*Exploratory Research and Hypothesis in Medicine* **2025** vol. 000(000) | 000–000 Epub
DOI: 10.14218/ERHM.2025.00014

## Original Article

# ATR-FTIR Spectroscopy for Differentiating Metaplastic Breast Carcinoma, Ductal Carcinoma *In Situ*, and Invasive Ductal Carcinoma: A Retrospective Study

Samuel T. Adeleke[1,2*] and Christopher Igbeneghu[2]

[1]*Department of Medical Laboratory Science, Bowen University, Iwo, Osun State, Nigeria;* [2]*Department of Medical Laboratory Science, Ladoke Akintola University of Technology, Ogbomosho, Oyo State, Nigeria*

**Received:** June 24, 2025 | **Revised:** July 09, 2025 | **Accepted:** July 14, 2025 | **Published online:** July 31, 2025

## Abstract

**Background and objectives:** Metaplastic breast carcinoma, a rare entity (<1% of breast neoplasms), lacks comprehensive spectroscopic characterization. This study aimed to address this gap by providing a qualitative and quantitative spectroscopic profile of metaplastic carcinoma in comparison to ductal carcinoma *in situ* (DCIS) and invasive ductal carcinoma (IDC).

**Methods:** A retrospective analysis was conducted on archival tissue blocks of metaplastic carcinoma (n = 10), DCIS (n = 12), and IDC (n = 31). Sections were stained with hematoxylin and eosin for histological confirmation. Attenuated total reflectance Fourier-transform infrared spectroscopy was performed on adjacent unstained sections, with normal breast tissue (n = 10) serving as the control. Spectral data were analyzed using t-tests to identify significant differences in peak intensities and ratios. Hierarchical clustering analysis and receiver operating characteristic curves were generated to assess the diagnostic potential of selected spectral features.

**Results:** Spectral analysis revealed that mean peak intensities were generally lower in all carcinoma subtypes compared to normal breast tissue. Specific ratios, including A1237/A1080 (phosphate; $p < 0.01$), A1043/1543 (glycogen; $p < 0.01$), and A1080/A1632 (nucleocytoplasmic index; $p < 0.03$), were significantly elevated in carcinomatous tissues. Receiver operating characteristic analysis identified peak 3,280 (area under the curve (AUC) = 0.93–0.96) as highly effective in differentiating normal from carcinomatous tissues. Peak 2,922 showed specificity for distinguishing normal tissue from IDC (AUC ≈ 0.7). Peak 1,744 effectively discriminated between DCIS and metaplastic carcinoma (AUC = 0.7). The ratio 1,080/1,632 (nucleocytoplasmic ratio) demonstrated exceptional diagnostic accuracy, distinguishing normal from carcinomatous tissues (AUC ≈ 1.0), DCIS from IDC (AUC ≈ 0.86), and DCIS from metaplastic carcinoma (AUC ≈ 0.8).

**Conclusions:** Attenuated total reflectance Fourier-transform infrared spectroscopy, particularly using peak 3,280 (Amide A) and the 1,080/1,632 ratio (nucleocytoplasmic index), offers a promising approach for discriminating between normal breast tissue and carcinoma, as well as differentiating pre-IDC from metaplastic carcinoma. These spectral markers demonstrate both statistical significance and diagnostic potential.

**Keywords:** Metaplastic carcinoma; Peak ratios; Tissue blocks; Dendrogram; Vibrational modes; Cut-off points; Biomarkers.

***Correspondence to:** Samuel T Adeleke, Department of Medical Laboratory Sciences, Bowen University, Iwo, Osun State 232102, Nigeria. ORCID: https://orcid.org/0000-0001-9912-9573. Tel: +234-8133114950; +234-8155781461, E-mail: Samuel.adeleke@bowen.edu.ng

**How to cite this article:** Adeleke ST, Igbeneghu C. ATR-FTIR Spectroscopy for Differentiating Metaplastic Breast Carcinoma, Ductal Carcinoma *In Situ*, and Invasive Ductal Carcinoma: A Retrospective Study. *Explor Res Hypothesis Med* 2025;000(000):000–000. doi: 10.14218/ERHM.2025.00014.

## Introduction

Metaplastic carcinoma (cancer), constituting less than 1% of all breast carcinoma cases,[1–3] is a rare breast tumor characterized by a mix of distinct epithelial types, predominantly squamous cells, and occasionally exhibiting features akin to adenocarcinoma and mesenchymal cells.[4] The tumor demonstrates significant histological and molecular heterogeneity, displaying histopathological characteristics comparable to those of invasive ductal carcinoma

(IDC).[3] The histopathological diagnosis is crucial for elucidating the clinicopathological traits associated with this aggressive breast carcinoma subtype, which often presents as high-grade and is prone to frequent recurrences.[1,5] Notably, immunohistochemical analyses often reveal a predominance of triple-negative characteristics, correlating with a poor prognosis.[6,7] Researchers have explored molecular characterization to identify unique abnormalities and potential drug-targetable biomarkers associated with this variant.[8,9] Evidence underscores its aggressive nature, leading to poorer prognosis and low response rates to targeted therapies.[1,6–9]

Recently, attenuated total reflectance (ATR) Fourier-transform infrared (FTIR) spectroscopy and other infrared-based methodologies have significantly advanced the understanding of carcinoma biochemistry. ATR-FTIR spectroscopy is a powerful analytical tool used to differentiate biological samples by detecting chemical signatures produced when infrared light induces vibrational transitions in IR-active atoms. The resulting spectral patterns provide insights into the biochemical composition of various samples. Notably, extensive investigations have documented the characterization of breast lesions utilizing infrared spectroscopy.[10–14] Various breast lesions, including fibroadenoma, fibrocystic changes, hyperplasia, intraductal and lobular carcinoma, IDC, and rare forms such as tubular, medullary, and apocrine mucinous breast carcinomas, have been analyzed, yielding unique spectral signatures for lipid, protein, nucleic acid, and carbohydrate components.[1,10,15–18] However, there remains a significant gap in the literature regarding the spectral characterization and biomarker identification of metaplastic breast carcinoma compared to normal breast tissue and ductal carcinomas.

Furthermore, most existing studies on the diverse spectrum of breast neoplasms have primarily focused on identifying spectral differences, with insufficient emphasis on the statistical significance of discriminative biomarkers such as nucleic acids, glycogen, proteins, and lipids.[13,18,19] Where statistical significance has been evaluated, diagnostic relevance, such as through receiver operating characteristic (ROC) curve analysis or logistic regression modeling, has been minimally addressed, particularly for breast tissue, with no specific studies focusing on rare subtypes like metaplastic carcinoma.[11,13,19–21] It is essential to assess the diagnostic potential of markers that demonstrate both spectral and statistical significance. Therefore, this study aimed to leverage ATR-FTIR spectroscopy to compare the intensity variations of various peaks and biomarkers related to phosphate, proteins, nucleocytoplasmic ratios, and glycogen, to effectively differentiate between metaplastic carcinoma, ductal carcinoma *in situ* (DCIS), IDC, and normal breast tissue from both statistical significance and diagnostic performance perspectives.

## Material and methods

### *Ethical considerations*

Ethical approval for this study was obtained from the Obafemi Awolowo University Teaching Hospital Ethics Review Committee, reference number: ERC/2025/01/01. The study was carried out with a waiver of informed consent for the use of archived, anonymized samples in accordance with the principles of the Declaration of Helsinki (as revised in 2024). The ethics committee agreed that this study did not require informed consent due to its retrospective nature and the minimal risk involved.

### *Tissue block retrieval*

A total of 53 formalin-fixed, paraffin-embedded tissue blocks from patients previously diagnosed with metaplastic carcinoma, preinvasive ductal carcinoma, and invasive ductal carcinoma between 2020 and 2024 were collected retrospectively from histopathology archives for re-review and subsequent preparation for ATR-FTIR analysis.

### *Tissue block analysis*

#### Histopathology staining

Tissue sections measuring 5 µm were prepared using Leica microtomes from blocks containing 10 metaplastic carcinoma cases, 12 DCIS cases, 31 invasive ductal carcinomas, and 10 samples of adjacent normal breast tissue from DCIS were included as control. The first set of sections underwent staining using standard hematoxylin and eosin protocols following the Carson and Capellano protocol.[22] The stained sections were reviewed by a pathologist, and the areas corresponding to metaplastic carcinoma, DCIS, and IDC (as shown in Fig. S1) were carefully marked for accurate point mapping with ATR-FTIR, as depicted in the flowchart (Fig. 1).

#### ATR-FTIR spectra acquisition and preprocessing

According to Cui *et al*.,[23] 20 µm-thick sections were sectioned with a Leica microtome and floated in a prewarmed water bath at 40–50°C, then picked up onto substrates. These sections were dried overnight, dewaxed in fresh xylene (Surgipath Medical Industries, Inc.), and dehydrated through descending grades of alcohol from 100% to 70%, then left to air-dry prior to ATR-FTIR analysis. Following established procedures, the second set of sections was placed on aluminum foil as a low-reflective substrate and brought into contact with an ATR crystal made of diamond for subsequent spectroscopic analysis using an Agilent Cary 630 spectrometer.[23] Blank measurements were taken with the substrate before recording measurements using point mapping on the marked carcinomatous areas of the tissue sections. On average, five spectra were recorded, preprocessed, and analyzed from each tissue section, with 200 scans collected at a resolution of 16 $cm^{-1}$. The spectral data were further subjected to normalization, smoothing, and baseline correction using Spectragryph software, as shown in the flowchart (Fig. 1). Key peaks and their corresponding intensities were documented in both tabular (Table 1) and graphical formats (Fig. 2).

### *Statistical analysis*

Intensity ratios from selected peaks were calculated. Statistical analyses, including Student's t-test (Bonferroni-corrected), were performed using SPSS version 26 to assess the significance of peak intensities and ratios (biomarkers), with a significance threshold of $p < 0.05$. Hierarchical clustering analysis (HCA) and ROC curve evaluations were also performed, providing area under the curve (AUC) values to determine the relatedness and diagnostic effectiveness of the observed peaks and ratios for metaplastic carcinoma and ductal carcinomas.

## Results

### *Qualitative analysis*

A total of 20 distinct peaks were identified and analyzed, enabling differentiation between normal tissue, ductal carcinomas, and metaplastic carcinoma (Fig. 2). The peaks at 3,280, 2,920, 1,744, 1,632, 1,543, 1,535, 1,528, 1,453, 1,446, 1,394, 1,386, 1,304, 1,274, 1,237, 1,230, 1,162, 1,155, 1,080, 1,073, and 1,043 $cm^{-1}$

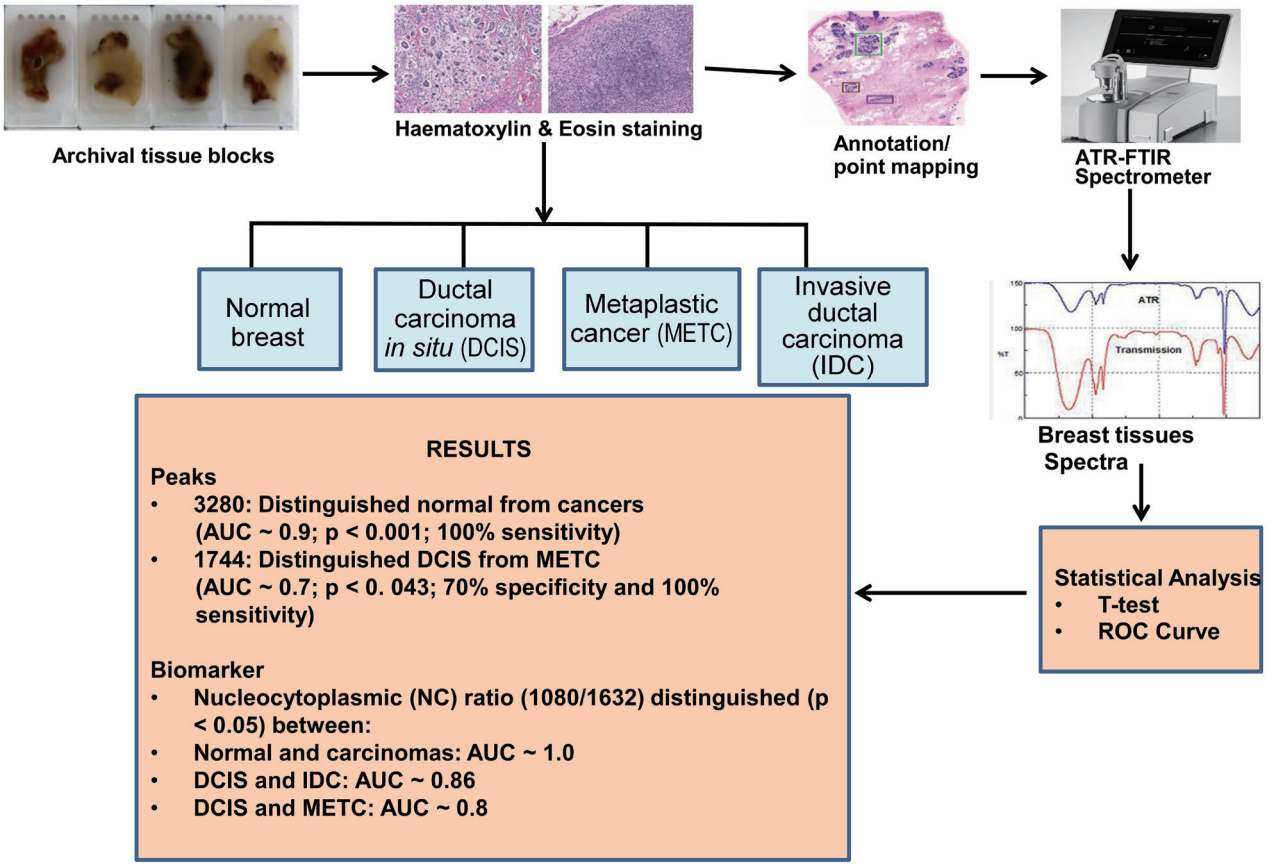


**Fig. 1. Flowchart of experimental procedure of ATR-FTIR spectroscopy of normal breast tissue, ductal and metaplastic carcinomas.** ATR-FTIR, attenuated total reflectance Fourier-transform infrared; AUC, area under the curve; ROC, receiver operating characteristic.

**Table 1. Peak shift differences between normal breast, ductal carcinoma *in situ*, invasive ductal carcinoma, and metaplastic carcinoma**

| Normal | Ductal carcinoma *in situ* | Invasive ductal carcinoma | Metaplastic carcinoma |
|---|---|---|---|
| 3,284 | 3,280 | 3,280 | 3,278 |
| 2,922 | 2,919 | 2,917 | 2,917 |
| 1,744 | 1,742 | 1,740 | 1,741 |
| 1,632 | 1,632 | 1,635 | 1,634 |
| 1,535 | 1,529 | 1,525 | 1,525 |
| 1,444 | 1,446 | 1,446 | 1,447 |
| 1,454 | 1,453 | 1,453 | 1,452 |
| 1,397 | 1,396 | 1,396 | 1,395 |
| 1,386 | 1,384 | 1,384 | 1,384 |
| 1,308 | 1,304 | 1,302 | 1,303 |
| 1,237 | 1,232 | 1,230 | 1,232 |
| 1,152 | 1,151 | 1,152 | 1,152 |
| 1,163 | 1,162 | 1,162 | 1,162 |
| 1,080 | 1,075 | 1,073 | 1,073 |
| 1,045 | 1,043 | 1,049 | 1,043 |

were characterized. Generally, most peaks corresponding to carcinomas exhibited a shift toward lower wavenumbers compared to normal breast tissue, indicating a rightward spectral shift (Table 1). This shift is consistent with specific molecular assignments presented in Table 2.[11,13–17,19,24–40] Table 1 reveals a distinct downward shift for peaks at 3,284, 2,922, 1,744, 1,535, 1,397, 1,308, 1,237, and 1,080 $cm^{-1}$ in ductal carcinoma *in situ*, invasive ductal carcinoma, and metaplastic carcinoma relative to normal breast tissue, suggesting conformational changes in the biomolecular assignments corresponding to these peaks. In contrast, peaks such as 1,632 and 1,444 $cm^{-1}$ shifted upward, with the former being more sensitive to protein conformation.

Normal breast tissue exhibited elevated absorbance in the higher wavenumber range (3,300–3,200 $cm^{-1}$), followed by ductal carcinoma *in situ*, compared with other breast carcinoma categories. Invasive ductal carcinoma and metaplastic carcinomas showed particularly increased absorbance within the 1,400–1,000 $cm^{-1}$ range compared to normal and *in situ* breast tissue.

### *Hierarchical clustering analysis (HCA)*

The a-fHCA conducted on the specified peaks illustrated varying degrees of relatedness and diversity, attributed to similarities and differences in the vibrational modes of various breast tissue types (Fig. 3a–f). Between normal and DCIS, the spectral features exhibited distinct groupings based on wavenumber proximity. Peaks at 1,162, 1,154, 1,304, 1,274, 2,922, 1,080, 1,073, and

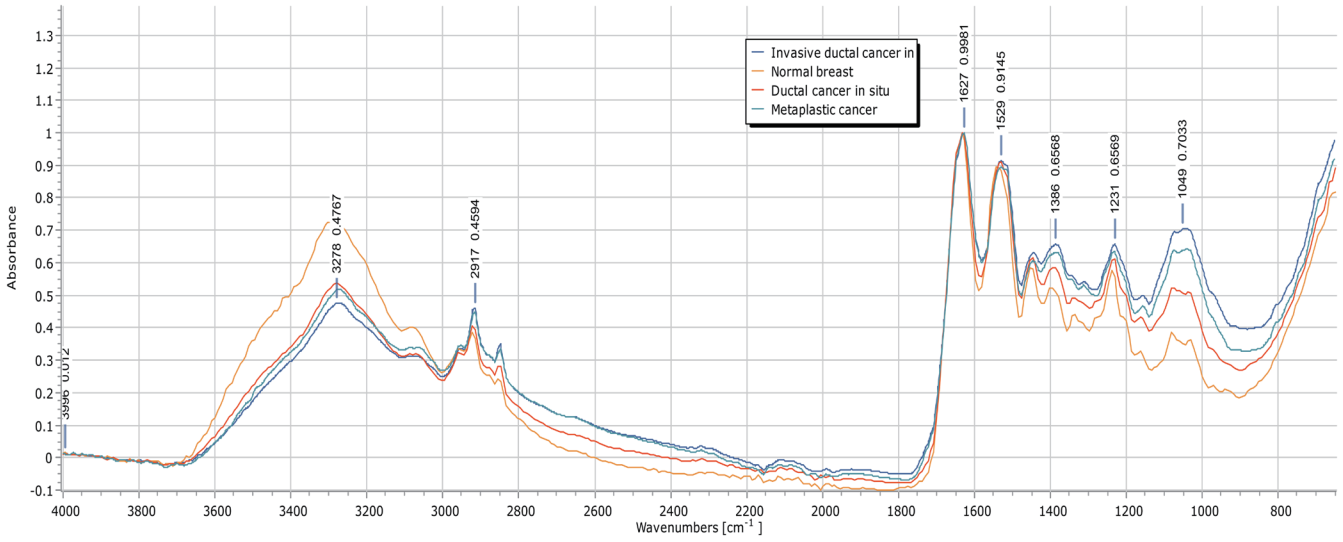


**Fig. 2. Spectral peaks discriminating between normal breast, ductal carcinoma *in situ*, invasive ductal carcinoma, and metaplastic carcinoma.**

1,043 $cm^{-1}$ formed a cohesive cluster, as did the sets 1,237/1,230/1,394/1,386/1,453/1,446 $cm^{-1}$ and 1,535/1,528/1,543/1,632 $cm^{-1}$. These groupings displayed significantly greater spectral relatedness than, for example, the relationship between peaks at 1,230 and 3,280 $cm^{-1}$ or 3,280 and 2,922 $cm^{-1}$. Notably, regions associated with protein vibrations were spectrally distant from clusters of lower-wavenumber peaks. In contrast, the peak at 1,744 $cm^{-1}$ demonstrated the most significant dissimilarity from all other spectral clusters examined. When comparing normal and metaplastic carcinomas, vibrational spectroscopy identified key spectral signatures (1,394–1,043 $cm^{-1}$) demonstrating homology within clusters. Specifically, peak intensities at 1,162–1,632 $cm^{-1}$ displayed the greatest discriminatory power between normal and metaplastic carcinomatous tissues. Secondary discriminatory potential was observed at 1,237 $cm^{-1}$ and 1,304 $cm^{-1}$, with diminished efficacy at 1,073 $cm^{-1}$, 1,155 $cm^{-1}$, and 1,543 $cm^{-1}$, correlating with increased Euclidean distances between the compared groups. Peaks such as 1,453, 1,386, and 1,543 $cm^{-1}$ exhibited marked dissimilarities when compared to clusters of peaks at 2,922, 1,744, 1,446, and 1,304 $cm^{-1}$, with a linkage distance of 10. Conversely, the amide peaks at 1,528, 3,280, 1,535, and 1,632 $cm^{-1}$, despite their conserved features, displayed significant diversity from other peaks, with a linkage distance of 25. Notably, peaks at 2,922 $cm^{-1}$ and 1,744 $cm^{-1}$, associated with lipid structures, were more closely grouped, indicating better clustering with normal rather than metaplastic carcinoma peaks. Similarly, spectral peak pairings exhibiting tight clustering and potential for discriminating between normal and IDC include 1,162/1,152, 1,304/1,274, 1,080/1,073/1,043, 1,394/1,386, 1,237/1,230, 1,453/1,446, and 1,535/1,528 (Cluster 1). Substantially more dissimilar yet related pairings include 1,162/1,304 and 1,535/1,543 (Cluster 2); 1,394/1,237, 1,237/1,453, and 1,394/1,453 (Cluster 3). Remaining clusters consist of 1,073/1,386 and 1,528/1,632 (Cluster 4); 1,274/1,043 (Cluster 5); 3,280/1,543 (Cluster 6); and 1,744/1,237 (Cluster 7). Furthermore, spectral analysis revealed distinct peak clusters (Cluster 1: 1,394, 1,386, 1,453, 1,446, 1,237, 1,230, 1,080, 1,073, 1,043; Cluster 2: 1,162, 1,155, 1,304, 1,274, 2,922, 3,280) alongside independently clustered peaks (1,535, 1,528, 1,543, 1,632). This configuration suggests potential biochemical relationships between DCIS and rarer metaplastic carcinomas. Additionally, similarities between Clusters 1 and 2 raise the hypothesis that peak 1,274 may exhibit characteristics analogous to peak 1,237, potentially due to concurrent biochemical modifications influencing its spectral signature. In differentiating DCIS and IDC, infrared spectroscopy reveals distinct peak clusters exhibiting equidistant spacing, suggesting underlying molecular regularity. Specifically, peaks at 1,162–2,922 $cm^{-1}$ and 1,043–1,632 $cm^{-1}$ demonstrate such patterns, predicting that a shift toward invasiveness may be accompanied by differential chemical modifications influencing peaks at 1,304 $cm^{-1}$ and 1,043 $cm^{-1}$, potentially altering their vibrational modes compared to previously characterized clusters. Euclidean distance analysis identifies spectral peak combinations of 1,394/1,386, 1,080/1,073, 1,073/1,043, 1,080/1,043, 1,237/1,230, 1,453/1,446, and 1,162/1,155 as highly discriminatory between IDC and metaplastic carcinoma. However, spectral overlap is observed in peak combinations such as 1,304/1,274, 1,162/1,304, 1,162/1,274, 1,535/1,543, 1,394/1,073, and 1,237/1,453, suggesting potential concurrent molecular alterations that blur distinct spectral signatures.

**Table 2. Peak ranges and their corresponding biochemical assignments**

| Peaks | Biochemical assignment | References |
|---|---|---|
| 3,273–3,284 | Amide A | 19,28,35,36 |
| 2,916–2,920 | Lipids | 13,26–29,35 |
| 1,529–1,535 | Amide II | 13–17,24,25,38 |
| 1,384–1,396 | Lipids, Proteins | 28,35,36,39 |
| 1,274, 1,304 | Collagen | 11,32,35,36 |
| 1,232–1,237 | Collagen, Phosphate | 28,35,37 |
| 1,073–1,080 | Carbohydrate, Phosphate | 28,33–35 |
| 1,038–1,045 | Glycogen | 27,28,35,39,40 |

### *ROC curve and peak analysis*

The selected peaks underwent ROC analysis, employing bootstrapping and Bonferroni correction to account for sample size and

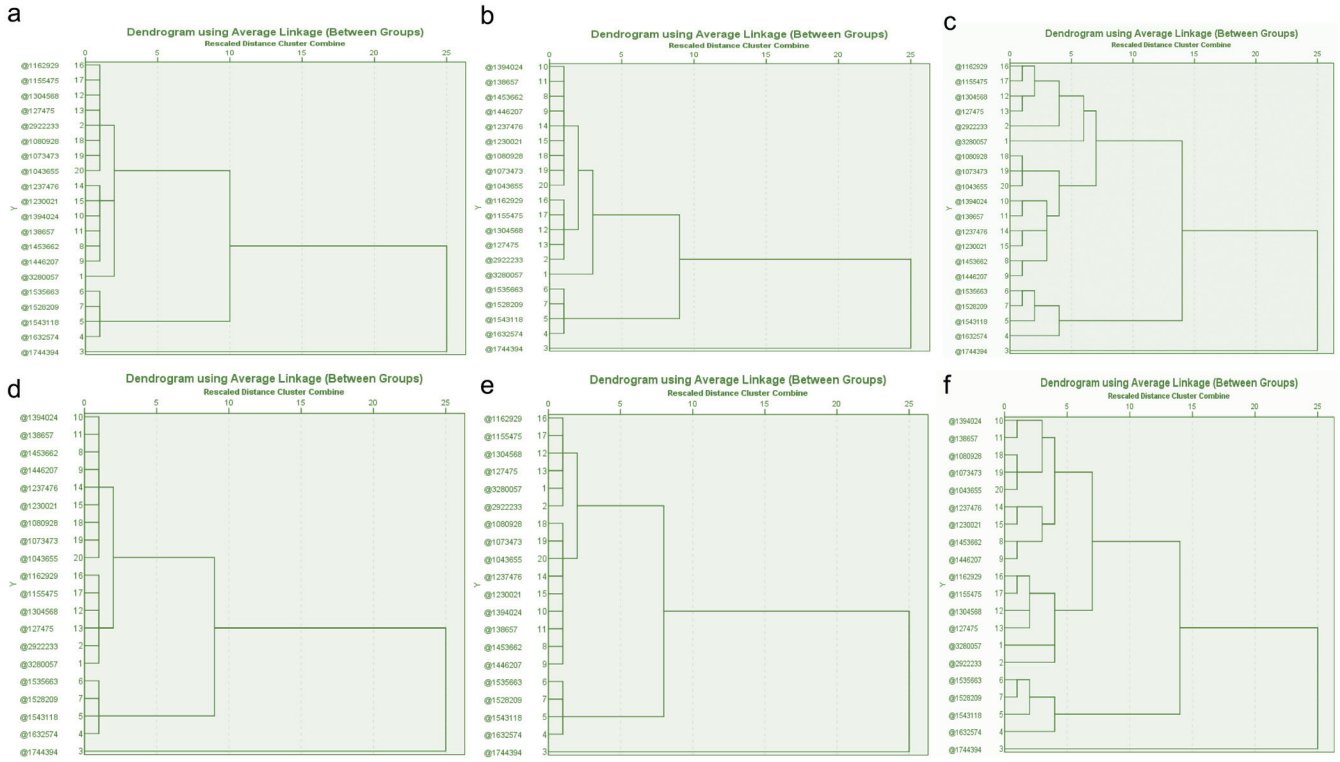


**Fig. 3. Hierarchical clustering analysis showing (a) notable peaks with close clustering patterns discriminating normal breast tissue and ductal carcinoma *in situ*. (b) Notable peaks with close clustering patterns discriminating normal breast tissue and metaplastic carcinoma. (c) Notable peaks with close clustering patterns discriminating normal breast tissue and invasive ductal carcinoma. (d) Notable peaks with close clustering patterns discriminating ductal carcinoma *in situ* and metaplastic carcinoma. (e) Notable peaks with close clustering patterns discriminating ductal carcinoma *in situ* and invasive ductal carcinoma. (f) Notable peaks with close clustering patterns discriminating metaplastic carcinoma and invasive ductal carcinoma.**

to evaluate parameters such as AUC, sensitivity, and specificity between normal breast tissue and breast carcinomas, as well as among the breast malignancies, as detailed in the figures and Tables 3 and 4. Peak 3,280 $cm^{-1}$, likely corresponding to Amide A,[19,26,28,35,36] along with peaks 1,543, 1,535, 1,528, and 1,632 $cm^{-1}$ (associated with beta-pleated sheets in Amide I and II) and 1543 (assigned to α-sheet of Amide II) (Table 2),[13–17,24,25,38] produced moderate to excellent AUC values (0.93) and exhibited high sensitivity (100%) with specificities ranging from 80% to 60% (Table 3 and Fig. S2).

### *Cut-off points and clustering patterns*

The identified cut-off points were congruent with hierarchical clustering patterns, facilitating the identification of peaks with similar vibrational characteristics and chemical properties. For instance, peaks at 1,043, 1,073, and 1,080 $cm^{-1}$ shared similar cut-off values (0.031, 0.032, and 0.032, respectively) between normal and carcinoma cases and among carcinoma categories, suggesting potential chemical similarity in their molecular structures, as shown in Table 5.

### *Peak ratio analysis*

With Bonferroni correction, comparing breast carcinomas to normal breast tissue showed that the peak ratio (A1632/A1543), corresponding to protein levels associated with beta-sheet structures, increased in malignancy but lacked statistical significance for discrimination. However, phosphate (A1237/A1080) and glycogen (A1043/A1543) levels demonstrated statistically significant increases in carcinoma compared to normal ($p < 0.01$), though these could not be validated as diagnostic markers. Additionally, the nucleocytoplasmic index peak ratio was significantly elevated ($p < 0.05$). However, ROC curve analysis for these peak ratios (A1632/A1543, A1237/A1080, and A1043/A1543) revealed sub-par AUC values and abysmal specificity (less than 1%), indicating limited diagnostic relevance despite statistical significance for differentiating normal tissue from breast lesions and among malignancies. Notably, the A1080/A1632 ratio achieved statistical significance ($p = 0.03$) with an AUC of 1.0, sensitivity (~100%), and specificity (~100%) in differentiating normal from carcinomas and DCIS from invasive breast carcinomas ($p < 0.001$), underscoring its potential diagnostic value as shown in Table 6 and Figure S3. However, glycogen-related peaks yielded AUC values of 0.6–0.7 with ~80% sensitivity and ~40% specificity, suggesting only cautious consideration for delineating invasive ductal and metaplastic carcinomas (Fig. S4).

## Discussion

Spectral analysis reveals distinct vibrational signatures differentiating normal breast tissue from carcinomatous counterparts, with normal tissue exhibiting a shift toward higher wavenumbers, suggesting diagnostic utility. Specifically, carcinoma samples displayed increased peak intensities at 1,045 $cm^{-1}$, 1,075 $cm^{-1}$, 1,080 $cm^{-1}$, 1,151 $cm^{-1}$, 1,162 $cm^{-1}$, 1,274 $cm^{-1}$, 1,237 $cm^{-1}$, 1,308

**Table 3. AUC, sensitivity, and specificity of different peaks for differentiation between normal breast tissue and malignant variants**

| Peaks | Normal | METC | AUC | *p*-value | Sensitivity% | Specificity% | Normal | DCIS | AUC | *p*-value | Sensitivity% | Specificity% | Normal | IDC | AUC | *p*-value | Sensitivity% | Specificity% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,045 | 0.0433 | 0.0611 | 0.267 | 0.24 | 100 | 10 | 0.0433 | 0.051 | 0.36 | 0.84 | 100 | 15 | 0.0433 | 0.064 | 0.16 | 0.06 | 66 | 7 |
| 1,073 | 0.0464 | 0.0597 | 0.267 | 0.24 | 100 | 10 | 0.0464 | 0.052 | 0.46 | 0.84 | 100 | 23 | 0.0464 | 0.063 | 0.24 | 0.14 | 66 | 7 |
| 1,080 | 0.0642 | 0.0607 | 0.333 | 0.17 | 100 | 10 | 0.0642 | 0.052 | 0.46 | 0.84 | 100 | 23 | 0.0642 | 0.063 | 0.24 | 0.14 | 66 | 7 |
| 1,155 | 0.04 | 0.047 | 0.367 | 0.5 | 100 | 10 | 0.04 | 0.044 | 0.39 | 0.54 | 67 | 15 | 0.04 | 0.049 | 0.29 | 0.24 | 66 | 7 |
| 1,162 | 0.041 | 0.047 | 0.433 | 0.74 | 100 | 20 | 0.041 | 0.045 | 0.41 | 0.64 | 100 | 15 | 0.041 | 0.048 | 0.31 | 0.26 | 67 | 10 |
| 1,230 | 0.0641 | 0.0605 | 0.567 | 0.89 | 100 | 40 | 0.0641 | 0.059 | 0.62 | 0.55 | 100 | 40 | 0.0641 | 0.062 | 0.52 | 0.93 | 100 | 29 |
| 1,237 | 0.065 | 0.06 | 0.567 | 0.74 | 100 | 40 | 0.065 | 0.059 | 0.64 | 0.46 | 100 | 46 | 0.065 | 0.062 | 0.57 | 0.69 | 100 | 35 |
| 1,274 | 0.051 | 0.05 | 0.5 | 1 | 100 | 30 | 0.051 | 0.049 | 0.54 | 0.84 | 100 | 30 | 0.051 | 0.051 | 0.47 | 0.88 | 100 | 26 |
| 1,304 | 0.048 | 0.047 | 0.467 | 0.87 | 100 | 30 | 0.048 | 0.047 | 0.44 | 0.74 | 100 | 15 | 0.048 | 0.053 | 0.34 | 0.38 | 67 | 16 |
| 1,386 | 0.059 | 0.06 | 0.467 | 0.87 | 100 | 30 | 0.059 | 0.057 | 0.51 | 0.95 | 100 | 30 | 0.059 | 0.063 | 0.39 | 0.52 | 67 | 20 |
| 1,394 | 0.06 | 0.06 | 0.5 | 1 | 100 | 30 | 0.06 | 0.057 | 0.54 | 0.84 | 100 | 30 | 0.06 | 0.062 | 0.44 | 0.74 | 100 | 23 |
| 1,446 | 0.0652 | 0.0585 | 0.6 | 0.61 | 100 | 40 | 0.0652 | 0.059 | 0.62 | 0.55 | 100 | 46 | 0.0652 | 0.06 | 0.6 | 0.61 | 100 | 45 |
| 1,453 | 0.583 | 0.596 | 0.633 | 0.5 | 100 | 50 | 0.583 | 0.058 | 0.62 | 0.55 | 100 | 46 | 0.583 | 0.059 | 0.61 | 0.56 | 100 | 45 |
| 1,528 | 0.0947 | 0.0802 | 0.633 | 0.5 | 100 | 50 | 0.0947 | 0.082 | 0.62 | 0.55 | 100 | 38 | 0.0947 | 0.082 | 0.65 | 0.41 | 100 | 48 |
| 1,535 | 0.0955 | 0.0789 | 0.733 | 0.24 | 100 | 60 | 0.0955 | 0.082 | 0.62 | 0.55 | 100 | 46 | 0.0955 | 0.082 | 0.65 | 0.41 | 100 | 45 |
| 1,543 | 0.0951 | 0.078 | 0.767 | 0.18 | 100 | 70 | 0.0951 | 0.08 | 0.67 | 0.38 | 100 | 46 | 0.0951 | 0.079 | 0.7 | 0.26 | 100 | 55 |
| 1,632 | 0.105 | 0.0879 | 0.733 | 0.24 | 100 | 60 | 0.105 | 0.09 | 0.64 | 0.46 | 100 | 46 | 0.105 | 0.088 | 0.68 | 0.35 | 100 | 52 |
| 1,744 | 0.0029 | 0.0105 | 0.133 | 0.06 | 0 | 60 | 0.0029 | 0.006 | 0.26 | 0.2 | 33 | 70 | 0.0029 | 0.011 | 0.16 | 0.06 | 33 | 42 |
| 2,922 | 0.0459 | 0.0463 | 0.533 | 0.87 | 100 | 40 | 0.0459 | 0.042 | 0.67 | 0.38 | 100 | 54 | 0.0459 | 0.047 | 0.54 | 0.83 | 100 | 45 |
| 3,280 | 0.0804 | 0.0512 | 0.93 | 0.028 | 100 | 80 | 0.0804 | 0.053 | 0.95 | 0.02 | 100 | 85 | 0.0804 | 0.049 | 0.96 | 0.01 | 100 | 83 |

Breast lesions: DCIS, ductal carcinoma *in situ*; METC, metaplastic carcinoma; IDC, invasive ductal carcinoma. Statistical term: AUC, area under the curve. Peak 3,280 consistently performed well in differentiating normal from all types of breast malignancies considered in this study, with 100% sensitivity and 80–85% specificity, as shown in Figure S2. Peaks 1,528–1,632 also offered above-average diagnostic utility in these comparisons.

**Table 4. AUC, sensitivity, and specificity of different peaks for differentiation among malignant variants**

| Peaks | DCIS | METC | AUC | *p*-value | Sensitivity % | Specificity % | DCIS | IDC | AUC | *p*-value | Sensitivity % | Specificity% | METC | IDC | AUC | *p*-value | Sensitivity% | Specificity% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,045 | 0.051 | 0.0611 | 0.65 | 0.22 | 70 | 46 | 0.051 | 0.064 | 0.31 | 0.05 | 62 | 23 | 0.0611 | 0.064 | 0.46 | 0.69 | 60 | 42 |
| 1,073 | 0.052 | 0.0597 | 0.65 | 0.24 | 70 | 46 | 0.052 | 0.063 | 0.35 | 0.11 | 62 | 26 | 0.0597 | 0.063 | 0.47 | 0.79 | 60 | 48 |
| 1,080 | 0.052 | 0.0607 | 0.62 | 0.32 | 70 | 46 | 0.052 | 0.063 | 0.36 | 0.13 | 62 | 39 | 0.0607 | 0.063 | 0.46 | 0.72 | 60 | 45 |
| 1,155 | 0.044 | 0.047 | 0.59 | 0.46 | 70 | 38 | 0.044 | 0.049 | 0.41 | 0.39 | 62 | 45 | 0.047 | 0.049 | 0.5 | 0.98 | 60 | 58 |
| 1,162 | 0.045 | 0.047 | 0.55 | 0.66 | 70 | 38 | 0.045 | 0.048 | 0.42 | 0.44 | 62 | 42 | 0.047 | 0.048 | 0.49 | 0.9 | 60 | 52 |
| 1,230 | 0.059 | 0.0605 | 0.5 | 1 | 70 | 23 | 0.059 | 0.062 | 0.44 | 0.54 | 69 | 29 | 0.0605 | 0.062 | 0.47 | 0.74 | 60 | 50 |
| 1,237 | 0.059 | 0.06 | 0.5 | 1 | 70 | 23 | 0.059 | 0.062 | 0.45 | 0.58 | 69 | 26 | 0.06 | 0.062 | 0.47 | 0.74 | 50 | 55 |
| 1,274 | 0.049 | 0.05 | 0.54 | 0.76 | 70 | 46 | 0.049 | 0.051 | 0.42 | 0.45 | 62 | 39 | 0.05 | 0.051 | 0.48 | 0.86 | 70 | 39 |
| 1,304 | 0.047 | 0.047 | 0.62 | 0.35 | 70 | 54 | 0.047 | 0.053 | 0.38 | 0.21 | 62 | 32 | 0.047 | 0.053 | 0.5 | 0.97 | 70 | 42 |
| 1,386 | 0.057 | 0.06 | 0.58 | 0.54 | 70 | 56 | 0.057 | 0.063 | 0.4 | 0.3 | 62 | 32 | 0.06 | 0.063 | 0.47 | 0.81 | 70 | 40 |
| 1,394 | 0.057 | 0.06 | 0.55 | 0.66 | 70 | 46 | 0.057 | 0.062 | 0.4 | 0.33 | 62 | 32 | 0.06 | 0.062 | 0.48 | 0.88 | 70 | 48 |
| 1,446 | 0.059 | 0.0585 | 0.48 | 0.85 | 70 | 30 | 0.059 | 0.06 | 0.47 | 0.77 | 60 | 45 | 0.0585 | 0.06 | 0.46 | 0.72 | 60 | 45 |
| 1,453 | 0.058 | 0.596 | 0.47 | 0.8 | 70 | 31 | 0.058 | 0.059 | 0.47 | 0.75 | 62 | 45 | 0.596 | 0.059 | 0.46 | 0.72 | 60 | 42 |
| 1,528 | 0.082 | 0.0802 | 0.45 | 0.66 | 70 | 31 | 0.082 | 0.082 | 0.47 | 0.77 | 62 | 48 | 0.0802 | 0.082 | 0.47 | 0.79 | 70 | 39 |
| 1,535 | 0.082 | 0.0789 | 0.45 | 0.66 | 70 | 31 | 0.082 | 0.082 | 0.48 | 0.85 | 62 | 45 | 0.0789 | 0.082 | 0.46 | 0.69 | 70 | 39 |
| 1,543 | 0.08 | 0.078 | 0.45 | 0.71 | 70 | 31 | 0.08 | 0.079 | 0.49 | 0.9 | 70 | 35 | 0.078 | 0.079 | 0.46 | 0.74 | 70 | 39 |
| 1,632 | 0.09 | 0.0879 | 0.45 | 0.66 | 70 | 31 | 0.09 | 0.088 | 0.52 | 0.83 | 70 | 46 | 0.0879 | 0.088 | 0.48 | 0.81 | 70 | 42 |
| 1,744 | 0.006 | 0.0105 | 0.72 | 0.04 | 70 | 100 | 0.006 | 0.011 | 0.24 | 0.008 | 54 | 13 | 0.0105 | 0.011 | 0.49 | 0.98 | 70 | 58 |
| 2,922 | 0.042 | 0.0463 | 0.54 | 0.76 | 70 | 31 | 0.042 | 0.047 | 0.41 | 0.36 | 70 | 35 | 0.0463 | 0.047 | 0.47 | 0.79 | 60 | 48 |
| 3,280 | 0.053 | 0.0512 | 0.48 | 0.85 | 70 | 31 | 0.053 | 0.049 | 0.58 | 0.45 | 70 | 60 | 0.0512 | 0.049 | 0.54 | 0.72 | 50 | 45 |

Breast lesions: DCIS, ductal carcinoma *in situ*; IDC, invasive ductal carcinoma; METC, metaplastic carcinoma. Statistical term: AUC, area under the curve. As shown above and in Figure S2, peak 1,744 demonstrated moderate diagnostic capability, with an AUC of 0.72, 71% sensitivity, and 100% specificity ($p = 0.04$) for discriminating between metaplastic carcinomas and ductal carcinoma *in situ*. In contrast, peak 3,280 (AUC ~0.6) displayed fair diagnostic potential for identifying pre-invasive ductal and invasive ductal carcinomas. No peak was found to reliably distinguish invasive ductal carcinoma from metaplastic carcinoma.

**Table 5. Discriminating peaks, cut-off points, and chemical assignments**

| Discriminating peaks | Average cut-off normal-carcinomas | Average cut-off carcinoma groups | Chemical assignment |
|---|---|---|---|
| 1,043 | 0.031 | 0.053 | Phosphate (asym)/Glycogen |
| 1,073 | 0.032 | 0.055 | |
| 1,080 | 0.032 | 0.055 | |
| 1,155 | 0.025 | 0.046 | Protein (amino acids) |
| 1,162 | 0.031 | 0.045 | |
| 1,230 | 0.052 | 0.053 | Phosphate (asym) |
| 1,237 | 0.053 | 0.053 | |
| 1,274 | 0.038 | 0.046 | Protein (Collagen, Amide III) |
| 1,304 | 0.037 | 0.047 | |
| 1,386 | 0.046 | 0.055 | Lipids, Protein |
| 1,394 | 0.046 | 0.055 | |
| 1,446 | 0.053 | 0.053 | Protein (methyl group) asym |
| 1,453 | 0.055 | 0.052 | |
| 1,528 | 0.077 | 0.074 | Protein (Amide II) |
| 1,535 | 0.079 | 0.073 | |
| 1,543 | 0.079 | 0.069 | |

asym, asymmetric vibration; sym, symmetric vibration.

$cm^{-1}$, 1,386 $cm^{-1}$, 1,394 $cm^{-1}$, and 1,453 $cm^{-1}$, while normal tissue showed enhanced intensities at 1,230 $cm^{-1}$, 1,446 $cm^{-1}$, 1,525 $cm^{-1}$, 1,535 $cm^{-1}$, 1,543 $cm^{-1}$, 2,919 $cm^{-1}$, 3,273 $cm^{-1}$, and 3,284 $cm^{-1}$.[12,13,19]

Normal tissue exhibited broader, more intense peaks at 3,273–3,284 $cm^{-1}$ (O–H and Amide A stretching vibrations),[19,24,25] while peaks at 2,916–2,919 $cm^{-1}$ ($CH_2$ asymmetric stretch of acyl lipids) in normal tissue showed a narrower band with lower intensity compared to carcinomas as shown in Table 2 This underscores the role of saturated lipids in membrane integrity and suggests carcinoma's potential for progression and metastasis.[20,26,41–43]

Conversely, higher expression of the lipid peak at 1,744 $cm^{-1}$ (C=O stretching vibrations from esters and phospholipids) was observed in malignant tissue,[19,28,41,42] positioning it as a key differential marker for metaplastic versus DCIS. This increased expression aligns with reduced hydrogen bonding, potentially augmenting de novo lipogenesis in carcinoma, indicating metastatic potential, and highlighting a redox environment in metaplastic carcinoma similar to aggressively proliferating ductal carcinoma.[28–30,39,42]

Normal tissue displayed sharper, more intense features at 1,627–1,632 $cm^{-1}$ (β-sheet Amide I stretching vibrations) and 1,529 (β-sheet Amide II) and 1,543 $cm^{-1}$ (α-sheet Amide II),[13–17,33,35–36,44–47] suggesting β-sheet proteins as markers distinguishing normal from breast carcinoma,[36] despite conflicting evidence regarding β-sheet protein reductions in metaplastic carcinoma. Protein peaks were fairly conserved across all malignant breast tissues.[10,16,32,42,46]

Normal breast tissue exhibited a more pronounced peak at 1,446 $cm^{-1}$ ($CH_2$ bending of lipids, triglycerides, and C–H vibrations),[21,48] supporting observations for unsaturated lipid bands and emphasizing membrane fluidity for cellular health. In contrast, the peak at 1,453 $cm^{-1}$ (asymmetric $CH_3$ vibrations and protein deformation) showed less distinct characteristics.[12,26,29,42]

Metaplastic carcinoma showed increased intensities in peaks at 1,386 $cm^{-1}$ and 1,394 $cm^{-1}$. Peaks at 1,394 $cm^{-1}$ is identified with fatty acids due to $COO^-$ symmetric stretching of amino acids side bonds vibration and 1,308 $cm^{-1}$ involving protein symmetric $CH_3$ bending assigned to collagen/Amide III vibrations,[28,35] highlighting fatty acids and amino acid involvement in neoplastic transformation.[29,30,33,35,36] Increased collagen deposition in carcinoma tissues suggests heightened tumor aggressiveness.[24,25,32,36,49,50]

Peaks at 1,230 $cm^{-1}$ and 1,232 $cm^{-1}$ (nucleic acid contributions) showed sharper features in normal tissue,[11,32,33,34–38,51] aligning with literature indicating differential intensities in carcinoma tissues due to genetic mutations and nuclear enlargement, particularly in metaplastic and ductal carcinomas, which exhibited increased intensity compared with normal breasts. Elevated peaks at 1,308 $cm^{-1}$ and 1,274 $cm^{-1}$ assigned to collagen- a matrix tissue due to metalloproteinases- were found prominently in carcinomas,[30,32,35–38,51] suggesting fibroblast activation and heightened collagen deposition in metastatic conditions.[16,31,32,41,50]

Peak 1,151 $cm^{-1}$ (C–O glycogen stretches) was more prevalent in malignant samples,[33,36,38,52] especially in invasive ductal carcinoma, followed by metaplastic carcinoma, pre-IDC, and normal breast tissue.[29] This is corroborated by other carbohydrate peaks, reflecting escalated cell activity in the G1 phase and heightened energy needs during DNA synthesis,[7,15,34,37,39,40] although interpretations are rare in the literature, especially concerning metaplastic carcinomas.

Peak 1,080 $cm^{-1}$ (nucleic acids) indicated shifts favoring normal tissue,[28,34,35,51] paralleling documentation of asymmetric phosphate vibrations linked to malignancy markers.[11,12,19,36,37,40] Meanwhile, peak 1,162 $cm^{-1}$ (collagen mechanics) showed intensified patterns in metaplastic as well as ductal carcinomas versus normal breast tissue.

Comparative evaluation of peak ratios revealed that nucleic acid and glycogen levels were statistically elevated in metaplastic carcinoma versus normal breast tissue, consistent with studies validating nucleic acid increases in carcinoma development.[38–40,47]

**Table 6. Biomarker mean differences and diagnostic performance between paired breast tissue types**

| Biomarkers | Breasts | | t-test | | AUC | | Sensitivity (%) | Specificity (%) | Cut-off |
|---|---|---|---|---|---|---|---|---|---|
| | Normal: Mean ± SEM | DCIS: Mean ± SEM | | *p*-value | | *p*-value | | | |
| NC | 0.44 ± 0.09 | 0.56 ± 0.07 | 3.11 | 0.008 | 0.97 | 0.013* | 100 | 92 | 2.10 |
| Phos | 0.71 ± 0.04 | 0.88 ± 0.04 | 1.90 | 0.08 | 0.08 | 0.03 | 100 | 0 | 0.03 |
| Glyc | 0.45 ± 0.03 | 0.64 ± 0.04 | −2.20 | 0.05 | 0.03 | 0.013* | 100 | 0 | 0.001 |
| | Normal: Mean ± SEM | METC: Mean ± SEM | | | | | | | |
| NC | 0.44 ± 0.09 | 0.67 ± 0.05 | 7.14 | 0.03 | 1.00 | 0.01* | 100 | 100 | 0.19 |
| Phos | 0.71 ± 0.04 | 1.01 ± 0.03 | 5.04 | 0.001 | 0.00 | 0.01 | 100 | 0 | 0.00 |
| Glyc | 0.45 ± 0.03 | 0.78 ± 0.04 | 4.92 | 0.001 | 0.00 | 0.01 | 100 | 0 | 0.00 |
| | Normal: Mean ± SEM | IDC: Mean ± SEM | | | | | | | |
| NC | 0.44 ± 0.09 | 0.71 ± 0.05 | 5.68 | 0.0001 | 0.98 | 0.03* | 100 | 94 | 1.68 |
| Phos | 0.71 ± 0.04 | 1.03 ± 0.02 | −5.05 | 0.0001 | 0.000 | 0.005 | 100 | 0 | −0.34 |
| Glyc | 0.45 ± 0.03 | 0.83 ± 0.03 | −4.85 | 0.0001 | 0.011 | 0.02 | 100 | 0 | −0.58 |
| | DCIS: Mean ± SEM | METC: Mean ± SEM | | | | | | | |
| Pro | 1.13 ± 0.01 | 1.13 ± 0.01 | −0.26 | 0.80 | 0.45 | 0.71 | 54 | 50 | 1.12 |
| NC | 0.56 ± 0.07 | 0.67 ± 0.05 | 2.98 | 0.007 | 0.83 | 0.008* | 85 | 80 | 1.53 |
| Pho | 0.88 ± 0.04 | 1.00 ± 0.03 | −2.27 | 0.03 | 0.23 | 0.03 | 92 | 0 | 0.73 |
| Gly | 0.64 ± 0.04 | 0.78 ± 0.04 | −2.62 | 0.02 | 0.22 | 0.02 | 92 | 0 | 0.49 |
| | DCIS: Mean ± SEM | IDC: Mean ± SEM | | | | | | | |
| Pro | 1.13 ± 0.01 | 1.14 ± 0.02 | −0.34 | 0.74 | 0.49 | 0.94 | 100 | 10 | 1.07: |
| NC | 0.56 ± 0.07 | 0.71 ± 0.05 | 0.06 | 0.0001 | 0.86 | 0.0001* | 100 | 49 | 1.31 |
| Pho | 0.88 ± 0.04 | 1.03 ± 0.02 | −3.58 | 0.001 | 0.21 | 0.003 | 100 | 0 | −0.28 |
| Gly | 0.64 ± 0.04 | 0.83 ± 0.02 | −4.22 | 0.0001 | 0.18 | 0.001 | 92 | 5 | 0.49 |
| | METC: Mean ± SEM | IDC: Mean ± SEM | | | | | | | |
| Pro | 1.13 ± 0.01 | 1.14 ± 0.02 | 0.18 | 0.86 | 0.49 | 0.93 | 65 | 50 | 1.11 |
| NC | 0.67 ± 0.05 | 0.71 ± 0.05 | −0.90 | 0.38 | 0.32 | 0.09 | 71 | 20 | 1.24 |
| Pho | 1.00 ± 0.03 | 1.03 ± 0.02 | 0.48 | 0.63 | 0.52 | 0.88 | 68 | 20 | 0.95 |
| Gly | 0.78 ± 0.04 | 0.83 ± 0.02 | 0.97 | 0.34 | 0.64 | 0.19 | 80 | 40 | 0.75 |

Biomarkers: Gly, glycogen; NC, nucleocytoplasmic ratio; Pho, phosphate; Pro, protein. Breast lesions: DCIS, ductal carcinoma *in situ*; IDC, invasive ductal carcinoma; METC, metaplastic carcinoma. Statistical terms: AUC, area under the curve; SEM, standard error of the mean; *significant difference for high AUC. The nucleocytoplasmic ratio demonstrates exceptional diagnostic accuracy (AUC ≈ 0.97–1.00, 100% sensitivity, >90% specificity) in distinguishing normal cells from a wide array of pre-invasive and invasive carcinomas. While its efficacy diminishes in differentiating ductal carcinoma *in situ* from invasive ductal carcinoma (AUC ≈ 0.83–0.86, ~80% sensitivity, ~85% specificity), it performs poorly in discriminating METC from IDC (AUC = 0.32). Select biomarkers, including glycogen, offer some utility in differentiating IDC from METC (AUC = 0.6–0.7), albeit with suboptimal sensitivity and specificity as indicated in Figure S4.

Elevated glycogen levels in other carcinomatous tissues are similarly documented,[36,52] although some studies suggest glycogen is relatively lower in normal tissues.[12,16,35] Elevated glycogen levels in metaplastic tissues could suggest rapid cell cycle shifts.[40,44] The nucleocytoplasmic index ratio also proved significantly relevant ($p = 0.03$) for discriminating between normal and all metaplastic carcinomas as well as between ductal carcinoma *in situ* and invasive breast carcinomas, reflecting carcinoma's need to compensate for its metabolic demands through excessive nuclear growth relative to the cytoplasm.[34,53]

Hierarchical clustering, one of the supervised learning models,[19,29,47,54–58] coupled with ROC curve analysis for cutoff determination, elucidates potential synergistic chemical relationships between spectral peaks and establishes diagnostic links between clustering patterns and optimal classification thresholds. This analysis revealed three distinct clusters capable of differentiating breast tissue types. Notably, hierarchical clustering associated less-utilized peaks (e.g., 1,394 and 1,386 $cm^{-1}$) with established carcinoma biomarkers, suggesting a shared chemical basis despite differing vibrational modes. The observed spectral proximity, quantified by linkage and Euclidean distances, implies closer functional relationships than conventionally recognized, potentially harboring diagnostic significance. Similarly, the co-clustering of the 1,162 $cm^{-1}$ peak (typically assigned to carotenoids) with other compounds suggests a more complex origin than previously considered.[54,55] Furthermore, the clustering of peaks at

1,535, 1,528, 1,543, and 1,632 $cm^{-1}$, all indicative of amide protein origins,[19,35,44,46] underscores their inherent similarities and highlights underlying chemical associations potentially relevant to metaplastic carcinogenesis.[31,33] Differential gene expression resulting in specific protein conformations likely contributes to the discriminatory power of these protein-related spectral features.[34,35,46]

Figure 3a–f delineate spectral clusters, pinpointing vibrational modes and enabling chemical assignments via spatial relationships. Phosphodiester vibrational peaks demonstrate a strong correlation, whereas Amide A and lipid peaks show marked divergence from other clusters. Spectral peak clustering analysis reveals disease-associated shifts in biochemical composition. Despite variability in peak patterns across matched breast carcinoma cases, the fundamental chemical identities and spectral signatures consistently differentiate breast carcinoma subtypes. Notably, initially dissimilar peaks converged within shared clusters upon refined discrimination. Moreover, protein peaks exhibit autonomous vibrational characteristics, making them robust markers for investigating protein conformational changes,[25,32,36,45,47] especially during the progression from DCIS to metaplastic carcinoma, and from DCIS to IDC.

ROC curve analysis revealed similar chemical properties across spectral peaks, suggesting their potential to differentiate normal, *in situ*, and invasive breast tissues. However, some peaks exhibited preferential elevation in specific carcinoma subtypes, indicating possible utility in subtyping. Despite achieving 100% sensitivity on average, low specificity and AUC values below 0.5 indicate limited diagnostic accuracy for most peaks in this study. Contrary to prior reports highlighting carbohydrates, their diagnostic relevance was diminished in our findings. Nucleic acid/phosphate peaks showed inconsistent diagnostic potential, likely reflecting tumor heterogeneity. Amide II peaks associated with β-sheet proteins and Amide I peaks demonstrated considerable diagnostic promise (AUC ≈ 0.7, sensitivity ≈ 100%).[14–17,35–38] The significant elevation of β-sheet proteins in normal tissue compared to metaplastic samples suggests their potential as diagnostic markers, although prior conflicting observations highlight complex metabolic influences. The peak ratio corresponding to the nucleocytoplasmic ratio (AUC = 1.0, 100% sensitivity and 100% specificity) compares well with ratios used in previous investigations, differentiating normal breast from tumorous phenotypes.[14,21,29,57]

Furthermore, peak 3,280 stood out for its high discriminatory power (AUC = 0.93), challenging assumptions about its limited value due to water interference,[29,35,56,57] and displaying the ability to distinguish normal from histological types and monitor treatment responses.[13,14,29,53,58] The increased levels of Amide A alongside other amide bands suggest a possible connection to oncogene-driven dysregulation, warranting further investigation.[24,33,39] Ongoing exploration of these spectral features is crucial for improving carcinoma diagnostics, with a focus on enhancing specificity while maintaining high sensitivity. Discrepancies with previous studies reporting exceptional diagnostic performance for certain peaks may stem from lipid peroxidation.[42,54] In contrast, peak 1,744's limited diagnostic power (AUC = 0.133) may have resulted from lipid loss during tissue processing.[26,42,48,53]

FTIR spectroscopy offers a powerful diagnostic alternative for breast tissue analysis, showing comparable efficacy to diffuse optical spectroscopy, intrinsic fluorescence spectroscopy, and diffuse reflectance spectroscopy. While diffuse optical spectroscopy, intrinsic fluorescence spectroscopy, and diffuse reflectance spectroscopy effectively quantify key biochemical markers—oxyhemoglobin, β-carotene, methemoglobin, tissue hemoglobin, collagen, lipids, and redox state —with exceptional diagnostic accuracy,[49,50,54,55] FTIR, particularly through analysis of the Amide A band, achieves similarly high levels of sensitivity and specificity in differentiating breast tissue types.[29,56,57] This confirms existing research and solidifies FTIR as a valuable tool in breast tissue characterization.[10,19,22,29,39,40,46,53,57,58]

Despite the value in investigating the unique characteristics of rare carcinomas, this study acknowledges its limitations. The small sample size, dictated by the carcinoma's infrequent occurrence during the study period, limits the broad applicability of the conclusions. Furthermore, the absence of comprehensive patient biodata, details regarding presentation patterns (days/year), and follow-up protocols prevented potentially valuable survival analysis. Future research should prioritize the inclusion of such data. Finally, the exclusion of other breast lesion subtypes beyond ductal carcinomas and normal tissue hinders a complete understanding of the specific biochemical and molecular differences defining metaplastic carcinoma within the spectrum of breast malignancies.

## Future directions

Future research should focus on expanding cohort sizes in studies of this rare carcinoma, incorporating a broader spectrum of breast tissue types, particularly fresh specimens to retain vital spectral data such as lipid profiles. Integrating complementary analytical techniques, such as nuclear magnetic resonance spectroscopy and Raman spectroscopy, alongside FTIR spectroscopy into standard carcinoma diagnostics is crucial for elucidating the biochemical alterations driving the transformation of normal breast tissue into invasive and rare subtypes, such as metaplastic carcinoma. Furthermore, implementing artificial intelligence and machine learning algorithms in clinical oncology workflows is essential for deriving predictive insights and identifying meaningful patterns within large spectroscopic datasets, ultimately facilitating biomarker discovery for improved diagnosis, prognosis, and treatment response.

## Conclusions

FTIR spectroscopy holds diagnostic promise, particularly for distinguishing metaplastic breast carcinoma from ductal carcinoma *in situ* and normal tissue. While statistically significant biomarkers are crucial for diagnostic integration, this study demonstrates that statistical significance does not guarantee diagnostic relevance. Protein peaks, particularly Amide A (3,280 $cm^{-1}$), Amide I (1,632 $cm^{-1}$), and β-sheet Amide II (1,543 and 1,535 $cm^{-1}$), exhibit strong diagnostic potential based on ROC analysis. Notably, the 1,080/1,632 ratio, reflecting the nucleocytoplasmic ratio, demonstrates an accurate diagnostic model. Cluster analysis reveals synergistic peak interactions that highlight chemical relationships among protein (1,453 and 1,386 $cm^{-1}$), lipid (1,446 and 1,394 $cm^{-1}$), and amide peaks (1,632 and 3,280 $cm^{-1}$). Cutoff point analysis quantitatively supports these findings, indicating similar biochemical signatures within clusters. The exceptional AUC of 1.0 for the nucleocytoplasmic ratio (A1080/A1632) underscores its clinical utility, aligning with conventional nucleocytoplasmic characterization in carcinoma diagnosis. These insights, while preliminary, advance understanding of the biochemical origins of metaplastic breast carcinoma and pave the way for refined diagnostic protocols and biomarker development. Further research is essential to validate and translate these findings into broader carcinoma research and clinical applications.

## Acknowledgments

The authors are grateful to Dr. Blessings Effiong (University of Uyo, Akwa Ibom, Nigeria) for advice on analysis of tissue sections with attenuated total reflectance Fourier-transform infrared spectroscopy, and to Dr. Iyiola (Achievers University, Owo, Nigeria) for technical support in sample retrieval and preparation for this study.

## Funding

None.

## Conflict of interest

The authors have no conflict of interest related to this publication.

## Author contributions

Study concept and design (STA, CI), acquisition of data (STA), analysis and interpretation of data (STA, CI), drafting of the manuscript (STA), critical revision of the manuscript for important intellectual content (STA, CI), administrative, technical, or material support (CI, STA), and study supervision (STA, CI). Both authors have made significant contributions to this study and have approved the final manuscript.

## Ethical statement

Ethical approval for the study was granted by the Obafemi Awolowo University Teaching Hospital Ethics Review Committee, registration number: ERC/2025/01/01. The study was carried out with a waiver of informed consent for the use of archived, anonymized samples in accordance with the principles of the Declaration of Helsinki (as revised in 2024). The ethics committee agreed that this study did not require informed consent due to its retrospective nature and the minimal risk involved.

## Data sharing statement

The photomicrographs and graphical formats (attached as supplementary materials) used in support of the findings of this study are included within the article.

## References

[1] Papatheodoridi A, Papamattheou E, Marinopoulos S, Ntanasis-Stathopoulos I, Dimitrakakis C, Giannos A, *et al*. Metaplastic Carcinoma of the Breast: Case Series of a Single Institute and Review of the Literature. Med Sci (Basel) 2023;11(2):35. doi:10.3390/medsci11020035, PMID:37218987.

[2] Abouharb S, Moulder S. Metaplastic breast cancer: clinical overview and molecular aberrations for potential targeted therapy. Curr Oncol Rep 2015;17(3):431. doi:10.1007/s11912-014-0431-z, PMID:25691085.

[3] Thapa B, Arobelidze S, Clark BA, Xuefei J, Daw H, Cheng YC, *et al*. Metaplastic Breast Cancer: Characteristics and Survival Outcomes. Cureus 2022;14(8):e28551. doi:10.7759/cureus.28551, PMID:36185859.

[4] Sinn HP, Kreipe H. A Brief Overview of the WHO Classification of Breast Tumors, 4th Edition, Focusing on Issues and Updates from the 3rd Edition. Breast Care (Basel) 2013;8(2):149–154. doi:10.1159/000350774, PMID:24415964.

[5] McKinnon E, Xiao P. Metaplastic carcinoma of the breast. Arch Pathol Lab Med 2015;139(6):819–822. doi:10.5858/arpa.2013-0358-RS, PMID:26030252.

[6] Tzanninis IG, Kotteas EA, Ntanasis-Stathopoulos I, Kontogianni P, Fotopoulos G. Management and Outcomes in Metaplastic Breast Cancer. Clin Breast Cancer 2016;16(6):437–443. doi:10.1016/j.clbc.2016.06.002, PMID:27431460.

[7] McCart Reed AE, Kalaw EM, Lakhani SR. An Update on the Molecular Pathology of Metaplastic Breast Cancer. Breast Cancer (Dove Med Press) 2021;13:161–170. doi:10.2147/BCTT.S296784, PMID:33664587.

[8] Reddy TP, Rosato RR, Li X, Moulder S, Piwnica-Worms H, Chang JC. A comprehensive overview of metaplastic breast cancer: clinical features and molecular aberrations. Breast Cancer Res 2020;22(1):121. doi:10.1186/s13058-020-01353-z, PMID:33148288.

[9] Vranic S, Stafford P, Palazzo J, Skenderi F, Swensen J, Xiu J, *et al*. Molecular Profiling of the Metaplastic Spindle Cell Carcinoma of the Breast Reveals Potentially Targetable Biomarkers. Clin Breast Cancer 2020;20(4):326–331.e1. doi:10.1016/j.clbc.2020.02.008, PMID:32197944.

[10] Khalil SKH, Khodeir MM, Abd El-Hakam R, Abd El-Monem Rezq R. Spectroscopic Study for Detection and Grading of Breast Carcinoma In vitro. Aust J Basic Appl Sci 2009;3(3):2419–2428.

[11] Eckel R, Huo H, Guan HW, Hu X, Che X, Huang WD. Characteristic infrared spectroscopic patterns in the protein bands of human breast cancer tissue. Vib Spectrosc 2001;27(2):165–173. doi:10.1016/S0924-2031(01)00134-5.

[12] Rehman S, Movasaghi Z, Darr JA, Rehman IU. Fourier transform infrared spectroscopic analysis of breast cancer tissues; identifying differences between normal breast, invasive carcinoma, and ductal carcinoma in situ of the Breast. Appl Spectrosc Rev 2010;45(5):355–368. doi:10.1080/05704928.2010.483674.

[13] Elshemey WM, Ismail AM, Elbialy NS. Molecular-Level Characterization of normal, benign, and malignant breast tissues using FTIR spectroscopy. J Med Biol Eng 2016;36(3):369–378. doi:10.1007/s40846-016-0133-0.

[14] Ghimire H, Garlapati C, Janssen EAM, Krishnamurti U, Qin G, Aneja R, *et al*. Protein Conformational Changes in Breast Cancer Sera Using Infrared Spectroscopic Analysis. Cancers (Basel) 2020;12(7):1708. doi:10.3390/cancers12071708, PMID:32605072.

[15] Dimitrova M, Ivanova D, Karamancheva I, Milev A, Dobrev I. Application of FTIR spectroscopy for diagnosis of breast cancer tumours. Journal of the University of Chemical Technology and Metallurgy 2009;44(3):297–300.

[16] Yousif ES, Fawaz NI, Mohamed BJ. Human breast tissue cancer diagnosis by FT-IR spectroscopy. Journal of University of Anbar for Pure Science 2015;9(3):19–26. doi:10.37652/juaps.2015.127546.

[17] Markouizou A, Anastassopoulou J, Kolovou P, Theophanides T, Tsekeris P. Fourier Transform Infrared Spectroscopy in the Study of Discrimination of Lobular Breast Cancers. Cancer Diagn Progn 2022;2(6):750–757. doi:10.21873/cdp.10170, PMID:36340465.

[18] Dovbeshko GI, Gridina NY, Kruglova EB, Pashchuk OP. FTIR spectroscopy studies of nucleic acid damage. Talanta 2000;53(1):233–246. doi:10.1016/s0039-9140(00)00462-8, PMID:18968108.

[19] Lazaro-Pachecoa D, Shaaban A, Baldenb G, Titiloye NA, Rehman S, Rehman IU. Deciphering the structural and chemical composition of breast cancer using FTIR spectroscopy. Appl Spectrosc Rev 2020;57(3):234–248. doi:10.1080/05704928.2020.1843471.

[20] Talari ACS, Evans CA, Holen I, Coleman RE, Rehman IU. Raman spectroscopic analysis differentiates between breast cancer cell lines. J Raman Spectrosc 2015;46(5):421–427. doi:10.1002/jrs.4676.

[21] Ferreira ICC, Aguiar EMG, Silva ATF, Santos LLD, Cardoso-Sousa L, Araújo TG, *et al*. Attenuated Total Reflection-Fourier Transform Infrared (ATR-FTIR) Spectroscopy Analysis of Saliva for Breast Cancer Diagnosis. J Oncol 2020;2020:4343590. doi:10.1155/2020/4343590, PMID:32104176.

[22] Carson FL, Cappellano CH. Histotechnology: A Self Instructional Text. 4th ed. Chicago, IL: American Society College of Pathology press, USA; 2019:264–282.

[23] Cui L, Butler HJ, Martin-Hirsch PL, Martin FL. Aluminium foil as a potential substrate for ATR-FTIR, transflection FTIR, or Raman spectrochemical analysis of biological specimen. Anal Methods 2016;8(3):481–487. doi:10.1039/C5AY02638E.

[24] Ji Y, Yang X, Ji Z, Zhu L, Ma N, Chen D, *et al*. DFT-Calculated IR Spectrum Amide I, II, and III Band Contributions of N-Methylacetamide Fine Components. ACS Omega 2020;5(15):8572–8578. doi:10.1021/acsomega.9b04421, PMID:32337419.
[25] Kluz-Barłowska M, Kluz T, Paja W, Sarzyński J, Łączyńska-Madera M, Odrzywolski A, *et al*. FT-Raman data analyzed by multivariate and machine learning as a new methods for detection spectroscopy marker of platinum-resistant women suffering from ovarian cancer. Sci Rep 2023;13(1):20772. doi:10.1038/s41598-023-48169-3, PMID:38008780.
[26] Moggio M, Errico S, Diano N, Lepore M. FTIR spectroscopy for evaluation and monitoring of lipid extraction efficiency for murine liver tissues analysis. Eng Proc 2021;10(1):9. doi:10.3390/ecsa-8-11321.
[27] Kumar S, Verma T, Mukherjee R, Ariese F, Somasundaram K, Umapathy S. Raman and infra-red microspectroscopy: towards quantitative evaluation for clinical research by ratiometric analysis. Chem Soc Rev 2016;45(7):1879–1900. doi:10.1039/c5cs00540j, PMID:26497386.
[28] Talari AC, Martinez MAG, Movasaghi Z, Rehman S, Rehman IU. Advances in Fourier transform infrared (FTIR) spectroscopy of biological tissues. Appl Spectrosc Rev 2017;52(5):456–506. doi:10.1080/05704928.2016.1230863.
[29] Mishra A, Zehra S, Bharti PK, Mathur SR, Ranjan P, Batra A, *et al*. Spectroscopic insight into breast cancer: profiling small extracellular vesicles lipids via infrared spectroscopy for diagnostic precision. Sci Rep 2024;14(1):9347. doi:10.1038/s41598-024-59863-1, PMID:38654096.
[30] Wei SC, Fattet L, Tsai JH, Guo Y, Pai VH, Majeski HE, *et al*. Matrix stiffness drives epithelial-mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. Nat Cell Biol 2015;17(5):678–688. doi:10.1038/ncb3157, PMID:25893917.
[31] Rasuleva K, Elamurugan S, Bauer A, Khan M, Wen Q, Li Z, *et al*. β-Sheet Richness of the Circulating Tumor-Derived Extracellular Vesicles for Noninvasive Pancreatic Cancer Screening. ACS Sens 2021;6(12):4489–4498. doi:10.1021/acssensors.1c02022, PMID:34846848.
[32] Stani C, Vaccari L, Mitri E, Birarda G. FTIR investigation of the secondary structure of type I collagen: New insight into the amide III band. Spectrochim Acta A Mol Biomol Spectrosc 2020;229:118006. doi:10.1016/j.saa.2019.118006, PMID:31927236.
[33] Depciuch J, Barnaś E, Skręt-Magierło J, Skręt A, Kaznowska E, Łach K, *et al*. Spectroscopic evaluation of carcinogenesis in endometrial cancer. Sci Rep 2021;11(1):9079. doi:10.1038/s41598-021-88640-7, PMID:33907297.
[34] Sahu RK, Argov S, Salman A, Huleihel M, Grossman N, Hammody Z, *et al*. Characteristic absorbance of nucleic acids in the Mid-IR region as possible common biomarkers for diagnosis of malignancy. Technol Cancer Res Treat 2004;3(6):629–638. doi:10.1177/153303460400300613, PMID:15560721.
[35] Movasaghi Z, Rehman S, Rehman IU. Fourier transform infrared (FTIR) spectroscopy of biological tissues. Appl Spectrosc Rev 2008;43(2):134–179. doi:10.1080/05704920701829043.
[36] Amjad M, Ullah H, Andleeb F, Batool Z, Nazir A, Gilanie G. Fourier transform infrared spectroscopy for investigation of human carcinoma and leukaemia. Lasers in Engineering 2021;51(1-5):217–233.
[37] Paraskevaidi M, Morais CLM, Ashton KM, Stringfellow HF, McVey RJ, Ryan NAJ, *et al*. Detecting Endometrial Cancer by Blood Spectroscopy: A Diagnostic Cross-Sectional Study. Cancers (Basel) 2020;12(5):1256. doi:10.3390/cancers12051256, PMID:32429365.
[38] Barnas E, Skret-Magierlo J, Skret A, Kaznowska E, Depciuch J, Szmuc K, *et al*. Simultaneous FTIR and Raman Spectroscopy in Endometrial Atypical Hyperplasia and Cancer. Int J Mol Sci 2020;21(14):4828. doi:10.3390/ijms21144828, PMID:32650484.
[39] Yu G, Xu JL, Niu Y, Zhang CZ, Zhang CP. Studies on breast tumor tissues with ATR-FTIR spectroscopy. In: Chance B, Chen M, Chiou AET, Luo Q (eds). Optics in Health Care and Biomedical Optics: Diagnostics and Treatment II. Vol. 5630. SPIE; 2005:796–801. doi:10.1117/12.576398.
[40] Yano K, Sakamoto Y, Hirosawa N, Tanooka S, Katayama H, Kumaido K, *et al*. Applications of fourier transform infrared spectroscopy, fourier transform infrared microscopy and near infrared spectroscopy to cancer research. Spectroscopy 2003;17:315–321. doi:10.1155/2003/329478.
[41] Sabtu SN, Sani SFA, Looi LM, Chiew SF, Pathmanathan D, Bradley DA, *et al*. Indication of high lipid content in epithelial-mesenchymal transitions of breast tissues. Sci Rep 2021;11(1):3250. doi:10.1038/s41598-021-81426-x, PMID:33547362.
[42] Surmacki J, Brozek-Pluska B, Kordek R, Abramczyk H. The lipid-reactive oxygen species phenotype of breast cancer. Raman spectroscopy and mapping, PCA and PLSDA for invasive ductal carcinoma and invasive lobular carcinoma. Molecular tumorigenic mechanisms beyond Warburg effect. Analyst 2015;140(7):2121–2133. doi:10.1039/c4an01876a, PMID:25615557.
[43] Denti V, Piga I, Guarnerio S, Clerici F, Ivanova M, Chinello C, *et al*. Antigen Retrieval and Its Effect on the MALDI-MSI of Lipids in Formalin-Fixed Paraffin-Embedded Tissue. J Am Soc Mass Spectrom 2020;31(8):1619–1624. doi:10.1021/jasms.0c00208, PMID:32678590.
[44] Yang Y, Sulé-Suso J, Sockalingum GD, Kegelaer G, Manfait M, El Haj AJ. Study of tumor cell invasion by Fourier transform infrared microspectroscopy. Biopolymers 2005;78(6):311–317. doi:10.1002/bip.20297, PMID:15898120.
[45] Perczel A, Gáspári Z, Csizmadia IG. Structure and stability of beta-pleated sheets. J Comput Chem 2005;26(11):1155–1168. doi:10.1002/jcc.20255, PMID:15952205.
[46] Faria RA, Leal LB, Thebit MM, Pereira SWA, Serafim NR, Barauna VG, *et al*. Potential Role of Fourier Transform Infrared Spectroscopy as a Screening Approach for Breast Cancer. Appl Spectrosc 2023;77(4):405–417. doi:10.1177/00037028231156194, PMID:36703259.
[47] Rygula A, Majzner K, Marze KM, Kaczor A, Pilarczyk M, Baranska M. Raman spectroscopy of proteins: a review. J Raman Spectrosc 2013;44(8):1061–1076. doi:10.1002/jrs.4335.
[48] Dannhorn A, Swales JG, Hamm G, Strittmatter N, Kudo H, Maglennon G, *et al*. Evaluation of Formalin-Fixed and FFPE Tissues for Spatially Resolved Metabolomics and Drug Distribution Studies. Pharmaceuticals (Basel) 2022;15(11):1307. doi:10.3390/ph15111307, PMID:36355479.
[49] Volynskaya Z, Haka AS, Bechtel KL, Fitzmaurice M, Shenk R, Wang N, *et al*. Diagnosing breast cancer using diffuse reflectance spectroscopy and intrinsic fluorescence spectroscopy. J Biomed Opt 2008;13(2):024012. doi:10.1117/1.2909672, PMID:18465975.
[50] Vasudevan S, Campbell C, Liu F, O'Sullivan TD. Broadband diffuse optical spectroscopy of absolute methemoglobin concentration can distinguish benign and malignant breast lesions. J Biomed Opt 2021;26(6):065004. doi:10.1117/1.JBO.26.6.065004, PMID:34189876.
[51] Sitnikova VE, Kotkova MA, Nosenko TN, Kotkova TN, Martynova DM, Uspenskaya MV. Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis. Talanta 2020;214:120857. doi:10.1016/j.talanta.2020.120857, PMID:32278436.
[52] Zelig U, Barlev E, Bar O, Gross I, Flomen F, Mordechai S, *et al*. Early detection of breast cancer using total biochemical analysis of peripheral blood components: a preliminary study. BMC Cancer 2015;15:408. doi:10.1186/s12885-015-1414-7, PMID:25975566.
[53] Talari ACS, Rehman S, Rehman IU. Advancing cancer diagnostics with artificial intelligence and spectroscopy: identifying chemical changes associated with breast cancer. Expert Rev Mol Diagn 2019;19(10):929–940. doi:10.1080/14737159.2019.1659727, PMID:31461624.
[54] Barreto AF, Stillwell R, Campbell C, Wei AY, Abdusalam O, Pllice AM, *et al*. Prospective study of quantitative broadband diffuse optical spectroscopy for differential diagnosis of suspicious breast lesions. In: Fantini S, Taroni P (eds). Optical Tomography and Spectroscopy of Tissue XVI. Vol. 13314. SPIE; 2025:133140U. doi:10.1117/12.3043695.
[55] Pakalniskis MG, Wells WA, Schwab MC, Froehlich HM, Jiang S, Li Z, *et al*. Tumor angiogenesis change estimated by using diffuse optical spectroscopic tomography: demonstrated correlation in women undergoing neoadjuvant chemotherapy for invasive breast cancer? Radiology 2011;259(2):365–374. doi:10.1148/radiol.11100699, PMID:21406632.
[56] Sahu RK, Argov S, Salman A, Zelig U, Huleihel M, Grossman N, *et al*. Can Fourier transform infrared spectroscopy at higher wavenumbers (mid IR) shed light on biomarkers for carcinogenesis in tissues? J Biomed Opt 2005;10(5):054017. doi:10.1117/1.2080368, PMID:16292977.
[57] Luo Y, Liu H, Wu C, Paraskevaidi M, Deng Y, Shi W, *et al*. Diagnos-

tic segregation of human breast tumours using Fourier-transform infrared spectroscopy coupled with multivariate analysis: Classifying cancer subtypes. Spectrochim Acta A Mol Biomol Spectrosc 2021;255:119694. doi:10.1016/j.saa.2021.119694, PMID:33799187.

[58] Tomas RC, Sayat AJ, Atienza AN, Danganan JL, Ramos MR, Fellizar A, *et al*. Detection of breast cancer by ATR-FTIR spectroscopy using artificial neural networks. PLoS One 2022;17(1):e0262489. doi:10.1371/journal.pone.0262489, PMID:35081148.