

## RESEARCH ARTICLE

# Detection of breast cancer by ATR-FTIR spectroscopy using artificial neural networks

Rock Christian Tomas<sup>1\*</sup>, Anthony Jay Sayat<sup>2</sup>, Andrea Nicole Atienza<sup>2</sup>, Jannah Lianne Danganan<sup>2</sup>, Ma. Rollene Ramos<sup>2</sup>, Allan Fellizar<sup>3,4,5</sup>, Kin Israel Notarte<sup>6</sup>, Lara Mae Angeles<sup>6,7</sup>, Ruth Bangaol<sup>3,4</sup>, Abigail Santillan<sup>3,4</sup>, Pia Marie Albano<sup>2,3,4</sup>

**1** Department of Electrical Engineering, University of the Philippines Los Baños, Los Baños, Laguna, Philippines, **2** Department of Biological Sciences, College of Science, University of Santo Tomas, Manila, Philippines, **3** Research Center for the Natural and Applied Sciences, University of Santo Tomas, Manila, Philippines, **4** The Graduate School, University of Santo Tomas, Manila, Philippines, **5** Mariano Marcos Memorial Hospital and Medical Center, Batac, Ilocos Norte, Philippines, **6** Faculty of Medicine and Surgery, University of Santo Tomas, Manila, Philippines, **7** University of Santo Tomas Hospital, Manila, Philippines

\* [rvtomas1@up.edu.ph](mailto:rvtomas1@up.edu.ph)



## OPEN ACCESS

**Citation:** Tomas RC, Sayat AJ, Atienza AN, Danganan JL, Ramos M.R, Fellizar A, et al. (2022) Detection of breast cancer by ATR-FTIR spectroscopy using artificial neural networks. PLoS ONE 17(1): e0262489. <https://doi.org/10.1371/journal.pone.0262489>

**Editor:** David Mayerich, University of Houston, UNITED STATES

**Received:** June 17, 2021

**Accepted:** December 27, 2021

**Published:** January 26, 2022

**Copyright:** © 2022 Tomas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper. Furthermore, we have uploaded and made our source code and minimal dataset available in GitHub. One may be able to access the files at <https://github.com/rvtomas1/breast-cancer-ftir-codes-and-data>.

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

In this study, three (3) neural networks (NN) were designed to discriminate between malignant ( $n = 78$ ) and benign ( $n = 88$ ) breast tumors using their respective attenuated total reflection Fourier transform infrared (ATR-FTIR) spectral data. A proposed NN-based sensitivity analysis was performed to determine the most significant IR regions that distinguished benign from malignant samples. The result of the NN-based sensitivity analysis was compared to the obtained results from FTIR visual peak identification. In training each NN models, a 10-fold cross validation was performed and the performance metrics—area under the curve (AUC), accuracy, positive predictive value (PPV), specificity rate (SR), negative predictive value (NPV), and recall rate (RR)—were averaged for comparison. The NN models were compared to six (6) machine learning models—logistic regression (LR), Naïve Bayes (NB), decision trees (DT), random forest (RF), support vector machine (SVM) and linear discriminant analysis (LDA)—for benchmarking. The NN models were able to outperform the LR, NB, DT, RF, and LDA for all metrics; while only surpassing the SVM in accuracy, NPV and SR. The best performance metric among the NN models was  $90.48\% \pm 10.30\%$  for AUC,  $96.06\% \pm 7.07\%$  for ACC,  $92.18 \pm 11.88\%$  for PPV,  $94.19 \pm 10.57\%$  for NPV,  $89.04\% \pm 16.75\%$  for SR, and  $94.34\% \pm 10.54\%$  for RR. Results from the proposed sensitivity analysis were consistent with the visual peak identification. However, unlike the FTIR visual peak identification method, the NN-based method identified the IR region associated with C–OH C–OH group carbohydrates as significant. IR regions associated with amino acids and amide proteins were also determined as possible sources of variability. In conclusion, results show that ATR-FTIR via NN is a potential diagnostic tool. This study also suggests a possible more specific method in determining relevant regions within a sample's spectrum using NN.

## Introduction

Breast cancer remains the most prevalent cancer among women. Biennial mammography has been highly recommended for women 50 to 74 years old for early detection of this disease. Sensitivity of mammography has been increased to 92.7% when combined with magnetic resonance imaging (MRI). Meanwhile, combination with ultrasound (US) can increase sensitivity to only 52%. Therefore, in high-risk women for whom supplemental screening is specified, MRI is recommended when possible [1]. Supplemental screening with US for women with intermediate risk and dense breasts is an option to increase cancer detection. The mammographic sensitivity for breast cancer in women with very dense breasts is 47.6% and increased to 76.1% with US screening [2]. Suspicious lesions detected during mammograms are usually biopsied to confirm or rule out breast cancer. The most common form of biopsy is the core needle biopsy (CNB), which involves the removal of a portion of a tumor for histologic evaluation. The remainder is removed later after a definitive diagnosis of cancer. However, since tissue samples are collected by three to nine passes with a biopsy needle, the incisional surgical procedures become associated with elevated incidence of lymph node metastasis and higher local recurrence rates [3]. CEA can also be analysed to screen for breast cancer. However, it lacks disease sensitivity and specificity, hence cannot be used for screening a subpopulation with high risk for malignancies, a general asymptomatic population, or for independently diagnosing cancer. CA 15–3, which are soluble forms of the transmembrane protein Mucin1 (MUC1), is said to be overexpressed in malignant breast tumors. It was suggested that CA 15–3 and CEA can be considered complementary in detecting recurrence of breast cancer. However, their sensitivity is low and independent of the majority of the prognostic parameters that may be considered before relapse [4].

The potential of using infrared spectroscopy, in particular Fourier transform infrared (FTIR) spectroscopy, has been gaining popularity for cancer diagnostics over the last few years. The distinctive spectral properties associated with the changes in chemical composition and structure of biomolecules can be recognized by FTIR spectroscopy, making it a potential diagnostic tool [5]. Hence, when cells or tissues undergo transformation from normal to cancerous, changes in the physico-chemical structures and properties in a variety of their biomolecules can be simultaneously and indiscriminately probed by FTIR spectroscopy [6, 7]. FTIR, as compared to traditional microscopic examination of hematoxylin and eosin (H&E) stained biopsies, is more rapid, cost-effective, and objective since reading is based on changes in biochemical properties instead of morphology [7]. It eliminates the possibility of intra- and inter-observer variability, which is often the problem with H&E staining. Moreover, this technology does not make use of dyes and other contrasting agents which may interfere or affect the reading. Hence, it provides more accurate and reproducible results.

The application of artificial intelligence (AI) in cancer diagnosis is no longer new; numerous studies have already been applied to address the most prevalent cancers such as lung cancer [8–10], thyroid cancer [11], ovarian cancer [12, 13] and breast cancer [14–17]. Most studies make use of image-based data such as MRI images, computed tomography (CT) scans, positron emission tomography (PET) scans, X-rays, and H&E-stained biopsy images, which is the gold standard [18]. The most successful AI implemented in these studies involves artificial neural networks (NN), in particular, convolutional neural networks (CNN) due to their proven effectiveness in processing images. Furthermore, the underlying architecture of a CNN makes it easily feasible to create visual representation, highlighting the site of malignancy within a medical image. What limits image-based AI diagnostics, however, is that they heavily rely on the detection of a visible abnormality within the scanned region. This implies that a patient may already be at an advanced stage of malignancy before possible detection.

Moreover, the presence of dyes and other contrasting agents may make it difficult to apply in other laboratories an AI model trained using the procedures in one laboratory, if protocols and procedures are not well standardized.

The appeal of using FTIR data in AI is that they come in less file size and are easier to process than images, while still providing sufficiently adequate information for samples [19]. Hence, data storage costs may be minimized significantly and models utilizing such data become easier and faster to train, hence minimizing training costs. Moreover, the use of FTIR data may be able to predict the onset of cancer even before evident morphological changes [5], hence addressing the limitation of image-based AI diagnostics. However, since NNs are essentially black boxes, the underlying process involving its decision-making is inherently unknown; thus making them less appealing to use in a clinical setting. Furthermore, FTIR data are less intuitive to interpret than images, even with the assistance of AI visualizations; making them difficult to interpret. In this study, a method was formulated to address this limitation by providing a novel process of determining the most prevalent biomarkers as seen by trained artificial NNs; hence providing a basis on decision-making process. The proposed method is a modification of NN perturbation-based sensitivity analysis [20–22], which probes a NN's sensitivity towards changes in an input variable.

Hence, this study showed the potential of artificial neural networks (ANN) in accurately diagnosing breast cancer through infrared spectroscopy. Specifically, it designed multiple ANN to diagnose malignancy from breast tissues using ATR-FTIR data. The classification performance of the NN models were compared to six (6) most widely-used machine learning models. Moreover, this study also proposed a novel method for determining the IR regions which may be significant in determining breast cancer malignancy, as seen by a NN design. This proposed method may serve as a baseline process in analysing spectral data, and more importantly provide new insights and directions for pathologists and medical practitioners.

## Materials and methods

### Ethical clearance

Ethical clearance was obtained from the Institutional Review Board (IRB) of the University of Santo Tomas Hospital (USTH) in Manila, Philippines (Ref. No.: IRB-2018-07-135-IS) and Mariano Marcos Memorial Hospital and Medical Center (MMMHC-MC) in Ilocos Norte, Philippines (Ref. No.: MMMHMC-RERC-15-006). Written informed consent from the participants or their legal guardians have been waived by the respective ethics review boards since the study was restricted to the use of archived formalin fixed paraffin embedded (FFPE) breast tissues and did not involve additional procedures nor pose risk of harm to subjects. All methods were carried out in accordance with the Declaration of Helsinki and its later amendments.

### Study population and sample preparation

Two hundred (200) FFPE breast biopsies obtained from 192 adult patients seen at USTH and MMMH between January 2016 to December 2016 were included in the study. The samples were diagnosed by the resident pathologists of the hospital study sites as either benign ( $n = 91$ ) or malignant ( $n = 101$ ) based on microscopic examination of H&E-stained biopsies. The malignant samples were further subclassified as invasive ductal carcinoma, residual ductal carcinoma-in-situ, or invasive lobular carcinoma; and the benign samples as fibroadenoma, fibrocystic disease, benign fibroadipose tissue, fibrocollagenous cyst wall, or intraductal papillomas.

The FFPE tissues were uniformly cut at 5- $\mu\text{m}$  thickness using a microtome (Leica Biosystems, Germany) and three (3) adjacent tissue sections were mounted on glass slides. The two (2) outer sections were stained with H&E for re-evaluation by a third-party pathologist who

was blinded of the original diagnosis. The pathologist was instructed to classify the biopsy sample as either benign or malignant and to mark the location of the cancer cells if the sample was malignant, to serve as guide in the ATR-FTIR analysis [23]. The inner or middle tissue section was deparaffinized with xylol, dehydrated with alcohol, rinsed in distilled water, dried overnight, and subjected to ATR-FTIR analysis [24, 25].

Only the samples with similar diagnosis by the resident pathologists of the hospital study sites and the third-party pathologist were considered for further analysis. In this case, out of the 200 archival samples, only 166 ( $n = 78$  benign;  $n = 88$  malignant) were taken for ATR-FTIR processing. Furthermore, each of the 166 samples corresponded to one patient each to satisfy a one-is-to-one correspondence between patients and specimens [11].

### ATR-FTIR spectral analysis

An FTIR mid-infrared spectrometer equipped with a platinum ATR single reflection diamond sampling module (Bruker Optics, Germany) was used to obtain spectra of the breast samples. A performance qualification (PQ) test using OPUS 8.0 software's fully automated validation program was initially performed to ensure quality and accuracy of spectral data. The deparaffinized breast tissue sections were positioned directly in contact with the ATR diamond crystal's surface (2 mm x 2 mm) and the mid-IR region of  $4000\text{ cm}^{-1}$  to  $600\text{ cm}^{-1}$  was passed to and from the ATR accessory. Spectra were generated at a spatial resolution of  $4\text{ cm}^{-1}$  and an average of 48 scans was co-added to obtain an adequate signal-to-noise ratio [26–28], which was further supported by the software's validation program as “acceptable”. Prior to scanning each tissue sample, the background spectrum was recorded, and this spectrum was systematically subtracted by the software to routinely eliminate atmospheric effects. The malignant samples were scanned along the area containing the cancer cells, while the benign samples were scanned at random spots throughout their entire tissue section. The spectral data associated with a benign or a malignant tissue sample was obtained by computing for the median spectrum of their respective 48 scans.

### Characterization and pre-processing of spectral data

The spectral data set,  $X_{SD}$ , consisted of  $N \mid N = 166$  spectral vectors  $\vec{x}_{SD}^{(i)} \in \mathbb{R}^{\mathcal{L}} \forall i \in \mathbb{N} \leq N$ , comprising of 78 malignant FTIR data,  $X_{SD}^{(M)} \subset X_{SD}$ , and 88 benign FTIR data,  $X_{SD}^{(B)} \subset X_{SD}$ , where  $\mathcal{L} \mid \mathcal{L} = 462$  denotes the length of each vector. An element of a vector  $\vec{x}_{SD}^{(i)}, \vec{x}_{SD}^{(j)} \forall j \in \mathbb{N} \leq \mathcal{L}$  corresponds to an absorbance reading within the fingerprint region of  $1800\text{ cm}^{-1}$  to  $850\text{ cm}^{-1}$  at  $2\text{ cm}^{-1}$  steps. Furthermore,  $X_{SD}$  can also be characterized as a matrix of  $\mathbb{R}^{N \times \mathcal{L}}$  dimension.

All obtained spectral data were internally normalized using z-score normalization, which is the recommended method of normalization for the FNN designs [29, 30]. The normalization was done to eliminate bias from  $y$ -value discrepancies among the IR samples. Here, normalization was done per  $\vec{x}_{SD}^{(i)}$  using the equation

$$\vec{x}_{SD}^{(i)} := \frac{\vec{x}_{SD}^{(i)} - \text{mean}(\vec{x}_{SD}^{(i)})}{\text{std}(\vec{x}_{SD}^{(i)})} \quad \forall i \in \mathbb{N} \leq N \quad (1)$$

where the  $\text{mean}()$  and the  $\text{std}()$  notations denote the mean and the standard deviation of the elements of the vector  $\vec{x}_{SD}^{(i)}$ . The implemented normalization scales the elements of  $\vec{x}_{SD}^{(i)}$  to have an overall mean of 0, and a standard deviation of 1.  $X_{SD}$  also underwent baseline correction via OPUS 8.0 software via “rubber band method” with 64 baseline points. This was done to approximate a polynomial fit based on the minima of  $y$ -values of each element vector

$\vec{x}_{SD} \in X_{SD}$ . The fitted polynomial was then deducted for all  $\vec{x}_{SD}$  to extract the baseline corrected spectrum [25, 31–34]. Finally, the corrected spectrum was scaled within the fingerprint region, from  $1800 \text{ cm}^{-1}$  to  $850 \text{ cm}^{-1}$  [32, 35]. Other than baseline correction using rubber band method, no further user intervention was done to assess the spectral data. To visualize the average spectrum of benign and malignant breast tissues, their respective median values for each wavenumber was plotted.

### Principal component analysis (PCA)

Principal component analysis (PCA) was performed to visualize the distribution of benign and malignant samples over two of the PCA’s most dominant components ( $F_1$  and  $F_2$ ). The process of translating  $X_{SD}$  to the reduced variable space,  $X_{PCA}$  ( $X_{SD} \rightarrow X_{PCA}$ ) is given by the equation

$$X_{PCA} = (X_{SD} - \bar{X}_{SD}) \times [\vec{s}_{F1}^T, \vec{s}_{F2}^T] \tag{2}$$

where  $X_{PCA} \in \mathbb{R}^{N \times 2}$  is the reduced sample space,  $\bar{X}_{SD}$  is the mean absorbance value of  $\vec{x}_{SD}^{(i)} \in X_{SD} \forall i \leq N$ , and  $\vec{s}_{F1}$  and  $\vec{s}_{F2}$  are the eigenvectors corresponding to the largest eigenvalue of the covariance matrix  $S_{SD} = (X_{SD} - \bar{X}_{SD})^T \times (X_{SD} - \bar{X}_{SD})$ . A PCA biplot of the malignant and the benign samples were drawn along the  $F_1$  and  $F_2$  axes to visualize the sample distribution.

### Classifiers

Three (3) feed forward neural networks (FNN) of different layer sizes were designed in the study. To benchmark the NN models, six of the most widely used machine learning models were also created, in particular, linear discriminant analysis (LDA), support vector machine (SVM), logistic regression (LR), decision tree (DT), random forest (RF), and Naïve Bayes (NB). The following subsections discuss in detail the design of each model.

**Cross-validation of models.** A 10-fold cross-validation procedure was used to evaluate all models; where 70% of the spectral data set  $X_{SD}$ , were used for the training set  $S_{TR} \subset X_{SD}$ , and the remaining 30% were equally partitioned for the validation set  $S_V \subset X_{SD}$  (15%) and the test set  $S_{TS} \subset X_{SD}$  (15%). The cross-validation procedure was repeated over 50 trials ( $T$ ) to ensure stability of results [36]. For each trial, the elements of the sets  $S_{TR}$ ,  $S_V$ , and  $S_{TS}$  were reselected from  $X_{SD}$  randomly. The sets satisfied the criteria  $S_{TR}^{(i)} \cup S_V^{(i)} \cup S_{TS}^{(i)} = X_{SD}$  and  $S_{TR}^{(i)} \cap S_V^{(i)} \cap S_{TS}^{(i)} = \emptyset \forall i \in \mathbb{N} \leq T$ . Moreover, the ratio of malignant and benign samples was preserved for each set. To evaluate the performance of each model, the metrics area under the curve (AUC), accuracy (ACC), positive predictive value (PPV), negative predictive value (NPV), recall rate (RR) and specificity rate (SR) were obtained. The overall mean and standard deviation of the metrics over the 50 trials were obtained using the formulas

$$M_m = \frac{1}{T} \sum_1^T \left( \frac{1}{N} \sum_1^N m_{n,t} \right) \tag{3.1}$$

$$\Sigma_m = \frac{1}{T} \sum_1^T \left( \sqrt{\frac{\sum_1^N (m_{n,t} - \bar{m}_{n,t})^2}{N - 1}} \right) \tag{3.2}$$

in which  $M_m$  and  $\Sigma_m$  are the overall mean and the overall standard deviation of a metric  $m$ , where  $m_{n,t}$  is the metric value of a metric  $m$ , for a trial  $t \in \mathbb{N} \leq T$  and a fold  $n \in \mathbb{N} \leq 10$ . Moreover, the variable  $\bar{m}_{n,t}$  in Eq 3.2 is the  $N$ -fold mean of a metric  $m$ , which is also equal to  $\frac{1}{N} \sum_1^N m_{n,t}$  from Eq 3.1.

All models were designed and implemented using MATLAB R2020b on an Intel i7-6700 3.40 GHz CPU and an Nvidia GeForce GTX 1050 Ti GPU over a 16 GB RAM.

**Feedforward neural networks (FNN).** Three (3) feed forward neural networks (FNN) were designed in the study. Each neural network has an input layer  $\vec{N}_i$ , consisting of  $\mathcal{L}$  nodes corresponding to the length of each spectral vector input  $\vec{x}_{SD}$ , and an output layer  $\vec{N}_o$ , consisting of 2 nodes which correspond to the sample's diagnosis of being benign or malignant. A neural network varies in hidden layer size ( $n = 2, n = 4, n = 8$ ) with respect to the other. For all FNNs, scaled exponential linear units (SELU) were used as activation functions within the hidden layers [37], and softmax function for the output layer.

**Linear discriminant analysis (LDA).** Here,  $S_{TR}$  was used to design an LDA model  $f_{LDA}(x)$  which returns the probability of a spectral vector input  $x := \vec{x}_{SD}$  from being benign or malignant.  $S_{TS}$  was used to measure the metric performance of the model. Before constructing a linear separator among the samples, principal component analysis (PCA) was performed to reduce the dimension of data from 462 to two variables, which are  $F_1$  and  $F_2$ . The reduction of dimensional space via PCA ( $X_{SD} \rightarrow X_{PCA}$ ) follows the discussion from Eq 2.

The LDA model was constructed following Fisher's criterion [38] where the probability density function describing the likelihood of a sample from being malignant or benign is given by the function  $f_{LDA}(x)$  [11]

$$f_{LDA}(x) = softmax([p_M(x), p_B(x)]) \tag{4.1}$$

in which  $p_M(x)$  and  $p_B(x)$  are normal probability density functions describing the probability distribution of malignant and the benign samples, respectively. The normal curves are defined as

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\gamma(x)-\mu)^2}{2\sigma^2}} \tag{4.2}$$

where  $\gamma \in \Gamma \mid \Gamma = w_{min}^T X_{PCA}$ , where  $\Gamma$  is the projection of  $X_{PCA}$  onto  $w_{min}$ , while  $\mu$  and  $\sigma$  are the class mean and standard deviation, respectively.

In evaluating an element of the test set  $\vec{x}_{SD}^{(TS)} \in S_{TS}$ , the  $\mathcal{L} \times 2$  matrix  $[\vec{s}_{F1}^T, \vec{s}_{F2}^T]$  obtained from the training set via Eq 2 was first used to translate  $\vec{x}_{SD}^{(TS)}$  from a dimension of  $\mathbb{R}^{\mathcal{L}}$  to  $\mathbb{R}^2$  before evaluation using Eq 4.2.

**Support vector machine (SVM).** The designed SVM is a linear SVM of input  $x := \vec{x}_{SD}$ . The SVM was designed from the elements of the training set by considering an unconstrained Lagrange optimization problem given by the equations [39, 40]

$$f(\vec{w}, b) = \frac{1}{2} |\vec{w}|^2 \tag{5.1}$$

$$g(\vec{w}, b) = y_i(x \cdot \vec{w} + b) - 1 \tag{5.2}$$

$$L_{\min(\vec{w}, b)}(\vec{w}, b) = f(\vec{w}, b) - \sum_{i=1}^{N_{TR}} \alpha_i g(\vec{w}, b) = \frac{1}{2} |\vec{w}|^2 - \sum_{i=1}^{N_{TR}} \alpha_i [y_i(x \cdot \vec{w} + b) - 1] \tag{5.3}$$

where  $f(\vec{w}, b)$  corresponds to the equation for maximizing support vector elements separation,  $g(\vec{w}, b)$  is the conditional function for clustering the malignant and the benign classes, and  $L_{\min(\vec{w}, b)}$  is the Lagrangian  $L$  of  $f(\vec{w}, b)$  and  $g(\vec{w}, b)$  of variables  $\vec{w}$  and  $b$ . Here,  $\vec{w}$  is the weight vector of  $\mathbb{R}^{\mathcal{L}}$  dimension,  $b$  is the bias of  $\mathbb{R}^1$  dimension, while  $\alpha_i \in \alpha \mid \alpha \in \mathbb{R}^{N_{TR} \times \mathcal{L}}$  is the SVM's  $\alpha$ -matrix where the  $\alpha_i \neq 0 \forall i \in \mathbb{N} \leq N_{TR}$  elements correspond to the SVM's support

vectors. To determine suboptimal values for  $\vec{w}$ ,  $b$ , and  $\alpha$ , stochastic gradient descent (SGD) was implemented by considering the gradients

$$\frac{\frac{1}{2}|\vec{w}|^2}{\partial \vec{w}} = \frac{\frac{1}{2}\vec{w} \cdot \vec{w}}{\partial \vec{w}} \tag{5.4}$$

$$\frac{\partial L(\vec{w}, b)}{\partial \vec{w}} = \vec{w} - \sum_{i=1}^{N_{TR}} \alpha_i y_i x \tag{5.5}$$

$$\frac{\partial L(\vec{w}, b)}{\partial b} = -\sum_{i=1}^{N_{TR}} \alpha_i y_i \tag{5.6}$$

To optimize the model, a grid search was performed from a series of learning rates  $\ell$  from 1 to  $5 \times 10^{-5}$  over 1000 epochs. The validation set accuracy was considered as the optimization metric which determined the superiority of one model from the other. The process was repeated for 50 trials to ensure stability. The average validation accuracy of a model over the 50 trials determined the overall metric of the model for a considered learning rate,  $\ell_i$ . The  $\ell_i$  which constituted the highest overall validation accuracy was considered as optimal learning rate to train and test the SVM model. The output probability diagnosis of the model for benign and malignant cases,  $p_{SVM}(x)$ , was computed using Platt’s method [41].

**Logistic regression (LR).** The designed LR model is a  $\mathcal{L}$ -input classifier with an output probability  $p_{LR}(x)$  quantifying the likelihood of an input  $x := \vec{x}_{SD}$  to be malignant.  $p_{LR}(x)$  is defined as

$$p_{LR}(x) = \frac{e^{b+\vec{w} \cdot x}}{1 + e^{b+\vec{w} \cdot x}} \tag{6.1}$$

where  $w \in \mathbb{R}^{\mathcal{L}}$  and  $b \in \mathbb{R}^1$  are the weights and bias characterizing the LR model. In training the model, SGD was used to minimize loss over the training set. The considered loss function was the binary cross-entropy loss function given by

$$H = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i)) \tag{6.2}$$

where  $p(y_i)$  denotes the probability of obtaining a malignant diagnosis given a theoretical output of  $y_i$ ; where  $y_i = 1$  for malignant cases, and  $y_i = 0$  for benign. To optimize the model, a grid search was performed with a design similar to that performed for the SVM.

**Decision tree (DT) and random forest (RF).** The classification and regression trees (CART) algorithm was used to generate DTs of binary splits. The Gini’s diversity index was used to find the best input  $\Phi_j | j \in \mathbb{N} \leq \mathcal{L}$  for splitting the training set for each iteration of branching; where  $\Phi_j$  is the  $j^{th}$  wavenumber from the fingerprint region  $1800 \text{ cm}^{-1}$  to  $850 \text{ cm}^{-1}$ . Gini’s diversity index [42] is given by

$$GINI(j)_t = \sum_{m=1}^2 \left( \frac{N_{m|\hat{\Phi}}}{N_t} \cdot (P(m|t)^2 + [1 - P(m|t)]^2) \right) \tag{7.1}$$

where the maximum bound for the summation operation denotes the binary characteristic of the splitting considered. Furthermore,  $N_{m|\hat{\Phi}}$  is the total number of class  $m$  separated by the value  $\hat{\Phi} \in \vec{\Phi}_j | \vec{\Phi}_j = \text{row}_j(X_{SD})$  at node  $t$ ,  $N_t$  is the total number of elements in  $\vec{\Phi}_j$  at node  $t$  and  $P(m|t)$  is the probability of class  $m$  for being either malignant or benign from happening at node  $t$ . Since the elements in  $\vec{\Phi}_j$  are continuous variables, the best value of separation  $\hat{\Phi}$  was

identified from  $\vec{\Phi}_j$  by considering the element having the least  $GINI(j)_t$  metric [43]. The branching was recursively performed for each newly created node until the performance in the validation set accuracy decreased.

The designed RF utilized the creation of trees following the previously discussed. The diagnosis of the RF was determined as the prevailing diagnosis made by its constituent bags of DTs. To determine the optimum number of trees  $N_{RF}$  for the RF, a grid search from 3 to 100 trees was performed. The validation set accuracy was considered as the optimization criteria of the search. Each simulation was repeated over 50 times for each iteration  $n | n \in \{3 \leq \mathbb{N} \leq 100\}$  to ensure stability. The average accuracy over the 50 trials served as the final performance metric of the RF for an  $N_{RF}$  equal to  $n$ . The final RF constituted the design with the highest average accuracy.

**Naïve bayes (NB).** The designed NB is a classifier of two classes of  $n | n \in \mathbb{N} \leq \mathcal{L}$  inputs. For each  $j^{th}$  input,  $\vec{\Phi}_j$ , the best  $\hat{\Phi}$  value of separating the elements of  $\vec{\Phi}_j$  between two sub-classes was determined. The algorithm for finding  $\hat{\Phi}$  is the same as that of the DT and RF designs where the Gini's index was used (Eq 7.1). The predictive value  $f_{NB}(x)$  of the NB is defined as the probability of a sample for being malignant, given an input  $x := \vec{x}_{SD}$ .  $p_{NB}(x)$  is given by

$$p_{NB}(x) = \frac{\prod_{j=1}^n P(m_j | malignant)}{\prod_{j=1}^n P(m_j)} \quad (8.1)$$

where the numerator corresponds to the total probability of an input  $x$  from happening, given  $n$ -inputs considered, with an  $x(j)$  value classified as class  $m_j$  for a determined separation  $\hat{\Phi}_j$  for malignant cases. On the other hand, the denominator is the total probability of the set of  $m_j$  from  $x$  for ever happening. The NB classifier outputs a malignant diagnosis when  $p_{NB} > 50\%$ ; otherwise, the diagnosis is benign. In order to determine the optimal  $n$ -value for the classifier, the number of inputs was increased from 3 to  $\mathcal{L}$ , where the inputs of the least  $GINI(j)_t$  value were considered first. The optimization was terminated at the  $n$ -value where the validation accuracy of the model started to decrease. Each iteration of  $n$  was repeated for 50 trials, where the average validation accuracy from the 50 trials was the considered optimization metric criterion. The final NB constituted to the design with the highest average validation accuracy.

## Identification of dominant spectral components

In order to identify the most significant wavenumbers which influenced a sample's diagnosis via the NN models, a novel sensitivity analysis was performed based on the optimized FNNs ( $n = 2, n = 4, n = 8$ ). It must be noted that a visual peak analysis of the obtained spectral data was also performed prior to sensitivity analysis to compare the identified significant wavenumbers from the NN.

**Visual peak analysis.** Significant peaks in the fingerprint region were identified through visual inspection of  $X_{SD}$ . Test of normal distribution using Shapiro-Wilk test and variance of homogeneity were performed for the identified peaks. Since all data followed a non-normal distribution, they were subjected to Mann-Whitney  $U$  test to assess if the absorbance peaks of malignant samples were significantly different ( $p$ -value  $< 0.05$ ) from the benign samples. Statistical analyses were performed using MATLAB 2020b.

**Sensitivity analysis of neural network.** A modified neural network committee-based (NNC) sensitivity analysis was considered using the input perturbation algorithm. In order to simplify the NNC sensitivity analysis, the committee of NNs were designed and trained following the design architecture of the optimized FNNs [20–22].

For the analysis, an experimental set  $S_{EXP} \subset X_{SD}$  was used to train and analyse the FNN design.  $S_{EXP}$  comprises of 70% randomly-selected elements from  $X_{SD}$ . The elements of  $S_{EXP}$  was randomly selected from  $X_{SD}$ . Moreover, the quantity of malignant and benign samples from  $S_{EXP}$  were equally proportioned. For each selected input  $\vec{x}_{SD}(j) \forall j \in \mathbb{N} \leq \mathcal{L}$ , a perturbation  $\Delta x_j$  from -50% to 50% of  $\vec{x}_{SD}(j)$  at 5% steps was added to the  $\vec{x}_{SD}(j)$ , and the mean square error (MSE) of the output was tabulated; where  $\overline{\vec{x}_{SD}(j)}$  is the mean value of  $S_{EXP}$  for the  $j^{th}$  input [21]. The MSE for the  $j^{th}$  input variable at its  $k^{th}$  step perturbation  $\Delta x_{j,k}$  ( $MSE_{j,k}$ ) is calculated using the formula

$$MSE_{j,k} = \sum_{l=1}^{N_{EXP}} \frac{\sum_{i=1}^2 \sqrt{O_{j,k}(i)^2 - \widehat{O}(i)^2}}{2} \tag{9.1}$$

where  $N_{EXP}$  is the total number of elements in  $S_{EXP}$ ,  $O_{j,k}$  is the output vector of the model for  $\Delta x_{j,k}$ , and  $\widehat{O}$  is the ideal output vector; where  $\widehat{O} = \langle 1, 0 \rangle$  for a benign sample, and  $\widehat{O} = \langle 0, 1 \rangle$  for a malignant one. The overall response of the network  $MSE_j$  for the  $j^{th}$  input was computed by averaging  $MSE_{j,k}$  for all  $\Delta x_{j,k}$   $k$ -steps given by

$$MSE_j = \frac{\sum_{i=1}^{N_k} MSE_{j,k}(i)}{N_k} \tag{9.2}$$

To ensure stability of the performed sensitivity analysis, the process was repeated for a committee of 50 NN (*i.e.*, 50 trials). The overall response of the  $j^{th}$  input, was computed as the average of the  $j^{th}$  input response over the 50 trials  $\overline{MSE_j}$ .

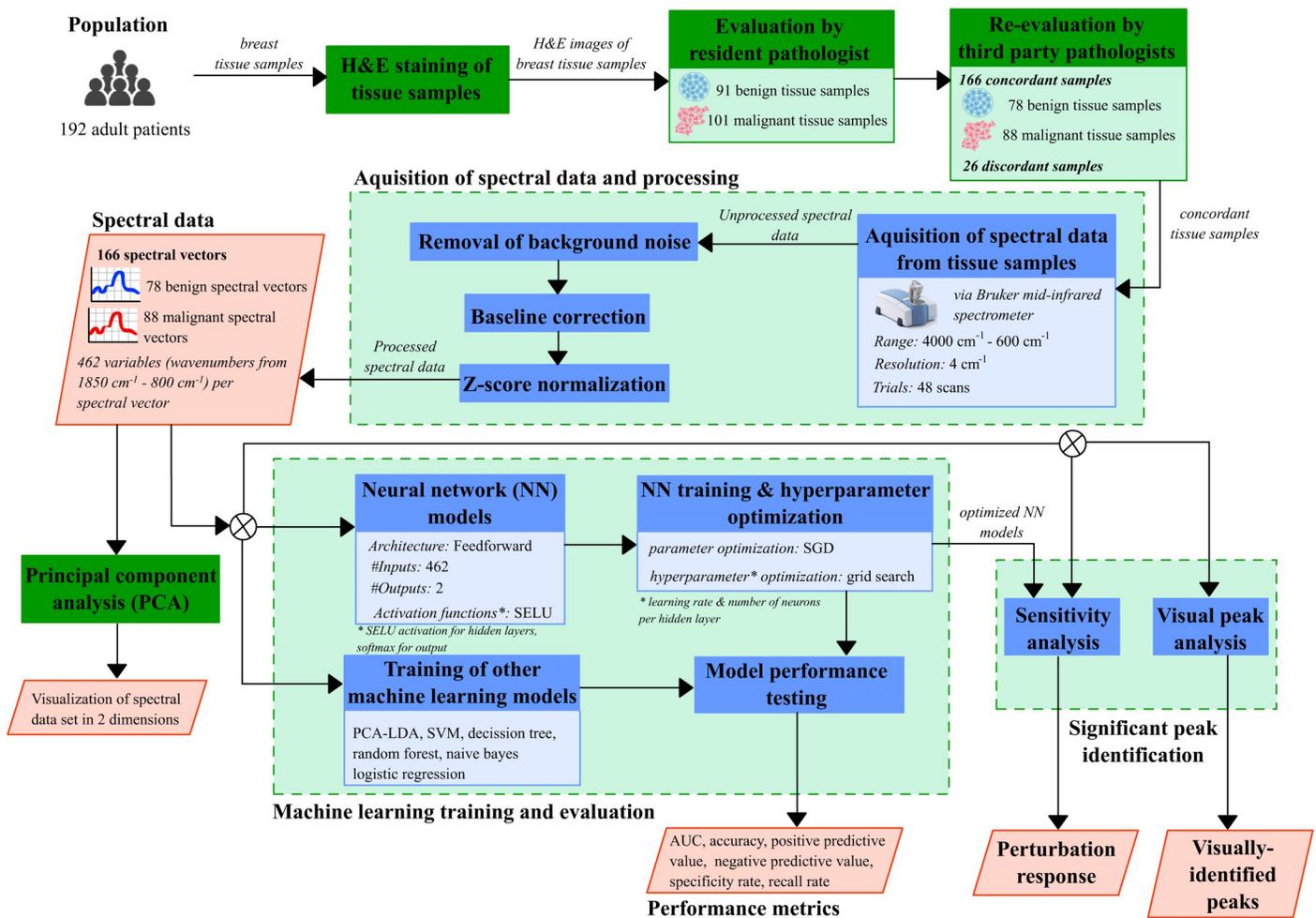
To visualize the perturbation response of the considered FNN,  $\overline{MSE_j}$  was plotted for all  $j \in \mathbb{N} \leq \mathcal{L}$ . The sensitivity analysis was performed for each FNN ( $n = 2, n = 4, n = 8$ ).

**Motivation and theory.** The diagnostic ability of often-used machine learning models such as SVM, LDA, and PCA greatly relies on the variation among and between data within a data set  $X$ . This variation is often quantified using a covariance matrix,  $S \in \mathbb{R}^{N^2}$ . By obtaining the eigenvectors associated with  $S$ , the characteristic *form* of the data may be easily represented. However, this approach becomes ineffective if the variation among the data set elements  $X_j \in X$  becomes significantly small (but not infinitesimally small to imply repetitive data). In such a scenario:  $S \rightarrow 0$ , the obtained eigenvector solutions,  $\vec{e}_j | j \in \mathbb{N} \leq N$ , become trivial where  $|\vec{e}_j| \rightarrow 0 \forall j \in \mathbb{N} \leq N$ . This makes it difficult to distinguish important variables which may prove significant in determining an accurate diagnosis. For artificial NNs, the weights determine the correlation of the input parameters toward the outputs under 0-bias condition. Here, the magnitude of a weight determines the magnitude of the correlation, while the *sign* of the weight determines the direction of influence [44]. This simplistic model does not however explain the contribution of biases and activation functions in a network’s decision-making process.

The proposed model of this study probes the significance of an input parameter  $\vec{x}_{SD}(j)$  based on the magnitude of change at the output for given ranges of input perturbations  $\Delta x_{j,k}$ . The magnitude of influence of an input variable is given by the  $MSE_{j,k}$ , which is a magnitude value at the range of 0 to 1 since the output are probabilities. In obtaining  $MSE_{j,k}$ ,  $X_{SD}^{(M)}$  and  $X_{SD}^{(B)}$  were assumed as a single set since analyzing them separately would not provide an overall determination of the significant variables considering both classification. Furthermore, since the spectral data set  $X_{SD}$  was normalized for each set element  $\vec{x}_{SD}^{(i)}$ ,  $MSE_{j,k}$  was expected to be less varied between samples for all  $\vec{x}_{SD}^{(i)} \in X_{SD}$  across all input variables. This assumption

makes it justifiable to denote the overall average,  $MSE_j$ , as the overall measure of an input response for a single neural network. The overall  $MSE$  magnitude,  $\overline{MSE}_j$ , was assumed as the final metric to measure the influence of a given variable, which is the average  $MSE_j$  measure considering multiple similar NNs. In such process, it was assumed that each NN had similar input responses since each followed the same architecture, training, and optimization. Lastly, in determining the most significant input variables,  $\overline{MSE}_j$  was no longer ranked in contrast to usual sensitivity analyses [20–22, 45, 46]. Since the functional groups and vibrational modes associated with the input variables are usually presented in ranges within the IR spectrum, it was more appropriate to identify and discuss significant peaks from the plot of  $\overline{MSE}_j$  rather than in a ranked form.

Overall, this study proposed that input variables associated with comparatively high  $MSE$  constitute to significant wavenumbers important in the NN's diagnosis. Since the NN models were assumed as the prevailing models and are highly accurate, the determined wavenumbers may thus provide insights in the associated changes in chemical composition and structure of biomolecules in cancerous breast tissues. The overall method implemented in the study is summarized in Fig 1.



**Fig 1. Experimental design process flowchart.** The figure shows the experimental design implemented in the study, from the acquisition of breast tissue samples, to the acquisition, processing and analysis of spectral data.

<https://doi.org/10.1371/journal.pone.0262489.g001>

## Results

### Samples

The clinical characteristics of the samples were retrieved from medical records and histopathology reports of the hospital study sites (Table 1). Among the malignant samples, majority were invasive ductal carcinoma. Meanwhile, the benign breast samples were mainly fibroadenoma and fibrocystic change (Table 1). The above classifications were based on microscopic examination of H&E-stained specimens and immunohistochemical staining (if needed or available) following the current WHO classification.

The variation between benign and malignant samples is shown in Fig 2. From the PCA plot, 90.28% of the variability was associated with the first principal component  $F_1$ , while only 5.12% was associated with the second principal component  $F_2$ . Most of the benign samples were scattered along the negative domain of the  $F_1$  axis while malignant samples were evenly scattered. Both sample classes followed a parabolic distribution across the determined principal axes. Overall, the PCA biplot suggests that the benign and malignant breast samples were highly similar in characteristics.

### Feedforward neural network designs

The NN input layer consisted of 462 nodes which corresponded to the defined IR absorbance of each sample in frequencies between  $1800\text{ cm}^{-1}$  to  $850\text{ cm}^{-1}$ . Three (3) FNN models were designed with varying layer sizes ( $n = 2$ ,  $n = 4$ ,  $n = 8$ ). The quantity of neurons per FNN hidden layer was kept constant for each repetition. The general NN architectures are summarized in Table 2.

Gaussian random initialization was assumed for weight initialization, while a zero-value initialization was used for the biases. Moreover, SELU activation was used for all neuron activation functions except for the respective output layers in which the softmax activation function was used. All NN were trained through backpropagation via AdaGrad stochastic gradient decent (SGD) [47] over 1000 epochs. The binary cross-entropy function was considered as the cost function for all NN designs during the training process. To avoid over-fitting, a dropout of 90% was used for each feed forward hidden layers as recommended for SELU activation [29, 37].

### Neural network optimization

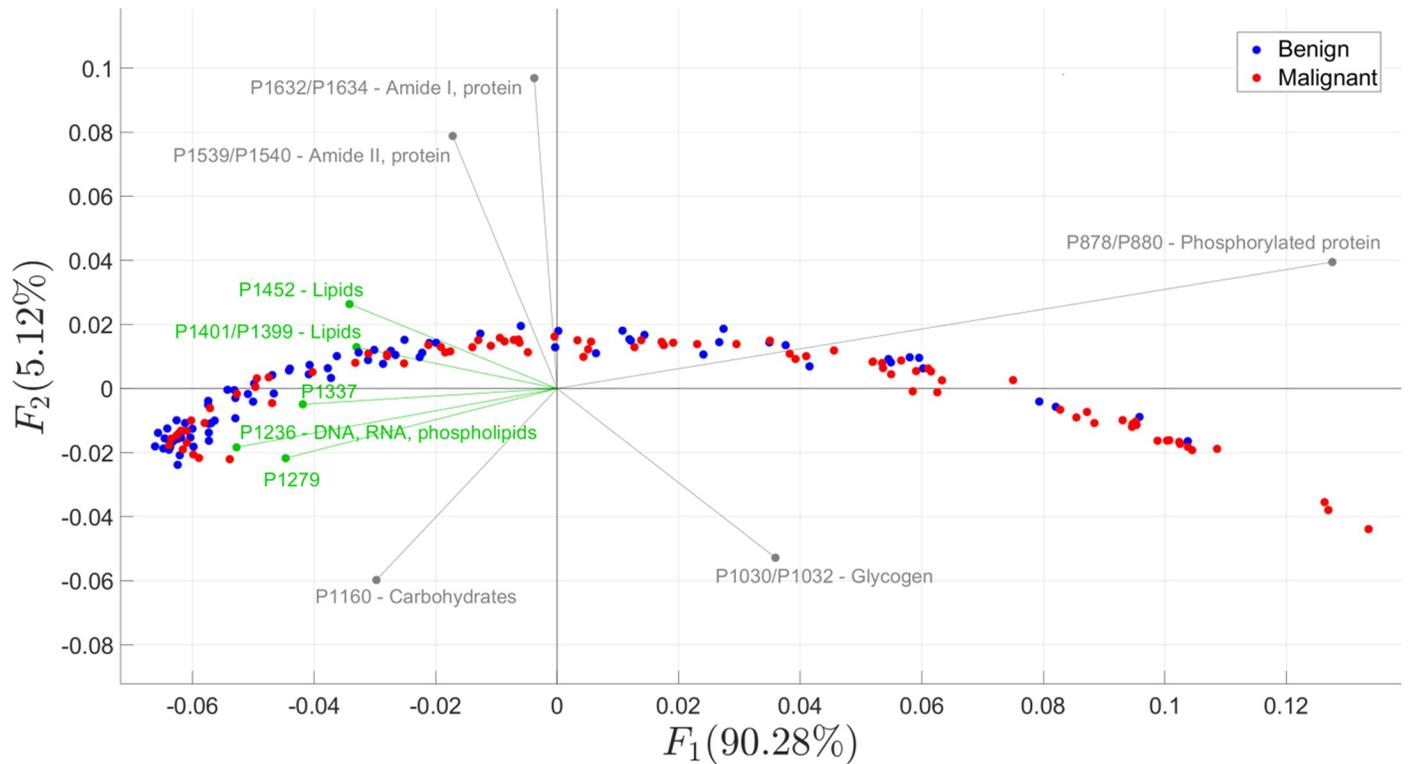
Pre-training, a grid search was performed to determine the optimal learning rate and the layer width. To limit the search space of the performed grid search, the explored learning rates  $L$ ,

**Table 1. Clinical data of the patients with breast lesions\*.**

Diagnosis	Classification*	Total no. of samples ( $n = 164$ )
Malignant	Invasive ductal carcinoma ( $n = 76$ )	88
	Others ( $n = 12$ )	
Benign	Fibroadenoma ( $n = 46$ )	78
	Fibrocystic change ( $n = 20$ )	
	Other ( $n = 12$ )	

\*Retrieved from hospital records and were based on microscopic evaluation of H&E-stained slides by resident pathologists of the hospital study sites.

<https://doi.org/10.1371/journal.pone.0262489.t001>



**Fig 2. PCA biplot showing data points of malignant and benign samples.** The red points denote malignant samples while blue points denote benign samples plotted across the two most dominant components ( $F_1 = 90.28\%$  and  $F_2 = 5.21\%$ ). The vectors show the wavenumbers associated with peak absorbance, where those highlighted in green were identified as significant wavenumbers in discriminating benign from malignant samples.

<https://doi.org/10.1371/journal.pone.0262489.g002>

and the considered layer width  $N_W$ , were limited to 10 and 20 elements, respectively, where

$$L \ni l_i = \frac{l_{i-1}}{k} | l_0 = 1; i \in \mathbb{N} \leq 9; k = \begin{cases} 2 \forall i \in \mathbb{N}_{Odd} \\ 5 \forall i \in \mathbb{N}_{Even} \end{cases}$$

**Table 2. Feedforward neural network architecture.**

	FNN2	FNN4	FNN8
<b>Overall architecture</b>	input layer– 462 neurons	input layer– 462 neurons	input layer– 462 neurons
	2 hidden FC layers –350 neurons	4 hidden FC layers– 400 neurons	8 hidden FC layers –300 neurons
	output layer– 2 neurons	output layer– 2 neurons	output layer– 2 neurons
<b>Learning rate</b>	0.01	0.01	0.01
<b>Network weight initialization</b>	Gaussian random: $\sigma = \sqrt{\frac{1}{input\ size}}$ , $\mu = 0$		
<b>Input normalization</b>	Z-score normalization: $\bar{x} := \frac{\bar{x} - mean(\bar{x})}{std(\bar{x})}$		
<b>Activation functions</b>	Input/Hidden layers–SELU		
	Output layer - softmax		
<b>Cost/fitness function</b>	Binary cross-entropy		
<b>Fully connected layer dropout</b>	90% SELU dropout		
<b>Training optimizer</b>	AdaGrad stochastic gradient descent (SGD)		
<b>Epoch</b>	1000		

and  $N_W = \{10, 15, 20, 25, 30, 40, 50, 60, 80, 100, 120, 140, 160, 180, 200, 250, 300, 350, 400, 462\}$ . For every combination of hyperparameters, each NN model was trained over 1000 epochs using the training set via a 10-fold cross-validation procedure over 50 trials. The best hyperparameter combination for each model was identified as the hyperparameter combination with the highest average validation accuracy.

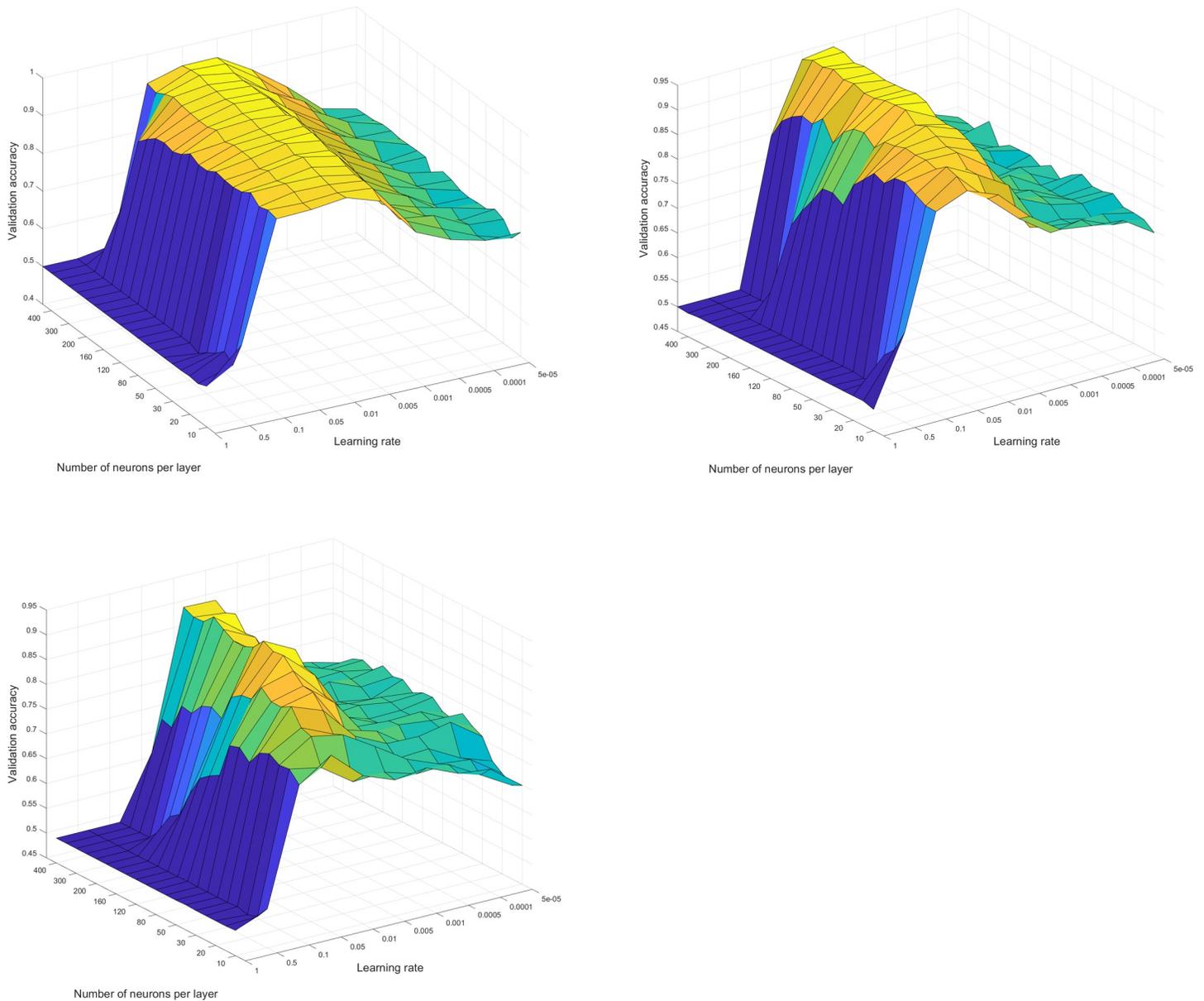
The result of the grid search for each NN design (Fig 3) shows that for all models, the designs peaked at a learning rate of about 0.005 to 0.1, with the FNN2 having the most stable deviation of performance across its range and the FNN8 as the most unstable. Regardless of the model, it can further be seen that at learning rates above 0.1, the models exhibited very poor performance in validation accuracy which may be due to divergence and large weight oscillations. Meanwhile, the validation accuracy stagnated at learning rates of below 0.005, which may be due to very small changes in the NN's weights and biases. The optimized learning rate for each model was identified to be all equal to 0.01, while the number of neurons per layer of each model to achieve best performance were 350, 400, and 300 for FNN2, FNN4, and FNN8, respectively.

### Diagnostic performance of models

Diagnostic performance of the NN models and the other machine learning models (NB, DT, RF, LR, LDA, SVM) was determined by comparison with the gold standard, which is the microscopic examination of H&E-stained tissues by pathologists. In terms of ACC, NPV and RR, all the NN models were able to significantly surpass all the other machine learning models (Tables 3 and 4). However, none of the models was able to outperform the SVM model as to AUC ( $95.72\% \pm 4.94\%$ ). Among the NN models, the best AUC was achieved by FNN4 ( $90.48\% \pm 10.30\%$ ) followed by FNN8 ( $90.35\% \pm 10.10\%$ ) then FNN2 ( $90.05\% \pm 9.95\%$ ). The relatively low AUC among NN models may be attributed to a lack of training time. In terms of PPV, the NN model performances did not significantly differ from the best benchmark model, which was SVM. As to NPV, the NN models were able to surpass the performance of the best benchmark model by an average of 4.42% for FNN2, 3.16% for FNN4 and 2.08% for FNN8. As to SR, the models performed significantly less than SVM, but were able to outperform the other benchmark models. The models were able to outperform SVM in terms of RR by an average of 2.94% for FNN2, 1.23% for FNN4 and 0.17% for FNN8. Note, however, that the RR value for FNN8 was not significantly differed from that of the SVM. The observed increase in metric performance from FNN8 to FNN2 may be due to a lack in training time for the deeper models. Overall, the significantly high ACC value of the FNN models may prove them viable classifiers in distinguishing malignant from benign samples using FTIR data. The results are summarized in Tables 3 and 4.

### Visual peak analysis of data

The fingerprint spectral region ( $1800\text{ cm}^{-1}$  to  $850\text{ cm}^{-1}$ ) showed that the absorbance spectra of the malignant and benign tissues were significantly distinct from each other, specifically at bands  $1452\text{ cm}^{-1}/1452\text{ cm}^{-1}$ ,  $1399\text{ cm}^{-1}/1401\text{ cm}^{-1}$ ,  $1337\text{ cm}^{-1}/1337\text{ cm}^{-1}$ ,  $1279\text{ cm}^{-1}/1279\text{ cm}^{-1}$ ,  $1236\text{ cm}^{-1}/1236\text{ cm}^{-1}$ , which represent lipids, DNA, RNA and phospholipids (Table 2). Other peaks tested, in particular  $1632\text{ cm}^{-1}/1634\text{ cm}^{-1}$ ,  $1539\text{ cm}^{-1}/1540\text{ cm}^{-1}$ ,  $1160\text{ cm}^{-1}/1160\text{ cm}^{-1}$ ,  $1032\text{ cm}^{-1}/1030\text{ cm}^{-1}$ , and  $880\text{ cm}^{-1}/878\text{ cm}^{-1}$ , were found as non-distinct among samples. These bands represent amide I proteins, amide II proteins, carbohydrates, glycogen, and phosphorylated proteins. It is worth mentioning that the tissues analyzed had uniform thickness ( $5\mu\text{m}$ ) achieved by sectioning with a microtome. The median absorbance spectra of benign and malignant breast tissue samples are shown in Fig 4.



**Fig 3. A. Grid search surface plot of FNN2.** The plot shows the grid search surface plot for optimizing the FNN2 model for learning rate and number of neurons per hidden layer. Validation accuracy peaked at a learning rate of 0.01 and 350 neurons per hidden layer. Low performance at high learning rates (blue region) may be due to divergence and high parameter oscillations, while stagnation of performance at low learning rates (green region) may be due to insufficient training time. **B. Grid search surface plot of FNN4.** The plot shows the grid search surface plot for optimizing the FNN4 model for learning rate and number of neurons per hidden layer. Validation accuracy peaked at a learning rate of 0.01 and 400 neurons per hidden layer. The same behavior in the low and high learning rate regions, observed in the FNN2 surface plot, is also evident here. **C. Grid search surface plot of FNN8.** The plot shows the grid search surface plot for optimizing the FNN8 model for learning rate and number of neurons per hidden layer. Validation accuracy peaked at a learning rate of 0.01 and 300 neurons per hidden layer. Among the FNN surface plots, the FNN8 constituted to the most unstable response in validation accuracy.

<https://doi.org/10.1371/journal.pone.0262489.g003>

Test of homogeneity showed that the characteristic IR absorbance peaks of the malignant cases were significantly different ( $p < 0.05$ ) from the benign samples. The visually identified peak positions and absorbances in the fingerprint IR region ( $1800 \text{ cm}^{-1}$  to  $850 \text{ cm}^{-1}$ ) that could significantly differentiate the malignant from benign samples are summarized in Table 5. Their corresponding functional group, vibrational mode, and molecular source assignments

Table 3. Diagnostic performance of the models.

Model	AUC (%)	Accuracy (%)	PPV (%)	NPV (%)	SR (%)	RR (%)
LDA	71.58±11.84	66.17±10.20	77.67±13.47	55.98±15.34	74.35±11.80	61.56±9.65
Logistic Regression	86.27±9.58	65.05±12.63	68.83±31.57	60.79±41.06	53.46±32.44	73.00±23.96
Naïve Bayes	66.90±10.90	64.69±10.41	54.44±16.24	77.40±15.09	60.23±9.62	74.22±15.28
Decision Trees	69.03±12.83	69.38±11.58	69.56±15.82	69.56±17.37	67.79±13.58	73.19±12.93
Random Forest	84.63±9.30	76.78±9.49	76.94±14.56	76.60±14.88	76.18±12.23	79.91±10.88
SVM	95.72±4.94	90.44±7.78	91.03±9.61	89.77±10.33	90.56±9.68	91.40±8.34
FNN2	90.05±9.95	96.06±7.07	89.83±12.57	94.19±10.57	85.56±18.33	94.34±10.54
FNN4	90.48±10.30	95.54±8.07	91.72±12.06	92.93±12.81	88.38±17.47	92.63±13.31
FNN8	90.35±10.10	95.46±7.55	92.18±11.88	91.85±12.67	89.04±16.75	91.58±13.47

**Abbreviations:** LDA—linear discriminant analysis; SVM—support vector machines, FNN—feed forward neural network; AUC—area under the curve; PPV—positive predictive value, NPV- negative predictive value; SR—specificity rate; RR—recall rate.

<https://doi.org/10.1371/journal.pone.0262489.t003>

are also listed in the aforementioned table. It was observed that peak absorbances representing lipids, DNA, RNA and phospholipids were significantly decreased in most malignant tissues.

The results of the performed test of significance is further backed-up in Fig 2, where wave-numbers associated with significant peak absorbances were more closely projected to  $F_1$  than those which were not. This implies that lipids, DNA, RNA and phospholipids are responsible for the high variability among the samples, suggesting further that these biomolecules highly varies between malignant and benign classes.

### Significant peaks identified by neural networks

The input response produced by each NN design is summarized in Fig 5. As shown, the response of each network is highly similar to one another. For all NN designs, the IR region associated with the functional groups, in particular within  $\sim 960\text{ cm}^{-1}$  to  $\sim 1050\text{ cm}^{-1}$ , showed the greatest input response within the considered spectrum. Meanwhile, the least response was evident within  $\sim 1250\text{ cm}^{-1}$  to  $\sim 1320\text{ cm}^{-1}$ .

Using the input response of NN from neural network committee-based sensitivity analysis, the significant peaks were identified from the fingerprint spectral region ( $1800\text{ cm}^{-1}$  to  $850\text{ cm}^{-1}$ ) as shown in Fig 5. The information on the protein content, including its secondary structures such as amide I and amide II, are observed in the region between  $1800\text{ cm}^{-1}$  to  $1500\text{ cm}^{-1}$  [48, 49]. The bands found at  $\sim 1635\text{ cm}^{-1}$  are associated with the amide I protein that arises from the C = O stretching vibrations of the amide groups of the protein backbone [49, 50]. The bands at  $1540\text{ cm}^{-1}$  results to N-H bending in the amide II groups, which is associated with aromatic

Table 4. Test of significance of neural network performance metrics relative to SVM.

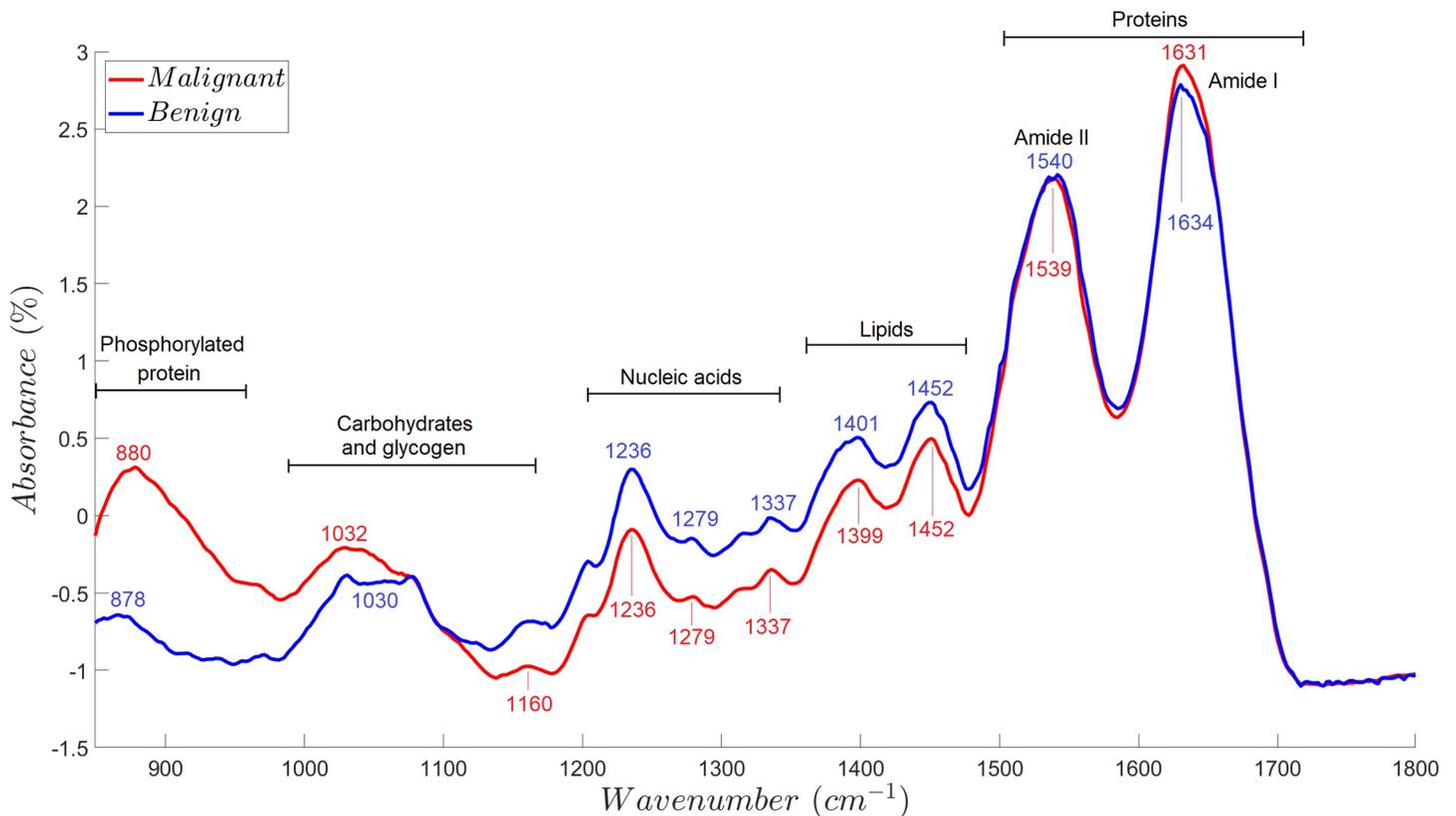
Model	Difference of average performance metric* (p-value**)					
	AUC (%)	ACC (%)	PPV (%)	NPV (%)	SR (%)	RR (%)
FNN2	-5.67 (<< 0.05)	5.62 (<< 0.05)	-1.20 (0.2137)	4.42 (<< 0.05)	-4.95 (0.0219)	2.94 (0.0006)
FNN4	-5.24 (<< 0.05)	5.09 (<< 0.05)	0.69 (0.5249)	3.16 (0.0001)	-2.18 (0.0161)	1.23 (0.0359)
FNN8	-5.37 (<< 0.05)	5.01 (<< 0.05)	1.15 (0.1042)	2.08 (0.0036)	-1.52 (0.1814)	0.17 (0.5832)

\*The differences of performance metrics were computed by subtracting the FNN performance metric to that of the SVM.

\*\* Mann-Whitney  $U$  test (two-tailed); significant when  $p < 0.05$ .

**Abbreviations:** : LDA—linear discriminant analysis; SVM—support vector machines, FNN—feed forward neural network; AUC—area under the curve; ACC—accuracy ; PPV—positive predictive value, NPV- negative predictive value; SR—specificity rate; RR—recall rate.

<https://doi.org/10.1371/journal.pone.0262489.t004>



**Fig 4. Median ATR-FTIR absorbance spectra of malignant ( $n = 88$ ) and benign ( $n = 78$ ) breast tissue samples.** The figure shows the median FTIR spectrum of malignant and benign breast tissue samples and their corresponding peaks identified via visual analysis. The plot shows almost similar absorbance among benign and malignant samples within wavenumbers associated with the amide proteins. Benign tissue samples, relative to malignant tissue samples, are shown to have increased absorbance within the region associated with lipids and nucleic acids while having decreased absorbance within the region associated with carbohydrates, glycogen, and phosphorylated proteins.

<https://doi.org/10.1371/journal.pone.0262489.g004>

amino acids. The bands  $1454\text{ cm}^{-1}$  and  $1393\text{ cm}^{-1}$  are associated with the  $\text{CH}_2$ ,  $\text{CH}_3$  deformation modes mainly from proteins and lipids [49, 51]. Peak position  $1393\text{ cm}^{-1}$  resulted from  $\text{COO}^-$  symmetric stretching of amino acids [49]. The region of  $1300\text{ cm}^{-1}$ – $800\text{ cm}^{-1}$  corresponds to the variations of functional groups that are present in proteins, nucleic acids, carbohydrates and phospholipids such as  $\text{PO}_2^-$ , C–O, and C–C [48, 52]. The band at  $1238\text{ cm}^{-1}$  is at the range of sugar-phosphate chain vibrations which is related to  $\text{PO}_2^-$  asymmetric stretching of nucleic acids [52]. Furthermore, this is also the range for Amide III ( $1299\text{ cm}^{-1}$ – $1200\text{ cm}^{-1}$ ) for its C–N stretching, N–H bending, C = O stretching, and O = C–N bending [53, 54]. The vibrations of C–O group from glycogen are observed in peak  $1077\text{ cm}^{-1}$  [51]. The band  $1030\text{ cm}^{-1}$ – $1045\text{ cm}^{-1}$  is due to C–O stretching and C–O bending of the C–OH groups of carbohydrates such as glucose, fructose [50, 54]. The band at  $990\text{ cm}^{-1}$  is related to phosphorylation of proteins and ribose-phosphate chain [50] while  $\sim 962\text{ cm}^{-1}$  is associated with symmetric phosphate stretching modes from phosphate diester groups in nucleic acids and phospholipids [51].

## Discussion

FTIR spectroscopy is a prospective novel diagnostic tool that is used to distinguish cancer samples from normal ones at high sensitivity, specificity, and accuracy [55, 56]. Considering the molecular complexity of biological specimens, chemometric techniques such as the

**Table 5. Comparison of the spectrum variables (peak positions and normalized absorbances) of malignant and benign breast samples in the fingerprint IR region (1800cm<sup>-1</sup> to 850cm<sup>-1</sup>) via visual peak identification.**

Malignant samples (n = 88)		Benign samples (n = 78)					p-value*
Peak position	Mean abs ±SD	Peak position	Mean abs ±SD	Functional Group	Vibrational Mode	Molecular Source[1–5]	
P1632	2.8311 ±0.2347	P1634	2.7934 ±0.1968	O = C-N-H	ν(CO), ν(CN)	Amide I, protein	0.0553
P1539	2.1152 ±0.2475	P1540	2.1533 ±0.1955	O = C-N-H	γ(N-H), ν(C-C), ν(C-N)	Amide II, protein	0.4146
<b>P1452</b>	<b>0.4797 ±0.2924</b>	<b>P1452</b>	<b>0.5920 ±0.2651</b>	-(CH <sub>2</sub> ) <sub>n</sub> , -(CH <sub>3</sub> ) <sub>n</sub>	δ <sub>as</sub> (CH <sub>3</sub> ), δ <sub>as</sub> (CH <sub>2</sub> ), δ <sub>s</sub> (CH <sub>3</sub> )	<b>Lipids</b>	<b>0.0031</b>
<b>P1399</b>	<b>0.2208 ±0.2822</b>	<b>P1401</b>	<b>0.3375 ±0.2481</b>	-(CH <sub>2</sub> ) <sub>n</sub>	δ <sub>s</sub> (CH <sub>3</sub> )	<b>Lipids</b>	<b>0.0026</b>
<b>P1337</b>	<b>-0.2794 ±0.3509</b>	<b>P1337</b>	<b>-0.1598 ±0.3269</b>				<b>0.0140</b>
<b>P1279</b>	<b>-0.4509 ±0.3683</b>	<b>P1279</b>	<b>-0.3397 ±0.3479</b>				<b>0.0231</b>
<b>P1236</b>	<b>-0.0416 ±0.4474</b>	<b>P1236</b>	<b>0.1160 ±0.4204</b>	RO-PO <sub>2</sub> -OR	ν <sub>as</sub> (PO <sub>2</sub> <sup>-</sup> )	<b>DNA, RNA, phospholipids</b>	<b>0.0129</b>
P1160	-0.8352 ±0.2566	P1160	-0.7759 ±0.2634	C-O-H	ν(CO), γ(COH)	Carbohydrates	0.0673
P1032	-0.1732 ±0.3355	P1030	-0.2386 ±0.2966	C-O-H	def(CHO)	Glycogen	0.1339
P880	0.1482 ±1.0361	P878	-0.0945 ±0.9590	C-O-P	ν(COP)	Phosphorylated protein	0.0773

\* Mann-Whitney *U* test (two-tailed); significant when  $p < 0.05$ .

\*\*values in bold refer to significantly higher peak absorbance ( $p < 0.05$ ).

**Abbreviations:** ν: stretching; δ: bending; γ: wagging, twisting and rocking; s: symmetric; as: asymmetric; def: deformation.

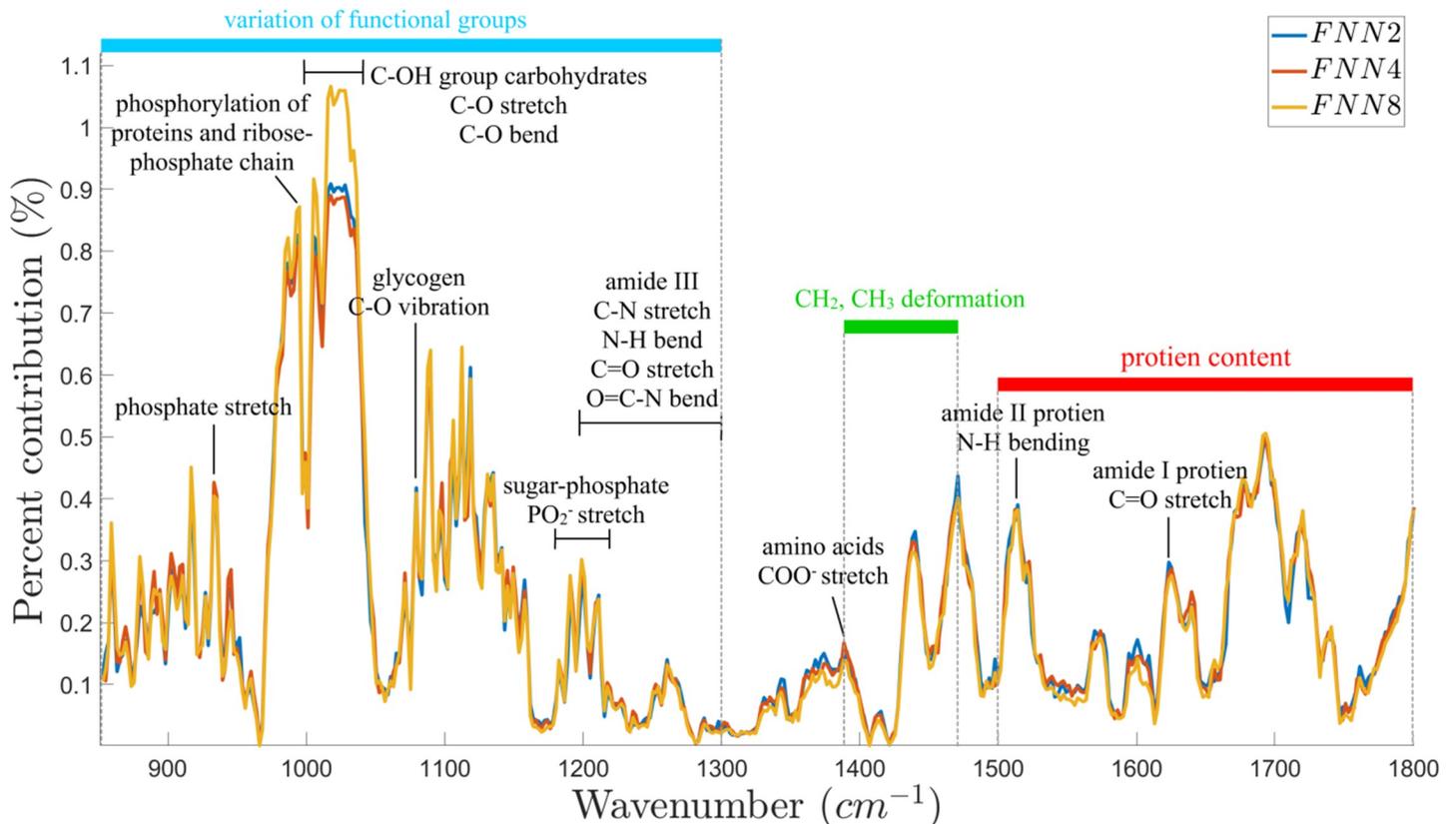
#### References:

- [1] Wu M, Zhang W, Tian P, et al. Intraoperative diagnosis of thyroid diseases by fourier transform infrared spectroscopy based on support vector machine. *Int J Clin Exp Med* 2016; 9: 2351–2358.
- [2] Dong L, Sun X, Chao Z, et al. Evaluation of FTIR spectroscopy as diagnostic tool for colorectal cancer using spectral analysis. *Spectrochim Acta—Part A Mol Biomol Spectrosc* 2014; 122: 288–294.
- [3] Simonova D, Karamancheva I. Application of Fourier transform infrared spectroscopy for tumor diagnosis. *Biotechnol Biotechnol Equip* 2013; 27: 4200–4207.
- [4] Lewis PD, Lewis KE, Ghosal R, et al. Evaluation of FTIR Spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC Cancer* 2010; 10: 640.
- [5] Zhang X, Xu Y, Zhang Y, et al. Intraoperative detection of thyroid carcinoma by fourier transform infrared spectrometry. *J Surg Res* 2011; 171: 650–656.

<https://doi.org/10.1371/journal.pone.0262489.t005>

principal component analysis (PCA) and artificial neural networks (ANN) that combine statistical and mathematical algorithms are utilized to generate chemo-physical evidence from spectral data [55].

The advent of computers with enhanced processing capabilities and enhanced memory capacity have led in the rise of computer-aided diagnosis (CAD), which combines algorithms or methods from pattern recognition and digital image processing [56]. Meanwhile, scientists have been drawn to the potential application of FTIR spectroscopy in the clinical setting to improve accuracy and reproducibility of cancer diagnosis, while omitting the need for complex and time-consuming clinical processing of tissue biopsy samples [55]. This is best exemplified by the study of Großerueschkamp, *et al.*, wherein they combined FTIR imaging and a novel trained random forest (RF) classifier for the automated marker-free histopathological annotation of lung tumor classes and subtypes of adenocarcinoma without further treatment of tissue samples. This study yielded greater reproducibility and accuracy of 97% for the annotation of lung tumor classes and 95% for the identification of prognostic adenocarcinoma subtypes [57]. Subsequently, FTIR reduced intra- and inter-operator variability through its objectivity, reproducibility, and improved accuracy over current methodologies for cancer diagnosis [55]. This also permits the standardization of spectral measurement and analysis, which is necessary for the construction of FTIR spectral databases with highly specific spectroscopic markers for the various stages and grades of different cancer types applicable to the clinical settings [58]. Additionally, an easy and objective data interpretation can be done by non-spectroscopists by incorporating powerful algorithms for automatic data analysis of large data sets [55].



**Fig 5. Input response of neural networks from the designed NN-based sensitivity analysis.** The line plots show the input response of each neural network design per change of absorbance value per wavenumbers. A high per cent contribution magnitude implies a high response to a change for the particular wavenumber, hence may serve as a marker in identifying malignant samples from benign samples. As evident from the figure, the response of each network is nearly the same, which only varies slightly in the magnitude contribution.

<https://doi.org/10.1371/journal.pone.0262489.g005>

The designed NNs exhibited superior accuracy (>90%) relative to the best benchmark model (SVM). These metrics prove them not only as excellent classifiers in distinguishing malignant from benign breast cancers using ATR-FTIR data, but also excellent classifiers in general. Overall, the FNN2 model was able to obtain larger metric values relative to the other NN models. The decrease in the performance metrics, in particular, accuracy, NPV, and RR of the models as a function of the layer quantity makes it evident that the deeper models may have lacked training time. Do note, however, that an opposite trend is evident for the SR and PPV metrics, implying that the designed architectures may approach a classifier that becomes increasingly more accurate in detecting malignant rather than benign samples as the model gets deeper. This observation presents a trade-off between the model's capability in confirming truly malignant from benign samples. This, further, implies that if a more accurate positive screening test is more in-demand, then deeper models may be assumed. Conversely, for more accurate negative screening tests, a less deep model may be more necessary.

Considering the metric comparison performed, the non-significant PPV metric of the designed NNs make them equally competent to SVM. However, the significantly higher NPV metric of the designed NN models make them more superior classifiers in terms of identifying benign samples as truly exhibiting non-malignancy. The significance of the SR and the RR are parallel to the significance of the PPV and NPV by definition, respectively. This characteristic makes the designed NN models more practical to use in situations where diagnosing non-

malignant patients as malignant becomes very costly. While administering an incorrect diagnosis is very detrimental for a patient in general, for financially non-capable individuals, an accurate diagnosis of non-malignancy may be of more importance since a false diagnosis of malignancy risks the individual of financial burden in chemotherapy, and a probable decline in health which further necessitates added costs. In developing countries such as the Philippines, the use of highly specific diagnostic tool such as the designed NNs in this study may prove more beneficial for patients undergoing cancer diagnosis. Regardless of the use, however, the models show their potential as highly specific tools to assist pathologists and medical practitioners in the field.

The designed neural networks that were used to analyze the FTIR spectra were able to identify significantly decreased peak absorbances characteristic of lipids, nucleic acids, and phospholipids in malignant tissues, which were similarly evident in the performed visual analysis (Table 5). Breast cancer is often characterized by the stimulated production of novo lipids which are essential for cell growth, proliferation, and oxidative stress resistance. The triacylglyceride storage in lipid droplets has been suggested to work as fuel source after re-oxygenation during intermittent hypoxia, whereas fatty acids promote redox balance supporting a high-glycolytic rate in malignant tissues. Lipids also form the structural basis of paracrine hormones and growth factors which stimulate tumor growth, neovascularization, invasion, and metastatic spread [59]. The decrease in lipids and phospholipids may reflect the utilization of these biomolecules for nutrition and energy source; and thus, prevent their accumulation in cancer cells during cancer progression [60]. A significant difference in the absorptive peak was also apparent in the DNA/RNA spectral region, with the malignant breast samples showing significantly lower peak absorbance than benign samples. This is in contrary to the findings of Lazaro-Pacheco *et al.* wherein higher contribution of nucleic acid bands was identified in cancerous samples in comparison with normal breast tissues in different spectral regions. They argued that high concentration of these biomolecules is expected since there is increased cellular content in response to an abnormal proliferation [61]. However, in a study involving ovarian cancer, the RNA/DNA absorption peaks were significantly lower in malignant tissues than in borderline and benign ovarian tissues [62]. The lowered peak absorbance among malignant samples may be due to fragment transfer of tumor DNA or cell-free RNA from the cancerous area to the bloodstream; thus, consequently decreasing the nucleic acid content at the primary tumor area [63].

Interestingly, there was no distinctive difference between malignant and benign breast tissues in the absorptive peaks of carbohydrates as perceived by the visual analysis performed (Table 5). In relation, the P1160 wavenumber vector that is associated with carbohydrates was projected significantly far from  $F_1$ , implying further that this biomolecule is not a possible cause of variability (Fig 2). However, the sensitivity analysis stated otherwise since the highest peak was evident across the IR region associated with the C–OH group of carbohydrates (Fig 5). Studies have shown that assessing glycogen levels is a good differentiation marker between malignant and benign tissues, with malignant samples generally consuming more glycogen to sustain survival during prolonged hypoxia and glucose deprivation as well as to sustain metastasis [64]. This ability of the sensitivity analysis to recognize the carbohydrates as differentiating factor, which in contrast was not detected by mere visual peak analysis, further proves the proposed method as a more discerning method to identify important spectral biomarkers. Possibly, the breast cancer cells could have already catabolized their glycogen stores as well as their subsequent by-products such as glucose for survival in nutrient-deprived environment [23]; hence, became relatively indistinguishable by the usual visual peak analysis. Given the proposed method's superior ability, the study suggests that the identified peaks within the higher IR wavelength region ( $\sim 1400\text{ cm}^{-1}$  to  $\sim 1800\text{ cm}^{-1}$ ) be given attention, particularly those associated with  $\text{CH}_2$ ,  $\text{CH}_3$  deformation modes and amide protein stretch and bends.

A variety of biological materials including blood, solid tissues, urine, and sputum have been studied using FTIR spectroscopy to develop better alternatives for cancer diagnosis and management. In the clinical setting, blood and tissue samples remain to be widely used as opposed to other specimens for diagnosing disease [55]. In less developed countries, the use of FFPE samples can provide technical ease and economic advantage for longitudinal tissue specimen storage as they can be easily retrievable from accredited repositories for further analysis [65]. Compared to immunohistochemical and molecular assays, FTIR can be a cheaper alternative for detecting biochemical markers in pathologic FFPE specimens based on unique vibrational patterns [60]. With the introduction of machine learning, FTIR spectroscopy in clinical diagnostic settings can reduce intra- and inter-operator variability and improve accuracy and reproducibility of cancer diagnosis, while omitting the need for complex and time-consuming clinical processing of clinical samples [56].

The generation of NN models from the FTIR fingerprint of benign and malignant FFPE breast tissues led to the identification of significant wavenumbers apart from those at peak absorbances, which can be used to discriminate malignant from benign tissues. Interestingly, unique peak absorbances distinctive of lipids, nucleic acids, and phospholipids were identified, showing that these biomolecules were significantly decreased in malignant tissues as compared to benign samples, and can, therefore, be used as biochemical fingerprints to aid in cancer diagnosis.

While the current study shows that NN models from FTIR spectra can be used as an adjunct tool for diagnosing breast cancer, additional clinical studies should be made to bring this technology into the clinical setting. Due to financial constraint, this study was conducted using only the basic type of FTIR spectrometer with limited spatial resolution. To acquire a comprehensive spectral data, additional FFPE samples may be analyzed using an FTIR coupled with an infrared microscope to detect vibrational motions of molecules within very restricted regions. Other spectroscopic techniques such as the Raman spectroscopy can also be used to further probe molecular vibrations to aid in the characterization and discrimination of tissue types [66]. The creation of spectral database and the generation of novel powerful algorithms for automatic data analysis of large data sets is another prospect to accelerate point-of-care decisions and improve therapeutic management for breast cancer patients [67]. Further studies also show that an alternative sample to tissues could be blood plasma, since the use of plasma is cheaper, less invasive, and easier to process [68, 69]. Through the integration of AI and FTIR, spectral biomarkers in plasma samples may be identified to monitor treatment response; a study which is already being investigated by the research team.

In summary, the present study generated NN models that led to the identification of unique infrared spectrum of absorption in the lipid, nucleic acid, phospholipid, and carbohydrates regions that could effectively discriminate malignant from benign breast tissues. To the researchers' knowledge, this is the first study to have used several machine learning tools to identify malignant breast tissues based on FTIR spectral data.

## Acknowledgments

We thank Mr. Patrick Jun Paul Lawan and Ms. Ericka Hidalgo for their technical assistance, and Ralph Christian Tomas for his scientific input.

## Author Contributions

**Conceptualization:** Rock Christian Tomas, Ruth Bangaoil, Abegail Santillan, Pia Marie Albano.

**Data curation:** Anthony Jay Sayat, Andrea Nicole Atienza, Jannah Lianne Danganan, Ma. Rollene Ramos, Allan Fellizar.

**Formal analysis:** Rock Christian Tomas, Anthony Jay Sayat, Andrea Nicole Atienza, Jannah Lianne Danganan, Ma. Rollene Ramos, Lara Mae Angeles.

**Funding acquisition:** Pia Marie Albano.

**Investigation:** Anthony Jay Sayat, Andrea Nicole Atienza, Jannah Lianne Danganan, Ma. Rollene Ramos, Allan Fellizar, Lara Mae Angeles.

**Methodology:** Rock Christian Tomas, Anthony Jay Sayat, Andrea Nicole Atienza, Jannah Lianne Danganan, Ma. Rollene Ramos, Kin Israel Notarte, Lara Mae Angeles, Ruth Bangaol, Abegail Santillan, Pia Marie Albano.

**Project administration:** Pia Marie Albano.

**Resources:** Allan Fellizar.

**Software:** Rock Christian Tomas.

**Supervision:** Ruth Bangaol, Abegail Santillan, Pia Marie Albano.

**Validation:** Rock Christian Tomas, Lara Mae Angeles.

**Writing – original draft:** Rock Christian Tomas, Anthony Jay Sayat, Andrea Nicole Atienza, Jannah Lianne Danganan, Ma. Rollene Ramos, Allan Fellizar, Kin Israel Notarte, Lara Mae Angeles, Ruth Bangaol, Abegail Santillan, Pia Marie Albano.

**Writing – review & editing:** Rock Christian Tomas, Kin Israel Notarte, Pia Marie Albano.

## References

1. Mainiero MB, Lourenco A, Mahoney MC, Newell MS, Bailey L, Barke LD, et al. ACR Appropriateness Criteria Breast Cancer Screening. *J Am Coll Radiol* [Internet]. 2013; 10:11–4. Available from: <https://doi.org/10.1016/j.jacr.2016.09.021> PMID: 23290667
2. Brem RF, Lenihan MJ, Lieberman J, Torrente J. American women having dense breast tissue. *AJR Am J Roentgenol*. 2015; 204(2):234–40. <https://doi.org/10.2214/AJR.13.12072> PMID: 25615743
3. Mathenge EG, Dean CA, Clements D, Vaghar-kashani A, Giacomantonio M, Malueth B, et al. Core Needle Biopsy of Breast Cancer Tumors Increases Distant Metastases in a Mouse Model. *Neoplasia* [Internet]. 2014; 16(11):950–60. Available from: <https://doi.org/10.1016/j.neo.2014.09.004> PMID: 25425969
4. Kabel AM. Tumor markers of breast cancer: New prospectives. *J Oncol Sci* [Internet]. 2017; 3(1):5–11. Available from: <https://doi.org/10.1016/j.jons.2017.01.001>
5. Bunaciu AA, Aboul-Enein HY, Fleschin S. FTIR Spectrophotometric Methods Used for Antioxidant Activity Assay in Medicinal Plants. *Appl Spectrosc Rev* [Internet]. 2012; 47(4):245–55. Available from: <http://www.tandfonline.com/doi/abs/10.1080/05704928.2011.645260>
6. Yang D, Castro DJ, el-Sayed IH, el-Sayed MA, Saxton RE, Zhang NY. A Fourier-transform infrared spectroscopic comparison of cultured human fibroblast and fibrosarcoma cells: a new method for detection of malignancies. *J Clin Laser Med Surg* [Internet]. 1995; 13(2):55–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10150573> <https://doi.org/10.1089/clm.1995.13.55> PMID: 10150573
7. Siqueira LFS, Lima KMG. A decade (2004–2014) of FTIR prostate cancer spectroscopy studies: An overview of recent advancements. *Trends Anal Chem* [Internet]. 2016; 82:208–21. Available from: <https://doi.org/10.1016/j.trac.2016.05.028>
8. Bhalerao RY, Jani HP, Gaitonde RK, Raut V. A novel approach for detection of Lung Cancer using Digital Image Processing and Convolution Neural Networks. 2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019. 2019;577–83.
9. Rossetto AM, Zhou W. Deep Learning for Categorization of Lung Cancer CT Images. Proceedings—2017 IEEE 2nd International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE 2017. 2017;272–3.

10. Kido S, Hirano Y, Hashimoto N. Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). 2018 International Workshop on Advanced Image Technology, IWAIT 2018. 2018;1–4.
11. Santillan A, Tomas RC, Bangaoil R, Lopez R, Gomez MH, Fellizar A, et al. Discrimination of malignant from benign thyroid lesions through neural networks using FTIR signals obtained from tissues. *Anal Bioanal Chem*. 2021; 413(8):2163–80. <https://doi.org/10.1007/s00216-021-03183-0> PMID: 33569645
12. Kaur B, Mann KS, Grewal MK. Ovarian cancer stage based detection on convolutional neural network. Proceedings of the 2nd International Conference on Communication and Electronics Systems, ICCES 2017. 2018;2018-Janua(Icces):855–9.
13. Rahman MA, Muniyandi RC, Islam KT, Rahman MM. Ovarian Cancer Classification Accuracy Analysis Using 15-Neuron Artificial Neural Networks Model. 2019 IEEE Student Conference on Research and Development, SCORed 2019. 2019;33–8.
14. Zou L, Yu S, Meng T, Zhang Z, Liang X, Xie Y. A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis. *Computational and Mathematical Methods in Medicine*. 2019; 2019(Dm). <https://doi.org/10.1155/2019/6509357> PMID: 31019547
15. Zuluaga-Gomez J, Masry Z Al, Benagoune K, Meraghni S, Zerhouni N. A CNN-based methodology for breast cancer diagnosis using thermal images. 2019;0–2.
16. Ragab DA, Sharkas M, Marshall S, Ren J. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*. 2019; 2019(1):1–23. <https://doi.org/10.7717/peerj.6201> PMID: 30713814
17. Mishra A, Cheng H. *Advanced CNN Architectures*. 2017.
18. Vert JP. Artificial intelligence and cancer genomics. *Healthcare and Artificial Intelligence*. 2020; 1(February):165–74.
19. Barth A. Infrared spectroscopy of proteins. *Biochim Biophys Acta—Bioenerg*. 2007; 1767(9):1073–101. <https://doi.org/10.1016/j.bbabi.2007.06.004> PMID: 17692815
20. Zeng X, Yeung DS. A quantified sensitivity measure for multilayer perceptron to input perturbation. *Neural Computation*. 2003; 15(1):183–212. <https://doi.org/10.1162/08997660321043757> PMID: 12590825
21. Cao M, Qiao P. Neural network committee-based sensitivity analysis strategy for geotechnical engineering problems. *Neural Computing and Applications*. 2008; 17(5–6):509–19.
22. Sysoev A, Ciurlia A, Sheglevatykh R, Blyumin S. Sensitivity analysis of neural network models: Applying methods of analysis of finite fluctuations. *Periodica polytechnica Electrical engineering and computer science*. 2019; 63(4):306–11.
23. Bangaoil R, Santillan A, Angeles LM, Abanilla L, Lim A, Ramos MC, et al. ATR-FTIR spectroscopy as adjunct method to the microscopic examination of hematoxylin and eosin-stained tissues in diagnosing lung cancer. *PLoS One [Internet]*. 2020; 15(5):e0233626. Available from: <https://doi.org/10.1371/journal.pone.0233626> PMID: 32469931
24. Podshyvalov A, Sahu RK, Mark S, Kantarovich K, Guterman H, Goldstein J, et al. Distinction of cervical cancer biopsies by use of infrared microspectroscopy and probabilistic neural networks. *Appl Opt [Internet]*. 2005; 44(18):3725. Available from: <https://www.osapublishing.org/abstract.cfm?URI=ao-44-18-3725> <https://doi.org/10.1364/ao.44.003725> PMID: 15989047
25. Salman A, Shufan E, Sahu RK, Mordechai S, Sebbag G. Insights on colorectal cancer relapse by infrared microscopy from anastomosis tissues: Further analysis. *Vib Spectrosc*. 2016; 83:17–25.
26. Zhang X, Xu Y, Zhang Y, Wang L, Hou C, Zhou X, et al. Intraoperative detection of thyroid carcinoma by fourier transform infrared spectrometry. *J of Surg Res*. 2011; 171(2):650–6. <https://doi.org/10.1016/j.jss.2010.05.031> PMID: 20828740
27. Wu M, Zhang W, Tian P, Ling X, Xu Z. Intraoperative diagnosis of thyroid diseases by fourier transform infrared spectroscopy based on support vector machine. *Int J Clin Exp Med*. 2016; 9(2):2351–8.
28. Bhosale JS. High signal-to-noise Fourier transform spectroscopy with light emitting diode sources. *Rev Sci Instrum*. 2011; 82(9). <https://doi.org/10.1063/1.3633944> PMID: 21974569
29. Klambauer G, Unterthiner T, Mayr A, Hochreiter S. Self-normalizing neural networks. In: 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017.
30. Costa F, Marques A, Arnaud-Fassetta G, Alonso J, Martins I, Guerra C. Self-Normalizing Neural Networks Günter. *European Continental Hydrosystems under Changing Water Policy*. 2013;(Nips):99–112.
31. Bogomolny E, Argov S, Mordechai S, Huleihel M. Monitoring of viral cancer progression using FTIR microscopy: A comparative study of intact cells and tissues. *Biochim Biophys Acta*. 2008; 1780(9):1038–46. <https://doi.org/10.1016/j.bbagen.2008.05.008> PMID: 18588944

32. Ghimire H, Venkataramani M, Bian Z, Liu Y, Perera AGU. ATR-FTIR spectral discrimination between normal and tumorous mouse models of lymphoma and melanoma from serum samples. *Sci Rep*. 2017; 7(1). <https://doi.org/10.1038/s41598-017-17027-4> PMID: 29209060
33. Liu H, Su Q, Sheng D, Zheng W, Wang X. Comparison of red blood cells from gastric cancer patients and healthy persons using FTIR spectroscopy. *J Mol Struct*. 2017; 1130:33–7.
34. Wang X, Shen X, Sheng D, Chen X, Liu X. FTIR spectroscopic comparison of serum from lung cancer patients and healthy persons. *Spectrochim Acta—Part A Mol Biomol Spectrosc* [Internet]. 2014; 122:193–7. Available from: <https://doi.org/10.1016/j.saa.2013.11.049> PMID: 24316532
35. Lewis PD, Lewis KE, Ghosal R, Bayliss S, Lloyd AJ, Wills J, et al. Evaluation of FTIR Spectroscopy as a diagnostic tool for lung cancer using sputum. *BMC Cancer* [Internet]. 2010; 10(1):640. Available from: <http://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-10-640> <https://doi.org/10.1186/1471-2407-10-640> PMID: 21092279
36. Philipp M, Rusch T, Hornik K, Strobl C. Measuring the Stability of Results From Supervised Statistical Learning. *Journal of Computational and Graphical Statistics*. 2018; 27(4):685–700.
37. Costa F, Marques A, Arnaud-Fassetta G, Alonso J, Martins I, Guerra C. Self-Normalizing Neural Networks Günter. In: 31st Conf Neural Inf Process Syst (NIPS). 2017. p. 99–112.
38. Gu Q, Li Z, Han J. Linear discriminant dimensionality reduction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2011; 6911 LNAI(PART 1):549–64.
39. Cristianini N, Schölkopf B. Support vector machines and kernel methods: The new generation of learning machines. *AI Magazine*. 2002; 23(3):31–41.
40. Murty MN, Raghava R. Linear support vector machines. *SpringerBriefs in Computer Science*. 2016; (9783319410623):41–56.
41. Platt JC. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. 2000;(June):1–11.
42. Gehrke J. Classification and Regression Trees. *Encyclopedia of Data Warehousing and Mining*. 2011;246–80.
43. Reza M, Miri S, Javidan R. A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset. *International Journal of Advanced Computer Science and Applications*. 2016; 7(6):1–33.
44. Guha R, Stanton DT, Jurs PC. Interpreting computational neural network quantitative structure-activity relationship models: A detailed interpretation of the weights and biases. *Journal of Chemical Information and Modeling*. 2005; 45(4):1109–21. <https://doi.org/10.1021/ci050110v> PMID: 16045306
45. Montavon G, Samek W, Müller KR. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing: A Review Journal*. 2018; 73:1–15.
46. Zhang Z, Beck MW, Winkler DA, Huang B, Sibanda W, Goyal H. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine*. 2018; 6(11):216–216. <https://doi.org/10.21037/atm.2018.05.32> PMID: 30023379
47. Duchi J, Hazan E, Singer Y. Randomized smoothing for (parallel) stochastic optimization. *J Mach Learn Res*. 2011; 12:2121–59.
48. Sitnikova VE, Kotkova MA, Nosenko TN, Kotkova TN, Martynova DM, Uspenskaya M V. Breast cancer detection by ATR-FTIR spectroscopy of blood serum and multivariate data-analysis. *Talanta*. 2020; 214 (October 2019):120857. <https://doi.org/10.1016/j.talanta.2020.120857> PMID: 32278436
49. Kar S, Katti DR, Katti KS. Fourier transform infrared spectroscopy based spectral biomarkers of metastasized breast cancer progression. *Spectrochim Acta Part A Mol Biomol Spectrosc*. 2019; 208:85–96. <https://doi.org/10.1016/j.saa.2018.09.052> PMID: 30292907
50. Blat A, Wiercigroch E, Smeda M, Wislocka A, Chlopicki S, Malek K. Fourier transform infrared spectroscopic signature of blood plasma in the progression of breast cancer with simultaneous metastasis to lungs. *Journal of Biophotonics*. 2019; 12(10):1–11. <https://doi.org/10.1002/jbio.201900067> PMID: 31265171
51. Depciuch J, Stanek-Widera A, Skrzypiec D, Lange D, Biskup-Frużyńska M, Kiper K, et al. Spectroscopic identification of benign (follicular adenoma) and cancerous lesions (follicular thyroid carcinoma) in thyroid tissues. *J Pharm Biomed Anal*. 2019; 170:321–6. <https://doi.org/10.1016/j.jpba.2019.03.061> PMID: 30954022
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* [Internet]. 2015; 15(1):408. Available from: <http://bmccancer.biomedcentral.com/articles/10.1186/s12885-015-1414-7> <https://doi.org/10.1186/s12885-015-1414-7> PMID: 25975566
53. Elishemey WM, Ismail AM, Elbially NS. Molecular-Level Characterization of Normal, Benign, and Malignant Breast Tissues Using FTIR Spectroscopy. *Journal of Medical and Biological Engineering*. 2016; 3(3):369–78.

54. 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. <https://doi.org/10.1155/2020/4343590> PMID: [32104176](https://pubmed.ncbi.nlm.nih.gov/32104176/)
55. Su KY, Lee WL. Fourier transform infrared spectroscopy as a cancer screening and diagnostic tool: A review and prospects. *Cancers*. 2020; 12(115):1–19. <https://doi.org/10.3390/cancers12010115> PMID: [31906324](https://pubmed.ncbi.nlm.nih.gov/31906324/)
56. Jothi AA, Rajam MA. A survey on automated cancer diagnosis from histopathology images. *Artificial Intelligence Review*. 2017; 48(1):31–81.
57. Großerueschkamp F, Kallenbach-Thieltges A, Behrens T, Brüning T, Altmayer M, Stamatis G, et al. Marker-free automated histopathological annotation of lung tumour subtypes by FTIR imaging. *Analyst*. 2015; 140(7):2114–20. <https://doi.org/10.1039/c4an01978d> PMID: [25529256](https://pubmed.ncbi.nlm.nih.gov/25529256/)
58. Baker M, Hussain Shawn, Lovergne L, Untereiner V, Hughes C, Lukaszewski R, et al. Developing and understanding biofluid vibrational spectroscopy: A critical review. *Chemical Society Reviews*. 2016; 45(7):1803–18. <https://doi.org/10.1039/c5cs00585j> PMID: [26612430](https://pubmed.ncbi.nlm.nih.gov/26612430/)
59. Santos CR, Schulze A. Lipid metabolism in cancer. *FEBS Journal*. 2012; 279(15):2610–23. <https://doi.org/10.1111/j.1742-4658.2012.08644.x> PMID: [22621751](https://pubmed.ncbi.nlm.nih.gov/22621751/)
60. Bénard A, Desmedt C, Durbecq V, Rouas G, Larsimont D, Sotiriou C, et al. Discrimination between healthy and tumor tissues on formalin-fixed paraffin-embedded breast cancer samples using IR imaging. *Journal of Spectroscopy*. 2010; 24:67–72.
61. Lazaro-Pacheco D, Shaaban A, Baldwin G, Titiloye N, Rehman S, Rehman I. Deciphering the structural and chemical composition of breast cancer using FTIR spectroscopy. *Applied Spectroscopy Reviews*. 2014; 6:29–32.
62. Li L, Wu J, Yang L, Wang H, Xu Y, Shen K. Fourier transform infrared spectroscopy: An innovative method for the diagnosis of ovarian cancer. *Current Management and Research*. 2021; 13:2389–99.
63. Zaporozhchenko IA, Ponomaryova AA, Rykova EY, Laktionov PP. The potential of circulating cell-free RNA as a cancer biomarker: challenges and opportunities. *Expert Review of Molecular Diagnostics*. 2018; 18(2):133–45. <https://doi.org/10.1080/14737159.2018.1425143> PMID: [29307231](https://pubmed.ncbi.nlm.nih.gov/29307231/)
64. Zois CE, Harris AL. Glycogen metabolism has a key role in the cancer microenvironment and provides new targets for cancer therapy. *J Mol Med*. 2016; 94(2):137–54. <https://doi.org/10.1007/s00109-015-1377-9> PMID: [26882899](https://pubmed.ncbi.nlm.nih.gov/26882899/)
65. Zhang P, Lehmann BD, Shyr Y, Guo Y. The Utilization of Formalin Fixed-Paraffin-Embedded Specimens in High Throughput Genomic Studies. *Int J Genomics*. 2017; Article ID. <https://doi.org/10.1155/2017/1926304> PMID: [28246590](https://pubmed.ncbi.nlm.nih.gov/28246590/)
66. Auner GW, Koya K, Huang C, Broadbent B, Trexler M, Auner Z, et al. Applications of Raman spectroscopy in cancer diagnosis. *Cancer and Metastasis Reviews*. 2018; 37(4):691–717. <https://doi.org/10.1007/s10555-018-9770-9> PMID: [30569241](https://pubmed.ncbi.nlm.nih.gov/30569241/)
67. Hughes C, Baker MJ. Can mid-infrared biomedical spectroscopy of cells, fluids and tissue aid improvements in cancer survival? A patient paradigm. *Analyst*. 2016; 141(2):467–75. <https://doi.org/10.1039/c5an01858g> PMID: [26501136](https://pubmed.ncbi.nlm.nih.gov/26501136/)
68. Meany DL, Sokoll LJ, Chan DW. Early detection of cancer: Immunoassays for plasma tumor markers. *Expert Opinion on Medical Diagnostics*. 2009; 3(6):597–605. <https://doi.org/10.1517/17530050903266830> PMID: [19966928](https://pubmed.ncbi.nlm.nih.gov/19966928/)
69. Park J, Shin Y, Kim TH, Kim DH, Lee A. Plasma metabolites as possible biomarkers for diagnosis of breast cancer. *PLoS ONE*. 2019; 14(12):1–12. <https://doi.org/10.1371/journal.pone.0225129> PMID: [31794572](https://pubmed.ncbi.nlm.nih.gov/31794572/)