MULTI-OMICS PROFILING OF PLASMA FOR BREAST CANCER DISCRIMINATION BY ATR-FTIR SPECTROSCOPY AND MACHINE LEARNING

Izabella Cristina Costa Ferreira[1]; Alinne Tatiane Faria Silva[1]; Lara Andrade Marques[1]; Mário Machado Martins[1]; Juliana Carvalho Penha Pereira[1]; Donizeti William Santos[1]; Paula Philbert Lajolo[1]; Carlos Eduardo Paiva[2]; Marcelo De Almeida Maia[1]; Yara Cristina de Paiva Maia[1]

[1] Universidade Federal de Uberlândia, Uberlândia, MG.
[2] Hospital de Câncer de Barretos, Barretos, SP.
E-mail do autor correspondente: yara.maia@ufu.br

INTRODUCTION

Early and accurate detection of breast cancer (BC) remains a clinical challenge, driving the search for sensitive and non-invasive diagnostic strategies.

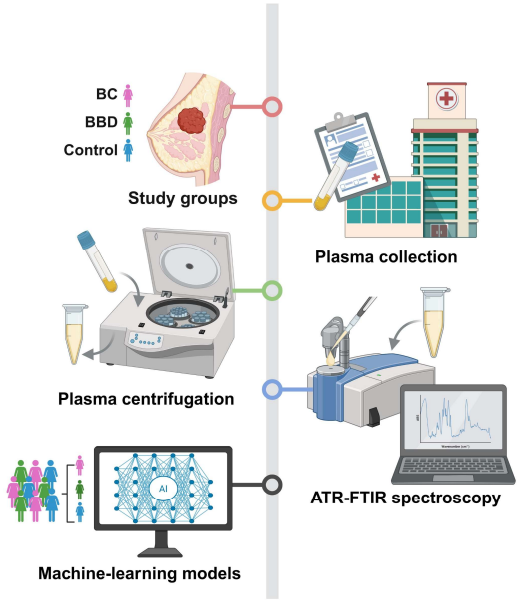
ATR-FTIR spectroscopy + Machine learning (ML)

promising multi-omics tool for analyzing biofluids, which provides a minimally invasive source of tumor biomarkers

OBJECTIVES

To evaluate the diagnostic potential of ATR-FTIR + ML algorithms in distinguishing healthy controls and benign breast diseases (BBD) from early-stage BC patients using plasma.

METHODS



Created with BioRender.com.

RESULTS

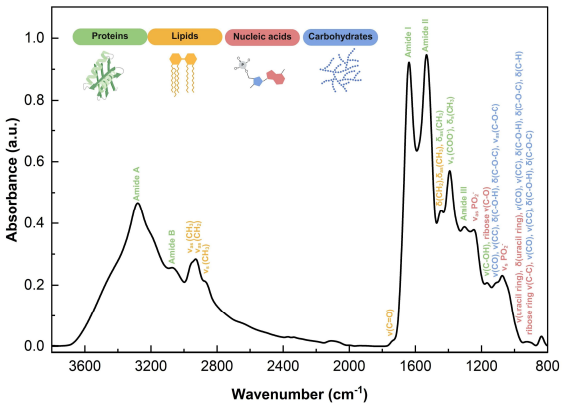


Figure 1. Representative ATR-FTIR spectrum of plasma from the BC group. Tentative of biomolecular peak assignments, where ν = stretching vibrations, δ = bending vibrations, s = symmetric vibrations and as = asymmetric vibrations. Created with BioRender.com.

Table 1. Machine learning performance using ATR-FTIR spectra of plasma for differentiating the study groups. The table shows the spectral regions for the best algorithm and the relative metrics. The best performance is in bold. BC= Breast cancer; BBD= Benign breast disease.

Region	Algorithm	ACC	SEN	SPE	AUC
CONTROL vs. BC					
3600-3000 cm^{-1}	AdaBoost	0.8	0.833	0.769	0.801
3000-2800 cm^{-1}	Naive Bayes	0.88	0.917	0.846	0.859
1800-800 cm^{-1}	AdaBoost	0.88	0.833	0.923	0.878
3000-2800 + 1800-800 cm^{-1}	Naive Bayes	0.88	0.917	0.846	0.897
1700-1480 cm^{-1}	Gradient Boosting	**0.92**	**0.917**	**0.923**	**0.846**
1200-800 cm^{-1}	Naive Bayes	0.8	0.833	0.769	0.811
BBD vs. BC					
3600-3000 cm^{-1}	Gradient Boosting	0.778	0.75	0.8	0.767
3000-2800 cm^{-1}	Gradient Boosting	0.667	0.667	0.667	0.644
1800-800 cm^{-1}	Gradient Boosting	0.741	0.75	0.733	0.728
3000-2800 + 1800-800 cm^{-1}	AdaBoost	0.778	0.75	0.8	0.775
1700-1480 cm^{-1}	AdaBoost	0.741	0.75	0.733	0.742
1200-800 cm^{-1}	AdaBoost	**0.889**	**0.917**	**0.867**	**0.892**
CONTROL vs. BBD					
3600-3000 cm^{-1}	Neural Network	0.821	0.8	0.846	0.892
3000-2800 cm^{-1}	Gradient Boosting	0.821	0.8	0.846	0.785
1800-800 cm^{-1}	Naive Bayes	0.893	0.867	0.923	0.928
3000-2800 + 1800-800 cm^{-1}	AdaBoost	**0.964**	**1**	**0.923**	**0.962**
1700-1480 cm^{-1}	Neural Network	0.929	0.867	1	0.938
1200-800 cm^{-1}	Neural Network	0.893	0.867	0.923	0.969

CONCLUSION

This study reinforces the potential of plasma as an informative source for liquid biopsy, which may contribute to earlier BC detection, improve clinical decision-making and reduce unnecessary invasive procedures in both benign and malignant breast alterations.

ACKNOWLEDGMENTS

FAPEMIG · CNPq · HOSPITAL DE AMOR · CAPES