsinki. All patients were informed of the purpose of the study, and written consent was obtained before recruitment and sampling.

### 2.2 Sample Collection

A total of 39 patients were enrolled in this study, including 15 with suspected SCC and 12 with suspected BCC. Normal samples were collected from the skin collection areas of 12 patients who received skin grafts. All samples were procured from the Institute of Photomedicine, Shanghai Skin Disease Hospital, School of Medicine, Tongji University, China. After surgical excision of the skin tissue, the residual blood on the tissue surface was cleaned with sterile gauze. Each sample had a diameter of ~5 mm. The skin tissues were placed in sterile sample tubes, stored in a portable medical cryostat (2°C to 8°C), and transported to the Institute of Acoustics of Tongji University laboratory for testing within 1 h.

### 2.3 Experiment Setup and PA Measurements

Figure 1(a) shows a schematic of the PA experimental setup. An optical parametric oscillator system pumped by a Nd:YAG laser (Phocus Mobile, OPOTEK, Carlsbad, California, United States) was used to provide laser pulses with wavelengths ranging from 1200 to 1700 nm in 10 nm intervals, covering the strong absorption ranges of lipid and collagen. The laser energy output over the entire wavelength range was controlled to 0.1 to 0.5 mJ per pulse, with a pulse duration of 5 ns and pulse repetition rate of 10 Hz. A laser beam with a diameter of 3 mm illuminated the skin tissue, leading to an optical energy density of 7 to 14 mJ/cm<sup>2</sup>, which was below the safety limit specified by the American National Standards Institute. As shown in Fig. 1(a), skin tissue was placed on the phantom to avoid strong scattering of the sound signal by any hard boundary. The PA signals generated by the entire laser irradiation of the skin tissues were received by a needle hydrophone with a bandwidth of 1 to 20 MHz (HNC1500, ONDA Corp., Sunnyvale, California, United States). The laser energy varies with the wavelength and fluctuates only slightly over time. To determine the laser energy variations during the PA measurements, 10% of the laser energy was projected onto a black body, and the PA signals generated by the black body were received by a 5 MHz focused transducer (V326, Immersion Transducers, Olympus Corp., Tokyo, Japan). After amplification (5072PR, Olympus Corp., Tokyo, Japan) to boost the PA signals of the skin tissue samples by 25 dB, they were averaged 64 times and recorded using a digital oscilloscope (HDO6000, Teledyne Lecroy, New York, United States) at a sampling rate of 2500 MHz. To improve the stability and reduce the measurement error, PA signals from each skin tissue sample were detected at two different positions. We developed an efficient automated experimental program that includes laser wavelength switching, triggering, and data acquisition. With this program, PA data acquisition of each skin tissue sample and the blackbody could be completed in <13 min, covering 51 wavelengths (1200 to 1700 nm), with an average of 64 times per wavelength.

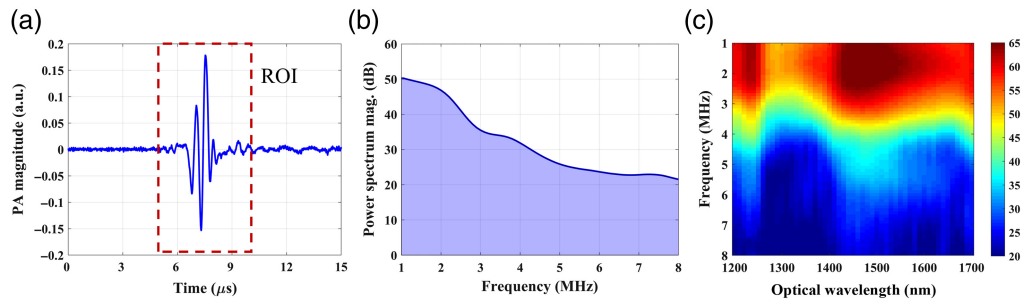


**Fig. 1** PA experimental system and experiment detection method. (a) Experimental setup for performing *ex vivo* PA measurements. (b) Block diagram of the experimental method. The red, blue, and green dotted boxes represent PA experiments and data processing, histopathology analysis, and machine learning, respectively.

## 2.4 PA Spectral Analysis

The PA signals were analyzed using MATLAB software (R2019b, MathWorks, Natick, Massachusetts, United States). First, the PA signal generated by the skin tissue was calibrated using the peak-to-peak value of the PA signal generated by the black body at each wavelength. Second, based on the calibrated PA signals for each skin tissue sample, the power spectra of the PA signals [Fig. 2(a)] were calculated using Welch's method with a  $5 \mu\text{s}$  moving Hamming window and 60% overlap, as shown in Fig. 2(b). Considering the signal-to-noise ratio, the ultrasound frequency was first analyzed in the 1 to 8 MHz range. The power spectra of the PA signals obtained in the 1200 to 1700 nm wavelength range were combined to form a PA physiochemical spectrogram (PAPCS), as depicted in Fig. 2(c). The horizontal axis of the PAPCS is the optical wavelength, representing the relative optical absorption of different endogenous chromophores in skin tissue, whereas the vertical axis is the ultrasonic frequency, revealing the structural distribution corresponding to different optical absorptions in skin tissue. The color bar represents the amplitude of the power spectrum. Differences in the lipid and collagen content and in the microstructure of the TMEs of different tumors will form unique PAPCSs.





**Fig. 2** (a) Representative PA signal from skin tissue. The red dotted box indicates the region of interest. (b) Example of the power spectrum density of a PA signal generated by skin tissue. The purple area indicates the APSD. (c) Example of a PAPCS for skin tissue.

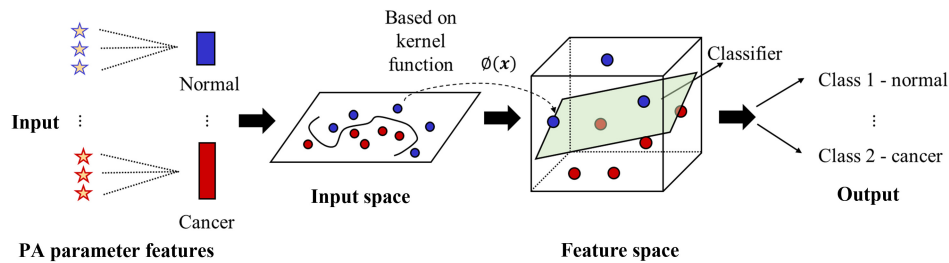
Furthermore, the changes of relative lipid and collagen content in skin tissues were quantitatively analyzed. Based on the PA power spectral analysis at a wavelength of  $\lambda$  nm [Fig. 2(b)], we calculated the relative area of power spectrum density (APSD) at the corresponding wavelength as follows:

$$\text{Relative APSD}_{\lambda} = \frac{\left( \int_{f_0}^{f_1} p(f) df \right)_{\lambda}}{\left( \int_{f_0}^{f_1} p(f) df \right)_{\lambda_0}}, \quad (1)$$

where  $p(f)$  is the power spectral density at each frequency,  $f_0 = 1$  MHz, and  $f_1 = 8$  MHz. The PA signal is affected by low-frequency system noise and high-frequency measurement noise. Based on the PA power spectrum analysis, setting a high cut-off frequency  $f_1$  can help avoid high-frequency noise, whereas setting a low cut-off frequency  $f_0$  can help minimize system noise. Considering the signal-to-noise ratio and the cluster size of lipid and collagen obtained from histology images (see Fig. 4), we set the low and high cutoff frequencies to 1 and 8 MHz, respectively. The PA power spectral density in the specified frequency range was then summarized as the PA absorption value at each wavelength. A reference wavelength  $\lambda_0$  (690 nm) was used to eliminate systematic errors. We then obtained the relative PA absorption of each skin tissue sample with reduced system noise. To improve the stability and reduce the measurement error, PA signals from each skin tissue sample were detected at two different positions. The relative APSD values from two different directions were then obtained and averaged for further analysis. The relative APSD obtained at wavelength  $\lambda$  (nm) reflects the relative optical absorption of the corresponding endogenous chromophore; thus, it is related to the relative endogenous chromophore content of the skin tissue. According to the literature, lipid exhibits strong absorption at  $\sim 1200$  to  $1240$  nm.<sup>25,53,54</sup> Some studies have also shown that at  $1600$  to  $1700$  nm excitation,<sup>55–57</sup> the lipid PA signal is enhanced compared to that at  $1200$  nm excitation. In addition, the selective detection of collagen at  $\sim 1300$  to  $1340$  nm can be achieved.<sup>46,58</sup> Thus, to reduce the measurement error caused by a single wavelength, the relative APSD obtained at the absorption wavelength range of lipid (or collagen) were averaged. To examine whether the changes in collagen and lipid content of TME in different types of skin tissues were statistically significant, unpaired  $t$ -tests were performed using GraphPad Prism 9.0.

## 2.5 Support Vector Machine Analysis

Support vector machine (SVM) analysis is one of the simplest machine learning classification method that is supported by rigorous mathematical theory, is highly interpretable, and can identify the key factors for classification tasks. In this study, a SVM classifier was applied to perform automatic different types of tumors classification by combining relative APSD values at different wavelengths. The APSD values obtained in the three wave bands ( $1200$  to  $1240$  nm,  $1600$  to  $1700$  nm, and  $1300$  to  $1340$  nm) were used as the input characteristic parameters for the classification of different types of skin tissues, thus realizing their automatic discrimination, as shown in Fig. 3. The input data comprised 39 datasets with 21 features. Considering the limited number of clinical skin tissue samples available for categorization, we used the C-type SVM



**Fig. 3** Classification of normal and skin cancer tissues using an SVM classifier. The APSD values obtained at three wave bands are used as the feature parameters for the SVM input. An optimal hyperplane constructed in the corresponding feature space determines the output.

model with a radial basis function (RBF) kernel, which is suitable for classifying small amounts of data. The RBF kernel can be defined as follows:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0, \quad (2)$$

where  $x$  is the variable matrix,  $i$  and  $j$  are the indices of the matrix, and  $\gamma$  is the kernel parameter. There are two parameters for an RBF kernel: the regularization parameter  $C$  and kernel parameter  $\gamma$ . Parameter selection for optimal SVM categorization was achieved by grid searching<sup>59</sup> using cross-validation. Various pairs of  $(C, \gamma)$  values were tested, and the pair with the best cross-validation accuracy was selected. To increase the classification reliability, a three-fold cross-validation approach was applied by rotating the datasets used to train and test the SVM model. The samples were randomly divided into two groups: two-thirds of the samples were used for training, and the remaining samples were used for prediction. This process was implemented thrice, with each group being tested once. Finally, the performance of the classifier model was evaluated using confusion matrices and receiver operating characteristics (ROCs) curves.

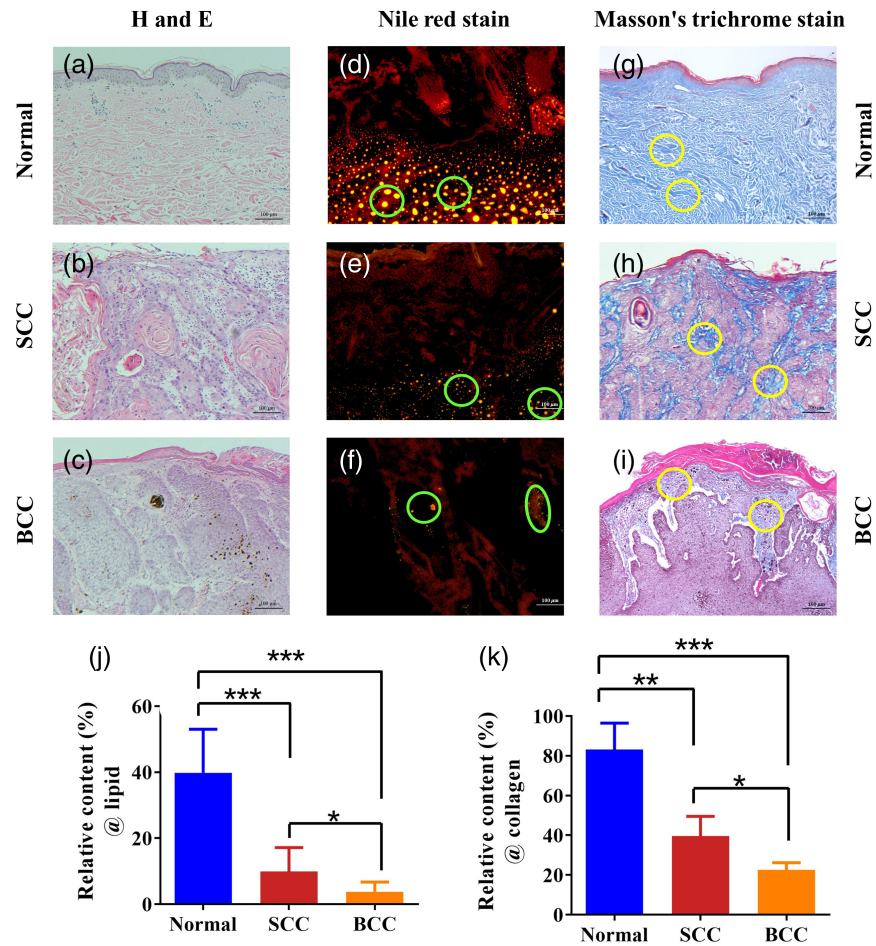
## 2.6 Histopathology

To validate the PA measurements, the samples were collected and prepared for histological analysis after performing the PA measurements. The pathological analysis was divided into two parts. Some tissue samples were fixed by immersion in a 10% formalin solution and subsequent sectioning after tissue dehydration and paraffin embedding, completing the production of paraffin sections. Next, the sections were stained with hematoxylin and eosin (H&E) and others stained with Masson's trichrome. Masson's trichrome can stain collagen with blue color, which can be used for collagen detection. The other tissue sample was treated with an optical coherence tomography embedding agent (optical cutting temperature compound, Sakura Americas, New Mexico, United States) and flushed with distilled water to obtain frozen sections. Serial longitudinal sections (5  $\mu\text{m}$  thick) were subsequently cut and stained with Nile red for lipid detection. Nile red can stain the lipid with yellow color. The staining results were observed by microscopy, as shown in Fig. 4. The relative lipid and collagen contents of the skin tissues were obtained from histological images. The ratio of the positively stained area to the whole area in the histological image, calculated using ImageJ software for each sample slice, was used as the gold standard for the relative content of collagen or lipid in the skin tissues, presenting a quantitative comparison with the relative APSD results.

## 3 Results

### 3.1 PAPCSs of Different Types of Skin Tissue

Figures 5(a)–5(c) show the averaged PAPCSs of the normal, SCC, and BCC tissues, respectively. Compared with the normal tissue, in the 1200 to 1240 nm and 1600 to 1700 nm wave bands, the PAPCSs of tumor (SCC and BCC) tissues both show lower spectral magnitudes, indicating decreases in lipid content. In addition to the changes in lipid content, the PAPCSs of tumors exhibit differences in the collagen absorption wavelength range (1300 to 1340 nm). The PAPCSs of SCC and BCC show lower PA spectral magnitudes, indicating reductions in collagen contents. Interestingly, compared with the PAPCS of the BCC tissues, that of the SCC tissues

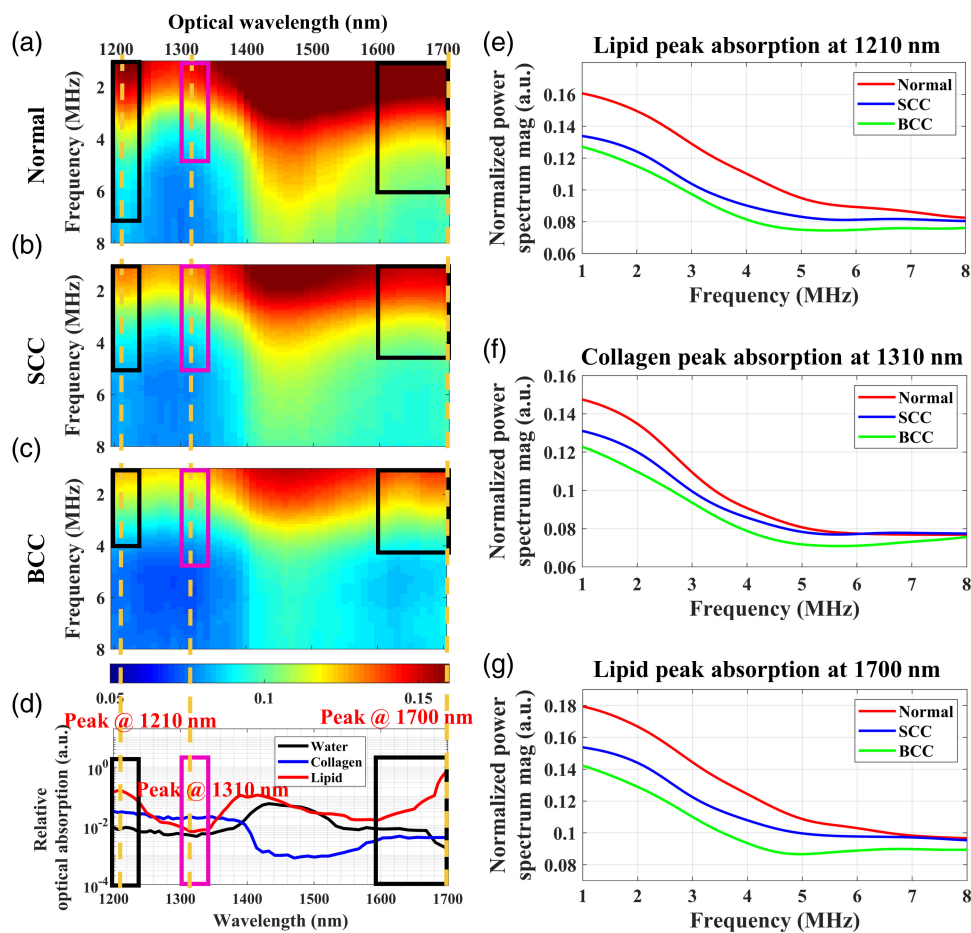


**Fig. 4** Histological results. (a)–(c) Histological examination of normal, SCC, and BCC tissues using H&E staining. (d)–(f) Histological examination of the lipid contents in normal, SCC, and BCC tissues using Nile red staining. (g)–(i) Histological examination of the collagen contents in normal, SCC, and BCC tissues using Masson's trichrome staining. The green and yellow circle indicate the lipid and collagen clusters, respectively. (j) and (k) Statistical results of the positive-stained areas for the relative lipid and collagen contents of skin tissues obtained by histological images (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ,  $t$ -test was conducted using GraphPad Prism 9.0).

shows higher spectral magnitudes in the lipid- and collagen-absorption wavelength ranges, implying that SCC tissues contain higher lipid and collagen content than BCC tissues. Specifically, Figs. 5(e)–5(g) show the PA power spectrum of different skin tissues at the absorption peaks of lipid (1210 and 1700 nm) and collagen (1310 nm), respectively. The pathological images of the skin tissues [Figs. 4(d)–4(i)] verify these differences in the lipid and collagen contents of the three types of skin tissues. The differences in PAPCSs between different skin tumors suggest that lipid and collagen in TMEs can be used as biomarkers of tumor diversity.

### 3.2 PASA Semi-Quantitative Results

Based on the PASA results for skin tissues obtained at different wavelengths, the relative APSD was calculated to semi-quantify the relative lipid and collagen contents in the skin tissues. We calculated the wavelength dependence of the relative APSD for each skin tissue sample. Figures 6(a)–6(c) show the content characterization fluctuations at different wavelength. Thus, we averaged the relative APSD values in the absorption band of lipid or collagen to eliminate the measurement error caused by a single wavelength. Figures 6(d)–6(f) show the statistical results for the averaged relative APSD values, which can be used to distinguish between different types of skin tissues. As shown in Figs. 6(d) and 6(e), the lower lipid content of tumors leads to a decrease in its spectral amplitude and, consequently, a lower relative APSD for the 1200 to 1240 nm and 1600 to 1700 nm wavebands. The relative APSD values of the tumors are also



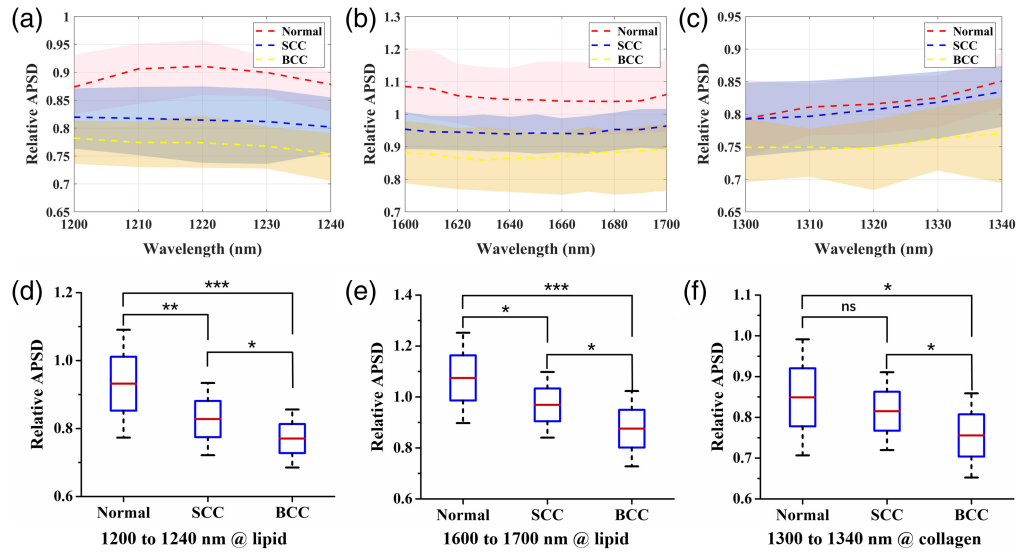
**Fig. 5** Averaged PAPCSs of different skin tissues: (a) normal skin tissue, (b) SCC, and (c) BCC. (d) Optical absorption spectra of lipid, collagen, and water in the 1200 to 1700 nm wavelength range.<sup>25,60</sup> The regions indicated by black boxes (1200 to 1240 nm and 1600 to 1700 nm) correspond to regions with strong lipid absorption. The regions indicated by magenta boxes (1300 to 1340 nm) correspond to strong collagen absorption. The yellow dashed lines indicate the peak lipid or collagen absorption within each region. (e)–(g) Averaged PA power spectrum of different skin tissues at the absorption peaks of lipid (1210 and 1700 nm) and collagen (1310 nm).

lower for the 1300 to 1340 nm wavelength band, corresponding to their decreased collagen content, as shown in Fig. 6(f). In addition, different types of tumors exhibit significant differences in their collagen and lipid contents ( $p < 0.05$ ). For example, as shown in Figs. 6(d)–6(f), compared to those of SCC tissues, the relative APSD values of BCC tissues are lower in the regions corresponding to the optical absorption of lipid (1200 to 1240 nm and 1600 to 1700 nm) and collagen (1300 to 1340 nm), indicating that the lipid and collagen contents in the BCC tissues are lower than those contents in SCC. Figures 4(j) and 4(k) depict the relative lipid and collagen contents of the three different skin tissues obtained from the histological images and reveal notable differences. The PA results were corroborated by the changes in the lipid and collagen contents revealed in the histological images.

### 3.3 SVM Classification Results

An SVM was employed to classify skin cancers based on the lipid and collagen contents in the skin tissues. The relative lipid and collagen contents in the TME were characterized by the relative APSD values at different wavelengths. The classification results of the three-fold cross-validation approach are presented in Table 1. As shown in Table 2, the average classification accuracies of the SVM trained based only on the parameter at a single wavelength are 69.2%, 64%, and 89%. However, combining the parameters of the three bands results in an





**Fig. 6** (a)–(c) PA-related APSD values with wavelengths in different skin tissues. The dashed lines indicate the averaged PA spectra, and the shaded areas overlapping the lines indicate the standard deviations. (d)–(f) Statistical results of the PA-related APSD values for the lipid and collagen contents in different skin tissues (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , ns: no statistical significance).

accuracy of 92.3%, the results show that the combination of multiple biomarkers can achieve more accurate intelligent diagnoses.

The classification performance of the SVM was evaluated using a confusion matrix, as shown in Fig. 7(b). The confusion matrix provides the classification accuracy of the SVM for individual classes. Each column of the matrix corresponds to a true label, and each row corresponds to a predicted label. The main diagonal shows the classification accuracy of the SVM for the individual classes. The off-diagonal values indicate the misclassification rates of the SVM for classifying the individual classes. The results show that SVM-based multiclass categorization can achieve diagnostic accuracies of 91.7% (normal), 93.3% (SCC), and 91.7% (BCC). The ROC curves yielded by the SVM for the normal, SCC, and BCC tissues are presented in Fig. 7(c), revealing areas under the ROC curves of 0.94, 0.96, and 0.92, respectively. Overall, SVM-based categorization proved effective in diagnosing tumors with a high level of accuracy.

## 4 Discussion

TME is a complex system in which collagen and lipid are the important components of TME. It has been reported that collagen and lipid can be used as biomarker to assess tumors.<sup>61,62</sup> Several techniques have been available to detect lipid and collagen, but they all have limitations, including: invasiveness, time-consumption, and limited detection depths. Based on the contrasting optical absorption spectra of different endogenous chromophores in tissue, PA detection technique can provide rich content and microstructure information of the endogenous chromophores in biological tissues, allowing more accurate tumor diagnoses. In addition, PA has a greater detection depth than optical detection technology, benefiting from the fact that the PA signal originates from the absorption of laser energy by biological tissue, independent of the phase of the light wave.

Experiments on different types of skin tissues showed that the PAPCS obtained by PASA contained rich diagnostic information. The loss of collagen and lipid content can be observed in the corresponding PAPCS and can be used to characterize different types of tumors. Owing to the differences in lipid and collagen content of different tumors, each type of tumor has a unique PAPCS. The PASA parameter APSD is correlated with the collagen and lipid contents. The decreases in the lipid and collagen contents in skin cancer tissues cause the PA spectral amplitude of the power spectrum to decrease, corresponding to a lower relative APSD value. Furthermore, SCC originates from higher differentiated epithelial cells than BCC, with higher amounts of

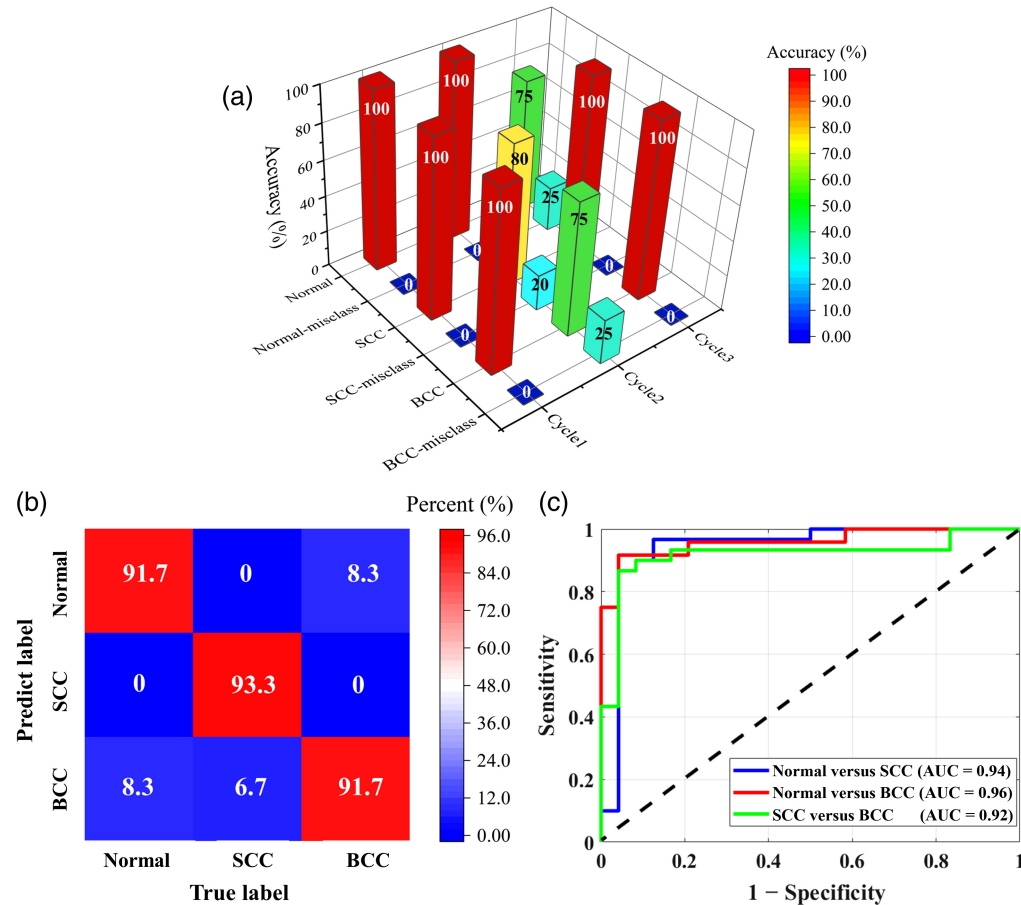


**Table 1** SVM categorization. A three-fold cross-validation approach was conducted in which the samples were divided into three groups at random. Each cycle used two groups for training and one group for testing. Cate. = Categorization. Samp. = Sample.

Test sample condition	Cate. #	Test Samp. #	Cycle1			Cycle2			Cycle3			All		
			1210 nm	1310 nm	1700 nm	1210 nm	1310 nm	1700 nm	1210 nm	1310 nm	1700 nm			
Normal tissue	1	1	2	3	3	1	1	1	1	3	2	2	3	3
	1	2	1	1	1	1	1	1	1	1	1	1	1	1
	1	3	1	1	1	1	1	1	1	1	1	1	1	1
	1	4	3	2	1	1	2	1	1	1	1	1	1	1
SCC	2	5	2	1	2	2	2	1	2	2	2	2	2	2
	2	6	2	3	2	2	2	2	2	2	2	2	2	2
	2	7	3	2	2	2	2	2	2	3	1	3	2	2
	2	8	2	2	2	2	3	1	2	2	2	2	2	2
	2	9	2	2	2	2	1	2	2	2	2	2	2	2
	3	10	1	2	3	3	3	1	2	1	3	3	3	3
	3	11	3	3	3	3	3	3	3	3	3	3	1	3
BCC	3	12	3	3	3	3	3	2	3	3	3	3	3	3
	3	13	2	3	3	3	1	2	3	3	3	3	3	3
	Training accuracy (%)		95.8	75.8	95.8	100	92.5	100	96.7	100	83.3	96.7	92.5	95.8
Testing accuracy (%)		61.5	61.5	92.3	100	76.9	53.8	84.6	69.2	76.9	76.9	76.9	92.3	

**Table 2** Mean SVM categorization accuracy in Table 1.

	1210 nm	1310 nm	1700 nm	All
Mean training accuracy (%)	90.5	90.8	95	98.6
Mean testing accuracy (%)	69.2	64	87.2	92.3



**Fig. 7** SVM classification results. (a) Classification results for normal, SCC, and BCC tissues at each fold, including the accurate classification and misclassification rates. (b) Confusion matrix for the normal, SCC, and BCC tissues. (c) Comparison of the ROC curves for the normal, SCC, and BCC tissues. AUC represents an area under ROC curve. Normal-misclass, normal-misclassification; SCC-misclass, SCC-misclassification; BCC-misclass, BCC-misclassification.

lipid and collagen,<sup>63</sup> resulting in their APSD values higher than those for BCC in the absorption band of lipid and collagen. Statistical results of the content and microstructure changes of endogenous chromophores related to histology can be obtained according to PASA parameters. Thus, PASA provides an objective method of classifying SCC and BCC independently of the experience of a physician or pathologist. In addition, the PASA eliminates the effects of noise and system errors, providing system-independent semi-quantitative results. Our results suggest that PASA can provide sensitive, semi-quantitative content of endogenous chromophore in tissues to achieve non-invasive diagnosis and identification of different types of tumors.

Frequency spectrum of PA signal is related to the size and content of the optical absorber. In addition, our previous work has investigated the frequency anisotropy caused by the direction of the structure<sup>35</sup>. There appeared well distributed and more parallel collagen in normal skin tissues. While in the tumor tissue (SCC and BCC), the collagen was obviously unorderly and the amount

of collagen appears reduced.<sup>64</sup> There was no statistical difference in the structure and distribution of collagen cluster between SCC and BCC. In the future, we can use multiple parameters to characterize the distribution and content of collagen and lipid simultaneously to better identify tumor tissues from normal tissues.

Machine learning is an effective feature extraction tool. Even in small samples, the characteristics of PA signals can be fully learned to achieve good disease diagnosis, and machine learning has been widely used in the detection of various diseases.<sup>65,66</sup> In this study, an SVM, one of the simplest machine learning tools, was utilized to distinguish automatically between different types of tumors based on the extracted PA parameters. The proposed method takes advantage of quantitative parameters of multiple wavelengths, contains rich diagnostic information, and improves the separation of high-dimensional datasets. As the number of clinical skin tissue samples was limited, we used the SVM algorithm with the RBF kernel, which is appropriate when the amount of data to be classified is small. Furthermore, to prevent overfitting based on a small number of training datasets, *K*-fold cross-validation was performed. The choice of the number of *K*-fold depends on the sample size, number of parameters, structure of data, and so on.<sup>67</sup> Due to the limitation of the sample size, three-fold cross-validation was used in our study. In future work, more training datasets should be included and a higher and appropriate *K*-fold cross-validation should be selected, which would improve the robustness of the SVM model and make the classification of skin cancer types using the SVM approach more accurate.

Considering the accessibility of human tumor samples, the object of this study is human skin tumor tissue *in vitro* for pathological examination. *In vitro* experiments significantly improved the operability of the samples and the stability of the experiments. However, the disadvantage is that the optical absorption of hemoglobin cannot be measured. The optical absorption of hemoglobin is dominant in the visible to near infrared wavelength range (i.e., 400 to 900 nm). Therefore, the results acquired at the wavelengths of 1200 to 1700 nm were less affected by the hemoglobin. In the future, we plan to use the PASA method *in vivo*. Considering the influence of water *in vivo* experiment, which has high absorption in the 1200 to 1700 nm, a spectral unmixing procedure should be used to enable more precise evaluation of the relative contents of lipid and collagen. In fact, this approach also has considerable potential to be extended to the diagnosis of other types of cancers with altered collagen or lipid contents, e.g., breast cancer,<sup>12</sup> prostate cancer.<sup>13,14</sup> Furthermore, there were clear differences in the PA power spectrum between skin cancers and normal tissues. This finding not only confirms that the PA signals of tumor and healthy tissue are different but also suggests the possibility of using the PASA to define tumor boundaries in the future. Currently, the primary treatment for skin cancers is surgical excision. However, the existing clinical techniques cannot accurately define the boundaries of tumors, leading to a postoperative residual tumor tissue, which may cause disease recurrence. In future studies, we also intend to apply the PASA *in vivo* to identify the boundaries of skin cancers more accurately and objectively, to help doctors remove such tumors completely and reduce complications.

## 5 Conclusion

In summary, we introduced the PASA method to characterize the content of endogenous chromophore in the TME to better understand the interaction and regulation of tumors and TME. The parameters extracted from PASA were used to semi-quantify the content of endogenous chromophore. It was found that the collagen and lipid content can be used as biomarkers of tumor diversity, and we used this to successfully diagnose different types of tumors with improved accuracy. Considering the PA technique is non-invasive and has greater detection depth than pure optical technology, the proposed method shows considerable potential for non-invasive and more accurate diagnosis of tumors *in vivo*.

---

## Disclosures

The authors declare no competing interests.

## Code, Data, and Materials Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Acknowledgments

This project was supported by the National Natural Science Foundation of China (Grant No. 12034015, 62088101, 81872212, and 82073013), National Key Research and Development Program of China (Grant No. 2017YFC0111400), Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0100), Shanghai Municipal Commission of Science and Technology Project (Grant No. 19511132101), Program of Shanghai Academic Research Leader (Grant No. 21XD1403600). We would like to thank Editage<sup>68</sup> for English language editing.

## References

1. Z. Tan et al., "The role of tumor inflammatory microenvironment in lung cancer," *Front. Pharmacol.* **12**, 688625 (2021).
2. X. Huang et al., "Regulation of tumor microenvironment for pancreatic cancer therapy," *Biomaterials* **270**, 120680 (2021).
3. N. M. Anderson and M. C. Simon, "The tumor microenvironment," *Curr. Biol.* **30**(16), R921–R925 (2020).
4. F. R. Balkwill, M. Capasso, and T. Hagemann, "The tumor microenvironment at a glance," *J. Cell Sci.* **125**(23), 5591–5596 (2012).
5. E. Brown, J. Brunker, and S. E. Bohndiek, "Photoacoustic imaging as a tool to probe the tumour microenvironment," *Dis. Model. Mech.* **12**(7), dmm039636 (2019).
6. J. Sapudom and T. Pompe, "Biomimetic tumor microenvironments based on collagen matrices," *Biomater. Sci.* **6**(8), 2009–2024 (2018).
7. M. Fang et al., "Collagen as a double-edged sword in tumor progression," *Tumor Biol.* **35**(4), 2871–2882 (2014).
8. P. Cirri and P. Chiarugi, "Cancer associated fibroblasts: the dark side of the coin," *Am. J. Cancer Res.* **1**(4), 482–497 (2011).
9. K. R. Levental et al., "Matrix crosslinking forces tumor progression by enhancing integrin signaling," *Cell* **139**(5), 891–906 (2009).
10. F. Mollinedo and C. Gajate, "Lipid rafts as signaling hubs in cancer cell survival/death and invasion: implications in tumor progression and therapy," *J. Lipid Res.* **61**(5), 611–635 (2020).
11. M. Martin-Perez et al., "The role of lipids in cancer progression and metastasis," *Cell Metab.* **34**(11), 1675–1699 (2022).
12. H. Abramczyk et al., "The role of lipid droplets and adipocytes in cancer. Raman imaging of cell cultures: MCF10A, MCF7, and MDA-MB-231 compared to adipocytes in cancerous human breast tissue," *Analyst* **140** (7), 2224–2235 (2015).
13. Y. Pu et al., "Changes of collagen and nicotinamide adenine dinucleotide in human cancerous and normal prostate tissues studied using native fluorescence spectroscopy with selective excitation wavelength," *J. Biomed. Opt.* **15**(4), 047008 (2010).
14. N. Burns-Cox et al., "Change in collagen metabolism in prostate cancer: a host response that may alter progression," *J. Urol.* **166**(5), 1698–1701 (2001).
15. S. L. Best et al., "Collagen organization of renal cell carcinoma differs between low and high grade tumors," *BMC Cancer* **19**(1), 490 (2019).
16. L. A. Austin, S. Osseiran, and C. L. Evans, "Raman technologies in cancer diagnostics," *Analyst* **141**(2), 476–503 (2016).
17. A. Zoumi, A. Yeh, and B. J. Tromberg, "Imaging cells and extracellular matrix *in vivo* by using second-harmonic generation and two-photon excited fluorescence," *Proc. Natl. Acad. Sci. U.S.A.* **99**(17), 11014–11019 (2002).
18. S. S. Pokharel et al., "Current MR imaging lipid detection techniques for diagnosis of lesions in the abdomen and pelvis," *Radiographics* **33**(3), 681–702 (2013).
19. C. A. Lieber et al., "*In vivo* nonmelanoma skin cancer diagnosis using Raman microspectroscopy," *Lasers Surg. Med.* **40**(7), 461–467 (2008).
20. A. M. Ferreira Lima et al., "Discrimination of non-melanoma skin cancer and keratosis from normal skin tissue *in vivo* and *ex vivo* by Raman spectroscopy," *Vib. Spectrosc.* **100**, 131–141 (2019).
21. L. V. Wang and J. Yao, "A practical guide to photoacoustic tomography in the life sciences," *Nat. Methods* **13**(8), 627–638 (2016).
22. S. Zackrisson, S. M. W. Y. van de Ven, and S. S. Gambhir, "Light in and sound out: emerging translational strategies for photoacoustic imaging," *Cancer Res.* **74**(4), 979–1004 (2014).

23. L. V. Wang and S. Hu, "Photoacoustic tomography: *in vivo* imaging from organelles to organs," *Science* **335**(6075), 1458–1462 (2012).
24. V. Ntziachristos et al., "Looking and listening to light: the evolution of whole-body photonic imaging," *Nat. Biotechnol.* **23**(3), 313–320 (2005).
25. H.-W. Wang et al., "Label-free bond-selective imaging by listening to vibrationally excited molecules," *Phys. Rev. Lett.* **106**(23), 238106 (2011).
26. C. Kim, C. Favazza, and L. V. Wang, "*In vivo* photoacoustic tomography of chemicals: high-resolution functional and molecular optical imaging at new depths," *Chem. Rev.* **110**(5), 2756–2782 (2010).
27. S. Hu and L. V. Wang, "Optical-resolution photoacoustic microscopy: auscultation of biological systems at the cellular level," *Biophys. J.* **105**(4), 841–847 (2013).
28. J. Yao et al., "High-speed label-free functional photoacoustic microscopy of mouse brain in action," *Nat. Methods* **12**(5), 407–410 (2015).
29. M. K. A. Singh, Ed., *LED-Based Photoacoustic Imaging: From Bench to Bedside*, Springer Singapore, Singapore (2020).
30. H. Lei et al., "Characterizing intestinal inflammation and fibrosis in Crohn's disease by photoacoustic imaging: feasibility study," *Biomed. Opt. Express* **7**(7), 2837 (2016).
31. K. E. Wilson et al., "Multiparametric spectroscopic photoacoustic imaging of breast cancer development in a transgenic mouse model," *Theranostics* **4**(11), 1062–1071 (2014).
32. H. F. Zhang et al., "Functional photoacoustic microscopy for high-resolution and noninvasive *in vivo* imaging," *Nat. Biotechnol.* **24**(7), 848–851 (2006).
33. X. Wang et al., "Noninvasive laser-induced photoacoustic tomography for structural and functional *in vivo* imaging of the brain," *Nat. Biotechnol.* **21**(7), 803–806 (2003).
34. T. Feng et al., "Characterization of bone microstructure using photoacoustic spectrum analysis," *Opt. Express* **23**(19), 25217 (2015).
35. M. Zhang et al., "Photoacoustic power Azimuth spectrum for microvascular evaluation," *Photoacoustics* **22**, 100260 (2021).
36. S. Wang et al., "Quantitative detection of stochastic microstructure in turbid media by photoacoustic spectral matching," *Appl. Phys. Lett.* **102**(11), 114102 (2013).
37. X. Gao et al., "Quantitative imaging of microvasculature in deep tissue with a spectrum-based photo-acoustic microscopy," *Opt. Lett.* **40**(6), 970 (2015).
38. G. Xu et al., "Photoacoustic spectrum analysis for microstructure characterization in biological tissue: a feasibility study," *Appl. Phys. Lett.* **101**(22), 221102 (2012).
39. G. Xu et al., "The functional pitch of an organ: quantification of tissue texture with photoacoustic spectrum analysis," *Radiology* **271**(1), 248–254 (2014).
40. T. Feng et al., "Detection of collagen by multi-wavelength photoacoustic analysis as a biomarker for bone health assessment," *Photoacoustics* **24**, 100296 (2021).
41. K. Ma et al., "Myocardial infarct border demarcation by dual-wavelength photoacoustic spectral analysis," *Photoacoustics* **26**, 100344 (2022).
42. W. Xie et al., "Wavelet transform-based photoacoustic time-frequency spectral analysis for bone assessment," *Photoacoustics* **22**, 100259 (2021).
43. W. Xie et al., "Photoacoustic characterization of bone physico-chemical information," *Biomed. Opt. Express* **13**(5), 2668 (2022).
44. Y. Yang et al., "Photoacoustic tomography of tissue subwavelength microstructure with a narrowband and low frequency system," *Appl. Phys. Lett.* **101**(3), 034105 (2012).
45. G. Xu et al., "Quantifying Gleason scores with photoacoustic spectral analysis: feasibility study with human tissues," *Biomed. Opt. Express* **6**(12), 4781 (2015).
46. S. Huang et al., "Interstitial assessment of aggressive prostate cancer by physio-chemical photoacoustics: an *ex vivo* study with intact human prostates," *Med. Phys.* **45**(9), 4125–4132 (2018).
47. Y. Chen et al., "Prostate cancer identification via photoacoustic spectroscopy and machine learning," *Photoacoustics* **23**, 100280 (2021).
48. S. Wu et al., "Photoacoustic spectrum analysis for quick identification and grading of prostate cancer," in *IEEE Int. Ultrasonics Symp.*, IEEE, Las Vegas, Nevada, United States, pp. 1–4 (2020).
49. S. Wu et al., "Quick identification of prostate cancer by wavelet transform-based photoacoustic power spectrum analysis," *Photoacoustics* **25**, 100327 (2022).
50. T. Feng et al., "Characterization of multi-biomarkers for bone health assessment based on photoacoustic physicochemical analysis method," *Photoacoustics* **25**, 100320 (2022).
51. J. Li et al., "Photoacoustics spectral analysis for *in vivo* detection of collagen contents in cancers," in *IEEE Int. Ultrasonics Symp.*, IEEE, Venice, pp. 1–4 (2022).
52. J. Li et al., "Molecular breast cancer subtype identification using photoacoustic spectral analysis and machine learning at the biomacromolecular level," *Photoacoustics* **30**, 100483 (2023).



53. B. Wang et al., "Detection of lipid in atherosclerotic vessels using ultrasound-guided spectroscopic intravascular photoacoustic imaging," *Opt. Express* **18**(5), 4889 (2010).
54. J. Hui et al., "Bond-selective photoacoustic imaging by converting molecular vibration into acoustic waves," *Photoacoustics* **4**(1), 11–21 (2016).
55. P. Wang et al., "Bond-selective imaging of deep tissue through the optical window between 1600 and 1850 nm," *J. Biophotonics* **5**(1), 25–32 (2012).
56. B. Wang et al., "Intravascular photoacoustic imaging of lipid in atherosclerotic plaques in the presence of luminal blood," *Opt. Lett.* **37**(7), 1244 (2012).
57. C. K. N. Patel and A. C. Tam, "Optical absorption coefficients of water," *Nature* **280**(5720), 302–304 (1979).
58. Y. Zhu et al., "Identifying intestinal fibrosis and inflammation by spectroscopic photoacoustic imaging: an animal study *in vivo*," *Biomed. Opt. Express* **9**(4), 1590 (2018).
59. C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," National Taiwan U, 2003, [www.csie.ntu.edu.tw/~cjlin/papers/guide/guide](http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide).
60. W. F. Cheong, S. A. Prael, and A. J. Welch, "A review of the optical properties of biological tissues," *IEEE J. Quantum Electron.* **26**(12), 2166–2185 (1990).
61. S. A. Arnold et al., "Lack of host SPARC enhances vascular function and tumor spread in an orthotopic murine model of pancreatic carcinoma," *Dis. Model. Mech.* **3**(1–2), 57–72 (2010).
62. X. Li et al., "The immunological and metabolic landscape in primary and metastatic liver cancer," *Nat. Rev. Cancer* **21**(9), 541–557 (2021).
63. S. Heuke et al., "Detection and discrimination of non-melanoma skin cancer by multimodal imaging," *Healthcare* **1**(1), 64–83 (2013).
64. S. Y. Xiong, J. G. Yang, and J. Zhuang, "Nonlinear spectral imaging of human normal skin, basal cell carcinoma and squamous cell carcinoma based on two-photon excited fluorescence and second-harmonic generation," *Laser Phys.* **21**(10), 1844–1849 (2011).
65. J. Rodrigues, "Exploring photoacoustic spectroscopy-based machine learning together with metabolomics to assess breast tumor progression in a xenograft model *ex vivo*," *Lab. Invest.* **101**(7), 952–965 (2021).
66. A. R. Rajanna et al., "Prostate cancer detection using photoacoustic imaging and deep learning," *Electron. Imaging* **2016**(15), 1–6 (2016).
67. Y. Jung, "Multiple predicting  $K$ -fold cross-validation for model selection," *J. Nonparametr. Stat.* **30**(1), 197–215 (2018).
68. <http://www.editage.com>.

**Mengjiao Zhang** received her BS degree in marine technology from Ocean University of China, in 2013. She is currently pursuing a PhD in physics at the Institute of Acoustics, School of Physics Science and Engineering, in Tongji University. From 2021 to 2022, she was a visiting student in the School of Biomedical Engineering and Imaging Sciences, King's College London. Her current research focuses on photoacoustic (PA) measurement of microvascular evaluation and PA imaging.

**Long Wen** received a master's degree in dermatology and venereology from Anhui Medical University, China, in 2019. Currently, he is a junior physician fellow at Shanghai Skin Disease Hospital affiliated with Tongji University. His research interests include the non-invasive diagnosis of skin diseases, especially skin tumors, and photodynamic therapy of skin diseases.

**Chu Zhou** received his master's degree in dermatology and venereology from Anhui Medical University, Hefei, China. His research focuses on optical coherence tomography and PA imaging of skin diseases.

**Jing Pan** received her master's degree in physics from Tongji University, China, in 2019. Her research interest includes the medical imaging and spectrum analysis, especially PA imaging and analysis of skin tumors and prostate cancer.

**Shiying Wu** is a PhD candidate student at Institute of Acoustics, School of Physics Science and Engineering of Tongji University. She received her bachelor's degree from Tongji University. During her undergraduate study, she won the scholarship for outstanding students and the second prize of mathematical modeling contest. During her postgraduate study, she won the honor of doctoral student scholarship. Her current research focuses on PA spectrum analysis and PA imaging.

**Peiru Wang** is a dermatologist in Shanghai Skin Disease Hospital and an associate professor of Tongji University. She is engaged clinical transformation of photomedical technology and

treatment and perform research on skin aging and skin cancer. She has published 32 peer-reviewed articles in English and 8 articles in Chinese about dermatology and ALA-PDT. She is now involved in 8 academic projects including the Natural Science Foundation program (NSFC) and several clinical trials regarding skin diseases and cosmetics.

**Haonan Zhang** received his PhD in physics from Tongji University, Shanghai, China, in 2019. His research focuses on PAs imaging and ultrasonic therapy.

**Panpan Chen** received her BS degree in photoelectric information science and engineering from Anhui Normal University, China, in 2018. She is currently pursuing her MS degree at the Institute of Acoustics, Tongji University, China. Her current research interests include PA image reconstruction, bone imaging and deep learning. Her most recent research focused on the PA imaging of numerical bone model based on the modified deep learning network.

**Qi Chen** received his bachelor's degree in medical imaging from Wenzhou Medical University, Wenzhou, China. Currently, he is a graduate student at Tongji University School of Medicine, Shanghai, China. His research focuses on the application of PA imaging in dermatology.

**Xiuli Wang** is a doctoral tutor, the director of the Institute of Photomedicine, Tongji University School of Medicine, the chief expert of the Photodynamic Therapy Research Center of the Dermatology Branch of the Chinese Medical Association, and the executive director of the International Photodynamic Medicine Association. She was the first to carry out photodynamic clinical treatment, basic research, and technology promotion in dermatology in China. She took the lead in establishing the China Photodynamic Clinical Treatment Center and formulated a Chinese plan for the photodynamic treatment of skin diseases.

**Qian Cheng** received her BS degree in physics, and MS and PhD degrees in acoustics from Tongji University, China, in 2000, 2003, and 2006, respectively. She is currently a professor at Tongji University. Since 2006, her research interests have been optoacoustic phenomena, near-field acoustic imaging technique, Schlieren imaging technique, and the development of the acoustic detecting instruments. Her most recent research has focused on the clinical translation of PA imaging and quantitative analysis, and in particular for tumor diagnosis and evaluation.