

RESEARCH

Open Access



Employing Raman Spectroscopy and Machine Learning for the Identification of Breast Cancer

Ya Zhang¹, Zheng Li¹, Zhongqiang Li¹, Huaizhi Wang¹, Dinkar Regmi¹, Jian Zhang², Jiming Feng³, Shaomian Yao³ and Jian Xu^{1*}

Abstract

Background Breast cancer poses a significant health risk to women worldwide, with approximately 30% being diagnosed annually in the United States. The identification of cancerous mammary tissues from non-cancerous ones during surgery is crucial for the complete removal of tumors.

Results Our study innovatively utilized machine learning techniques (Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN)) alongside Raman spectroscopy to streamline and hasten the differentiation of normal and late-stage cancerous mammary tissues in mice. The classification accuracy rates achieved by these models were 94.47% for RF, 96.76% for SVM, and 97.58% for CNN, respectively. To our best knowledge, this study was the first effort in comparing the effectiveness of these three machine-learning techniques in classifying breast cancer tissues based on their Raman spectra. Moreover, we innovatively identified specific spectral peaks that contribute to the molecular characteristics of the murine cancerous and non-cancerous tissues.

Conclusions Consequently, our integrated approach of machine learning and Raman spectroscopy presents a non-invasive, swift diagnostic tool for breast cancer, offering promising applications in intraoperative settings.

Keywords Raman spectroscopy, Machine learning, Late-stage breast cancer; Cancerous and non-cancerous tissue classification

Background

Breast cancer is one of the most prevalent cancers diagnosed in females in the United States; new breast cancer accounts for 31% of estimated various cancers, and mortality of all types of breast cancer is the second out of all female cancerous diseases [1]. Regular breast cancer screening, including clinical breast exams and mammograms, has enhanced early detection rates and is crucial for the prognosis and treatment planning of surgery, radiation therapy, chemotherapy, and the latest targeted therapies [2]. The diagnosis of breast cancer is not only via imaging but also associated with histopathological analysis of the patient's tissue, which is invasive and painful [3].

*Correspondence:

Jian Xu

jianxu1@lsu.edu

¹Division of Electrical and Computer Engineering, College of Engineering, Louisiana State University, Baton Rouge, LA 70803, USA

²Division of Computer Science & Engineering, College of Engineering, Louisiana State University, Baton Rouge, LA 70803, USA

³Department of Comparative Biomedical Science, School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA 70803, USA



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Haka and colleagues compare the pathological reports to the breast cancer Raman spectra classification, achieving over 90% sensitivity and specificity [4]. For late-stage breast cancer and non-surgical breast cancer cases, accurately assessing prognosis becomes particularly critical in guiding therapeutic decisions, personalizing treatment plans, and improving patient outcomes [5].

The most common imaging modalities for breast cancer detection are mammograms, magnetic resonance imaging (MRI), ultrasound, and fluorescence imaging [6–11]. However, those approaches have various limitations, which impact their diagnostic efficacy. In particular, mammography exhibits reduced sensitivity in the presence of dense mammary tissue, a condition that can mask potential malignancies. MRI, while offering detailed tissue visualization, incurs excessive costs and often necessitates contrast agents, posing potential risks and discomfort for patients. Ultrasound, despite its non-invasive nature, is heavily dependent on the operator's skill and experience, leading to variability. Indocyanine Green (ICG)-assisted near-infrared imaging illustrated the ability to identify cancerous and non-cancerous tissues. Researchers applied ICG to the sentinel lymph node for early breast cancer detection, achieving solid diagnostic results [12–14]. However, fluorescence imaging is exogenous imaging in nature, which brings contamination to the tissues and the fluorescent dye may cause side effects to the patients [15–18]. The Raman imaging system can not only avoid this problem by directly detecting the tissues without any processes but also provide chemical component information [19]. Moreover, compared to conventional imaging methods, optical imaging and spectroscopy methods like Raman spectroscopy are minimally invasive and offer the advantage of quicker, more specific, and more sensitive cancer detection [20]. Raman spectroscopy serves as a molecular fingerprinting technique, evaluating vibrational and rotational energies to examine intermolecular functional groups and their molecular structures. This method facilitates rapid molecular analysis of tissues *in vivo* and *ex vivo*, suitable for biopsy evaluations or laboratory investigations, owing to its non-destructive approach [19]. It can discern variations in molecular compositions and structures between normal and breast cancer tissues, making it a powerful tool for identifying cancerous changes with precision [4, 21–25]. Traditional approaches to analyzing Raman Spectra mainly involve manual feature selection and linear statistical models, which may not capture the high-dimensional, nonlinear relationships inherent in the data. These limitations can hinder the accuracy and robustness of cancer diagnostics based on Raman spectroscopy, making it difficult to differentiate between various types and stages of cancer effectively.

Machine learning is currently being explored in diverse cancer diagnosis and classification fields. By feeding large amounts of biomedical data (e.g., cancerous Raman spectra), machine learning algorithms can autonomously deliver diagnostic outcomes and rapidly and effectively explore hidden valuable features related to cancerous tissues [26]. Machine learning applications have been used for breast cancer analysis, achieving effective results in previous research [27]. Kneipp et al. utilized PCA and K-means algorithms to differentiate between secretions from normal and cancerous breast duct epithelial cells [28]. Wu et al. achieved over 90% accuracy in classifying luminal and basal-like breast cancer subtypes using SVM-based algorithms that analyzed pathway-based biomarkers linked to specific genes [29]. There are few studies about late-stage breast cancer diagnosis using Raman spectroscopy, especially in the mouse model. Kast et al. applied principal component analysis - discriminant function analysis (PCA-DFA) for breast cancerous tissue and normal tissue classification [30]. Though some human breast cancer tissues were studied with the Convolutional Neural Networks (CNN) model [31, 32], there were few animal breast cancer studies reported. Animal models may also provide useful insights for clinical diagnoses; chemical components and contents could provide a consistent comparison to the human model. To our knowledge, this is the first study of CNN-enhanced signal processing for Raman spectroscopy-assisted animal breast cancer diagnosis for classification and feature extraction.

Evaluating the efficacy of machine learning-assisted Raman spectroscopy in diagnosing late-stage breast cancer in mouse models is imperative [33]. Raman spectroscopy can be performed to provide detailed molecular-level information about the tissues' chemical composition, enabling precise differentiation between cancerous and non-cancerous mammary tissues [21]. To the best of our knowledge, this study marks a pioneering effort in employing Raman spectroscopy, enhanced with machine learning algorithms—specifically, Random Forest, Support Vector Machine, and Convolutional Neural Networks—for the detailed analysis of the stage IV breast cancer tissues.

Materials and Methods

Cell Lines and Medium

4T1 (ATCC[®], CRL-2539[™]), a mouse breast cancer cell line that mimics human Stage IV breast cancer, was applied in this study to perform the nonsurgical model of breast cancer. The 4T1 cells were cultured in RPMI-1640 Medium (ATCC[®], 30-2001[™]) with 1% Penicillin-Streptomycin (Fisher Scientific, Gibco[™] 15-140-122), 10% Fetal Bovine Serum (Fisher Scientific, Gibco[™] A5256801) in

T75 flasks in the sterile and humidity incubator, setting 37 °C, 5% Carbon Dioxide.

Animal, Raman Spectroscopy System, and Raman Measurement

Animal Model

To generate the allograft animal model, the 4T1 cells were subcutaneously injected into 20 six- to eight-week-old athymic nude Nu/J female mice (IMSR_JAX:002019). Each mouse was injected with 2×10^6 4T1 cells resuspended in 100 μ L phosphate-buffered saline (PBS). When the tumor volume was about 50 mm³, the mice were euthanized via isoflurane as the first form and cervical dislocation as the second form. This study was approved by the Institutional Animal Care and Use Committee of Louisiana State University (The protocol number: IACUC#23–061), and all operations followed the guidelines on animal research.

Raman Spectroscopy System and Raman Measurement

The Raman Spectroscopy system used in this study consists of Raman Endoscopic Probe (EmVision LLC, Loxahatchee, Florida, United States), QE Pro spectrometer (Ocean Optics, Inc. Orlando, Florida, United States), and 785 nm laser diode source (Turnkey Raman Laser-785 Series, Ocean Optics Inc., Orlando, Florida, United States) connected with a desktop computer to perform

Raman data acquisition via OceanView Software with 3 s exposure time [34, 35]. Once the mice were euthanized, the tumor was resected for collecting Raman spectra (Fig. 1). Eighteen female mice were used for data acquisition, 959 Raman spectra were collected from the tumor, and 1075 Raman spectra were collected from the breast (Fig. 1a and b). Breast cancer specimens and normal mammary tissues were examined histologically after hematoxylin and eosin (H&E) staining (Fig. 1c and d).

Data Processing

Preprocessing of the collected raw Raman data is of chief importance as the data contain multiple noises [36–38]. The preprocessing was guided by the following steps: autofluorescence backgrounds of raw data were removed by asymmetric least squares fitting [39, 40]; Savitzky-Golay smoothing filter was applied to remove the noise without changing the main peak intensity [41]; normalizing the Raman data from 0 to 1 via mapping the minmax function. The procedures were implemented in MATLAB (version R2022a, MathWorks Inc., Natick, MA, USA).

Data Analysis via Random Forest (RF) Model, Support Vector Machine (SVM) Model, and Convolutional Neural Network (CNN)

Figure 2 demonstrated the structures of RF, SVM, and proposed CNN models [35, 42, 43]. Random Forest is

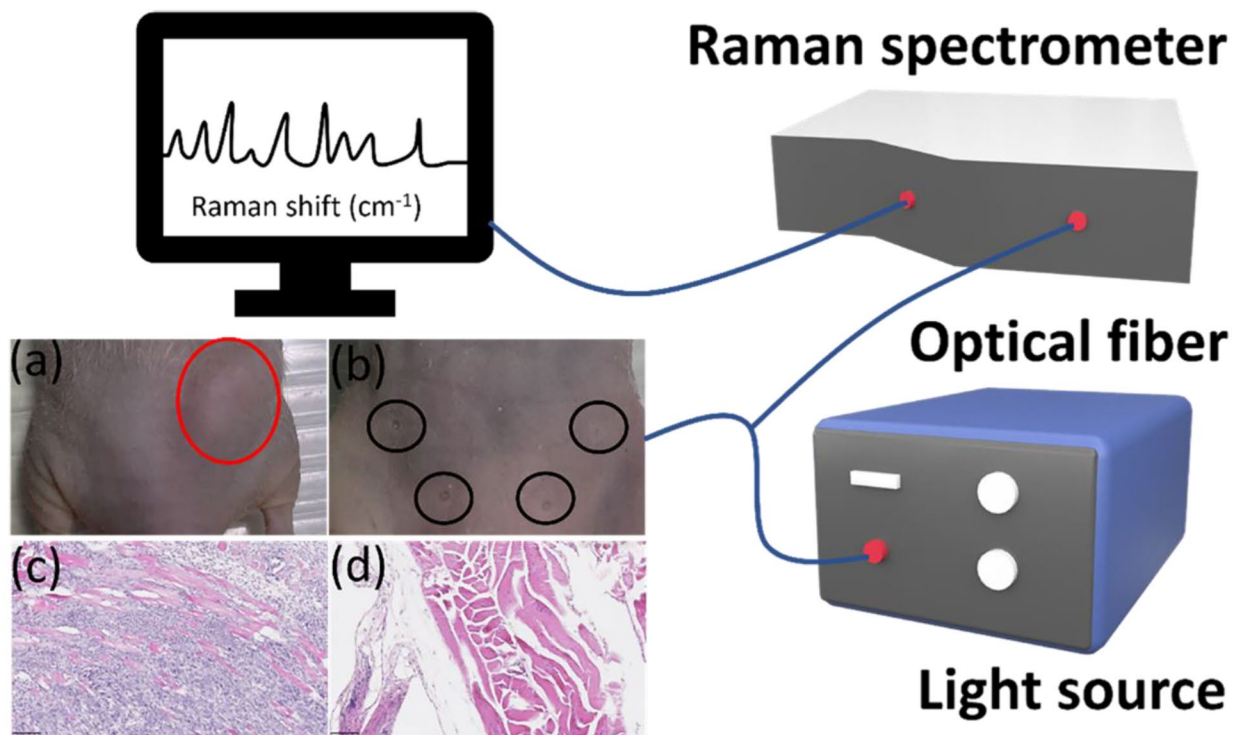


Fig. 1 Schematic diagram of Raman system in a murine cancer model. (a) tumor; (b) normal breast; (c) tumor with H&E staining; (d) normal breast with H&E staining

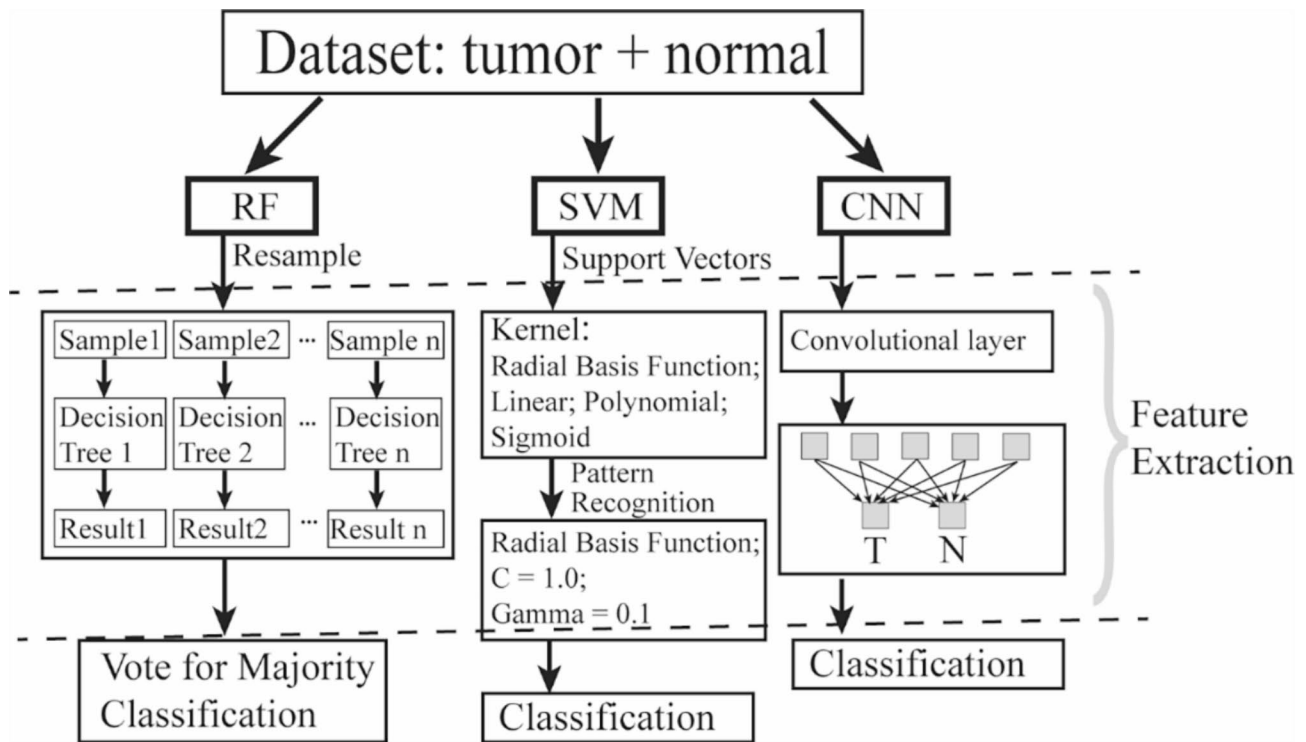


Fig. 2 Structures of RF, SVM, and CNN models

based on decision trees, and each decision would achieve a result after resampling. The majority vote finalizes the classification performance. SVM uses support vectors/margins and kernel of radial basis function/ linear/ polynomial/ sigmoid for the classification. The CNN model has one convolutional layer and one fully connected layer. The CNN model utilized a kernel size of five coupled with a stride of two. It employed the binary cross-entropy loss function (specifically BCEWithLogitsLoss) and was configured with a learning rate of 0.01, a momentum setting of 0.9, and a weight decay parameter set to 0.00004. For optimization, the model leveraged Stochastic Gradient Descent (SGD).

Results

H&E Staining

Figure 1c reveals invasive cancerous cells penetrating the normal muscle tissue, accompanied by an increased density of blood vessels. In contrast, Fig. 1d depicts normal mammary tissue characterized by regular arrangements of muscle and fat tissue, illustrating the typical structure and composition of healthy mammary tissue. This heightened vascular presence suggests a greater consumption of nutrients by the cancerous tissue compared to normal tissue, indicative of the aggressive nature of invasive breast cancer. The observed pathological features in these images have been validated by experienced pathologists, confirming the diagnostic significance of these findings.

Raman Spectra

Figure 3 shows the normalized average Raman spectra of healthy and cancerous mammary tissues in the range of 600–1800 cm^{-1} . The pronounced lipid content (e.g., 968, 1442, and 1738 cm^{-1}) exists in the healthy mammary tissue (Fig. 3a); conversely, breast cancer tissue's elevated intensity of proteins (e.g., 890 cm^{-1} and 1104 cm^{-1}) and decreased 1442 cm^{-1} band owing to the contribution of lipids [44] (Fig. 3b). The increased protein content and altered lipid profiles in cancerous tissues indicate the metabolic reprogramming associated with cancer progression [3, 30].

Classification Performances of Machine Learning Models

For RF classification, our study allocated 80% of the data to train the model and reserved the remaining 20% for testing its efficacy. The model demonstrated an average accuracy rate of 94.47%. It achieved a specificity of 96.73% and a sensitivity of 92.4%. The receiver operating characteristic (ROC) curve of RF is shown in Fig. 4a. The area under the curve (AOC) was 0.9849.

In the exploration of SVM classification utilizing the Radial Basis Function (RBF) kernel, our study partitioned the data, allocating 80% for training and the remaining 20% for testing. The RBF-SVM model demonstrated a commendable average accuracy of 96.76%, with an impressive specificity of 98.74% and a sensitivity of 94.90%. The receiver operating characteristic (ROC)

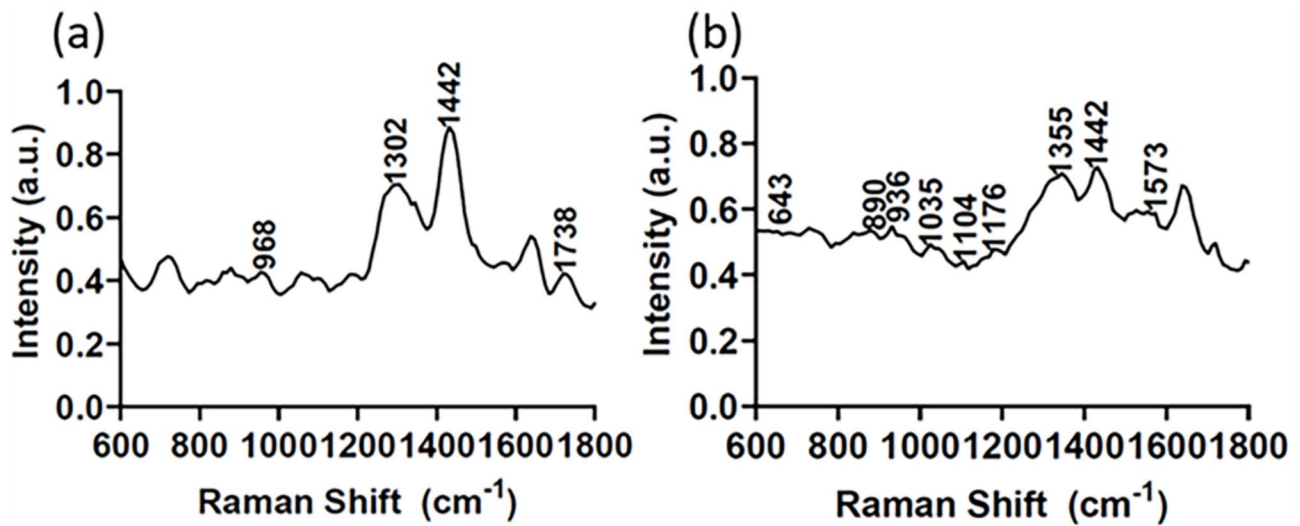


Fig. 3 Averaged Raman spectra of normal tissues (a) and breast cancer (b) with their respective remarkable peaks

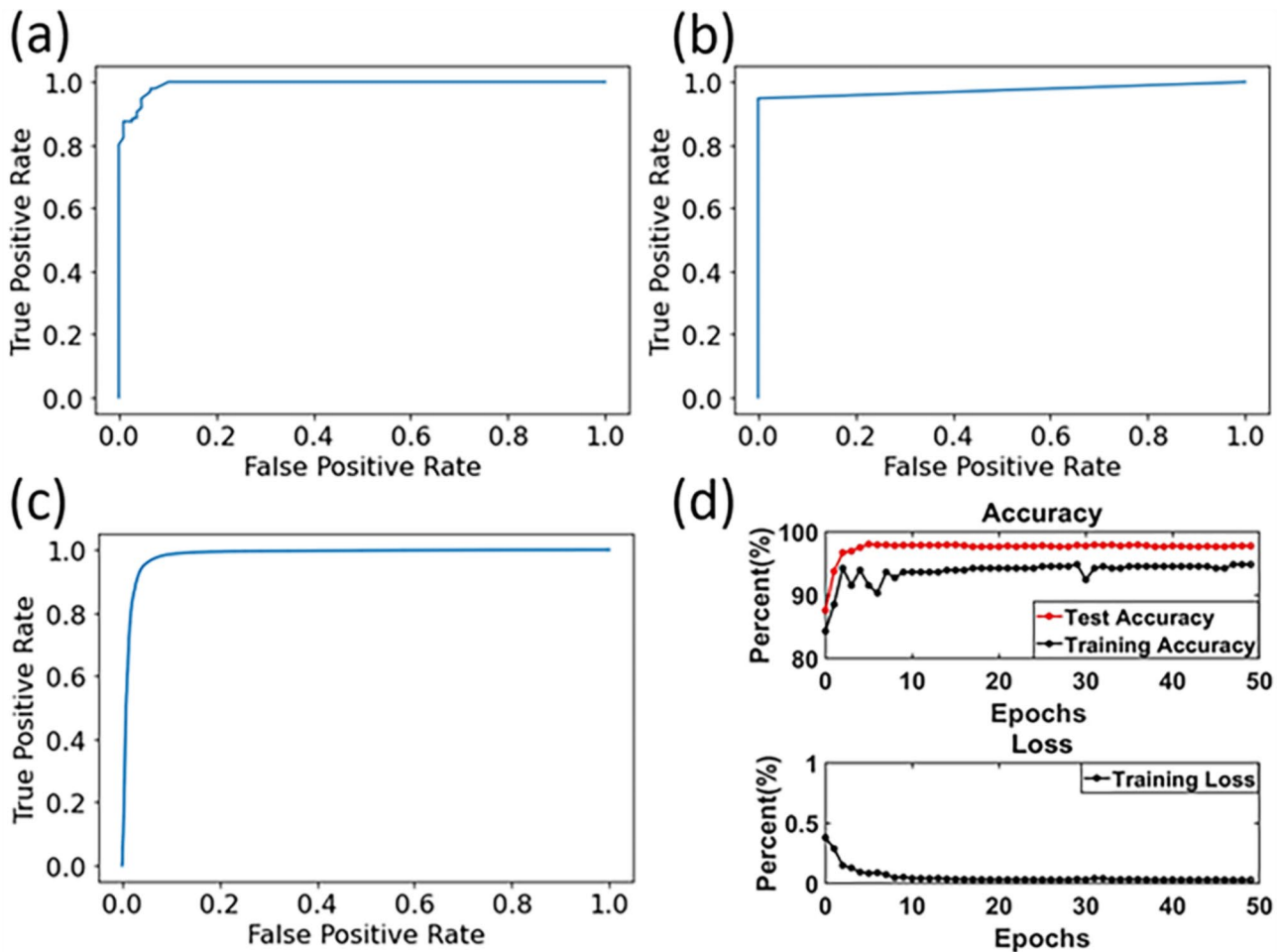


Fig. 4 ROC curves of RF (a), RBF-SVM (b), and CNN (c); (d) accuracy/loss to epochs curve of CNN model

Table 1 Classification performances of RF, SVM, and CNN models

Algorithm	Kernel function type	Sensitivity	Specificity	Accuracy
RF		92.40%	96.73%	94.47%
SVM	RBF	94.90%	98.74%	96.76%
	Linear	94.38%	98.60%	96.60%
	Polynomial	94.53%	98.56%	96.76%
	Sigmoid	0.73%	99.77%	45.99%
CNN		95.65%	99.51%	97.58%

curve of RBF-SVM is shown in Fig. 4b. The area under the curve (AOC) was 0.9722. We also test other kernels of SVM models. The model's performances are shown in Table 1.

For CNN classification, our study applied 80% of the data for training with 50 epochs and the rest for testing. The CNN model achieved an average accuracy of 97.58%, with unparalleled specificity and sensitivity of 99.51% and 95.65%, respectively. The receiver operating characteristic (ROC) curve of CNN is shown in Fig. 4c. The area under the curve (AOC) was 0.9842. The accuracy became convergent after ~35 epochs (Fig. 4d).

Discussion

The application of machine learning-assisted Raman spectroscopy extends beyond breast cancer, showcasing its versatility across different cancer types. Our group has performed this approach on pancreatic cancer and laryngeal cancer [35, 36], which have been successfully verified in both mice and humans. These applications have further validated the technique's efficacy, demonstrating high accuracy and sensitivity in detecting late-stage cancers, including murine breast cancer models. Despite these promising results in many other cancers, the application of machine learning-assisted Raman spectroscopy for the diagnosis of breast cancer has not yet been explored in late-stage human subjects. Our study is the first work of late-stage breast cancer in the mouse model, which could be a prior exploration of the human model before the clinical trials.

The range of 600–1800 cm^{-1} is notably responsive to molecular alterations, offering insights into the intricate molecular interactions among various bonds [45]. Such spectral analysis is instrumental in identifying changes in biochemical components across different tissues, facilitating the differentiation between normal and pathological states [45]. In Fig. 3, the intensity of the Raman spectrum of health tissue is significantly different than that of cancerous tissue: at the beginning of the Raman shift (600–1200 cm^{-1}), the normal has a lower intensity than tumor; from 1200 to 1500 cm^{-1} , the tumor has higher intensity, especially the peak of 1442 cm^{-1} ; after that, the tumor has slightly higher intensity again. The difference in spectral intensity between healthy and

Table 2 Peak assignments of chemical components and bonds

Peak (cm^{-1})	Tentative assignment	References
643	C-C twisting (tyrosine)	[48–50]
719	$\text{N}^+(\text{CH}_3)_3$, choline group, stretch band, C-N (membrane phospholipid head)/nucleotide peak, characteristic for Phosphatidylcholine, sphingomyelin	[51]
740	Adenine/nucleic acid	[52]
847	C-O-C skeletal mode, Monosaccharides	[53]
869	Proline	[50, 54]
890	Structure protein modes of tumors	[55]
936	C-C protein backbone α -helix(proline/glycogen), collagen	[30, 56]
968	Lipids, $\delta(=\text{CH})$ wagging	[30, 50]
1035	Collagen	[50, 57]
1066	Proline (collagen)	[23]
1104	Phenylalanine(proteins)	[50, 58]
1117	C-C stretch	[50, 54]
1176	C-H bending tyrosine(proteins)	[55, 56]
1264	$\delta(=\text{CH})$ of triglycerides	[30, 50, 59]
1302	$\delta(\text{CH}_2)$ twisting, wagging, phospholipids(lipid)	[30, 54, 56]
1355	Guanine	[47, 60]
1442	Triglycerides	[3, 50]
1573	Guanine, adenine, TRP (protein)	[50, 55]
1645	α -helix (Amide I)	[50]
1656	α -helix (Amide I)	[30, 50]
1720	Amide I	[61]
1738	Lipids	[50]

cancerous tissues is largely due to the change in lipid content and proteins. Lipid, the major composition in the mammary tissue, has a big Raman cross-Sect. [46]. The Raman bands of lipids at 1302, 1442 cm^{-1} weaken cancerous tissues, which suggests a depletion of lipid reserves during the cancer transformation process. The Raman band of 890 cm^{-1} reflects the structural protein modes of tumors [47]. Compared with healthy tissues, the bands at 936, 1176, and 1573 cm^{-1} are visible in cancerous tissues. These observations imply a dominant protein contribution, underscoring the molecular changes that occur as tissue becomes cancerous.

Several literature sources identify specific molecular structures with specific Raman peaks. The relevant peak assignments for our data are noted in Table 2. The Raman spectrum of normal mammary tissue (Fig. 3a) is dominated by contributions from lipids. The peaks at 1302 and 1442 cm^{-1} reflect the lipid-rich composition of the tissue. The Raman spectra of mammary gland tumors (Fig. 3b) reflect increased protein and reduced lipid compared to normal mammary gland tissue. This shift is evidenced by the presence of more pronounced protein peaks at frequencies of 643, 890, 936, 1035, 1104, 1176, 1355, and 1573 cm^{-1} . The variation in peak intensities between normal and cancerous tissues underscores the molecular

changes accompanying the transition to a cancerous state, with a marked reduction in lipid content and a concomitant increase in proteins.

While these results are promising, they may misconstrue the actual diagnostic ability of Raman spectroscopy. The operating environment in which the spectra were collected, with minimal variability between samples and under the guidance of experienced pathologists, may not fully represent the complexities and challenges encountered in-vivo studies. The introduction of greater variability in sample collection methods and the potential lack of detailed pre-measurement information about the sample in less controlled settings could diminish the diagnostic accuracy. In addition, machine learning algorithms increase the probability of identifying pathological changes owing to their excellent data analysis properties. This study systemically compared three algorithms (RF, SVM, and CNN) with the Raman spectra of mouse breast cancer, which was the first effort in the field of spontaneous-Raman-scattering-aided breast cancer diagnosis [4, 44, 45, 54, 62, 63]; the CNN model typically outperformed the RF and SVM models in the cancerous and non-cancerous tissue classification. Multiple algorithms help validate the results.

Before Raman can be translated for clinical use, many barriers must be overcome. (i) A human tissue Raman data should be collected. (ii) Bio-clean equipment should be designed. To meet this challenge, human samples should be collected under the supervision of trained pathologists and fiber should be sterilizable. Ideally, accompanying algorithms must preprocess and classify the data in near real-time to give an immediate diagnosis in the operating room. In addition, we will also build a benign tumor label since this work only focuses on the normal tissue and malignant tumor.

Our study underscores the distinct Raman spectral features of normal and cancerous tissues and their utility in machine-learning models for diagnosing late-stage breast cancer. These findings pave the way for further research and development to overcome the challenges of translating Raman spectroscopy from a highly controlled research tool to a practical, real-time diagnostic instrument in clinical settings. Emphasizing these aspects can provide a balanced view of the technology's current achievements and the steps needed to realize its full potential in improving late-stage breast cancer diagnosis and treatment outcomes. The state-of-the-art cancer identification approach is the post-surgery histopathology analysis. In this work, we also did the histopathological test after the resection (Fig. 1c and d). However, there are some disadvantages of the intraoperative pathology analysis, such as high cost and long waiting time (compared to the Raman system). We'll also try to apply other

traditional methods (e.g., MRI) when we do clinical trials in the future.

Conclusion

This study represented the first effort to systematically compare the effectiveness of three machine learning algorithms (RF, SVM, and CNN) in classifying late-stage animal breast cancer and normal mammary tissue based on their spontaneous Raman scattering signals. The integration of Raman spectroscopy with machine learning techniques enables the automation of tissue classification processes. In particular, the proposed CNN demonstrated the best performance among machine learning approaches used in this study, with an average accuracy, specificity, and sensitivity of 97.58%, 99.51%, and 95.65%, respectively, which is significantly higher performance than previous studies in the field of breast cancer. The differentiation between normal and cancerous mammary tissues was primarily attributed to variations in lipid and protein concentrations, which are critical in the machine learning-based classification of mammary tissues. This underscores the pivotal role of molecular composition, particularly lipids and proteins, in distinguishing between healthy and pathological tissue states through Raman spectroscopy. Overall, the machine learning-assisted Raman spectroscopy demonstrated remarkable accuracy, sensitivity, and specificity in identifying late-stage breast cancerous tissues from non-cancerous tissues, which has the potential to be applied in human diagnosis in the future.

Through the identification of characteristic Raman peaks associated with advanced breast cancer, our approach has successfully demonstrated the potential of this hybrid technology in the accurate diagnosis of this disease stage. This innovation represents a significant leap forward, introducing a novel, efficient method for investigating late-stage breast cancer, which could revolutionize diagnostic practices and potentially improve patient outcomes by facilitating painless and more accurate detection.

Acknowledgements

We thank Sherry Ring from LSU Department of Comparative Biomedical Science for her preparation of the histopathology slides.

Author Contributions

Y.Z. did data curation, visualization, and writing – original draft. Z.L., Z.Q. L. did investigation, methodology, and validation. H.W. and D. R. did the investigation and methodology. J.Z.: did model modification. J.F., S.Y. did data interpretation and revision of the draft. J.X. did writing – review & editing, project administration, funding acquisition, and supervision.

Funding

This research was supported by NSF CAREER award (2046929) and LSU Collaborative Cancer Research Initiative (010163).

Data Availability

No datasets were generated or analysed during the current study.

Declarations

Ethics Approval and Consent to Participate

This study was approved by the Institutional Animal Care and Use Committee of Louisiana State University (The protocol number: IACUC#23–061), and all operations followed the guidelines on animal research.

Competing Interests

The authors declare no competing interests.

Received: 23 June 2024 / Accepted: 4 September 2024

Published online: 12 September 2024

References

1. Siegel RL, et al. Cancer statistics, 2023. *Ca Cancer J Clin*. 2023;73(1):17–48.
2. Huynh PT, Jarolimiek AM, Daye S. The false-negative mammogram. *RadioGraphics*. 1998;18(5):1137–54.
3. Hanna K, et al. Raman spectroscopy: current applications in breast cancer diagnosis, challenges and future prospects. *Br J Cancer*. 2022;126(8):1125–39.
4. Haka AS, et al. Diagnosing breast cancer by using Raman spectroscopy. *Proc Natl Acad Sci*. 2005;102(35):12371–6.
5. Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA*. 2019;321(3):288–300.
6. Ganesan K, et al. Computer-aided breast cancer detection using mammograms: a review. *IEEE Rev Biomed Eng*. 2012;6:77–98.
7. Menezes GL, et al. Magnetic resonance imaging in breast cancer: a literature review and future perspectives. *World J Clin Oncol*. 2014;5(2):61.
8. Guo R, et al. Ultrasound imaging technologies for breast cancer detection and management: a review. *Ultrasound Med Biol*. 2018;44(1):37–70.
9. Li Z, et al. Detection of pancreatic cancer by indocyanine green-assisted fluorescence imaging in the first and second near-infrared windows. *Cancer Commun*. 2021;41(12):1431.
10. Xu J, et al. New horizons in intraoperative diagnostics of cancer in image and spectroscopy guided pancreatic cancer surgery. *New Horizons Clin Case Rep*. 2017;1:2.
11. Veys I, et al. ICG fluorescence imaging as a new tool for optimization of pathological evaluation in breast cancer tumors after neoadjuvant chemotherapy. *PLoS ONE*. 2018;13(5):e0197857.
12. Sugie T, et al. Sentinel lymph node biopsy using indocyanine green fluorescence in early-stage breast cancer: a meta-analysis. *Int J Clin Oncol*. 2017;22:11–7.
13. Kitai T, et al. Fluorescence navigation with indocyanine green for detecting sentinel lymph nodes in breast cancer. *Breast Cancer*. 2005;12(3):211–5.
14. Murawa D, et al. Sentinel lymph node biopsy in breast cancer guided by indocyanine green fluorescence. *J Br Surg*. 2009;96(11):1289–94.
15. Robson A-L, et al. Advantages and limitations of current imaging techniques for characterizing liposome morphology. *Front Pharmacol*. 2018;9:80.
16. Zhang RR, et al. Beyond the margins: real-time detection of cancer using targeted fluorophores. *Nat Reviews Clin Oncol*. 2017;14(6):347–64.
17. Orosco RK, Tsien RY, Nguyen QT. Fluorescence imaging in surgery. *IEEE Rev Biomed Eng*. 2013;6:178–87.
18. Lassailly F, Griessinger E, Bonnet D. Microenvironmental contaminations induced by fluorescent lipophilic dyes used for noninvasive in vitro and in vivo cell tracking. *Blood J Am Soc Hematol*. 2010;115(26):5347–54.
19. Auner GW, et al. Applications of Raman spectroscopy in cancer diagnosis. *Cancer Metastasis Rev*. 2018;37:691–717.
20. Zhang L, et al. Raman spectroscopy and machine learning for the classification of breast cancers. *Spectrochim Acta Part A Mol Biomol Spectrosc*. 2022;264:120300.
21. Hanlon E, et al. Prospects for in vivo Raman spectroscopy. *Phys Med Biol*. 2000;45(2):R1.
22. Redd DC, et al. Raman spectroscopic characterization of human breast tissues: implications for breast cancer diagnosis. *Appl Spectrosc*. 1993;47(6):787–91.
23. Frank CJ, McCreery RL, Redd DC. Raman spectroscopy of normal and diseased human breast tissues. *Anal Chem*. 1995;67(5):777–83.
24. Bitar RA, et al. Biochemical analysis of human breast tissues xpp qa? Using Fourier-transform Raman spectroscopy. *J Biomed Opt*. 2006;11(5):054001–054001.
25. Haka AS, et al. Identifying microcalcifications in benign and malignant breast lesions by probing differences in their chemical composition using Raman spectroscopy. *Cancer Res*. 2002;62(18):5375–80.
26. Kim KG. Book review: deep learning. *Healthc Inf Res*. 2016;22(4):351.
27. Ozer ME, Sarica PO, Arga KY. New machine learning applications to accelerate personalized medicine in breast cancer: rise of the support vector machines. *OMICS*. 2020;24(5):241–6.
28. Kneipp J, et al. Characterization of breast duct epithelia: a Raman spectroscopic study. *Vib Spectrosc*. 2003;32(1):67–74.
29. Wu T, et al. A pathways-based prediction model for classifying breast cancer subtypes. *Oncotarget*. 2017;8(35):58809.
30. Kast RE, et al. Raman spectroscopy can differentiate malignant tumors from normal breast tissue and detect early neoplastic changes in a mouse model. Volume 89. *Biopolymers: Original Research on Biomolecules*; 2008. pp. 235–41. 3.
31. Fuentes AM, et al. Raman spectroscopy and convolutional neural networks for monitoring biochemical radiation response in breast tumour xenografts. *Sci Rep*. 2023;13(1):1530.
32. Shang L, et al. Polarized micro-Raman spectroscopy and 2D Convolutional Neural Network Applied To Structural Analysis and discrimination of breast Cancer. *Biosensors*. 2022;13(1):65.
33. Kourou K, et al. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
34. Li Z, et al. Indocyanine green-assisted dental imaging in the first and second near-infrared windows as compared with X-ray imaging. Volume 1448. *Annals of the New York Academy of Sciences*; 2019. pp. 42–51. 1.
35. Li Z, et al. Machine-learning-assisted spontaneous Raman spectroscopy classification and feature extraction for the diagnosis of human laryngeal cancer. *Comput Biol Med*. 2022;146:105617.
36. Li Z, et al. Detection of pancreatic cancer by convolutional-neural-network-assisted spontaneous Raman spectroscopy with critical feature visualization. *Neural Netw*. 2021;144:455–64.
37. Mazet V, et al. Background removal from spectra by designing and minimising a non-quadratic cost function. *Chemometr Intell Lab Syst*. 2005;76(2):121–33.
38. Cordero E, et al. Evaluation of shifted excitation Raman difference spectroscopy and comparison to computational background correction methods applied to biochemical Raman spectra. *Sensors*. 2017;17(8):1724.
39. He S, et al. Baseline correction for Raman spectra using an improved asymmetric least squares method. *Anal Methods*. 2014;6(12):4402–7.
40. Vickers TJ, Wambles RE Jr, Mann CK. Curve fitting and linearity: data processing in Raman spectroscopy. *Appl Spectrosc*. 2001;55(4):389–93.
41. Radzol A et al. *Optimization of Savitzky-Golay smoothing filter for salivary surface enhanced Raman spectra of non structural protein 1. in TENCON 2014–2014 IEEE Region 10 Conference*. 2014. IEEE.
42. Platt J, Cristianini N, Shawe-Taylor J. Large margin DAGs for multiclass classification. *Adv Neural Inf Process Syst*. 1999. 12.
43. Zheng B, Yoon SW, Lam SS. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst Appl*. 2014;41(4):1476–82.
44. Lazaro-Pacheco D, et al. Raman spectroscopy of breast cancer. *Appl Spectrosc Rev*. 2020;55(6):439–75.
45. Ma D, et al. Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network. *Spectrochim Acta Part A Mol Biomol Spectrosc*. 2021;256:119732.
46. Shafer-Peltier KE, et al. Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo. *J Raman Spectrosc*. 2002;33(7):552–63.
47. Talari ACS, et al. Raman spectroscopy of biological tissues. *Appl Spectrosc Rev*. 2015;50(1):46–111.
48. Cheng WT, et al. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma. *Microsc Res Tech*. 2005;68(2):75–9.
49. Contorno S, Darienzo RE, Tannenbaum R. Evaluation of aromatic amino acids as potential biomarkers in breast cancer by Raman spectroscopy analysis. *Sci Rep*. 2021;11(1):1698.
50. Movasaghi Z, Rehman S, Rehman IU. Raman spectroscopy of biological tissues. *Appl Spectrosc Rev*. 2007;42(5):493–541.
51. Krafft C, et al. Near infrared Raman spectra of human brain lipids. *Spectrochim Acta Part A Mol Biomol Spectrosc*. 2005;61(7):1529–35.
52. Staniszewska-Slezak E, Malek K, Baranska M. Complementary analysis of tissue homogenates composition obtained by Vis and NIR laser excitations

- and Raman spectroscopy. *Spectrochim Acta Part A Mol Biomol Spectrosc.* 2015;147:245–56.
53. Shetty G, et al. Raman spectroscopy: elucidation of biochemical changes in carcinogenesis of oesophagus. *Br J Cancer.* 2006;94(10):1460–4.
 54. Koya SK, et al. Accurate identification of breast cancer margins in microenvironments of ex-vivo basal and luminal breast cancer tissues using Raman spectroscopy. Volume 151. *Prostaglandins & Other Lipid Mediators*; 2020. p. 106475.
 55. Stone N, et al. Raman spectroscopy for identification of epithelial cancers. *Faraday Discuss.* 2004;126:141–57.
 56. David S, et al. In situ Raman spectroscopy and machine learning unveil biomolecular alterations in invasive breast cancer. *J Biomed Opt.* 2023;28(3):036009–036009.
 57. Andrus PG, Strickland RD. Cancer grading by Fourier transform infrared spectroscopy. *Biospectroscopy.* 1998;4(1):37–46.
 58. Lakshmi RJ, et al. Tissue Raman spectroscopy for the study of radiation damage: brain irradiation of mice. *Radiat Res.* 2002;157(2):175–82.
 59. Silveira L Jr, et al. Correlation between near-infrared Raman spectroscopy and the histopathological analysis of atherosclerosis in human coronary arteries. *Lasers Surg Medicine: Official J Am Soc Laser Med Surg.* 2002;30(4):290–7.
 60. Ruiz-Chica A, et al. Characterization by Raman spectroscopy of conformational changes on guanine–cytosine and adenine–thymine oligonucleotides induced by aminoxy analogues of spermidine. *J Raman Spectrosc.* 2004;35(2):93–100.
 61. Ogruc Ildiz G, et al. Raman spectroscopic and chemometric investigation of lipid–protein ratio contents of soybean mutants. *Appl Spectrosc.* 2020;74(1):34–41.
 62. Grajales D, et al. Towards real-time confirmation of breast Cancer in the OR using CNN-Based Raman Spectroscopy classification. *MICCAI Workshop on Cancer Prevention through early detection.* Springer; 2023.
 63. Fallahzadeh O, Dehghani-Bidgoli Z, Assarian M. Raman spectral feature selection using ant colony optimization for breast cancer diagnosis. *Lasers Med Sci.* 2018;33(8):1799–806.

Publisher's Note

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