

**COMBINING NEAREST NEIGHBOR CLASSIFIERS VERSUS
CROSS-VALIDATION SELECTION**

by

Minhui Paik and Yuhong Yang
Iowa State University

April 2004

ABSTRACT

Various discriminant methods have been applied for classification of tumors based on gene expression profiles, among which the nearest neighbor (NN) method was reported to perform relatively well. Usually cross-validation (CV) is used to select the neighbor size as well as the number of genes for the NN method. However, CV can perform poorly when there is considerable uncertainty in choosing the best candidate classifier. As an alternative to selecting a single “winner”, in this work, we propose a weighting method to combine the multiple NN rules. Three gene expression data sets are used to compare its performance with CV methods. The results show that when the CV selection is unstable, the combined classifier performs much better.