

Proceedings of the 7th Annual World Conference  
of the Society for Industrial and Systems Engineering,  
Binghamton, NY, USA  
October 11-12, 2018

## **Predicting Post-Operative Survivability of Lung Cancer Patients: A Data Mining Approach**

**A.F. Khalaf and Y. Wang**

Department of Systems Science and Industrial Engineering,  
Binghamton University,  
Binghamton, New York, USA

Corresponding Author's Email: [yongwang@binghamton.edu](mailto:yongwang@binghamton.edu)

**Abstract:** Thoracic surgery is one of the treatment options for lung cancer patients. Risk measurement of such surgery is critical to choose the best option among treatments. In this paper, one-year survivability of lung cancer patients after thoracic surgery is investigated using data mining techniques. The study deals with imbalance dataset, and the synthetic minority over-sampling technique (SMOTE) is applied to resolve the problem. Feature selection is also performed by using information gain as the criterion when selecting significant attributes. Next, the performance of five classification methods, namely naïve Bayes, radial basis function network, support vector machine, logistic regression, and random forest are applied to predict one-year survivability of lung cancer patients after the thoracic surgery and the results are compared. It shows that after applying the over-sampling technique to the dataset, all the classifiers work well according to performance measurements, but random forest outperforms other classifiers.

*Keywords:* Lung cancer, Data mining, Imbalance data, Feature selection, Classification