Multimodal machine learning model prediction of pathological complete response to neoadjuvant chemotherapy in triple-negative breast cancer

D. Groheux*1, L. Ferrer, J. Vargas, A. Martineau1, L. Teixeira1, P. Menur, P. Bertheau1, O. Gallinato2, T. Colin1, J. Lehmann C11

*Saint-Louis Hospital, APHP, Paris, France; **SOPHIA GENETICS, Pessac, France; †Université de Paris, INSERM U976, Paris, France; ‡SOPHIA GENETICS, Saint Sulpice, Switzerland.

*Corresponding author: David Groheux, MD, PhD, dgroheux@yahoo.fr

Methods & Materials

2 Methods & Materials

Predicted probability of non-pCR status was generated for each patient through a full machine learning (ML) pipeline integrating real-life multimodal data.

Interpretability tools were proposed to bridge the gap towards a clinical adoption of a ML-based predictive signature of non-pCR status.

3 Literature and model building & evaluation

Baseline multimodal data could help predict long-term outcomes.

Interpretability tools

Permutation feature importance measures the decrease in the AUC of the model when a single feature is removed from the model.

Local (patient-specific) interpretability

Patient PMOL222 had a predicted probability of non-pCR at 22% (output value) whereas the averaged prediction in the dataset was 63.2% (base value). Features decreasing the risk level were notably sphericity (low value) and rGGI (high value) while clinical T-stage (high value) was the main feature causing increased risk level.

Global interpretability

Permutation feature importance measures the decrease in the AUC of the model when a single feature is removed from the model.

4 Predictive performances

The aggregation of multiple data modalities (clinical, biological, imaging) could help improve prediction of pathological complete response (pCR) status after neoadjuvant chemotherapy in non-metastatic triple-negative breast cancer (TNBC) patients.

SOPHiA DDM™ for Radiomics Platform (Research Use Only; SOPHIA GENETICS SA, Switzerland) enabled tumor segmentation and radiomics features extraction.

Predicted probability of non-pCR status was generated for each patient through a full machine learning (ML) pipeline integrating real-life multimodal data.

Interpretability tools were proposed to bridge the gap towards a clinical adoption of a ML-based predictive signature of non-pCR status.

5 Conclusion and perspectives

Knowledge gain

Proof of concept study suggesting that machine learning applied to baseline multi-modal data can help predicting pCR status after neoadjuvant chemotherapy for TNBC at the individual patient level, as well as stratifying patients to inform long-term outcomes.

Clinical applications

Patients that would be predicted as non-pPCR could benefit from concomitant treatment with immunotherapy, or dose intensification.

Global democratization

This algorithm will be further validated in a larger, multicentric cohort.