Abstract

Microarrays are novel biotechnological technology that is being used widely in cancer research. By allowing the monitoring of expression levels in cells for thousands of genes simultaneously, microarray experiments may lead to a more complete understanding of cell’s function. This is due to the fact that the physiology of an organism is generally associated with changes in gene expression patterns, thus leading to a finer and more reliable classification.

Microarray data is an arrangement of points in rows and columns. Out of the various techniques of data mining, classification and clustering are two processes that have great potential in microarray data analysis. This research work focuses on using machine learning classification algorithms for predicting the presence or absence of cancer. A classification model for microarray data analysis consists of three major steps, namely, preprocessing, gene selection and identification or prediction of genetic defect. The preprocessing step consists of cleaning algorithms like normalization, missing value handling routines which enhance the quality of the gene microarray data and help to improve the subsequent steps. Gene selection is a process
where a set of informative genes is selected from the gene expression data in a form of microarray dataset. This process helps improve the performance of the classifier. The third step, classification, is a process to classify microarray data into several predefined classes that have its own characteristics.

References


Index Terms

Computer Science

Algorithms
Keywords

SVM, Microarraydata, classification, Gene Selection