

CCSU
DEPARTMENT OF MATHEMATICAL SCIENCES

COLLOQUIUM

Friday, November 20

2:00 – 3:00 PM

Maria Sanford, Room 101

A NOVEL APPROACH TO IDENTIFICATION OF STABLE AND BIOLOGICALLY INTERPRETABLE GENOMIC AND PROTEOMIC BIOMARKERS

DARIUS M. DZIUDA

CENTRAL CONNECTICUT STATE UNIVERSITY

Abstract: Various supervised learning algorithms can be used to identify parsimonious multivariate biomarkers. While such biomarkers may contain information sufficient for quite efficient classification of new cases, they are often unstable and do not necessarily provide insight into biological processes underlying class differentiation.

In this talk, we will define two concepts – the *Informative Set of Genes* and the *modified bagging schema* – and present a method utilizing them for identification of robust biomarkers and for elucidation of biological processes associated with class differences.

Using the *modified bagging schema*, hundreds or thousands of classifiers are built from the *Informative Set of Genes*. Each of these classifiers is a result of independent feature selection process. By examining the distribution of genes among the classifiers and by analyzing gene expression patterns represented in the *Informative Set of Genes*, we identify primary sets of informative genes. These primary sets can be used to identify robust multivariate biomarkers. They also represent the optimal amount of information allowing biological interpretation of class differences.

Keywords: *Biomarker Discovery, Personalized Medicine, Multivariate Biomarkers, Ensemble Classifiers, Data Mining, Bioinformatics, Genomics, Proteomics, Differential Diagnosis, Drug Discovery, Classification and Prediction Systems.*

For further information:

gotchevi@ccsu.edu 860-832-2839