A DNA MICROARRAY IMAGE ANALYSIS APPROACH FOR CANCER DETECTION AND CLASSIFICATION

A thesis submitted to the Department of Electronics, Communication and
computer Engineering, Faculty of Engineering, Helwan University
in partial fulfillment of the requirements for the degree of
Doctor of philosophy
In
Computer Engineering

Submitted by

Eng. Mary Monir Saeid Heness

Supervised by

A.Prof. Dr. Zaki Bassyouni Nossair
Department of Electronics, Communication,
Engineering
Faculty of Engineering
Helwan University

A.Prof. Dr. Mohamed Ali Saleh
Department of Computers and systems
Engineering
Faculty of Engineering
Helwan University

DEPARTMENT OF ELECTRONICS, COMMUNICATION
AND COMPUTER ENGINEERING
HELWAN UNIVERSITY
2020

ملخص الرسالة باللغة الإنجليزية:

In recent years, Deoxyribonucleic Acid (DNA) microarray technology has been the most widely used tool to monitor the expression level of many thousands of genes simultaneously. Analyzing large amount of gene expression data from microarray images can play a very important role in biology medicine, especially in cancer diagnosis. However, microarray images are usually contaminated with noise and artifacts, for example, low intensity and poor quality spots. A crucial step in the microarray image processing is gridding, which consists of global gridding (sub-array detection) and local gridding (individual spot detection). Therefore, the first proposed technique in this thesis is a fully automated gridding approach for cDNA microarray images. The proposed technique includes both global gridding and local gridding. Our technique is developed based on multi-resolution analysis and a new adaptive threshold method. The proposed gridding approach is fully automated in the sense that it does not require any user intervention and the only input needed is the microarray image. The presented technique can be applied to images with different specifications such as resolution, number of sub-arrays, number of spots in each sub-array, and noise levels. The experimental results show that the proposed method is highly accurate when compared with existing software tools as well as with recently published techniques. Our results also show that the presented approach is very effective for gridding microarray images with low intensity, poor quality spots, and missing/irregular spots. The spot detection accuracy of the proposed method is improved by up to 5.48% compared with that of the other published algorithms.

The second proposed technique in this thesis is a microarray cancer classification method which is developed using discrete wavelet transform (DWT) for data reduction and genetic algorithm (GA) for feature selection. One of the most widely used applications for microarray data is the classification of different types of cancer according to their gene expression profiles. Microarray cancer data classification has been typically a challenging task for machine learning researchers. This is mainly due to the complexity and the high dimensionality of the associated feature space despite the fact of limited availability of training data size. Accordingly, dimensionality reduction of data is significant methodology to greatly improve the classification performance of microarray data. Hence, this work presents a new methodology for the reduction of the number of features employing DWT and a new technique for feature selection of microarray data using GA. This proposed methodology is used to identify the minimum set of features with the most relevant impact within a large population of features. In order to benchmark the quality and performance of our newly proposed method, we adopted several models of wavelet decomposition levels and various types of wavelet algorithms. In addition, we compared our method with other published cancer classification methods using standard microarray data sets. The experimental results show that the proposed technique outperforms the other published cancer classification algorithms in most of the cases that have been investigated.