



12TH INTERNATIONAL CONFERENCE ON CLINICAL SPECTROSCOPY

Breast Cancer Subtypes Diagnostic via High Performance Supervised Machine Learning

Sajid Farooq¹, Matheus del-Valle¹, Moises Oliveira dos Santos^{1,2}, Sofia Nascimento³, Emerson Soares Bernardes³, Denise Maria Zzell^{1,*}

¹Instituto de Pesquisas Energéticas e Nucleares, IPEN-CNEN, Av. Prof. Lineu Prestes, 2242, CEP: 05508-000, São Paulo-SP, Brasil

²Escola Superior de Tecnologia, Universidade do Estado do Amazonas, 69050-030, Brasil

Aim: Breast cancer molecular subtypes are being used to improve clinical decision. The Fourier transform infrared (FTIR) spectroscopic imaging, which is a powerful and non-destructive technique, allows performing a non-perturbative and labelling free extraction of biochemical information towards diagnosis and evaluation for cell functionality. However, methods of measurements of large areas of cells demand a long time to achieve high quality images, making its clinical use impractical because of speed of data acquisition and dearth of optimized computational procedures. In order to cope with these challenges, Machine learning (ML) technologies can facilitate to obtain accurate prognosis of Breast Cancer (BC) subtypes with high action ability and accuracy.

Methods: Here we propose a ML algorithm based method to distinguish computationally BC cell lines. The method is developed by coupling K neighbors Classifier (KNN) with Neighborhood Component Analysis (NCA) and NCA-KNN methods enables to identify BC subtypes without increasing model size as well additional parameters.

Results: By incorporating FTIR imaging data, we show that using NCA-KNN method, the classification accuracies, specificities and sensitivities improve up to 97%, even at very low co-added scan (S₄). Moreover, a clear distinctive accuracy difference of our proposed method was obtained in comparison with other ML supervised models.

Conclusion: For confirming our model results performance, the cross validation (k fold = 10) and receiver operation characteristics (ROC) curve were used and found in great agreement, suggest a potential diagnostic method for BC subtypes, even with small co-added scan < 8 at low spectral resolution (4 cm⁻¹).