Signal Discrimination in Fluorescence In-Situ Hybridization Images

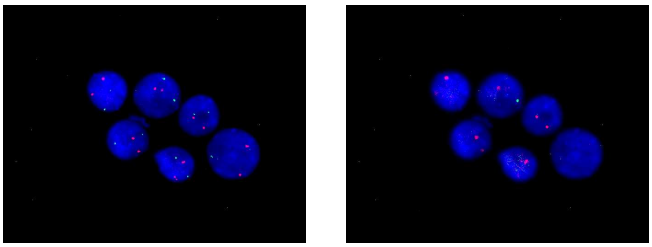
Boaz Lerner BOAZ@EE.BGU.AC.IL
Department of Electrical & Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel

Abstract

Fast and accurate analysis of signals in fluorescence in-situ hybridization (FISH) images is required for detecting genetic abnormalities. The analysis depends upon a classifier to discriminate between artifacts and valid signals of several fluorophores (colours), and well discriminating features to represent the signals. We evaluate feature sets by illustrating the probability density functions and scatter plots for the features, and using two class separability criteria: a scatter criterion and the classification accuracy. The same classification accuracy is the ultimate criterion in evaluating the classifier. Represented by the recommended features, up to 90% of valid signals and artifacts of two fluorophores are correctly classified using a neural network (NN) based hierarchical strategy. The classifier accuracy is found to be comparable with that of the support vector machine and slightly inferior to that of the Bayesian NN. Although applied to cytogenetics, the paper presents a comprehensive, unifying methodology of qualitative and quantitative evaluation of pattern feature representation that may be applicable to other real-world pattern recognition problems.

1. Introduction

In recent years, fluorescence in-situ hybridization (FISH) has emerged as one of the most significant new developments in the analysis of human chromosomes. FISH offers numerous advantages compared with conventional cytogenetic techniques since it allows numerical chromosome abnormalities to be detected during normal cell interphase. One of the most important applications of FISH is dot counting, i.e., the enumeration of signals (also called dots) within the nuclei, as the dots in the image represent the inspected chromosomes. For an accurate estimation of the distribution of the number of chromosomes over cell population, especially in applications involving a relatively low frequency of abnormal cells, large numbers of cells are needed to be examined. As visual evaluation of large numbers of cells and enumeration of hybridization signals is very tedious, laborious and time-consuming, FISH analysis for dot counting can be expedited by using an automatic procedure (Netten *et al.*, 1997).



(a) (b)

Figure 1. Two FISH images used in dot counting taken at the same field-of-view but at different focal planes: (a) in and (b) out-of-focus. In the original coloured image, chromosomes 13 and 21 were indicated by green and red signals, respectively, while the nuclei were coloured in blue.

To perform dot counting, an automatic system usually selects the 'sharpest' image along the Z-axis of each field-of-view (FOV) using auto focusing mechanism. Employing this mechanism, however, extends the length of the analysis and is prone to errors (Netten *et al.*, 1997). We suggest to dispense with autofocusing mechanism and to base FISH dot counting on a neural network (NN) classifier discriminating between in and out-of-focus images taken at different focal planes of the same FOV (Figure 1). Each image of a FOV is analysed and each image signal is classified by the NN as real (focused) signal or artifact, which is the result of out-of-focusing. Following the discrimination of valid signals and artifacts in each image, the image that contains no artifacts is selected as the in-focus image to represent the FOV. Proportion estimation of cells having different numbers of signals can be then performed using in-focus images captured from the sample. This estimation provides the cytogeneticist indication for deviation from normality,

e.g., excess or deficit of a chromosome. The suggested method shortens the length of image acquisition, as images are captured coarsely without the necessity to find the exact location of the in focus image and since selected image objects rather than the whole image are analysed. Combining with multi-spectral analysis (Section 2), the suggested methodology also shortens the length of image analysis. However, as the system is required to classify real signals and artifacts, its ability to discriminate between focused and unfocused signals should be more accurate than that of the discriminating element of a system employing an auto-focusing mechanism, as the latter encounters only valid signals. Therefore, the proposed system depends upon two interrelated components: a highly-accurate classifier to distinguish between valid and artifact signal data, and well-discriminating multi-feature signal representation. Accurate classification will be achieved when both components are successful. Indeed, we study here feature representation for FISH signals and the use of a classifier to discriminate between focused (real) signals and unfocused (artifact) signals. Different classification strategies, some of them hierarchical, based on the two-layer perceptron neural network (NN) are employed for the classification.

Section 2 of the paper describes FISH image analysis and feature measurement, while Section 3 presents signal classification. The results of experiments to evaluate the features and different classification strategies are given in Section 4, while Section 5 concludes the paper with a discussion.

2. FISH image analysis

2.1. FISH multi-spectral image processing and segmentation

The process of preparing, hybridizing and screening FISH samples, as well as the procedure of capturing FISH images are described in Lerner *et al.* (2001a). A total of 400 images are collected from five slides and stored in TIFF format.

Generally, special multi-stage (e.g., TopHat-based) procedures that rely on heuristically-derived thresholds and parameters are conventionally applied to the intensity image in order to segment nuclei and signals (Netten *et al.*, 1997). Multi-spectral image processing and segmentation, however, avoids the use of these procedures. By analysing each of the three colour channels – red, green and blue (RGB) of a FISH image separately, processing and segmentation can be facilitated. Nuclei are analysed using the blue channel, whereas red and green signals are analysed using the red and green channels, respectively. Finding 'optimal' global thresholds in the RGB image is almost trivial compared with thresholding an intensity image since the channels contain no background and only blue (red, green) objects are found in the blue (red, green) channel. Also, for these reasons, moderate changes in the threshold values barely affect the overall accuracy of image analysis. In this work, threshold values of 0.5 and 0.8 of the maximum channel intensity are found suitable for the segmentation of signals and nuclei, respectively. Following thresholding, noise reduction, boundary smoothing of the nuclei by morphological operations and spatio-spectral correlation between nuclei and signals are implemented to complete the segmentation of the nuclei and signals.

Finally, multi-spectral image analysis also yields hue-based features, which are found here efficient for FISH signal representation and classification. Furthermore, it allows the analysis of multiple fluorophores which is not commonly possible.

2.2. Signal feature measurement and evaluation

Following segmentation, sets of features are measured to represent the signals. The features include area, eccentricity and a number of spectral features. We compute, at the specific colour plane, three RGB intensity-based measurements: the total and average channel intensities and the channel intensity standard deviation. Following the conversion of RGB to HSI (hue, saturation, intensity) colour format, we can also compute four HSI based measurements: maximum hue, average hue, hue standard deviation, and delta hue, which are more appropriate for signal discrimination than RGB-based features. Delta hue is the difference between the maximum and average hue normalized by the average hue. This feature takes up values near zero for real signals and larger values for artifacts. Two additional features of the set are the two coordinates of the eigenvector corresponding to the largest eigenvalue of the red and green intensity components of the signal. The last feature is the signal average grey intensity, $I_1 = (R + G + B)/3$, where R, G and B are the intensities in the red, green and blue channels, respectively. Table 1 lists and numbers the twelve features to facilitate their identification in the rest of the paper.

Visual evaluation can provide preliminary insight into the relative merit of the features to the classification procedure, dependencies between the features and potential causes of misclassification. Feature evaluation using a class separability criterion can elaborate the visual analysis and quantify the importance of features

Table 1. The set of features studied in the work. Numbers are used in the rest of the paper to identify the features. Texture indicates standard deviation of intensity (5) or hue (8). Eig. 1, 2 are abbreviations for the two coordinates of the eigenvector corresponding to the largest eigenvalue of the red and green intensity components of the signal.

Number	Feature
1	Area
2	Eccentricity
3	Total Channel Intensity
4	Average Channel Intensity
5	Texture
6	Maximum Hue
7	Average Hue
8	Hue Texture
9	Delta Hue
10	Eig. 1
11	Eig. 2
12	Average Grey Intensity

and sets of features. The criterion can also be applied to feature selection, i.e., the selection of a (small) subset of the feature set yielding an accurate classification in minimal computational cost. In practical problems and for a not very large feature set, we can search among all the possible feature sub-sets and evaluate each one of them using the criterion. The sub-set that achieves the highest value of the criterion is then selected to represent the patterns to the classifier. For moderate and large feature sets, optimal (e.g., branch and bound algorithm) and sub-optimal (e.g., sequential forward selection algorithm) search methods, respectively, are generally employed (Devijver & Kittler, 1982).

The criterion of separability that is considered here, called J_1 (Fukunaga, 1990), is based on the within-class scatter matrix, $S_w = \sum_{i=1}^{L} P_i E\{(X - M_i)(X - M_i)^T|\omega_i\} = \sum_{i=1}^{L} P_i \Sigma_i$, and the between-class scatter matrix, $S_b = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T$, where $M_0 = E\{X\} = \sum_{i=1}^{L} P_i M_i$ is the mean pattern of the mixture distribution. $X|\omega_i$ are patterns of class ω_i $(i = 1, L)$ with mean M_i, covariance matrix Σ_i and a priori probability P_i. The criterion $J_1 = tr(S_w^{-1} S_b)$ is expected to be larger when the between-class scatter matrix is larger and/or the within-class scatter matrix is smaller.

Qualitative feature evaluation using conditional probability density functions and scatter plots, and quantitative feature evaluation using criterion J_1 and the NN classification accuracy are presented in Section 4.

3. Signal discrimination

One of the main purposes of this work is to demonstrate automatic signal discrimination in in and out-of-focus FISH images. Although the application of the research is to dot counting, we are not interested for the moment in estimating the proportions of cells having specific numbers of signals, but in the ability to accurately distinguish between real signals and artifacts. This ability forms the basis of the proposed dot counter.

The neural network (NN) classifier, which has been considered here, predicts a posteriori probability of class membership. We describe bellow a two-class NN which is appropriate for two of the three classifiers which have been used in this study. The third classifier, a multi-class NN is described in Bishop (1995).

Consider a training dataset D which consists of N data points with binary class labels $\{t_1 \dots t_N\}$ and vectors of inputs $\{\mathbf{x}_1 \dots \mathbf{x}_N\}$. We assume that the data was generated by some true underlying function $y(\mathbf{x})$. Our objective is to learn the parameters $\boldsymbol{\theta}$ of some approximating function $f(\mathbf{x}, \boldsymbol{\theta})$ whose form is dependent on our model choice, $\mathcal{M}$, so that we may make 'good predictions' about our class labels.

We define a *likelihood* function as Bishop (1995)

$$p(D|\boldsymbol{\theta}) = \prod_{n=1}^{N} f(\mathbf{x}_n, \boldsymbol{\theta})^{t_n} \left[1 - f(\mathbf{x}_n, \boldsymbol{\theta})\right]^{(1-t_n)}. \quad (1)$$

The approximating function, f, is represented by the output of an NN with H hidden nodes in its single hidden layer

$$f(\mathbf{x}_n, \boldsymbol{\theta}) = \sigma\left(\sum_{h=1}^{H} v_h g(\mathbf{u}_h^{\mathrm{T}} \mathbf{x}_n)\right). \quad (2)$$

The parameters $\boldsymbol{\theta}$ have been split into the input to hidden weights represented by H vectors $\mathbf{u}_h$, each vector being the weights that 'fan-in' to hidden node h, and $\mathbf{v}$, the vector of the hidden to output weights, consisting of H elements v_h. We have omitted biases for notational simplicity. We take the activation function g to be a hyperbolic tangent, and $\sigma(\cdot)$ is the logistic sigmoid function

$$\sigma(z) = \frac{1}{1 + \exp(-z)} \quad (3)$$

which constrains the output of the network to be between 0 and 1 allowing us to interpret f as the probability $P(C_1|\mathbf{x})$ that an input vector $\mathbf{x}$ belongs to class C_1.

We may now define an 'error function' as the negative log likelihood leading to the cross entropy error function

$$-\ln p(D|\boldsymbol{\theta}) = -\sum_{n=1}^{N}\{t_n \ln(f(\mathbf{x}_n,\boldsymbol{\theta})) + (1-t_n)\ln(1-f(\mathbf{x}_n,\boldsymbol{\theta}))\} \quad (4)$$

This error function may be minimised by a gradient-based optimisation method, e.g., the scaled conjugate gradient algorithm as in our case.

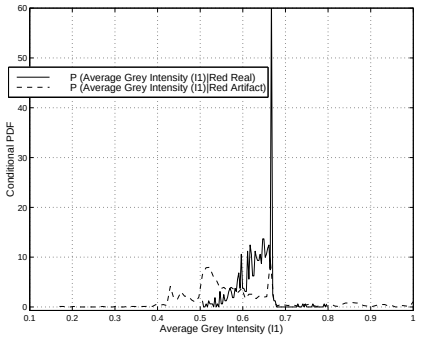
In the suggested methodology, signals (representing the red and green fluorophores) are classified into four classes: 'real red', 'artifact red', 'real green' and 'artifact green'. Within the 'artifact' classes we expect to find unfocused and overlap signals. Labels for the patterns, as belonging to one of the four classes, are needed to train and evaluate the classifier, and they are obtained by an expert cytogeneticist. Following the normalisation of the features to zero mean and unit variance, patterns of signals extracted from all the images are divided randomly into training and test sets and classification into one of the four classes is implemented using cross-validation. A validation set which is drawn from the training set assures that the classifier is not over-trained, and the 'optimal' configuration is selected. This guarantees rapid training and improved generalisation capability of the classifier.

Three NN-based classification strategies are examined here. In the first, called the 'monolithic strategy', patterns are classified into the four classes using a single NN. The two other strategies are hierarchical and based on the assumption that the classification problem can be considered as a two times two-class problem rather than a four-class classification problem. In the second strategy, termed the 'independent', patterns are classified into 'red' and 'green' classes using the 'colour network' and independently by a second network, the 'real network', into reals and artifacts. Classification of a pattern into the four classes is achieved by a common decision of both networks. In the third strategy, called 'combined', patterns are first classified into 'red' and 'green' classes using the 'colour network' and then based on the results of this network they are classified by two other networks, the 'real-red network' and the 'real-green network', into reals and artifacts of the two colours.

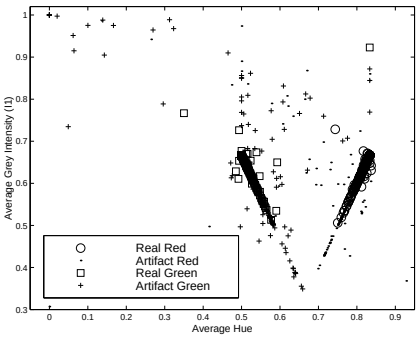
4. Experiments and results

We established a database of 400 in and out-of-focus FISH images, which were captured from five slides. Following segmentation, 3,144 objects within the nuclei were identified as signals and features were measured for them. Based on labels provided by expert inspection, 1,145 of the signals were considered as 'reals' (among them 551 were red) and 1,999 as 'artifacts' (among them 1,224 were red).

First, experiments to evaluate signal feature representations were held. Figure 2 demonstrates the application of two methods of visual evaluation, one using histogram estimates of conditional probability density functions (pdfs) and the second using scatter plots, to the features of Table 1. Figure 2a and Figure 2b depict, respectively, the conditional pdf of the average grey intensity (I_1) for red signals and a scatter plot of the average grey intensity (I_1) and the average hue. Similar graphs have also been derived for other combinations of classes and features. The large extent of overlap between histogram estimations for different classes demonstrates some of the expected difficulties of the application domain.



(a)

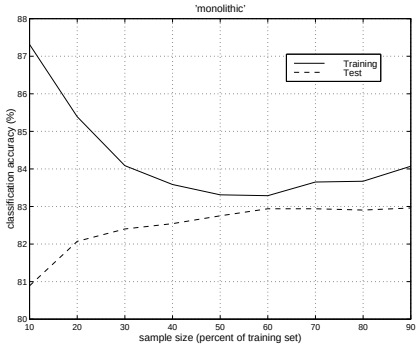


(b)

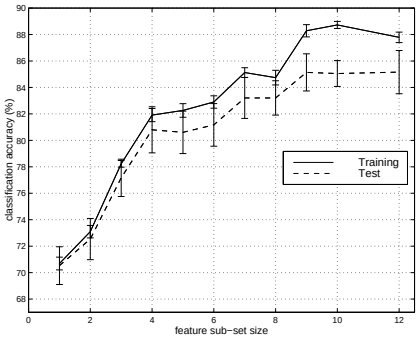
Figure 2. (a) A histogram estimate of the one-dimensional conditional pdf of the average grey intensity (I_1) for red signals, and (b) a scatter plot of the average grey intensity (I_1) vs. average hue.

The complete visual analysis (Lerner *et al.*, 2001b) has showed that in order to achieve accurate signal classification multi-feature representations are needed. Therefore, we employed two class separability criteria, scatter criterion J_1 and the probability of misclassification of the NN, in evaluating such representations. For the latter criterion we had to optimise the network configuration and study the network learning curves. Thus, experiments to find suitable configurations for the NNs of each of the three strategies were performed. Input and output dimensions of each of the NN-based classification strategies were determined by the feature space dimension and the number of classes, respectively. The number of hidden units was determined such that the network had the highest generalisation capability measured on a validation set drawn from the training set. Training of each of the networks, in each of the experiments reported here, was continued for 200 epochs and using three random network initialisations (a committee). The results were averaged over these initialisations as part of a cross-validation (CV-5) experiment. Each of the strategies, classifying signals represented by different feature sub-sets, was checked using its own optimal configuration. In addition, we examined the sensitivity of the classification accuracy of each strategy against the sample size by repeating the experiment for training sets of different sizes. The size of the training set was increased from 10% to 90% of the data, where the same unseen 10% of the data was used for the test. The results in Figure 3a demonstrate that the classification accuracy of the 'monolithic strategy' on the test set follows, as expected, the increase of the training sample size until its maximum level. However, the classification accuracy on the training set has a minimum. The explanation is that for a very small sample size, training is very simple and classification of a few training patterns can be very accurate. It is, however, more difficult to maintain this accuracy as the sample size increases and more variants of the training patterns are added. The classification accuracy, hence, decreases until it reaches a minimum for a 'critical mass' of learned patterns. After this point, as sample size continues to grow, the additional patterns are not so different from those of the 'critical mass'. Thus, learning of the patterns of the (extended) 'critical mass' is intensified, while at the same time the fraction of misclassified patterns becomes lower. The result of both trends is towards the improvement of the classification accuracy on the training set as is shown in Figure 3a.

Table 2 shows the results of feature evaluation using the two class separability criteria, J_1 and the classification accuracy, for several manually-selected three



(a)



(b)

Figure 3. (a) Classification accuracy of the 'monolithic' strategy for increasing sample sizes. (b) Classification accuracy (mean and standard deviation) of the NN-based 'monolithic strategy' for increasing sizes of feature sub-sets. Each sub-set includes the 'best' (according to J_1) features. The largest sub-set includes the entire feature set.

feature sub-sets. These sub-sets are based on the most discriminative single features found in previous experiments. Unseen signals represented by different combinations of features are classified as reals or artifacts of two colours with accuracies of up to 87.5% depending on the classification strategy. This accuracy increases to 89.2% when four feature sub-sets are examined (Lerner *et al.*, 2001b). The table also shows values of, and ranks according to, J_1 of the above feature sub-sets. Feature sub-sets consisting of hue features (maximum or average) and intensity features (average channel intensity or average grey intensity) are found to provide the best representations of the signals.

Table 2 demonstrates arbitrary selection of the size of the feature sub-set (3). We extended this examination by optimising the size of the sub-set and its content simultaneously. The 'optimal' size was determined by the highest classification accuracy of the 'monolithic strategy' evaluated for networks of different input sizes. The 'optimal' content for each sub-set size was set by criterion J_1. Figure 3b shows the

Table 2. The accuracy on the training (Tr.) and test (Tst.) sets of the three NN-based strategies classifying FISH signals represented by different combinations of features (numbered according to Table 1). This accuracy is also employed as a feature evaluation criterion compared with the rank a feature sub-set achieves (among 120 sub-sets) according to scatter criterion J_1.

Feature Combination	'monolithic'		'independent'		'combined'		J_1	Rank
	Tr.	Tst.	Tr.	Tst.	Tr.	Tst.		
4, 7, 12	79.0	78.2	78.5	77.7	81.9	81.4	1.7543	1
4, 5, 6	79.0	77.3	79.1	77.3	82.4	81.3	1.6789	2
6, 9, 12	80.4	79.3	79.9	79.0	83.8	83.4	1.2218	56
1, 4, 7	84.3	83.0	82.1	81.5	88.3	87.5	1.4958	14

classification accuracy obtained using the optimal features for each size of sub-set, and the corresponding NN configuration having the 'optimal' number of hidden units required to classify this sub-set, as determined on the validation set. For small sub-set sizes, the classification accuracy increases almost linearly with size. However, employing larger sub-sets only improves the accuracy moderately until the 'curse-of-dimensionality' deteriorates the results. The effect of the curse-of-dimensionality is even more evident after comparing the test highest accuracy (85.1%) achieved by the 'monolithic strategy' on the entire set (Figure 3b) with accuracies (higher than 86%) obtained using several manually-selected sub-sets of three and more features (Table 2). The success of the latter sets also hints to the inferiority of criterion J_1 compared with the classification accuracy in selecting optimal feature sub-sets for classification.

Finally, we compared the accuracy of the NN classifier with those of three other state-of-the-art techniques: Bayesian neural network (BNN), support vector machine (SVM) and naive Bayesian classifier (NBC), as well as with that of a linear classifier. We divided the classification task into two: classification of signals into colour (red or green) and classification of signals as 'real' or 'artifact'. This simplification determined indirectly the 'independent strategy' as the technique of choice to represent the NN classifier in the comparison. The configuration and parameters of each of the classifiers were determined on a validation set using a cross-validation experiment to enable peak performance of each of the techniques (Lerner & Lawrence, 2001). The entire feature set was employed without applying feature selection. The comparison shown in Table 3 reveals that the BNN is the most accurate (although not always significantly) classification technique for both tasks, and the NN and SVM are comparable and second best. The inferiority of the NBC compared with the other techniques is attributed to the relatively large amount of dependency among features of the set. This dependency violates the independence assumption of the NBC (John & Langley, 1995), and thereby decreases its accuracy. This result emphasizes the vital role of preliminary feature selection (or feature extraction performed as part of the classification process) in removing correlated features in order to facilitate pattern classification especially for the NBC.

5. Discussion

Auto-focusing is a long and critical step required for dot counting in FISH image analysis that upon failure undermines the whole analysis and leads to unreliable results. The alternative methodology suggested here, based on multi-spectral image analysis, well-discriminative features and signal classification, is an accurate and efficient screening mechanism for obtaining in-focus images necessary for FISH dot counting applied in the detection of genetic abnormalities.

Features have been evaluated by conditional pdfs and scatter plots providing preliminary visual insight into the classification procedure. Feature selection enabled the choice of feature sub-sets, which maximised scatter criterion J_1 measuring class separability. However, the ultimate and most reliable criterion for evaluating features for class separability in problems with non-parametric class conditional pdfs is the probability of misclassification. Mismatches in selecting optimal sub-sets by the two criteria can be attributed to the inferiority of J_1 for class patterns that have not equal covariance matrices and to its sensitivity to the relative locations of the classes in the feature space. In addition, the hidden layer of the NN classifier performs an additional feature extraction stage, which expands the classifier discrimination power beyond that of a scatter criterion.

As part of the classification process, extensive feature evaluation has been performed. This evaluation demonstrated the superiority of hue and intensity-based features for FISH signal classification. When features of the two families were combined together,

Table 3. Classification accuracies of five techniques measured on the FISH data discriminating real signals from artifacts (Real/Artifact) and red from green signals (Colour).

model	Real/Artifact (%)	Colour (%)
Neural Network (NN)	86.4	98.1
Bayesian Neural Network (BNN)	88.2	98.8
Support Vector Machine (SVM)	87.2	98.4
Naive Bayesian Classifier (NBC)	83.0	94.0
Linear Classifier	84.1	94.6

even a single hue feature could separate entirely signals of two fluorophores, leaving the task of discriminating real signals from artifacts to intensity features. Consequently, feature sets consisting of features of both families enabled an NN-based hierarchical strategy to classify nearly 90% of FISH signals as reals or artifacts of two fluorophores. The NN, accomplishing high classification accuracy with low computational requirements, provided performance comparable with those of other state-of-the-art classification techniques. As almost only the measured features are specific to this classification problem, the methodology presented here for a complete qualitative and quantitative evaluation of feature representations can also be applied to other real-world pattern classification problems.

Acknowledgment

This work was supported in part by the Paul Ivanier Center for Robotics and Production Management, Ben-Gurion University, Beer-Sheva, Israel.

Reference

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Clarendon: Oxford.

Devijver, P. A. and Kittler, J. (1982). *Pattern Recognition: A Statistical Approach.* New Jersey: Prentice-Hall International.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition (2nd ed.).* San Diego: Academic Press.

John,G. H. & Langley, P. (1995). Estimating Continuous Distributions in Bayesian Classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 338-345). San Francisco, CA:Morgan Kaufmann Publishers.

Lerner, B., Clocksin, W. F., Dhanjal, S., Hultén, M. A. & Bishop, C. M. (2001a) Automatic Signal Classification in Fluorescence In-Situ Hybridization Images. *Cytometry,* 43, 87-93.

Lerner, B., Clocksin, W. F., Dhanjal, S., Hultén, M. A. & Bishop, C. M. (2001b) Feature Representation and Signal Classification in Fluorescence In-Situ Hybridization Image Analysis. *IEEE Transactions on Systems, Man, and Cybernetics A*, 31, 655-665.

Lerner, B. & Lawrence, N. D. (2001). A Comparison of State-of-the-Art Classification Techniques with Application to Cytogenetics. *Neural Computing & Applications*, 10, 39-47.

Netten, H., Young, I. T., van Vliet, L. J., Vroljik, H., Tanke, H. J. & Sloos, W. C. R. (1997). FISH and Chips: Automation of Fluorescent Dot Counting in Interphase Cell Nuclei. *Cytometry*, 28, 1-10.