

Breast Cancer Detection Using Two-Step Reduction of Features Extracted From Fine Needle Aspirate and Data Mining Algorithms

Razieh Sheikhpour: School of Electrical and Computer Engineering, Yazd University, Yazd, Iran

Mahdi Agha Sarram: School of Electrical and Computer Engineering, Yazd University, Yazd, Iran

Mohammad Reza Zare Mirakabad: School of Electrical and Computer Engineering, Yazd University, Yazd, Iran

Robab Sheikhpour: Yazd Medical Science Branch, Islamic Azad University, Yazd, Iran

Corresponding Author: Mahdi Agha Sarram, mehdi.sarram@yazd.ac.ir

Abstract

Introduction: Early detection of breast cancer plays an important role in the treatment and survival of patients. Today, the features extracted from Fine Needle Aspirate (FNA) and data mining algorithms can be applied for designing intelligent systems in detection of breast cancer. The aim of the study is diagnosis of breast cancer using two-step reduction of features extracted from fine needle aspirate and data mining algorithms.

Method: In this study, Wisconsin Diagnostic Breast Cancer Data (WDBC) taken from the UCI machine learning repository was used. This dataset contains 569 samples of benign and malignancy with 30 characteristics extracted from FNA testing. In this study to improve the performance of breast cancer detection systems, two-step dimension reduction was proposed and the performance of data mining algorithms such as J48 decision tree, NaïveBayes, quadratic classifier, SVM and KNN was considered on the reduced features by proposed two-step dimension reduction.

Results: Two-step dimension reduction can improve the accuracy of data mining algorithms in detection of breast cancer. Proposed method has achieved the accuracy of 97.54% using KNN classifier based on Euclidean distance and two-step dimension reduction using feature selection based on correlation coefficient and PCA algorithm.

Conclusion: The results of this study showed that data mining algorithms and two-step dimension reduction can be very helpful in assisting the physicians to detect breast cancer.

Keywords: Breast Cancer, Data Mining, Two-step Dimension Reduction, Classification.