Abstract

Breast cancer represents the second important cause of cancer deaths in women today and it is the most common type of cancer in women. Disease diagnosis is one of the applications where data mining tools are proving successful results. Data mining with decision trees is popular and effective data mining classification approach. Decision trees have the ability to generate understandable classification rules, which are very efficient tool for transfer knowledge to physicians and medical specialists. In fundamental truth, they provide trails to find rules that could be evaluated for separating the input samples into one of several groups without having to state the functional relationship directly. The objective of this paper is to examine the performance of recent invented decision tree modeling algorithms and compared with one that achieved by radial basis function kernel support vector machine (RBF-SVM) on the diagnosis of breast cancer using cytological proven tumor dataset. Four models have been evaluated in decision tree: Chi-squared Automatic Interaction Detection (CHAID), Classification and Regression tree (C&R), Quick Unbiased Efficient Statistical Tree (QUEST), and Ross Quinlan new decision tree model C5.0. The objective is to classify a tumor as either benign or malignant based on cell descriptions compound by microscopic examination using decision tree models. The proposed algorithm imputes the missing values with C&R tree. Then, the performances of the five models are measured by three statistical measures; classification
Diagnosis of Breast Cancer using Decision Tree Models and SVM

accuracy, sensitivity, and specificity.

References


**Index Terms**

Computer Science  
Artificial Intelligence

**Keywords**

Breast cancer; classification  
decision tree algorithms; SVM; missing data imputation