ATTENUATED TOTAL REFLECTANCE FOURIER TRANSFORM INFRARED (ATR-FTIR) SPECTROSCOPY FOR DETECTION OF BREAST CANCER.

Panchala Dhanushi Walpita [1], Supipi Amanda Manage[1], Dr. A. A. P. Keerthi[1, *], Dr. P. R. Haputhanthri[2], Dr. Mahendra Perera[3]

College of Chemical Sciences, Institute of Chemistry, Ceylon, Rajagiriya[1], The Open University of Sri Lanka, Nawala, Nugegoda [2], Asiri Surgical Hospital, Colombo 05[3]

ABSTRACT

Breast cancer is a significant health condition affecting millions of women worldwide. Early detection of breast cancer is crucial for its successful treatment and cure. In this study, we developed a non-invasive, rapid, inexpensive, and accurate method for detecting potential breast cancer patients using Attenuated Total Reflection Fourier Transform Infrared (ATR-FTIR) spectroscopy coupled with multivariate analysis. The Infrared (IR) spectrum of blood serum has a diagnostically significant range of 2800-3000 cm^{-1} for breast cancer, which corresponds to the C-H asymmetric stretching vibration of CH_3 and CH_2 of lipid acyl chains. Application of Principal Component Analysis (PCA) and Partial Least Squares Regression (PLS-R) from the multivariate analysis shows the change in intensity of the bands between the control group and the patient group. Using PLS-R, we constructed a predictive linear regression model within the above wavenumber range of the obtained IR spectra to quantify lipids to correlate to breast cancer. The model found linear regression graphs of spectral data that were most predictive of breast cancer status, which could be used to predict the probability of breast cancer. The best-fitting PLS-R models are evaluated according to their determination coefficient (R^2) and the Root Mean Square Error of Calibration (RMSEC). After performing PLS-R analysis, it generated a mathematical model with an RMSEC value of 0.5778 and an R^2 value of 0.7546, proving that this approach may offer a less discomfiting and time-consuming method for screening breast cancer, which may encourage more women to undergo this lifesaving screening test.

Keywords: Breast Cancer, Blood serum, ATR-FTIR spectroscopy, Principal Component Analysis, Partial Least Squares Regression

** Corresponding author: aapkeerthi1976@yahoo.com*

ATTENUATED TOTAL REFLECTANCE FOURIER TRANSFORM INFRARED (ATR-FTIR) SPECTROSCOPY FOR DETECTION OF BREAST CANCER.

Panchala Dhanushi Walpita [1], Supipi Amanda Manage[1], Dr. A. A. P. Keerthi[1, *], Dr. P. R. Haputhanthri[2], Dr. Mahendra Perera[3]

College of Chemical Sciences, Institute of Chemistry, Ceylon, Rajagiriya[1], The Open University of Sri Lanka, Nawala, Nugegoda [2], Asiri Surgical Hospital, Colombo 05[3]

INTRODUCTION

Breast cancer affects millions of women worldwide. In 2020, 2.3 million women were diagnosed with breast cancer and 685,000 died globally (World Health Organization, 2023). Early detection of breast cancer is crucial for its successful treatment and cure. Mammography is the widely used procedure for the detection of breast cancer which uses X-rays. Exposure to X-rays carries potential health risks, including an increased likelihood of developing cancer later in life. The process of obtaining a mammogram may be an uncomfortable and painful experience for most women, prompting most to not get them (Ashkar & Zaki, 2017).

We report a method for rapidly detecting and quantifying breast cancer using Attenuated Total Reflectance Fourier Transform Infrared spectroscopy (ATR-FTIR) combined with multivariate analysis, namely Principal Component Analysis (PCA), and Partial Least-Squares Regression (PLS-R). This is a non-invasive, inexpensive, rapid, and accurate method for detecting breast cancer patients.

ATR-FTIR is an excellent vibrational spectroscopic technique for the analysis of biofluids (e.g., serum) due to its rapidity and ease of translation to the clinical environment, that is, ATR-FTIR spectroscopy requires minimal or no sample preparation when analyzing blood (Sharma et al., 2023).

Multivariate analysis is a statistical study of data in which multiple measurements are taken on each experimental unit and the relationships and structure of the multivariate measurements are calculated. PCA reduces dimensionality by reducing the number of variables in use so that the new set of reduced variables represents most of the total variance in the existing set of variables. The PLS-R method reduces the number of predictor variables to a smaller set. The predictors are then used in a regression analysis. Some applications differentiate between PLS 1 and PLS 2. When there is only one dependent variable, PLS 1 is appropriate (Yang et al., 2021).

The wavenumber range of 2800 to 3000 cm^{-1} of the Infrared (IR) spectrum of a blood serum sample corresponds to the C-H asymmetric stretching vibration of CH_3 and CH_2 of lipid acyl chains. This region shows a significant change in band intensities that can correlate with breast cancer.

METHODOLOGY

Blood samples were collected from 20 clinically diagnosed breast cancer patients and 20 healthy volunteers. The collected blood samples were centrifuged to obtain the serum samples. These serum samples (5 µL) were spotted on a Polyethylene Terephthalate (PET) microscopic slide and dried for 40 minutes at room temperature. Then the inverted slide was placed on the ATR diamond and spectra were acquired using The ABB MB3000 spectrometer (ABB Ltd., Switzerland).

After acquiring IR spectra for control and patient samples, we applied PCA and PLS-R to identify the spectral bands that differed in intensity between the control group and the patient groups. Using PLS-R, we constructed a predictive linear regression model within the 2800-3000 cm^{-1} range of the obtained IR spectra to quantify the lipid amount that can correlate with breast cancer (Perera et al., 2021). In order to quantify the lipid amount we ran a lipid profile for each participant as the reference method.

RESULTS AND DISCUSSION

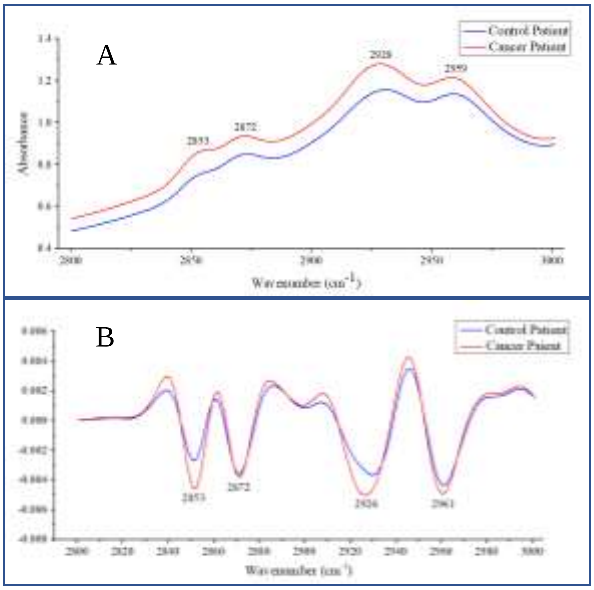

Figure 1:(A) Averaged ATR-FTIR absorbance and (B) calculated second derivative (2D) spectra of serum for the lipid (2800- 3000 cm-1) region (red and blue colour represents cancer and control patients respectively).

Figure 1 depicts the average IR spectra of serum samples obtained from breast cancer and control patients after undergoing baseline correction, standard normal variate (SNV), and Savitzky-Golay (SG) smoothing with a 9-point filter and the calculated 2D spectra of serum. The symmetric C-H stretching is at 2853 and 2872 cm^{-1}, and the asymmetric C-H stretching is at 2926 and 2961 cm^{-1}. An increase in the absorbance of a spectral band is proportional to an increase in the concentration of the corresponding functional group. Thus, the observed increase in the absorbance of lipid molecules in the serum samples of breast cancer patients compared to control samples suggests a higher concentration of lipid molecules in the serum of patients with breast cancer.

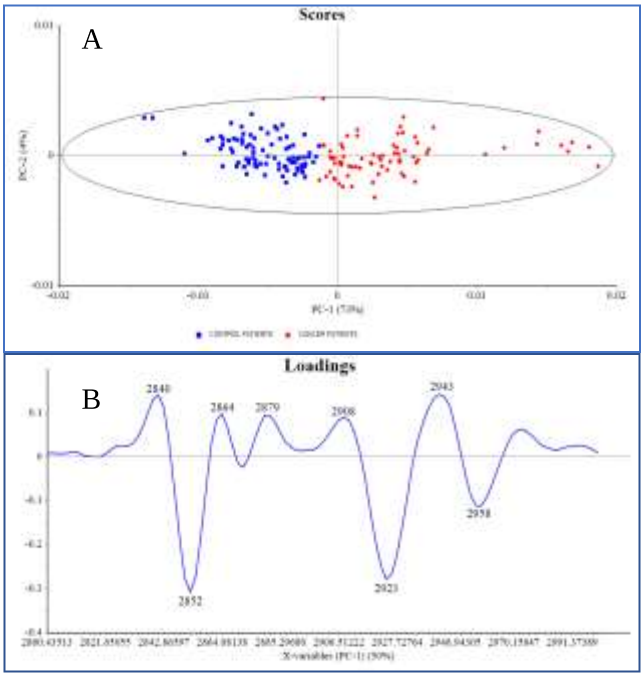

Figure 2: Calculated PCA scores plots of control and patient samples, second derivative spectra in the lipid regions, and their (B) PC1 loadings.

Figure 2 displays the scores obtained from the PCA conducted on the second derivative spectra. The PCA serves as a superior method for visualizing samples that are represented by numerous variables by projecting their original coordinates onto a new set of axes called principal components (PCs). These axes possess several properties that simplify the visualization of samples. The PCA algorithm generates score points for each spectrum along the first and second principal components (PC1 and PC2). Notably, the score plot in Figure 2 demonstrates clustering into groups based on the score point of each sample. It was observed that this clustering is based on the control sample and the patient.

The loading plot explains the bands that are responsible for clustering. With PCA clustering, we will be able to detect the presence of cancer by grouping patterns of the spectra of a patient once the model is developed. The Scores and Loadings in PCA are related such that negative Loadings correspond to positive Scores, and positive Loadings correspond to negative Scores.

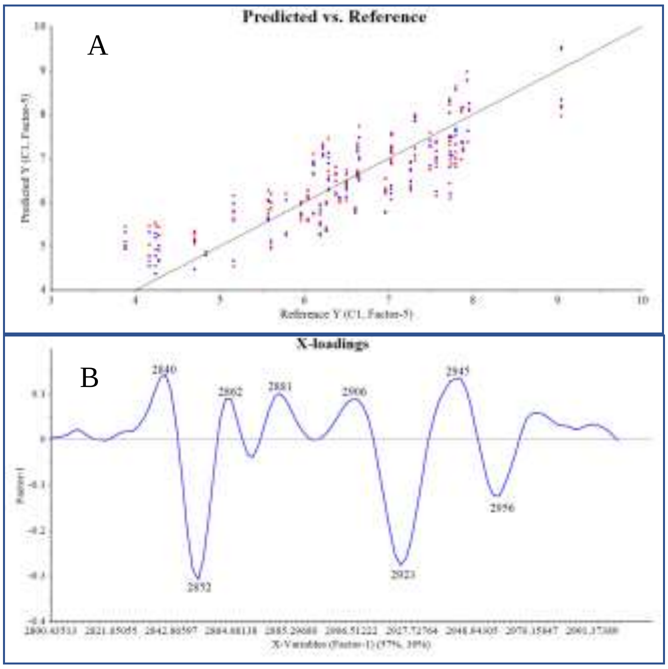

Figure 3: Partial least squares result in the lipid region (A) Regression plot for calibration and validation set. (B) Predicted vs. reference.

Figure 3 illustrates the regression plots for both the calibration and validation sets, accompanied by the predicted vs. reference plots. In this study, PLS-R was utilized as a multivariate method to develop a linear regression model that used the concentration of serum lipids as the predicted variable and the observable variable (IR spectra) projected onto a new multidimensional space. The model demonstrated a Root Mean Square Error of Calibration (RMSEC) value of 0.5778 and a determination coefficient (R^2) value of 0.7546. These findings indicate that the integration of ATR-FTIR spectroscopy with chemometrics can effectively differentiate between the serum samples of breast cancer patients and healthy individuals.

CONCLUSIONS

In this study, ATR-FTIR spectroscopy combined with chemometrics was utilized to investigate the serum of breast cancer patients and healthy individuals. The spectral band intensity comparison revealed that the concentration of lipids in the serum of breast cancer patients was significantly higher than that of healthy individuals. Using this finding, by carrying out the PCA analysis for the IR spectra of both cancer and control patients for the wavenumber range of 2800 to 3000 cm^{-1}, which represents the lipid region of IR spectra of human blood serum, we can confirm the presence or absence of a breast cancer within a patient. Whereas PLSR analysis of the IR spectra allowed us to conclude the severity of the breast cancer. Notably, our study is the first attempt in Sri Lanka to apply ATR-FTIR spectroscopy coupled with multivariate analysis as a screening method for detecting breast cancer. While other studies have investigated the use of PCA analysis and ATR-FTIR spectroscopy for detecting breast cancer globally, our study stands out for incorporating the PLS-R chemometric technique as part of the multivariate analysis. We hope that the findings of this study will be a vital pivotal point in developing a method of breast cancer diagnosis that will be much more affordable and accessible to the majority of the country's population.

REFERENCES

1. World Health Organization: WHO & World Health Organization: WHO. (2023). *Breast cancer*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/breast-cancer#:~:text=Scope%20of%20the%20problem&text=As%20of%20the%20end%20of,increasing%20rates%20in%20later%20life.
2. Ashkar, L. K., Zaki, Y. H. (2017). Female patients' perception of pain caused by mammography in the Western Region of Saudi Arabia. *Saudi Medical Journal*, 38(7), 768–771. https://doi.org/10.15537/smj.2017.7.19298
3. Sharma, C. P., Sharma, S., Singh, R. (2023). Species discrimination from blood traces using ATR FT-IR spectroscopy and chemometrics: Application in wildlife forensics. *Forensic Science International: Animals and Environments*, 3, 100060. https://doi.org/10.1016/j.fsiae.2022.100060.
4. Yang, X., Ou, Q., Qian, K., Yang, J., Bai, Z., Yang, W., Shi, Y., Liu, G. (2021). Diagnosis of lung cancer by ATR-FTIR spectroscopy and chemometrics. *Frontiers in Oncology*, 11, 753791. https://doi.org/10.3389/fonc.2021.753791.
5. Perera, K., Weragoda, G. K., Haputhanthri, R., Rodrigo, S. (2021). Study of concentration dependent curcumin interaction with serum biomolecules using ATR-FTIR spectroscopy combined with Principal Component Analysis (PCA) and Partial Least Square Regression (PLS-R). *Vibrational Spectroscopy*, 116, 103288. https://doi.org/10.1016/j.vibspec.2021.103288.

ACKNOWLEDGMENTS

The authors wish to acknowledge the College of Chemical Sciences, Institute of Chemistry Ceylon for the funding provided. They would also like to thank all patients who donated their serum as without these samples this research would not have been possible.