

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

## **Decision Tree-based Rules Extraction to Predict Breast Cancer Using Clinical Stages as a Dependent Variable**

**H. Saad and N.N. Nagarur**

Department of Systems Science and Industrial Engineering,  
Binghamton University  
Binghamton, NY, USA

Corresponding Author's email: [hsaad1@binghamton.edu](mailto:hsaad1@binghamton.edu)

**Author Note:** Hamza Saad is at present pursuing his doctoral degree in the Department of Systems Science and Industrial Engineering, Binghamton University. Nagen Nagarur is a professor of Systems Science and Industrial Engineering in the Department of Systems Science and Industrial Engineering, Binghamton University. His research interests include supply chains, operations management, and healthcare delivery.

**Abstract:** This study involves various factors that affect breast cancer in women, negatively or positively. Data has been collected about various biological and sociological factors and the clinical stage at which the disease was collected. Mining association rules is a useful technique to extract the critical relationship between variables by generating some rules used to build prediction. However, association rules did not give useful results or supported rules for this type of data. Hence a decision tree approach is applied in the study. A detailed tree graph including 19 nodes and 10 leaves was utilized to construct rules based on each decision node. Each leaf node in the tree has a value of accuracy; the highest accuracy is selected to confirm the strong rules. From the 10 rules, only five rules were selected to build prediction. The process of rules selection is supported by the accuracy of each leaf and impact of each variable in the output. Predictor screening was used to predict robust variables based on P-value and Chi-square value. Tumor size variable exists in all extracted rules with a strong impact.

**Keywords,** Classification, Decision tree, Predictor screening, Rules extraction. Breast cancer.