AutoFlow: A Novel Method for Assessing Minimal Residual Disease in Breast Cancer Patients by Identifying Bone Marrow Disseminated Tumor Cells Using Flow Cytometry Data

Mohanraj, Dheepthi (School: North Carolina School of Science and Mathematics)

Breast cancer (BC) is the leading cause of cancer death among women worldwide. While BC has favorable survival rates for localized cases, 26% of BC survivors will experience an incurable cancer recurrence within twenty years. Recurrent, metastatic breast cancer develops exclusively from disseminated tumor cells (DTCs) that persist after treatment, migrate to other tissues, and develop into tumors. The presence of bone marrow DTCs in BC patients is associated with a poor prognosis, but secondary treatment may improve outcomes for DTC-positive patients. Flow cytometry (FC) can characterize DTCs and surrounding bone marrow cells (BMCs) in an aspirate, but manual FC analysis is tedious, time-consuming, and variable. This research is the first investigation to produce a model that distinguishes between DTCs and benign BMCs when trained and tested on human FC data. To develop the final machine-learning framework, innovative regularization and thresholding methods were leveraged to combat class imbalance within a dataset of 225 million ‘benign’ data points and only 45 DTCs. This project's procedural pipeline utilizes a multilayer perceptron or XGBoost classifier to distinguish between live, healthy cells and debris/dying cells with up to 98% accuracy. An Isolation Forest algorithm isolates DTCs among the healthy cells with an overall classification accuracy of 99%, detecting at least 90% of the rare DTC data points. Further research can use the algorithms developed in this study to identify DTCs originating from other cancers or cells of hematopoietic lineage.

Awards Won:

Third Award of $1,000