

## NIH Summer Research Program 2007

**POSTER TITLE:** The Importance of Prevalence and Misclassification Cost Considerations in Computerized Clinical Decision Support Systems

**MENTOR:** Jim DeLeo, Chief, Scientific Computing Section, DCRI, NIHCC

**NIH AFFILIATION:** Scientific Computing Section, DCRI, NIHCC (2007 Summer Intern)

**NON-NIH AFFILIATION:** Arizona State University, Computer Science PhD Student

**THE IDEA IN BRIEF:**

Computational classifier methodologies often do not consider prevalence (expected or prior probability of events) or ignore the potentially dire consequence of misclassifying an event in candidate classes. Using a basic supervised learning no hidden layer back error propagation artificial neural network (ANN) computational architecture, we set out to demonstrate the importance of considering prevalence and misclassification cost in designing computational classifiers. The work presented here explores the same.

**ABSTRACT:**

We demonstrate the importance of considering prevalence,  $p_p$  and misclassification cost when designing computer-based clinical classification decision-support systems. We do this by first expressing misclassification cost as a probability,  $p_{mc}$  and then expressing the joint probability,  $P$  associated with  $p_p$  and  $p_{mc}$ . Next we produce a library of bivalent classifier artificial neural networks (ANNs) in which each ANN is trained with a data set reflecting one of a range of  $P$ -values. With each trained ANN, we process test data sets each reflecting the same set of  $P$ -values. Our results show near-optimum performance when test data set and training data set  $P$ -values are closely matched.

**PURPOSE:**

To demonstrate enhanced performance in computerized clinical classification decision support systems when knowledge of prevalence and misclassification cost is considered in the design of these systems.