

This document outlines the software used for each step of the methods. The sections below correspond to the sections in our manuscript.

Microarray Datasets: N/A

Data Preprocessing:

Cleaning and Imputation: cleanAndImpute.m, nnimpute.m, nearestneighbour.m

Quantile Normalization: normalize.m, findLCSAll.m, findLCS.m, averageSimilarProbesAll.m, averageSimilarProbes.m, geneMatchAll.m, geneMatchFast.m, quantNorm.m, rid_extensions.m

Biochemically-Inspired Gene Selection: N/A

mRMR Ranking: used the Matlab/C++ code written by the authors of the original mRMR paper; downloaded from <https://www.mathworks.com/matlabcentral/fileexchange/14608-mrmr-feature-selection--using-mutual-information-computation->

Support Vector Machine Learning: We used the fcecoc function of MATLAB's Statistics and Machine Learning Toolbox (2017a) as described in the manuscript.

Feature Selection Algorithms:

Complete: CSFS_strat_kfold.m, CSFS_strat_kfold_gridsearch.m

Backward: BSFS_strat_kfold.m, BSFS_strat_kfold_gridsearch.m

Forward: FSFS_strat_kfold, FSFS_strat_kfold_gridsearch.m

Validating Signatures (K-Fold Validation): kfoldvalidation_multiclassSVM.m, kFoldValidation.m

Validating Models (Traditional Validation): regularValidation_multiclassSVM.m

Performance Metrics: N/A –integrated into the validation functions