

<https://doi.org/10.29289/259453942025V35S1090>

# Artificial intelligence and machine learning for breast cancer recurrence prediction: a data-centric approach

Mariana Zuliani Theodoro de Lima<sup>1</sup>, Jaqueline Alvarenga Silveira<sup>2</sup>, Alexandre Ray da Silva<sup>2</sup>, Daniela Gregolin Giannotti<sup>2</sup>, André Luiz da Silva<sup>2</sup>, Rodrigo Macedo da Silva<sup>3</sup>, Jefferson Mazzei<sup>3</sup>, Fabio Francisco Oliveira Rodrigues<sup>3</sup>

<sup>1</sup>OncoAI – São Paulo (SP), Brazil.

<sup>2</sup>Universidade Presbiteriana Mackenzie – São Paulo (SP), Brazil.

<sup>3</sup>Instituto de Câncer Dr. Arnaldo – São Paulo (SP), Brazil.

**Introduction:** Breast cancer recurrence remains a significant clinical challenge, particularly in resource-constrained environments where long-term patient monitoring is often limited. **Objective:** Leveraging recent advances in artificial intelligence and machine learning, this study proposes a data-driven methodology to predict recurrence in breast cancer patients based on clinical and histopathological data extracted from electronic medical records. **Methods:** Through the development of a structured data processing pipeline, unstructured medical records were transformed into a high-quality dataset suitable for predictive modeling. The study introduced a stratified modeling approach, beginning with unsupervised clustering to identify distinct patient subgroups based on tumor aggressiveness. Supervised learning models were subsequently applied to each group, aiming to tailor predictions according to subgroup-specific characteristics. In parallel, survival analysis methods were employed to enhance interpretability and assess long-term outcomes. The ethics committee at Dr. Arnaldo Cancer Institute approved this study. **Results:** The models developed demonstrated promising performance in predicting recurrence, with results supporting the clinical applicability of the proposed stratified modeling approach. Notably, the integration of survival analysis enhanced the interpretability of predictions, allowing the identification of patterns related to long-term outcomes. **Conclusion:** These findings highlight the potential of artificial intelligence-driven tools to support clinical decision-making and personalized follow-up strategies for breast cancer patients. Ongoing work includes expanding validation across multiple institutions and exploring the integration of molecular biomarkers and advanced deep learning techniques.

**Keywords:** breast neoplasms; artificial intelligence; machine learning; survival analysis; medical oncology.