

Abstract

The high dimensionality and complex structure of DNA microarray data pose significant challenges for accurate cancer detection, as redundant and irrelevant features may lead to overfitting and reduced classification accuracy. To address these issues, dimensionality reduction is essential for eliminating non-informative features while preserving relevant information. This study proposes a Hybrid PCA-Autoencoder method that combines the linear feature extraction strength of Principal Component Analysis (PCA) with the non-linear representation learning capability of Autoencoders. The proposed approach is evaluated on five microarray datasets using a Support Vector Machine (SVM) classifier. Experimental results show that the hybrid method consistently outperforms standalone PCA and Autoencoder approaches, achieving classification accuracies of 100% for ovarian cancer, 98.18% for lung cancer, 84.21% for colon cancer, 66.67% for CNS tumors, and 73.33% for breast cancer. These findings highlight that the hybrid method offers a promising strategy for enhancing microarray-based cancer detection by combining complementary strengths of linear and non-linear dimensionality reduction.

Keywords

Cancer Detection, Dimensionality Reduction, Microarray, PCA, Autoencoder