SOPHiA GENETICS™

FDG-PET/CT and Multimodal Machine Learning Model Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer

Groheux D, et al. Cancers (Basel). 2025; 17(7):1249.
doi: 10.3390/cancers17071249

KEY OUTCOMES

Applying machine learning (ML)-based method to baseline multimodal data helps predict pathological complete response (pCR) status after neoadjuvant chemotherapy (NAC) for triple-negative breast cancer (TNBC) patients and may identify correlations with improved long-term outcomes.

The integration of clinical, histopathological, molecular, imaging, and radiomic features resulted in a robust AI-powered predictive model, that can allow for precision medicine and effective treatment decisions.

BACKGROUND

TNBC is a heterogeneous and aggressive subtype of breast cancer associated with poor outcomes

pCR status after NAC strongly correlates with favourable long-term outcomes

Predicting pCR based on patient baseline multimodal data is of clinical interest to guide the treatment regimen

OBJECTIVES

1. Evaluate a multimodal machine learning-based algorithm predictive of pCR
2. Predict response to NAC based on baseline data from breast cancer patients

METHODS & APPROACH

Data collection & curation

DATASET

57 patients with stage II-III TNBC treated with NAC

Treated at AP-HP Hôpital Saint Louis between 2008 and 2015

DATA MODALITIES

The data aggregated resulted in a total of 241 predictors collected for each patient.

CLINICAL	GENOMICS	HISTO-PATHOLOGICAL	IMAGING	RADIOMICS
11 features	13 features	11 features	206 PET features	195 features

Statistical Analysis

DESCRIPTIVE ANALYSIS

Feature selection

17 features selected for final analysis

- Wilcoxon rank-sum test for continuous covariates
- Fisher's exact test for binary covariates
- Missing values in predictors were handled with a single imputation

MODEL DEVELOPMENT

Training of different machine learning algorithms

- Logistic regression
- Binary decision trees
- Random forest
- **Support vector machine (SVM)**

MODEL LEARNING & EVALUATION

Pipeline hyperparameter optimization using:

- Grid search algorithm
- Nested leave-pair-out cross-validation

FINDINGS

The best results in **the prediction of pCR status after NAC** were obtained with **a support vector machine (SVM) algorithm with a linear kernel, aggregating baseline clinical data, histopathological and molecular features, and PET data, including radiomic features.**

AUC 0.82

95% CI [0.74; 0.90]

Excluding a specific data modality resulted in a **decrease of almost 10% in the AUC.**

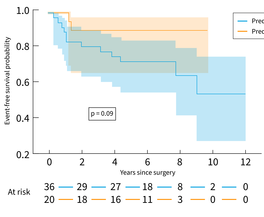
Datasets	AUC (95%CI)
Set 1	0.63 [0.51; 0.73]
Set 2	0.70 [0.60; 0.80]
Set 3	0.82 [0.74; 0.90]

Set 1: Clinical, histopathological, and PET non-radiomic features; Set 2: Set 1 + genomic data; Set 3: Set 2 + whole set of radiomic features

The **three features with the highest weight in the algorithm** were:

- Tumor SUVmax, 2 radiomic features
- GGIr
- Clinical T-stage

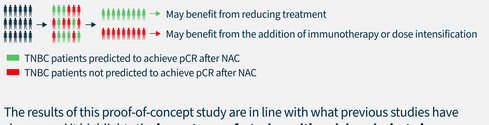
SUV: standardized uptake value; GGIr: genomic grade index



Patients with predicted pCR showed a longer event-free survival (EFS) than patients with predicted non-pCR (p=0.09, likely due to the small size of the cohort).

CLINICAL IMPACT OF AI PREDICTIVE APPROACHES

Predictive AI-based approaches have shown the ability to improve breast cancer treatment management by allowing an earlier identification of patients' response to chemotherapy, leading to a more personalized treatment.



The results of this proof-of-concept study are in line with what previous studies have shown, and it highlights the **importance of a truly multimodal analysis, to improve patient outcomes.**

Hôpital Saint-Louis AP-HP
Inserm

This project was executed by SOPHiA GENETICS in collaboration with **AP-HP Hôpital Saint Louis and INSERM.**

Disclaimer: Technology and concepts in development. May not be available for sale.

QL-MKT-250055