Evaluation of machine learning models for the classification of breast cancer hormone receptors using micro-FTIR images

Matheus del Valle
Center for Lasers and Applications
Nuclear and Energy Research Institute (IPEN-CNEN/SP)
São Paulo, Brazil
matheus.valle@usp.br

Moisés Oliveira dos Santos
Center for Lasers and Applications
Nuclear and Energy Research Institute (IPEN-CNEN/SP) and Technology School (EST) Amazonas State University (UEA)
Amazonas, Brazil
mosantos@uea.edu.br

Sofia Nascimento dos Santos
Radiopharmacy Center
Nuclear and Energy Research Institute (IPEN-CNEN/SP) and Technology
São Paulo, Brazil
snsantos85@gmail.com

Emerson Soares Bernardes
Radiopharmacy Center
Nuclear and Energy Research Institute (IPEN-CNEN/SP) and Technology
São Paulo, Brazil
emerson.bernardes@gmail.com

Denise Maria Zezell
Center for Lasers and Applications
Nuclear and Energy Research Institute (IPEN-CNEN/SP)
São Paulo, Brazil
zezell@usp.br

2021 SBFoton International Optics and Photonics Conference (SBFoton IOPC) | 978-1-6654-1948-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/SBFOTONIOPC50774.2021.9461946

***Abstract*—The breast cancer is the most incident cancer in women. Evaluation of hormone receptors expression plays an important role to outline treatment strategies. FTIR spectroscopy imaging may be employed as an additional technique, providing extra information to help physicians. In this work, estrogen and progesterone receptors expression were evaluated using tumors biopsies from human cell lines inoculated in mice. FTIR images were collect from histological sections, and six machine learning models were applied and assessed. Xtreme gradient boost and Linear Discriminant Analysis presented the best accuracies results, indicating to be potential models for breast cancer classification tasks.**

Keywords— FTIR image, breast cancer; hormone receptor; machine learning; classification

I. INTRODUCTION

Cancer is a group of diseases where abnormal cells grow uncontrollably, go beyond their usual boundaries to invade adjoining parts of the body and/or spread to other organs [1]. The breast cancer is the most incident cancer in women with 24.2%, or 2.1 million, of new cases in the same year, aside from a 6.6% mortality. The estimate for 2040 is 29.4 million new cases of cancer, where 2.8 million will be related to the breast cancer [2]. In Brazil, the estimate for 2020 is 66 thousand of new breast cancer cases, which correspond to 29.7% of the total new cancer cases in women [3].

Breast cancer classification can follow different parameters as molecular subtypes, type, stage, and grade. The molecular classification is performed by the evaluation of three hormone receptors expression: estrogen receptor (ER), progesterone receptor (PR), and Human epidermal growth factor receptor 2 (Her2). The assessing of these expressions plays an important role to outline treatment strategies[4].

The gold-standard method for classification is the immunohistochemistry, where semiquantitative analysis are employed, adding the operator subjectivity [5]. Fourier Transform Infrared (FTIR) spectroscopy images has demonstrated remarkable results when applied machine learning analysis [6]. In this way, this work aims to compare different machine learning models to classify two breast cancer hormone receptors expression.

II. MATERIALS AND METHODS

A. Sample Preparation

Two human breast cell lines were inoculated in Balb/c nude mice: the BT474, which is positive for the three receptors (ER, PR, and Her2), and the SKBR3 cell line, that is also positive HER2 expression, but negative for ER and PR [4].

Tumors of approximately 0.5 cm^3 were biopsied and preserved using the formalin fixation and paraffin embedding technique. Ten sections of 5 µm were prepared from each cell line, resulting in twenty sections, and placed in low-e microscope slides (MirrIR, Kevley Technologies).

B. FTIR images acquisition

The FTIR spectroscopy system used was a Cary Series 600 (Agilent Technologies, EUA), containing a Cary 660 FTIR spectrometer and a Cary 620 FTIR microscope. Hyperspectral images were acquired by the focal plane array (FPA) detector of 32x32 elements. Each element provided a spatial resolution of 5.5 µm, thus collecting a 176x176 µm image per sample.

The maximum wavenumber window allowed by the equipment of 3950 to 900 cm^{-1} was used. Spectral resolution was set to 4 cm^{-1}. As the samples were fixed in low-e slides, the system was set to operate in transflection mode. Background spectra were acquired using 256 co-added scans, while 64 scans were used for sample spectra.

C. Analysis

Spectra preprocessing was performed by following the protocol [7]:

- Outlier removal: Hotelling's T^2 versus Q residuals method.
- Fingerprint selection: spectra region cut from 1800 to 900 cm^{-1}.
- Smoothing: Savitzky–Golay filter applied with window size of 7 and polynomial order of 2.

Funding agencies: FAPESP (CEPID 05/51689-2, 17/50332-0), CAPES/PROCAD (88881.068505/2014-01), CAPES (Finance Code 001) and CNPq (INCT-465763/2014-6, PQ-309902/2017-7, 142229/2019-9).

- Second derivative: obtained by the same filter smoothing algorithm.
- Normalization: extended multiplicative signal correction (EMSC).
- Removal of substrate contribution: wax contribution modeled in the same EMSC algorithm.

Six machine learning models were implemented: Linear Discriminant Analysis (LDA); Partial Least Squares Discriminant Analysis (PLS-DA); K-Nearest Neighbors (KNN); Support Vector Machine (SVM); Random Forest (RF); Xtreme Gradient Boost (XGB).

The models were trained using a stratified cross-validation of 5 folds, hence using 80% of the data as training and 20% for testing. The cross-validation was performed 10 times, resulting in a total of 50 trainings, varying the random stated of the train-test indexes.

Statistical comparison was performed using the accuracies scores with Friedman and Nemenyi test [8]. Algorithms for all the analysis were developed by the authors using Python and R languages.

III. RESULTS

The Fig. 1 shows the test accuracies boxplot by each model. The XGB model presented the highest accuracies, with a mean of 0.995. LDA, PLS, and RF exhibit values close to XGB, while SVM and KNN models demonstrated worse results.

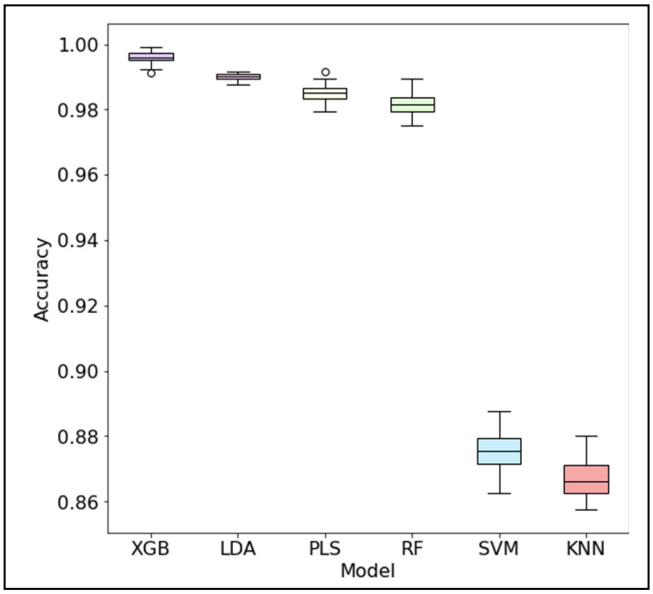

Fig. 1. Boxplot of the test accuracies presented by each model.

Friedman test presented significant statistical difference between the models. Therefore, Nemenyi statistical comparison was performed, where the results are shown in Fig 2. XGB and LDA critical values difference was lower than the critical distance of 1.066, resulting in no significant statistical difference between the best models. PLS did not presented difference from LDA and RF, but was different from XGB. SVM and KNN were the worst models, similar between each other, but with difference from the others.

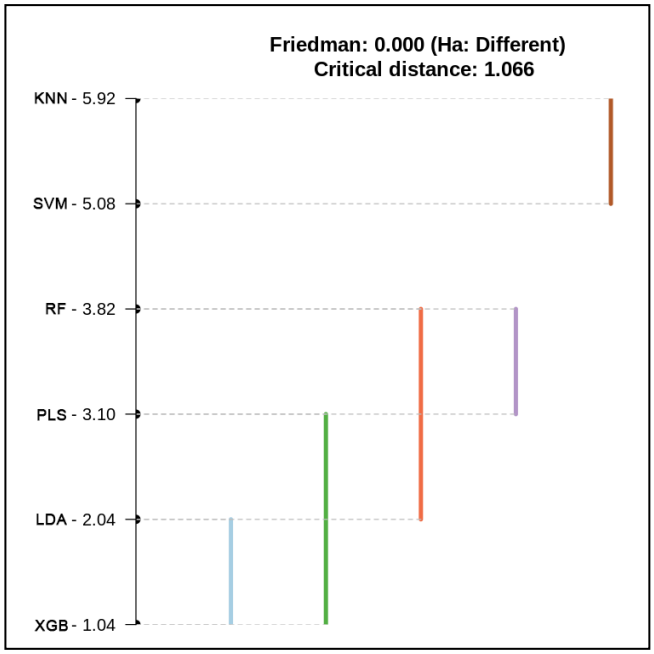

Fig. 2. Friedmand + Nemenyi test.

IV. DISCUSSION

The XGB algorithm applies decision trees with gradient boosting optimization [9]. In this way, each individual tree learns from the previous one, instead of lots of independent tree as in the RF model, improving its overall accuracy. In addition, regularization terms in the XGB model assist to prevent overfitting, decreasing the standard deviation (SD) of the folds' accuracy. Tree-based models exhibits good performance in datasets with a large number of features, as in this work using 467 wavenumbers, due to their subsampling techniques

LDA and PLS uses linear combinations of the features, resulting in dimensionality reduction [10]. While the LDA algorithm searches for a subspace to maximizes classes separation, calculating the inter and intraclasses distances, the PLS maximizes the covariance, finding the direction of the feature subspace to explain the higher variance of classes subspace. Thus, 467 features were decreased to ten components for PLS and to one for the LDA, as it is limited to the number of classes minus one.

Lower accuracies of SVM and KNN model may be related to the absence of subsampling and dimensionality reduction techniques, bringing difficulties for these models to handle with large data. Feature extraction and selection methods may improve the results of these models, and should be tested in future works. Larger SD in comparison to the other 4 models also indicates an inconsistent classification, probably due to overfitting the training data and making the test accuracy unstable.

V. CONCLUSION

Predictions with mean accuracies higher than 98%, as presented by XGB, LDA, PLS and RF, demonstrates the potential of FTIR spectroscopy when allied with machine learning techniques for breast cancer hormone receptors classification, where XGB and LDA were the best tested models. This analysis may provide additional information for the prognostic of breast cancer.

REFERENCES

[1] World Health Organization, "Cancer." https://www.who.int/health-topics/cancer (accessed May 22, 2021).

[2] World Health Organization, "WHO Report on Cancer: setting priorities, investing wisely and providing care for all," Geneva, 2020.

[3] Instituto Nacional De Cancer José Alencar Gomes Da Silva, "Estimativa 2020: Incidencia de Câncer no Brasil," Rio de Janeiro, 2020.

[4] N. Eliyatkin, E. Yalcin, B. Zengel, S. Aktaş, and E. Vardar, "Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way," *J. Breast Heal.*, vol. 11, no. 2, pp. 59–66, Apr. 2015, doi: 10.5152/tjbh.2015.1669.

[5] H. G. Russnes, O. C. Lingjærde, A.-L. Børresen-Dale, and C. Caldas, "Breast Cancer Molecular Stratification," *Am. J. Pathol.*, vol. 187, no. 10, pp. 2152–2162, Oct. 2017, doi: 10.1016/j.ajpath.2017.04.022.

[6] K.-Y. Su and W.-L. Lee, "Fourier Transform Infrared Spectroscopy as a Cancer Screening and Diagnostic Tool: A Review and Prospects," *Cancers (Basel).*, vol. 12, no. 1, p. 115, Jan. 2020, doi: 10.3390/cancers12010115.

[7] C. L. M. Morais, K. M. G. Lima, M. Singh, and F. L. Martin, "Tutorial: multivariate classification for vibrational spectroscopy in biological samples," *Nat. Protoc.*, vol. 15, no. 7, pp. 2143–2162, Jul. 2020, doi: 10.1038/s41596-020-0322-8.

[8] Janez Demsar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *J. Mach. Learn. Res.*, vol. 17, pp. 1–10, 2006.

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Mar. 2016, doi: 10.1145/2939672.2939785.

[10] Ruiping Wang, Huimin Guo, L. S. Davis, and Qionghai Dai, "Covariance discriminative learning: A natural and efficient approach to image set classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 2496–2503, doi: 10.1109/CVPR.2012.6247965.