Probabilistic forecast for breast cancer using distributional random forest to predict ODX based on clinico-pathological data

Romain Pic*[1], Clément Dombry[2], Zena Al Masry[3], and Christine Devalland

[1]Laboratoire de Mathématiques de Besançon (UMR 6623) (CNRS/LMB) – Université de Bourgogne : UMR6623, Centre National de la Recherche Scientifique : UMR6623, Université de Franche-Comté – UFR Sciences et techniques 16 route de Gray 25 030 Besançon cedex, France
[2]Université de Franche-Comté – Laboratoire de Mathématiques de Besançon – France
[3]FEMTO-ST – Université de Franche-Comté, Université de Franche Comté – France

Résumé

Oncotype DX (ODX) is a commercially molecular test for breast cancer assay (genomic health) that provides prognostic and predictive breast cancer recurrence information for hormone positive, HER2 negative patients. This test can help practicians assess if a chemotherapy treatment could be effective. However, this test is expensive and can be hard to interprete. The result of the test is a score between 0 and 100 and guides clinicians regarding prescription of radiotherapycan (e.g. for score higher than 25).
In order to overcome the limitations of this test, we used Distribution Regression Forests (DRF) to predict the ODX score. The cohort is a retrospective study between 2012 and 2019 with 334 cases that underwent ODX assay from three hospitals: Besançon, Belfort and Dijon. All patients have ER-positive and HER2-negative early breast cancer.
The DRF methodology allows to obtain different outputs: a predictive distribution of the ODX score for a given patient, the predicted probability of belonging in a given class, the uncertainty of the prediction (e.g. confidence intervals) and patients that are similar to the one of interest. The DRF method was compared to numerous state-of-the-art classification methodologies and reaches comparable or better results in terms of standard metrics. Moreover, the performance of the DRF methodology was evaluated in a distributional regression framework and three cases representative of the full range of predictions – good, average or bad – were outlined in order to present how the outputs of the DRF methodology can be interpreted to help practicians make a better-informed decision for their patient.

*Intervenant