Applying Logistic Regression and Artificial Neural Networks with Effective Feature Selection for Breast Cancer Detection

H Duan, Y Han, and SS Lu

Department of Systems Science and Industrial Engineering
State University of New York at Binghamton
Binghamton, NY 13902, USA

Corresponding author's Email: hduan2@binghamton.edu

Author Note: The authors are from the Department of Systems Science and Industrial Engineering (SSIE) and working in the Watson Institute for Systems Excellence (WISE) of State University of New York at Binghamton. Ying Han is a PhD Candidate, Dr. Shuxia, S. Lu is an associate professor of Department of Systems Science and Industrial Engineering, corresponding author Haoliang Duan currently study in SSIE master program.

Abstract: Breast cancer is one of the most common cancers affecting many females' life quality. Accurate diagnosis based on medical examination results allows patients receive proper treatment, which will be beneficial for patients, especially at the early stage. As such, diagnosis tool handling large number of decision parameters with high accuracy and efficiency is required. This study attempts to develop a mathematical model for an automated breast cancer detection system that can reduce complexity of analyzing medical examination results. Stepwise selection and RELIEF algorithm are applied to select important decision parameters for logistic regression and artificial neural networks (ANNs), respectively. Logistic regression (LR) model with seven (out of thirty) selected features is identified with 94.1% overall accuracy, 91.5% sensitivity, 95.5% specificity; ANN with three features, trained by backpropagation, achieves 95.3% overall accuracy, 91.8% sensitivity, and 95.9% specificity. The results from the proposed method showed relatively high accuracy and efficiency in analysis of cell nucleus features.

Keywords: Breast Cancer Detection, Logistic Regression, Artificial Neural Network, Feature Selection