Abstract

A major area of current research in data mining is the field of medical diagnosis. In the present study using the Breast cancer Wisconsin data sets, a feature selection algorithm Modified Correlation Rough Set Feature Selection (MCRSFS) predicts both diagnosis and prognosis by comparing several data mining classification algorithms. In the proposed approach, in level 1 of feature selection, features are selected based on rough set with different starting values of reduct. In level 2 features are selected from the reduced set based on the Correlation Feature Selection (CFS). Experiments show the proposed method is effective by comparing with others in terms of number of selected features and classification performance.

Index Terms

Computer Science
Pattern Recognition

Keywords
Data mining feature selection rough set correlation breast cancer.