



[Print this Page for Your Records](#)

[Close Window](#)

Control/Tracking Number: 11-A-7695-AACR  
 Activity: Abstract Submission  
 Current Date/Time: 11/16/2010 4:53:31 PM

#### Comparative study on feature selection methods in ovarian cancer diagnosis using chemometrics tools

Short Title:  
 Ovarian cancer diagnosis

Author Block: Seongkyu Yoon<sup>1</sup>, Hae Woo Lee<sup>1</sup>, Jay Liu<sup>2</sup>, Young Jung Na<sup>3</sup>. <sup>1</sup>University of Massachusetts, Lowell, MA; <sup>2</sup>Pukyong National University, Busan, Korea, Republic of; <sup>3</sup>CHA University, Seoul, Korea, Republic of

#### Abstract:

Comparative study on feature selection methods in ovarian cancer diagnosis using chemometrics tools

Hae Woo Lee<sup>1</sup>, Seongkyu Yoon<sup>1</sup>, Jay Liu<sup>2</sup>, and YoungJeong Na<sup>3</sup>

<sup>1</sup>Massachusetts Biomanufacturing Center, Chemical Engineering Department, University of Massachusetts, Lowell, MA, USA, <sup>2</sup>Department of Chemical Engineering, Pukyong National University, Busan, South Korea, <sup>3</sup>Comprehensive Gynecologic Cancer Center, Bundang CHA General Hospital and CHA University, Seoul Korea

Ovarian cancer has the poorest prognosis of the gynecological cancers. 75% of the patients are diagnosed with stage III-IV on their diagnosis. The 5 year survival rate of ovarian cancer patients whose disease was diagnosed at stage III or IV is less than 20%, but that of the patients diagnosed with stage I is higher than 90%. The early detection and diagnosis of cancer is one of the major application areas in omics technology and has received tremendous attention in last few years.

The main focus of this study is aimed at the investigation of robust formulations for the feature selection algorithms which are utilized during the statistical analysis of omics profiles in diagnosing ovarian cancer. The common difficulty in the feature selection for cancer-specific biomarker discovery is the inconsistency of the feature sets selected among different samples or by different selection methods. This issue is particularly crucial in biological applications where subsequent biomedical analysis of selected features requires relatively considerable time and costs. In this study, we analyzed the reproducibility of different feature selection algorithms usually employed in chemometrics and investigated their stability to provide valuable insight for designing proper feature selection algorithms in the domain of cancer-specific biomarker discovery. Two data sets of MALDI-TOF-MS (proteomics) and LC/TOF-MS (metabolomics) spectra measured from serum samples to diagnose ovarian cancer were analyzed. To evaluate the feature selection methods in terms of their consistency, we systematically quantified a measure of similarity (similarity<sub>S1</sub>) for the outcome of each method by running each selection procedures repetitively.

The results show that, in general, the multivariate feature selection methods outperformed the univariate methods, in that they selected more consistent features regardless of training set perturbations. The results clearly show that some feature selection methods can suffer severely from the irreproducibility issue and this might deflate the potential benefits of omics technology for cancer diagnosis. A possible remedy might be multi-objective formulation of the feature selection problem by simultaneously optimizing stability and accuracy of classification models, and this will be investigated further in future research.

:

Author Disclosure Information: S. Yoon: ; Univeristy of Massachusetts. H. Lee: ; University of Massachusetts. J. Liu: ; Pukyung National Un iversity. Y. Na: ; CHA University.

#### Sponsor (Complete):

Category and Subclass (Complete): CH03-03 Biological mass spectrometry and systems biology

Keywords/Indexing (Complete): Bioinformatics ; Ovarian cancer ; Diagnosis ; Mass spectrometry

Research Type (Complete): Basic research

#### Submission Details (Complete):

\*Primary Organ Site: Gynecological cancers: ovarian

\*Special Consideration: Not Applicable

\*Choose Chemical Structure Disclosure: NOT APPLICABLE. No compounds with defined chemical structures were used.

\*Please explain (maximum 250 characters with spaces): : N/A

#### Financial Support for Attendance (Complete):

**The presenter of this abstract meets the eligibility criteria for the AACR Scholar-in-Training Awards and will submit the abstract and application before the November 15 deadline.** : True

**The presenter of this abstract meets the eligibility criteria for the AACR-GlaxoSmithKline Outstanding Clinical Scholar Award and will submit the abstract and application before the November 15 deadline.** : True

**The presenter of this abstract meets the eligibility criteria for the AACR Minority-Serving Institution Faculty Scholar Awards and will submit the application before the December 10 deadline.** : True

**The presenter of this abstract meets the eligibility criteria for the AACR Minority Scholar in Cancer Research Awards and will submit the abstract and application before the December 10 deadline.** : True

Payment (Complete): Your credit card order has been processed on Tuesday 16 November 2010 at 4:52 PM.

Status: Complete

\*\*\*To log out, simply close your browser window. All information will be saved if you hit the Continue button after each step.

For all log-in problems or technical questions, please contact [OASIS Helpdesk](#) or call (217) 398-1792. If you have any policy questions related to the 102nd AACR Annual Meeting 2011, please contact AACR at (215) 440-9300 or (866) 423-3965.

[Leave OASIS Feedback](#)

Powered by [OASIS](#), The Online Abstract Submission and Invitation System<sup>SM</sup>  
© 1996 - 2010 [Coe-Truman Technologies, Inc.](#) All rights reserved.