AbstractID: 9809 Title: Classifier Design and Sample Size Effects for Computer-Aided Diagnosis

Classifier is an important component of CAD algorithms. The accuracy of a CAD algorithm depends on the ability of the classifier(s) to differentiate true and false lesions or malignant and benign lesions. Classifier design generally involves two stages: feature selection and parameter estimation. Case samples drawn from the patient population are required for designing the classifier for a given CAD task. Two important considerations are involved in collecting a database for CAD algorithm development: whether the distribution of the case samples adequately represents the general population, and whether the sample size is large enough to adequately train the classifier. These factors are often inter-related, and both affect the generalizability of the trained classifier to unknown cases.

Classifier performance is often evaluated by the resubstitution and the hold-out methods. In the resubstitution method, the same sample set is used for training and testing. In the hold-out method, some available samples are withheld from being used for training so that the trained classifier can be tested with the independent hold-out samples. Resampling techniques such as leave-one-out or N-fold cross validation can be used to achieve hold-out testing with efficient use of the available samples.

Without including the feature selection stage, Fukunaga derived that the bias of the estimated probability of misclassification (PMC) depends on $1/N_t$ (N_t = design sample size), when the higher order terms are negligible. In medical imaging, classifier performance is more commonly described by the area under the ROC curve, A_z,. We have extended Fukunaga's analysis to A_z for linear classifiers and multivariate normal distributions and found that A_z has the same general dependence on N_t as PMC. The resubstitution method optimistically biases A_z whereas the hold-out method pessimistically biases A_z, relative to the performance of a classifier trained with infinite sample size. As N_t increases, the resubstitution and the hold-out performances converge towards the unbiased estimate.

If feature selection is included in the classifier design process, the bias on the classifier performance becomes more complicated. Our studies indicated that the resubstitution estimate was always optimistically biased, except when too few features were selected. When feature selection was performed using the design samples alone, the hold-out estimate was always pessimistically biased. If feature selection was performed using the entire sample set while resubstitution and hold-out performances were evaluated during the parameter estimation stage, the hold-out performance could be pessimistically or optimistically biased, depending on the number of features available for selection, the number of available samples and their statistical distributions.

We will discuss issues related to classifier design in CAD. The performance of linear and non-linear classifiers such as neural networks or quadratic classifiers in multi-dimensional feature spaces with different class distributions will be compared. Simulation studies will be used to demonstrate the effects of sample size, dimensionality, and class distributions on classifier performance.

Educational Objectives:

1. Understanding the process of classifier design in CAD.
2. Understanding the effects of sample size on classifier design.
3. Understanding the tradeoffs between classifier complexity and performance with finite sample size.