

A NEW RBF NEURAL APPROACH FOR SPOT CLASSIFICATION IN DNA-MICROARRAY IMAGES

S. COCO, G. CAPIZZI, A. LAUDANI* and S. PELLEGRINO

*DIEES, University of Catania,
Catania, I-95125, Italy*

**E-mail: alaudani@diees.unict.it*

E. ALESSI

STMicronics

Abstract.

Spots in DNA-Microarray images exhibit several characteristics able to uniquely identify them for diagnostic purposes. In order to perform efficiently their classification a new tool based on Radial Basis Functions Artificial Neural Networks is presented. Classification is performed according to the following predefined spot classes: saturated spot, normal spot, donut spot, noisy saturated spot, noisy gaussian spot, noisy donut spot and background. These classes have been selected according to the more frequently observed morphology of spots. The neural system shows very good performance in terms of positive classification in typical DNA-Microarray images.

Keywords: DNA-Microarray; image processing; RBF neural networks; spot classification.

1. Introduction

DNA Microarray technology gives the possibility of identifying the genes involved in specific pathologies. A huge number of images are generated during DNA analysis experiments. DNA Microarray images have a recursive morphological structure characterized by a number of spots lined up in rows and columns. Spots represent results of DNA hybridization experiments and ideally should have a circular shape. Moreover microarray images from different laboratories may have different formats; spots dimensions, signal intensity and background features may vary from one image to another, in addition spots distribution may be different from one grid to another. In particular, an error occurring during the experiment or the acquisition phase due to several imperfections causes a variation of the typical aspect of images, see fig. 1.

For this reason a preliminary image analysis is required in order to properly determine spots within the DNA Microarray image.^{1,7,8,6,2,3,5,4} After the spots are extracted classification is performed according to a suitable number of predefined classes. Standard spot classification procedures, currently used for DNA microarray images, often do not lead to satisfactory results, due to significant cases of misclassifications. Hereafter a new system for the spot classification of DNA microarray images is presented based on a Radial Basis Functions Artificial Neural Network (RBF ANN). Classification is performed according

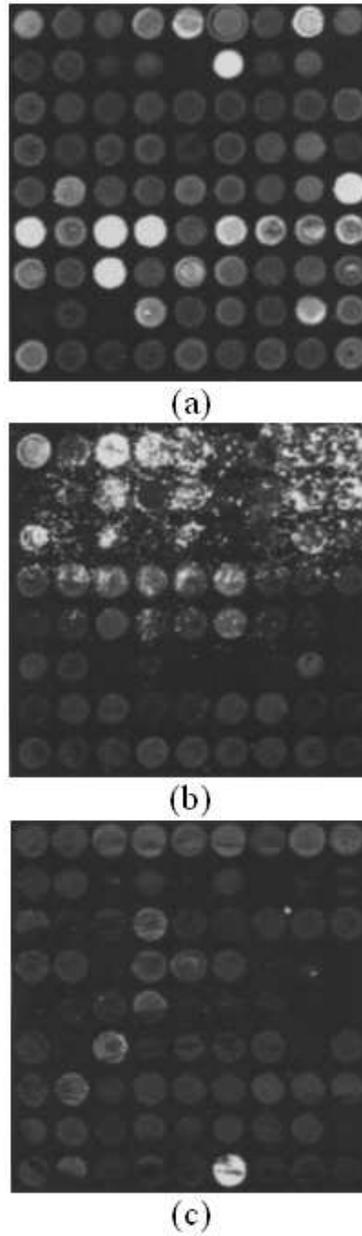


Fig. 1. DNA-Microarray images: (a) good quality; (b) noisy image; (c) striped image

to the following predefined spot classes: saturated spot, normal spot, donut spot, noisy saturated spot, noisy gaussian spot, noisy donut spot and background. These classes have been selected according to the more frequently observed morphology of spots in DNA Microarray images. The RBF classifier features have been tailored in order to balance requirements of efficiency and moderate computational effort.

2. The RBF Classifier

Before classification, an DNA microarray image analysis procedure is executed in order to determine spots within the image. The image analysis procedure extracts spots parameters such as topological data and structure data. In particular spots are localized

and segmented; moreover the spot intensity level and the background intensity level are determined. The analysis procedure is articulated into the following steps:

- Array localization: the localization consists in the extraction of the spots from the original image.
- Intra-spot segmentation: in this step, signal pixels of the extracted spots (true signal) are recognized and separated from the background signal (false signal)
- Spot quality measurement: for every spot it extracts the intensity value that represents a measure of the concentration of the interbred cDNA situated on the DNA chip. The quality measure considers the intensity median values of the pixels belonging to the class True Signal for a particular spot.
- Background treatment: In many Microarray images, the background takes a large quantity of memory space. But the information processed to extract DNA's information is contained inside the spots area. So it is not necessary to represent the background accurately, with all details about the noise, but it is sufficient to consider a mean background value.

After the image analysis has been completed, classification begins; it is articulated into two phases. The Classifier first operates identification, which is performed by resorting to the morphological structure. For this purpose a certain number of candidate classes are defined according to user choice. For the DNA Microarray images analyzed it has been found that seven classes are the best trade off between complexity and performance. The classes are: saturated spot, normal spot, donut spot, noisy saturated spot, noisy gaussian spot, noisy donut spot and background and are shown in fig. 1.

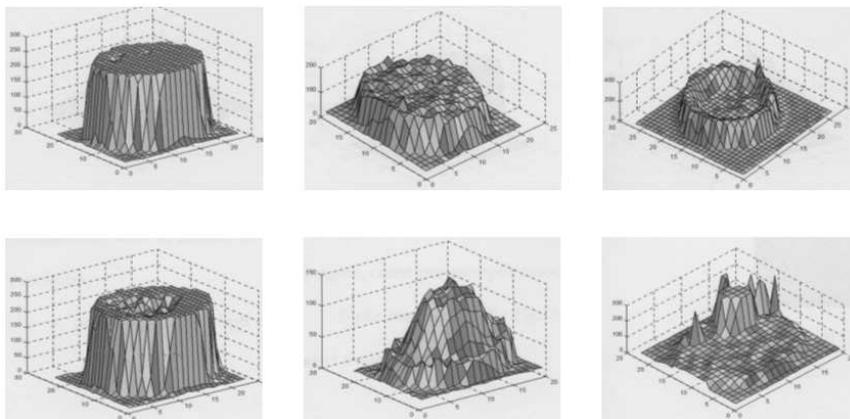


Fig. 1. The six classes of spots used for classification

The saturated spots exhibit a morphological structure similar to a frustum of cone, intensity values are higher than a given threshold. Gaussian spots also have a similar shape but intensity values are lower than a given threshold. They have an uniform profile. The donut spots exhibit a cavity near the center, with values of intensity lower than a given threshold, while they are saturated at the spot edge with higher values of intensity so that their structure is similar to a torus. The noisy saturated spots are morphologically similar to the saturated one, although the structure is jagged by the presence of noise added to the useful signal. The noisy gaussian spots are morphologically similar to the gaussians, although the structure is jagged by the presence of noise added to the useful

signal. The noisy donut spots are morphologically similar to the donuts, although the structure is jagged by the presence of noise added to the useful signal. The background spots introduce a constant profile, uniform and with low intensity. The peaks represent additive noise. The classification system is based on a RBF ANN having gaussian radial basis functions for its neurons. The number of neurons employed is about 1/3 of the number of the training patterns. The set of features necessary to identify the spots has been created on the basis of spot morphology; it is composed of the following data: Mean grey levels of the circular ring of the spot, Hausdorff dimension of the spot. The circular ring of the spot is the set of all pixels having equal distance from the center of the spot. Several values of distances are considered. In our classifier 15 values of distances have been used according to the spot morphology. The training set for the RBF NN is composed of 70 spots, 10 for each class. The RBF NN training is articulated into two phases: in the first one the centroids of the RBF are individuated by using the well-known K-means clustering algorithm, in the second one the ANN weights are calculated by resorting to the pseudo-inverse matrix.

3. An example of spot classification

In order to evaluate the classifier performance we have considered several TIFF DNA microarray images. Hereafter the spot classification of four DNA-Microarray images are reported; each image contains 608 spots. Two kinds of classifications have been performed: in the first kind the spots have been classified according to the 7 predefined classes, obtaining a correct classification for the 95%, of spots while in the second kind of classification the spots have been classified as valid and not valid reaching a correct classification of 99% of spots.

4. Conclusions

The presented system of classification for DNA microarray image based on a RBF neural network analyzes the spot matrix and classifies each spot by assigning it to a pre-defined class in an efficient way. In this classifier seven different spots classes have been defined according to their morphological structure. Nevertheless, this classification can be extended adding other spot classes. The performance of the developed system is very satisfactory obtaining more than 90% of correct spot classifications.

REFERENCES

1. K. Falconer. *The Geometry of Fractal Sets*. Cambridge University Press, Cambridge, 1985.
2. M.T.Musavi and al. On the training of radial basis functions classifiers. *Neural networks*, 5:595–6039, 1992.
3. J.Chandra Patra and al. Neural network-based analysis of dna microarray data. In *Proceedings of IEEE International Joint Conference on Neural Networks, 2005 (IJCNN 2005)*, pages 503–508, August 2005.
4. D. Peterson and C.H. Lee. A dna-based pattern recognition technique for cancer detection. In *Proc. of 26th Annual International Conference of the Engineering in Medicine and Biology Society, (EMBC 2004)*, pages 2956–2959, 2004.

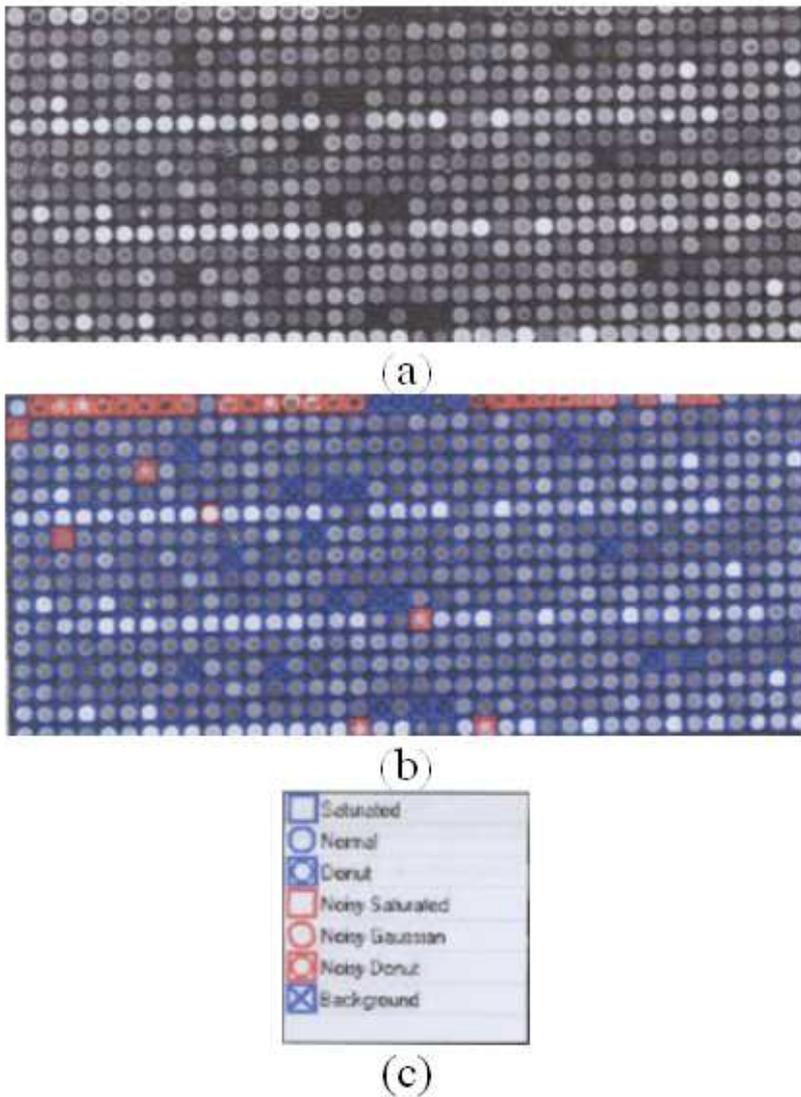


Fig. 1. . DNA-Microarray images: (a) an image used for the test; (b) output of classifier (c) legend

5. L.E. Peterson and al. Artificial neural network analysis of dna microarray-based prostate cancer recurrence. In *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, (CIBCB 2005)*, pages 1–8, November 2005.
6. S. L. Salzberg. *Gene Discovery in DNA Sequences*. The Institute for Genomic Research, IEEE, 1999.
7. M. Schena. *DNA Microarrays: A practical approach*. Oxford University Press, Oxford, 1999.
8. J. Schurmann. *Pattern Classification: A Unified View of statistical and Neural Approaches*. Jonh Willey & Sons, 1996.