ELSEVIER

Pattern Recognition Letters 18 (1997) 1205–1210

Pattern Recognition Letters

# Application of neural networks and genetic algorithms in the classification of endothelial cells [1]

Sameh M. Yamany [*], Kamal J. Khiani, Aly A. Farag [2]

*Computer Vision and Image Processing Lab, Department of Electrical Engineering, University of Louisville, Louisville, KY 40292, USA*

**Abstract**

This paper deals with developing automatic classification algorithms for identifying images of F-actin distribution in endothelial cells with different treatment of drugs. A novel combination of genetic algorithm based feature selection, used to obtain an optimal feature set, and artificial neural network methods including the multilayer feedforward network and the functional link network was used in the classification. © 1997 Elsevier Science B.V.

*Keywords:* Neural networks classification; Genetic algorithms; Endothelial cells; Feature selection; Functional link; Neural networks

## 1. Introduction

Endothelial cells are thin, flattened cells that line the vasculature in a single layer. They have a cytoskeleton that is composed of F-actin. A primary function of vascular endothelium is the mediation and control of water and solute exchange between blood plasma and the interstitial fluid. During inflammatory and immune reactions, the endothelial cells undergo a number of functional and morphological alterations caused by the agonists. Many of these agonists are mediators during inflammation and serve to increase permeability of the endothelial cell. This increase in permeability is associated with cytoskeletal F-actin changes in the endothelial cells along with the formation of intercellular gaps (Khiani et al., 1996). Fig. 1 shows the results of the agonists on the endothelial cells.

Developing automatic classification of the effects of agonists on endothelial cell F-actin distribution could be used as a screening process for other agonists. Current techniques use conventional methanolic extraction method to accurately classify the agonists effects. These techniques are usually expensive and time consuming. However it was noticed that each agonist can be characterized by the distribution of the F-actin fibers within the cell. Another challenge for the automatic classifier was the sensitivity of the image acquisition process to some parameters such as the current temperature, percentage of the agonist dosage in the image area, light intensity, randomness of agonist distribution in the plate, the photobleaching effect of the fluorescent dye and the aging effect of the agonist. Khiani et al. (1996), compared the performance of many of the classical classifiers to solve such a problem and they concluded that the neural network was found to perform best due to its generalization capability and robustness. The specific agonists used in this study are

[*] Corresponding author. E-mail: yamany@cairo.spd.louisville.edu.

[1] This work has been partially supported by grants from the Whitaker Foundation and the Center for Applied Microcirculatory Research, the University of Louisville.

[2] E-mail: farag@cairo.spd.louisville.edu.


*PII* S0167-8655(97)00140-2

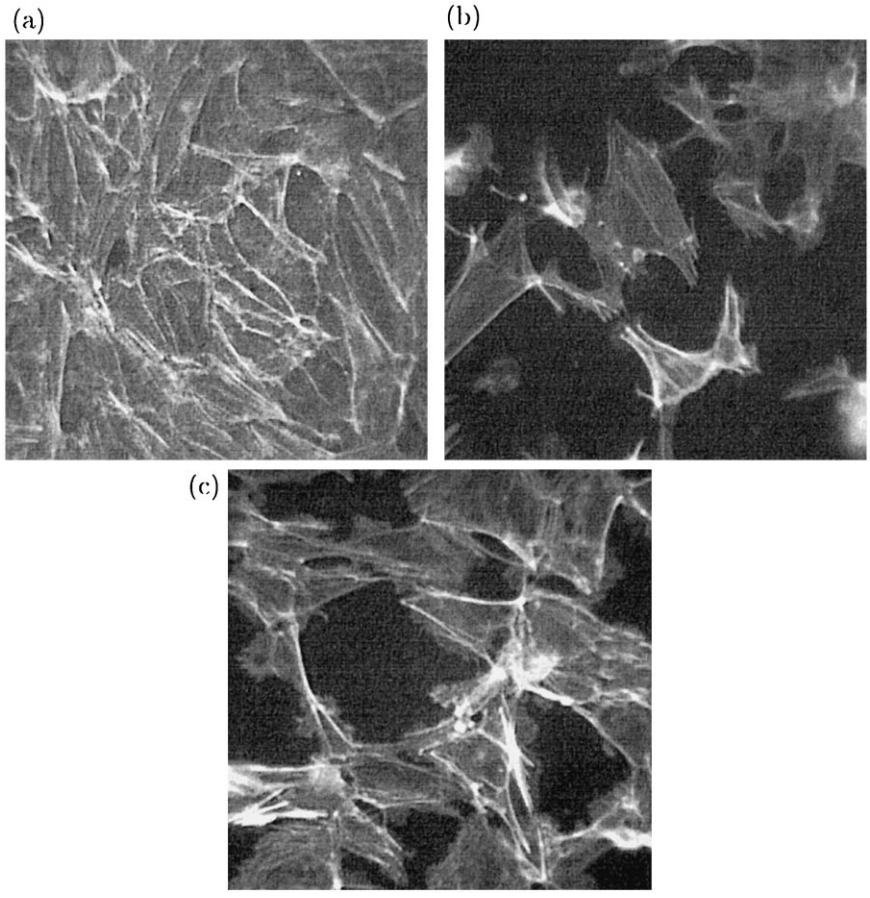


Fig. 1. Samples of control (a), $\alpha$-thrombin (b) and histamine (c) induced endothelial cells.

$\alpha$-thrombin and histamine. Many features, including first-order and second-order texture features, were extracted from the image and a genetic algorithm (G.A.) based feature selector was used to obtain the optimal set of features. Multilayer feedforward neural networks and functional link neural networks were compared in the classification.

## 2. Feature extraction

Human umbilical cord vein endothelial cells (HUVEC) were grown to confluence on Transwell membranes. The endothelial cells were plated with $\alpha$-thrombin and histamine. BODIPY-Phalloidin, which is a fluorescence dye, binds to intracellular F-actin and can therefore be used as an indicator of the relative distribution of the cytoskeletal protein (Ehringer et al., 1996). Cellular F-actin was photographed using a Carl Zeiss Axiovert 100 Microscope equipped with a mercury lamp and a rhodamine filter set. Thirty plate were used for each class for a total of 90 plates. From each plate five $256 \times 256$ gray images were digitized for a total of 450 images. The five images were digitized from random areas in the plate. These images were obtained over several weeks and at different temperature, light and agonists concentration. We used 225 images for training and 225 for testing.

The agonists effect on the endothelial cell is characterized by the migration of the F-actin fibers from within the cell and from the cell boundaries.

For each agonist, this redistribution of F-actin is a key feature for classification. The migration of these F-actin fibers results in the formation of textured regions in some parts of the image. First-order and second-order texture features were considered.

The second-order features describe the gray level spatial inter-relationships and measure gray level texture homogeneity (Kadah et al., 1996). These features were obtained from the spatial gray level dependence method (SGLDM) described by Haralick et al. (1973). The SGLDM is based on the estimation of the second-order joint conditional probability density function, $\Psi(i,j \mid d,\theta)$. Each $\Psi(i,j \mid d,\theta)$ is the probability of going from gray level $i$ to gray level $j$, given the intersample spacing $d$ and angle $\theta$. The estimated values can be written in matrix form, the so-called co-occurrence matrix. The unnormalized co-occurrence matrix, $\Psi$, is a function of the image $f$, a displacement vector $d = [\Delta i, \Delta j]$ and an angle $\theta$ over the domain $D$ and is represented in the following form:

$$\begin{aligned}\Psi_{ij}(f,d) = \#\{&((k_1,l_1),(k_2,l_2)) \mid \\ &(k_1,l_1),(k_2,l_2) \in \mathscr{D}, \\ &f(k_1,l_1) = i, f(k_2,l_2) = j, \\ &[k_2,l_2] - [k_1,l_1] = d\}, \end{aligned} \quad (1)$$

where # indicates the number of elements in the set. The value of $d$ used was 16. In general, symmetric co-occurrence matrices are considered. The joint probability densities of the co-occurring gray levels are approximated by normalizing the co-occurrence matrices. This normalization is achieved by dividing each entry of the co-occurrence matrix by the total number of paired occurrences in the image. Formally the angles are quantized to 45° intervals.

Haralick et al. (1973), has defined 14 statistical measures of texture which can be extracted from the co-occurrence matrices. These features contain information of image texture characteristics such as homogeneity, gray-tone linear dependencies, contrast, number and nature of boundaries present, and the complexity of the image.

First-order features such as the mean and variance of the normalized gray level histogram, the concentration, and other histogram parameters were also considered.

A total of 16 features including both the first-order and the second-order features, were extracted from the images. Some of these features were sensitive to the image acquisition process (e.g. temperature, light, photobleaching effect, etc.), however a combination of these features can be insensitive to the acquisition process. In order to obtain the minimum subset that gives the best classification while being insensitive to the image acquisition process, we used a G.A. feature selection algorithm. This algorithm is demonstrated in the next section.

## 3. Feature selection using genetic algorithms

Feature selection is one of the important stages in the design of a classifier. As the number of features used determines the performance of the classifier, also the sensitivity of these features determines the robustness of this classifier. Finding the best feature subset requires searching in the space formed by all the possible feature combinations. To obtain the optimal combination, this search could be very complex and time consuming.

According to Goldberg (1989), Genetic Algorithms are best known for their ability to efficiently search large spaces about which little is known, they seem to be an excellent choice for a search method for selecting features used in a recognition system as proposed by Vafaie and De jong (1992). The proposed GA-based method for transforming an initial feature set into a more useful one is shown in Fig. 2.

The novelty in our approach is the representation of the features in the GA process and the choice of the evaluation function. The features are represented as a long binary string where each bit corresponds to the presence/absence of this feature in the classifier. Each generation in the GA is formed of a population of 50 strings with crossover of 70% and mutation of 1%. The order of features in the string is also an important factor. As described by Goldberg (1989), the GA is based on the concept of schemata. A schemata $H$ is a similarity template describing a subset of strings with similarities at certain string positions. In general for string of length $l$ and $k$ alphabet there are $(k+1)^l$ schemata. The survival of any schemata depends on the average fitness, the order of the schemata $O(H)$, which is the number of

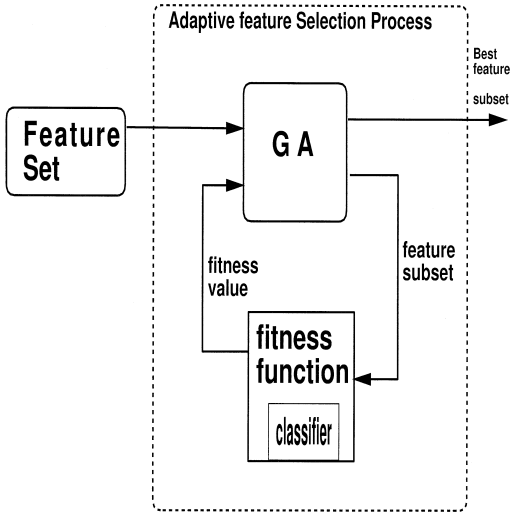


Fig. 2. GA based feature selection diagram.

fixed positions in the schemata, and the length of the schemata $d(H)$, which is the difference between the locations of the first and the last specific string position.

Suppose at time $t$, in a population $A(t)$ we have $m$ examples of a particular schemata, $H$; we will denote it by $m_t(H)$. A string $A_i$ is selected according to a probability $p_i$ defined as $p_i = f_i/\sum f_i$. Hence, if a non overlapping population of size $n$ is selected then we expect to have $m_{t+1}(H) = m_t(H) n f(H)/\sum f_i$ where $f(H)$ is the average fitness of $m_t(H)$. Let $f_a = n^{-1}\sum f_i$, then $m_{t+1}(H) = m_t(H) f(H)/f_a$. A schemata $H$ is destroyed by crossover with probability $p_d$ defined as $p_d = d(H)/(l-1)$, or it survives with $p_s = 1 - p_d$. If crossover is random with probability $p_c$, then $p_s \geq 1 - p_c d(H)/(l-1)$. So after crossover and mutation we expect

$$m_{t+1}(H) \geq m_t(H)\frac{f(H)}{f_a} \times \left(1 - p_c\frac{d(H)}{(l-1)} - O(H)\,p_m\right), \qquad (2)$$

where $p_m$ is the mutation probability. Therefore, the probability that a certain schemata will survive is determined by the fitness of the individuals in this schemata. If this fitness is high, which means it contains good solution to the problem, the probability of surviving is high. Also if the length of the schemata is small, the probability of surviving is higher. For this reason, the features are ordered in the string according to their importance, otherwise it would take a longer time for the GA to converge. The conversion of the GA in our approach is determined by the average fitness of each generation and the stopping criteria is when this average reaches above 90%.

Our proposed GA structure uses the following evaluation function, $f = (1 - E_c)\cdot(X - F_s/F_t)$, where $F_t$ is the total number of features, $F_s$ is the number of subset features, $E_c$ is the classifier error when using the feature subset $F_s$, and $X$ is any number greater than one. The parameter $X$ is used to tune the process to compromise between minimizing the number of features in the subset and maximizing the classification rate. Using this method with the sixteen features mentioned in the previous section and the 225 training images, we obtained a feature subset containing only four features: two first-order features and two from the second order statistics.

One of the first-order features used was the mean gray level and the second was the tenth percentile representing the gray level below which lies 10% of the total number of pixels in the image. Correlation was a second-order feature that was selected with the GA optimization

$$\text{Correlation} = \frac{\sum_i \sum_j (ij)\,p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}, \qquad (3)$$

where $\mu_x$, $\mu_y$, $\sigma_x$ and $\sigma_y$ are the means and standard deviation of $p_x(i)/p_y(j)$ which is the $i$th/$j$th entry in the marginal-probability matrix obtained by summing the rows/columns of $P(i,j)$. This feature measures the gray tone linear-dependencies in the image. The other second-order feature was the *Entropy* which is defined as

$$\text{Entropy} = -\sum_i \sum_j P(i,j)\log P(i,j). \qquad (4)$$

The entropy is a measure of the randomness of the image texture.

This GA feature selection algorithm is also very useful when the initial set contains a huge number of features. In such a case the algorithm reached the

best subset in just a few iterations. Using an exhaustive search, it was confirmed that the subset obtained from the GA feature selection algorithm was the optimum.

## 4. Multilayer and functional link neural network

One of the popular implementations of neural networks in pattern recognition is the multilayer feedforward neural network with an error back-propagation training algorithm (EBPTA) which is described by Zurada (1992). When using multilayer feedforward neural networks it is always difficult to know exactly how many hidden layers and how many neurons in each hidden layer will lead to the best classification results. Instead of this scheme, input/output mapping may also be achieved with an artificially augmented single-layer network. The advantage of this is that the number of neurons is forced to be the number of classes and the simple delta learning rule is used instead of the multilayer general delta learning rule. This is the concept of what Pao and Takefuji (1992) called *functional link network*. Functional link neural networks are single-layer networks that are able to handle linearly non-separable classes due to the dimensions of the inputs being increased by using nonlinear combinations of the input features. In the so-called *tensor model* the additional input terms are obtained for each of the input feature vectors as the product $x_i x_j$ for $\forall i, j \in [1, D]$ where $D$ is the dimension of the input vector. In our approach we only used the additional six inputs formed by multiplying all possible pairs. Due to its intrinsic mapping properties, the functional link network in some cases can perform better than the multilayer network and its distinct advantage is the easier training.

## 5. Results

The classifiers were trained with 225 images and tested with the other 225 images. The classification rate is based on correctly classifying the agonist in the plate and this is done by choosing the agonist with higher number of classified images in each plate. Different multilayer and functional link topologies were investigated with different training parameters. For the multilayer neural network, hidden layers of 3, 4 and 5 neurons were considered with different $\eta$ learning constant values. The 4-3-3 network obtained correct classification of 91% with $\eta$ values of 0.4 and 0.8. The difference with $\eta = 0.8$ was the reduction in the number of training cycles. The functional link neural network using the tensor model was a 10-3 (input,output) architecture.

Table 1
Results of classifying (MSE: Mean Square Error)

| Topology | $\eta$ | MSE | Correct plates (out of 45) | % correct images |
|---|---|---|---|---|
| F.F. 4-3-3 | 0.4 | 0.00118 | 41 | 81.3 |
| F.F. 4-3-3 | 0.8 | 0.00092 | 41 | 84.4 |
| F.F. 4-4-3 | 0.4 | 0.00069 | 42 | 91.5 |
| F.F. 4-4-3 | 0.8 | 0.00024 | 41 | 90.6 |
| F.F. 4-5-3 | 0.4 | 0.00024 | 42 | 93.7 |
| F.F. 4-5-3 | 0.8 | 0.00024 | 42 | 93.7 |
| F.L. 10-3 | 6 | 0.04560 | 43 | 95.5 |

Table 1 shows the results of the classifying the 45 test plates. The classification error resulted from some overlap between classes. This is because a few of the images obtained from the well plates had variations due to the non-homogeneous effects of the agonists on the cytoskeleton of the endothelial cells. Although the agonists are applied to the entire well, it cannot be assumed that the entire area was equally affected and this leads to some variations of the images. However, it is clear that the neural network has the power to learn and generalize the recognition of the F-actin distribution.

The features presented in this work reflect the global texture of the acquired image. Extensions to this work involve the use of features that represent the local texture characteristics of individual cells in the image (e.g. (Steier and Farag, 1997)).

## 6. Conclusion

Automatic classification of the F-actin distribution in endothelial cells treated with agonists was accomplished using neural network classification techniques. The GA feature selection process was successful in selecting the features subset for better

classification. These histogram and texture features were able to characterize the F-actin distribution and other phenomena such as pores and gaps. This enables accurate discrimination of the F-actin distribution on endothelial cells as a result of the agonists. Future work involves other agonists which may have similar F-actin distribution like the agonists used in the present experiment.

## References

Ehringer, W.D., Edwards, M.J., Miller, F.N., 1996. Mechanisms of $\alpha$-thrombin, histamine, bradykinin induced endothelial permeability, J. Cellular Physiology.

Goldberg, D.E., 1989. Genetic Algorithms: In Search, Optimization and Machine Learning. Addison-Wesley, Reading, MA.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Systems Man Cybernet. 3 (6).

Kadah, Y.M., Farag, A.A., Zurada, J.M. Badawi, A.M., Youssef, A.M., 1996. Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. IEEE Trans. Medical Imaging 15 (4).

Khiani, K.J., Yamany, S.M., Farag, A.A., Ehringer, W.D., Miller, F.N, 1996. Classification of the effects of F-actin under treatment of drugs in endothelial cells. In: Proc. ANNIE-96, Rolla, MO.

Pao, Y.H., Takefuji, Y., 1992. Functional-link net computing: Theory, system architecture, and functionalities. Computer 25 (5), 76–79.

Steier, K.L., Farag, A.A., 1997. F-actin distribution in endothelial cells. Tech. Rept. TR-CVIP 97-2.

Vafaie, H., De Jong, K., 1992. Genetic algorithms as a tool for feature selection in machine learning. In: Proc. 4th Internat. Conf. on Tools with Artificial Intelligence, Arlington, VA, pp. 200–204.

Zurada, J.M., 1992. Introduction to Artificial Neural Systems. West Publishing Company, St. Paul, MN.