Machine Learning Analysis of Cancer Cachexia

Roman Eisner*, Jianguo Xia**, D. Duong Hau, Cynthia Stretch, Russell Greiner, Vickie Baracos, David Wishart

*eisner@cs.ualberta.ca **jianguox@ualberta.ca

Introduction
Cancer cachexia is a significant and rapid loss of adipose and skeletal muscle tissue. Cachexia is actually quite difficult to detect and quantify, and weight loss is often not a symptom of cachexia until the final phases of the disease. Current methods to detect muscle loss are expensive and time-consuming, requiring whole body CT scans at two time intervals and careful, manual coding of muscle and adipose tissue. The development of a quick, non-invasive blood or urine test to detect the earliest stages of cachexia onset could have significant implications for treating this metabolic disorder. Such a test could potentially improve patient quality of life and ensure longer term survival. In an effort to develop such a test we have conducted a metabolomic study on urine samples of 74 cancer patients. Metabolites in these samples were quantified using 1H NMR. As shown here, the application of advanced machine learning techniques allows the detection of several characteristic metabolic profiles that are strongly indicative of the degree of cachexia as measured using conventional CT methodologies.

Data Processing

**Task A:** Classification of Cachectic Status
- Each sample is labeled as “Cachectic” or “Non-Cachectic” based on their urinary metabolic profile
- Leave One Out Cross-validation
- Random Forest (RF): 82% accuracy
- Naïve Bayes (NB): 86% accuracy

**Task B:** Regression Analysis (ASM):
- Regression is performed on the raw ASM data and Metabolite data
- Un-normalized metabolite concentrations are used
- Individual metabolite concentrations that are not confidently assigned using the Chenomx NMR Suite are ignored
- Learn classifier from (labeled) training data, to use on data from novel patients
- Important metabolites identified using Support Vector Machine (SVM) weights

Discussion
- The linear classifiers used implicitly assumes independence between metabolites. This works well because correlations between metabolites are generally low.
- Combining results from several machine learning approaches identifies many urinary metabolites associated with cancer cachexia
- Achieved over 80% accuracy using these metabolites (random forest or naive Bayesian network)

These identified metabolites are involved in a wide variety of metabolic pathways:
- gluconeogenesis
- methane metabolism
- glycine biosynthesis, and others
- Affected pathways reflect the overall deregulation of metabolic processes in cancer cachexia
- Cachexia refers to not only a tissue loss but a rapid rate of tissue loss. Further studies could include the rate at which muscle is lost, which could be an important factor in metabolic analysis.