

<https://doi.org/10.29289/259453942025V35S1022>

# Utilizing machine learning to identify biomarkers of chemoresistance in breast cancer: a complementary analysis with in vitro resistance platforms

Martina Lichtenfels<sup>1</sup>, Antônio Luiz Frasson<sup>2</sup>, Matheus Dalmolin<sup>3</sup>, Alessandra Borba Anton de Souza<sup>2</sup>, Julia Caroline Marcolin<sup>1</sup>, Camila Alves da Silva<sup>1</sup>, Caroline Brunetto de Farias<sup>1</sup>, José Luiz Pedrini<sup>4</sup>

<sup>1</sup>Ziel Biosciences – Porto Alegre (RS), Brazil.

<sup>2</sup>Pontifícia Universidade Católica do Rio Grande do Sul, Hospital São Lucas, Breast Cancer Center – Porto Alegre (RS), Brazil.

<sup>3</sup>Universidade Federal do Rio Grande do Norte, Laboratory of Machine Learning and Intelligent Instrumentation – Natal (RN), Brazil.

<sup>4</sup>Grupo Hospitalar Conceição, Breast Center – Porto Alegre (RS), Brazil.

**Objective:** This study aimed to use a machine learning algorithm to identify biomarkers of resistance to neoadjuvant chemotherapy (NACT) in breast cancer and validate these findings in a preliminary patient cohort, comparing results with an in vitro resistance platform. **Methods:** Clinicopathological data from breast cancer samples before NACT were analyzed, using public datasets and a proprietary database. Differential analyses compared patients with residual disease versus pathological complete response (pCR) to NACT. The XGBoost algorithm, a tree-based machine learning technique, and the SHAP tool were employed for interpretation. Additionally, tumor samples from patients with primary invasive breast cancer referred to NACT were collected and cultured in a chemoresistance platform. The samples were tested with cytotoxic drugs to classify the tumors based on cell viability. **Results:** These datasets included 1,012 patients exhibiting heterogeneous data. The XGBoost algorithm achieved 82% accuracy in classifying samples into pCR and residual disease, with SHAP analysis highlighting age, estrogen receptor status, and grade as key resistance predictors. Among ten patient samples, five achieved pCR, one had a good response with microinvasion, and four presented poor responses. In the chemoresistance platform, patients with pCR exhibited low resistance to the drugs used in chemotherapy, and those with poor responses demonstrated high rates of intermediate-to-high resistance to the drugs already used. Distinct resistance patterns to treatments not used in clinics were observed, suggesting these drugs could be alternative treatment options. The algorithm predicted NACT response with 81.8% accuracy in this cohort. **Conclusion:** These findings highlighted the capacity of the XGBoost algorithm in predicting breast cancer resistance, and in combination with the chemoresistance platform, allow the development of personalized therapeutic strategies.

**Keywords:** breast neoplasms; neoadjuvant chemotherapy; drug resistance; machine learning.