Artificial intelligence-driven synthetic radiogenomic applications in breast cancer

Lianghong Chen[1], Zi Huai Huang[2], Yan Sun[1,2], Qian Liu[2], Pingzhao Hu[1,2,3,4,5]

[1] Department of Computer Science, Western University, London, ON; [2] Department of Biochemistry, Western University, London, ON; [3] Department of Epidemiology and Biostatistics, Western University, London, ON; [4] Department of Oncology, Western University, London, ON; [5] The Children's Health Research Institute, Lawson Health Research Institute, London, ON. phu49@uwo.ca

Background: Breast cancer (BC) exhibits significant heterogeneity, making it crucial to analyze its phenotypic diversity at a multi-omics level for early detection and personalized treatment. The integration of medical images and genomics offers a novel perspective on the study of BC heterogeneity. However, the absence of paired medical images and genomics data could pose a significant challenge. We hypothesize that a well-trained conditional probabilistic diffusion model (CPDM) can address this unpaired data problem by generating BC magnetic resonance images (MRIs) based on genomic information. Generated MRIs can be used to predict clinical attributes including gene mutations, estrogen receptor (ER) status, human epidermal growth factor receptor 2-positive (ER+/HER2+) status and have survival significance.

Methods: A CPDM was trained on sagittal MRI projections, conditioned on multi-omic features extracted by using the Bayesian tensor factorization method. Frechet's Inception Distance (FID) was used to evaluate the CPDM performance. The well-trained CPDM was then used to impute the missing MRIs for 726 TCGA-BRCA patients from their multi-omic profiles. The generated MRIs were further used to train XGBoost models to predict clinical attributes such as ER and HER2 status, TP53 mutation status, etc. The performance of the XGBoost was evaluated using Receiver Operating Characteristic - Area Under the Curve (ROC-AUC), Precision-Recall-AUC (PR-AUC), and F1 score. The survival significance of the generated MRIs was estimated by their performance (C-index) in training survival models like DeepSurv and CoxPHFilter models. In addition, to evaluate the CPDM's generalizability in single genomics, we repeated our experiments using only the gene expression data of 123 ER+/HER2+ BC patients.

Results: The FID scores of two CPDMs are between 0 and 4. As shown in Figure 1, the generated MRIs achieved a ROC-AUC of 0.63 for predicting TP53 mutation status, a PR-AUC of 0.93 in predicting ER status, a PR-AUC of 0.89 in predicting ER+/HER2+ subtypes, and a C-index score of 0.92 in survival analysis.

Discussion: We explored the potential of CPDMs in generating MRIs from genomic profiles. The synthesized MRIs were well performed in predicting different clinical outcomes. This study offers valuable references for further radiogenomic research and precision medicine advancements.

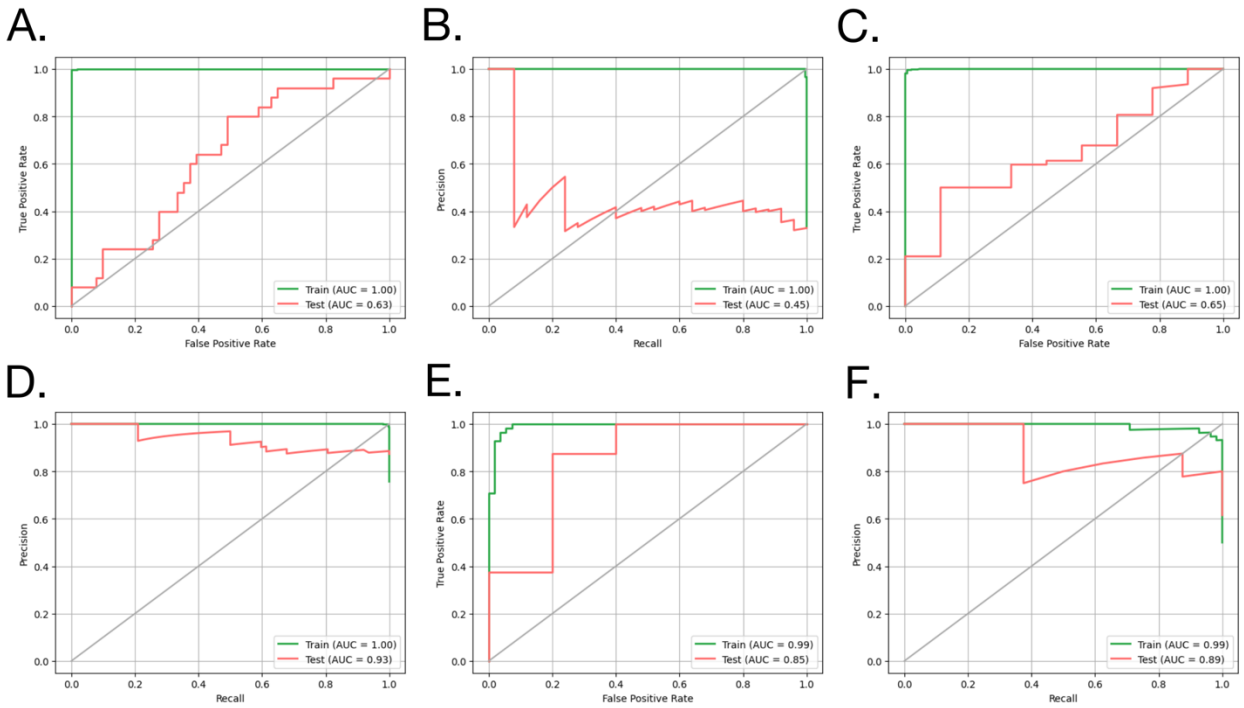

Fig. 1. ROC and precision-recall curve for classification tasks. **A.** ROC curves of the gene TP53 mutation status prediction based on the XGBoost model trained on the MRI (multi-omic version) features extracted by the ResNet50 model. **B.** Precision-recall curves of the gene TP53 mutation status prediction based on the XGBoost model trained on the MRI (multi-omic version) features extracted by the ResNet50 model. **C.** ROC curves of the ER status prediction based on the XGBoost model trained on the MRI (multi-omic version) features extracted by the PyRadiomics tool. **D.** Precision-recall curves of the ER status prediction based on the XGBoost model trained on the MRI (multi-omic version) features extracted by the PyRadiomics tool. **E.** ROC curves of the ER+/HER2+ subtypes prediction based on the XGBoost model trained on the MRI (gene expresssion version) features extracted by the PyRadiomics tool. **F.** Precision-recall curves of the ER+/HER2+ subtypes prediction based on the XGBoost model trained on the MRI (gene expresssion version) features extracted by the PyRadiomics tool.