Improved Detection of Breast Cancer Nuclei Using Modular Neural Networks

Analysis of Nuclei in Histopathological Sections with a System that Closely Simulates Human Experts

The evaluation of immunocytochemically stained histopathological sections presents a complex problem due to many variations that are inherent in the methodology. In this respect, many aspects of immunocytochemistry remain unresolved, despite the fact that results may carry important diagnostic, prognostic, and therapeutic information. In this article, a modular neural network-based approach to the detection and classification of breast cancer nuclei stained for steroid receptors in histopathological sections is described and evaluated. The system, named biopsy analysis support system (BASS), was designed so that it simulates closely the assessment procedures as practiced by histopathologists.

Overview: Classification and Detection of Breast Cancer Nuclei

Breast cancer is the major malignancy affecting the female population in industrialized countries; it is estimated that one-third of these patients eventually die because of this disease. There is, however, a wide variation in the length of survival of individual patients, with some surviving several years following diagnosis. It has been widely recognized that patient survival is associated with clinicopathological assessment of several factors, better known as prognostic factors, which enable clinicians to predict individual patient survival and also choose appropriate modes of treatment.

Of prominent importance among prognostic factors is the hormonal status of tumor cells, which involves analysis for estrogen and progesterone receptors. Traditionally, these analyses were performed using biochemical methods [1, 2], but more recently rapid advances in immunocytochemistry allow the direct detection of receptors in routinely processed histological sections of tumor tissue [2]. In such preparations, however, the results are subjective and at best only be semi-quantitative; thus, immunocytochemical methods are disadvantageous as compared to biochemical methods, which render a quantitative result in units of receptors per weight of tissue [3,4]. Despite this drawback, immunocytochemical methods have gained wide acceptance because they are less costly, easier to perform, need small amounts of tissue, and most importantly, can be carried out on routine histological sections. In this respect, immunocytochemistry allows the simultaneous assessment of tumor morphology and hormonal staining to be performed on serial sections. In addition, this technique has revealed the existence of heterogeneous staining within tumor nuclei—a finding that was not previously apparent by biochemical analysis [2,5].

In order to improve the predictive accuracy of immunohistochemical data with regard to estrogen and progesterone receptors, several investigators have devised diagnostic schemes, such as the H-score [2] or the diagnostic index [6]. These schemes are based on the combined evaluation of two variables; namely, the staining intensity of individual tumor nuclei and the percentage of cells that are stained at each intensity class. The aim of these manual diagnostic schemes is to enable a semi-quantitative assessment of the microscopical images, the interpretation of which is only subjective when carried out routinely. Moreover, there is currently a major effort in standardization and quality assurance in histopathology [7]. In this context, commercial image-analysis systems...
tems, such as SAMBA and CAS, have been developed and are being applied for the quantitation of immunocytochemical images [8-12]. These systems, although dedicated, do not actually simulate the manual procedure employed by human experts, which involves the classification of individual tumor nuclei. CAS and SAMBA use a global approach to classify stained nuclei by introducing threshold levels that distinguish between specific staining and background (nonspecific staining). In this respect, the development and application of computer-aided systems such as BASS [13,14], which reproduce and enhance the experts’ ability to detect objects of interest (stained nuclei in this case) on an individual rather than a global basis, may enable a more quantitative assessment of immunohistochemical results, and therefore improve their predictive accuracy.

In a previous study [13], BASS was shown to be more sensitive than the combined experts in the detection of breast cancer nuclei (sensitivity, SS, 75.5%), although its positive predictive value (PPV, 62.8%) was somewhat lower than that of the individual experts (PPV, 64.6 and 66.6%, respectively) (SS and PPV terms are explained later in this article). In an effort to improve the performance of BASS in the present work, both in terms of predictive value and accuracy of assigning a diagnostic index, a feedforward neural network module (FNN) based on a singular value decomposition (SVD) of image blocks was combined with the pre-existing receptive field/squashing function module (RFS). In addition, the database was expanded and a human expert was used to validate BASS diagnostic performance. This resulted in a marked improvement in the predictive value (PPV, 83.6%), while the sensitivity value (SS, 61.1%) decreased. Furthermore, the ability of BASS to accurately assign a diagnostic index was found to be 69%, compared to that of the two experts, which was 68% and 77%, respectively.

This study shows that detection and classification of individual nuclei in histopathological sections can be consistently performed by a modular neural network-based system. Moreover, since the system simulates the detection and grading strategies employed routinely by human experts, and uses a modular approach, the human experts’ control over the process can be applied at various stages during the assessment procedure. Thus, in addition to providing a basis for more accurate assessments, BASS can also contribute to the formulation of standardizing diagnostic scoring schemes for immunocytochemical data.

Material

Cryostat sections from frozen biopsies of breast cancer patients, 41 in total, were cut at 6 μm and placed on poly-L-lysine coated slides. The sections were fixed and immunolabelled using specific antibodies to estrogen and progesterone receptors (ER-ICA/PgR-ICA kits, Abbott, Germany). Positive nuclear staining (brown color) was visualized using the strept ABC kit linked to peroxidase (DAKO, Denmark). Subsequently, sections were counterstained with hematoxylin to highlight unlabeled nuclei which stained blue.

The immunocytochemically labeled slides were microscopically and independently evaluated by two experts using a five-point grading diagnostic scheme (Appendix A). Differences in grading certain cases were resolved by re-evaluation and consensus agreement between the two experts. Following manual scoring, a medical expert selected up to two regions of interest from each specimen for subsequent digitization. In total, 57 images from slides stained for either estrogen or progesterone receptors were digitized for subsequent analysis by BASS. The images were captured at 400x magnification (Zeiss Axioshot microscope, SONY DXC-980P camera) in 24-bit color and 640 x 480 pixel spatial resolution.

Modular Neural Network System

To assign a diagnostic index to a biopsy specimen, BASS implements a modular approach that resembles the algorithm used by human experts. BASS
proceeds to grade a biopsy image by first finding the location of nuclei in the image (see Fig. 1, Module I, Nuclei Detection) and then classifying them into one of five classes according to staining intensity (see Fig. 1, Module II, Nuclei Classification and Biopsy Scoring). Once the nuclei are graded, a five-class nuclei proportion vector is computed, which is used to determine the diagnostic index according to the manual grading scheme (Appendix A). BASS also contains an image database retrieval interface (see Fig. 1, Module III, Retrieval Interface), which can be used for content-based biopsy image retrieval from a database of cases [15].

In this study, the BASS detection module has been expanded to include an FNN operating on a block-based SVD of the biopsy images, besides the pre-existing RFS module [13]. Both approaches utilize localized operating principles that feature rotational invariance and insensitivity to noise. Due to the SVD transformation of the image blocks, these properties strongly resemble image energy distribution features that are related to texture and edges. The RFS module with its receptive field filter is designed to enhance or suppress relative differences of nuclei and background, and thus it is more related to average properties of local image structure. Since the upgraded BASS system is based on two different detection modules, methodologies for combining the results from both modules need to be investigated. The individual modules were evaluated as well as the logical combinations OR and AND. In the following sections, the RFS module, the FNN module, the combination of RFS and FNN modules, the classification module, and system validation methods are described.

Detection of Nuclei: The Receptive Field-Squashing Function (RFS) Module

The design of the RFS algorithm [13] is motivated by the observation that nuclei are localized objects occurring at different spatial scales in a noisy image context, and that on average they lack significant spatial axes. The RFS algorithm uses a combination of a difference of a Gaussian on-center off-surround receptive field and a squashing function to detect nuclei in biopsy images. Receptive fields, well-known from the fields of biological vision [16-18] and image processing [19, 20], appear to be a particularly good choice as part of a detector mechanism for nuclei. A particular advantage of receptive fields based on Gaussians is the fact that the filters are localized both in the spatial and frequency domains. Moreover, the filters can be conveniently scaled to accommodate optimal response in a desired range of spatial scales. Here, the receptive fields are applied as general templates to enhance matching local image structure and suppress the rest. Due to the typical center-surround structure of Gaussian receptive fields, the matching process also depends on the immediate neighborhood of nuclei that are supposed to be matched by the filter. This feature gives control over how well a nucleus has to be separated from other image structures in order to be detected. A detector always needs to decide whether an event is "to be detected" or "not to be detected." A squashing function, initialized using image statistics as part of the iterative process of the RFS algorithm, acts as a soft threshold. The function divides the image pixels into background pixels and pixels belonging to nuclei by gradually transforming both sets closer to the extreme values.

The RFS nuclei detection algorithm automatically adapts to local and global imaging conditions. The user may interact with the algorithm via two parameters, which indicate the approximate nuclear size in the biopsy image and the distance to neighboring image structures such as other nuclei. However, these parameters were fixed for all images at the same value to avoid user interaction in the experiments. In the following subsections, the major steps of the RFS detection algorithm are presented, while a more detailed account can be found in [13].

Step 1: Convert Color Image to Optical Density Image
The original RGB color image is transformed into a scalar array \( I \) using the Y channel (optical density) of the RGB-YIQ transform [21].

\[
I(x,y) = \begin{bmatrix}
R(x,y) \\
G(x,y) \\
B(x,y)
\end{bmatrix} = \begin{bmatrix}
0.299 & 0.587 & 0.114
\end{bmatrix} \cdot \begin{bmatrix}
R(x,y) \\
G(x,y) \\
B(x,y)
\end{bmatrix}
\]

(1)

where \( 0 \leq I, R, G, B \leq 255 \). The \( Y \) channel captures the sensitivities of the human visual system with respect to perceived optical density of each color. The scalar array \( I \) consists of the inverted \( Y \) channel to accommodate the detection characteristics of the on-center off-surround-type receptive field.

Step 2: Compute the Receptive Field Filter
The receptive field array is given by the following equation:

\[
Rf(x,y) = B \cdot \alpha \cdot \exp \left( -\frac{x^2 + y^2}{2\sigma_1^2} \right) - J \cdot \exp \left( -\frac{x^2 + y^2}{2\sigma_2^2} \right)
\]

(2)

where \( B, \alpha, \sigma_1, \sigma_2, \) and \( J \) are constants. The parameters \( \rho_1 \) and \( \rho_2 \) determine the sensitivity of the filter regarding the range of object sizes and were set to 2.5 and 7.5, respectively; \( B, \alpha, \) and \( J \) are automatically determined as given in [13].

Step 3: Apply Iteratively the Receptive Field and the Squashing Function

\[
\text{WHILE}(k < 3)\text{DO}
\]

\[
\text{ADJUST } scale, offset, incl
\]

\[
I_{n+1} = S(g(Rf \otimes I_n))
\]

ENDWHILE

(3)

where \( I_0 \) and \( S(g) \) depict, respectively, the image array and the squashing function at the \( k \)-th iteration. At the beginning of each iteration, the squashing function parameter...
2. PNN nuclei detection module: (a) block raster scanning of input image, (b) SV expansion and feedforward neural network classification of image blocks, (c) calculation of the exact nuclei locations (morph. shrink = morphological shrink operator).

Step 4: Threshold Bimodal Histogram
The histogram vector of \( I_3 \) (i.e., the image array after the third iteration) is smoothed with a moving average filter. The threshold value \( T \) is set equal to the histogram bin with the minimum amount of pixels between the two modes of the histogram:

\[
T = \min_{w=1}^{w_1} (\text{hist}(I) \odot [a_{0.7}, a_{0.3}, \text{scale}^{-1}])
\]

where \( \text{hist}(I) \) returns the histogram of \( I \), \( a_{0.7}, a_{0.3}, \text{scale}^{-1} \) define the moving average filter coefficients, and \( \min_{w} \) returns the minimum between two maxima, which is used to obtain the corresponding threshold intensity \( T \). \( T \) is used to segment the sample image into background and candidate nuclei. The candidate nuclei in the image (i.e., all connected sets of pixels) are recorded in an object list.

Step 5: Revise the List of Detected Nuclei
The nuclei center locations are computed by determining the center of gravity of each candidate nucleus and then returned.

Detection of Nuclei: The Feedforward Neural Network (FNN) Module
The algorithm presented in this section detects the locations of nuclei in biopsy images, based on a supervised neural network. Block-based processing of the images is adopted, followed by an SVD of each block. The most important singular values are fed as inputs to a neural network classifier, which, in turn, determines the likelihood that the original image block contains a nucleus. The neural network is trained in a supervised mode, thus allowing the knowledge provided by the experts to be included in its design.

In general, features are extracted by applying a transformation to the image blocks, so that the image space is mapped onto another space, which is assumed to be more suitable for classification. Commonly used features include the mean and variance of the values of pixels in each block, moments, and other correlation-type or higher order statistical parameters, coefficients of the Fourier or the discrete cosine transform (DCT) of the block, as well as deterministic parameters related to the size, color, and connectivity of the block [22, 23]. Generally, the above mapping is not intended to be one-to-one; i.e., there is some loss of information when moving from the original image to the feature space, mainly to simplify the classification task. In particular, those features that do not significantly contribute to improving separability of the
The singular values of a matrix have a very important property—that of energy compaction [24]. This property makes them suitable as input to the neural network classifier investigated in this study [25].

Let \( A_{m,n} \) be a real rectangular matrix, and \( k \) denote its rank. Without loss of generality, we can assume that \( m \geq n \). Then, there exist two orthogonal matrices, \( U_{m,n} \), \( V_{n,n} \), and a diagonal matrix, \( \Lambda_{m,n} \), for which the following formula holds:

\[
A = U \Lambda V^T
\]  

(6)

where \( \Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_m, 0, 0, ..., 0) \), \( \lambda_1 > \lambda_2 > ... > \lambda_m \), and \( (\cdot)^T \) denotes the transpose of a matrix. Each \( \lambda_i, i = 1, 2, ..., m \), is an eigenvalue of \( AA^T \), or equivalently \( A^T A \), and \( \lambda_i \) are the singular values of matrix \( A \). The eigenvectors \( u_i \) of \( AA^T \), related with the eigenvalues \( \lambda_i \), are the columns of matrix \( U \), and the eigenvectors \( v_i \) of \( A^T A \) are the columns of matrix \( V \).

Given the diagonal matrix and using a column vector \( e = (1, 1, ..., 1)^T \) of size \( n \), we can generate the singular value vector \( (svv) \) of matrix \( A \) by postmultiplying matrix \( A \) with vector \( e \):

\[
svv = A \cdot e
\]  

(7)

Under the constraint \( \lambda_1 > \lambda_2 > ... > \lambda_m \), matrix \( A \) and vector \( svv \) are unique for a given matrix \( A \) [25].

Apart from the above, it should be mentioned that singular values are insensitive to small changes in matrix \( A \). Assuming that matrix \( A \) denotes an image block, \( svv \) and consequently the neural network detector are insensitive to small changes of pixel values caused by noise or different illumination conditions. Moreover, singular values remain the same even if the image block is rotated, translated, or transposed. The above properties are highly desirable in the detection task, where the position, and not the orientation, of the nuclei is required. The proposed scheme is composed of the following steps.

Step 1: Color Image to Optical Density Image Conversion

This step is identical to that of the RFS module, except that the optical density image is not inverted [see Eq. (1)].

Step 2: Histogram Stretching and Thresholding

The histogram stretching and thresholding step aims at smoothing the noisy background and increasing (i.e., normalizing) the contrast between the nuclei and the background of the optical density image. Suppression of the noisy background is desirable in any detection task. In addition, contrast varies between the images, potentially affecting the efficiency of the neural network detector. Thus, contrast variations should be minimized.

If \( X \) is an optical density image whose histogram values are limited in the interval \([a, b]\) where \( a \geq 0, b \leq c \), then histogram stretching generates an image \( Y \) whose histogram lies in the extended interval \([0, c]\). Image \( Y \) is derived by the following transformation:

\[
Y(i,j) = c \cdot \frac{X(i,j) - a}{b - a}.
\]  

(8)

After histogram stretching, the noisy background is smoothed using an appropriate threshold, \( T_c \), which is a function of image \( Y \). The thresholded image \( Z \) is given by:

\[
Z(i,j) = \begin{cases} 
Y(i,j), & Y(i,j) < T_c \\
Y(i,j), & Y(i,j) \geq T_c.
\end{cases}
\]  

(9)

Step 3: SV Expansion and Feedforward Neural Network Classification of Image Blocks

This step is based on a block-oriented architecture, as illustrated in Fig. 2. In particular, the preprocessed images are raster scanned both horizontally and vertically using a scanning step of \( k \) pixels and a block size of \( N \times N \) pixels. Thus, the image is separated into overlapping blocks, \( B(i,j) \), of \( N \times N \) pixels [see Fig. 2(a)]. These blocks are subsequently transformed using SVD, producing singular value (SV) feature vectors [see Fig. 2(b)]. The SV feature vectors are then subjected to dimensionality reduction to decrease further the complexity of the algorithm and to drop those singular values, which do not significantly contribute to the separability of the dataset. The truncated feature vectors are fed into the neural network [see Fig. 2(b)]. Consequently, classification is performed in the singular-value domain. The analog neural network output values for all blocks form an output intensity image \( O \) the size of which is \((k \times k)\) times less than that of the original image, where \( k \) is the scanning step, as shown in Fig. 2(a).

Pixel values in the output image that are close to unity show that the corresponding blocks in the input image have a high likelihood to belong to a nucleus.

Step 4: Calculation of the Exact Nuclei Locations

Ideally, the output of the neural detector would include isolated points indicating
the nuclei positions. However, this is not the case, due to the different nuclear sizes and to overlapping nuclei. Instead of isolated points, the output image generally contains clusters of points (i.e., blocks) within, and possibly around, the area of the nuclei. In order to achieve the final classification of the blocks, selection of the value of a global threshold $T$ is needed. On the other hand, the exact positions of the nuclei centers must be calculated.

First, the global threshold $T$, which is actually a parameter indicating how conservative the detector is, can be selected by the user, allowing external control of the algorithm performance. Then, the exact positions of nuclei are computed as the local maxima of the output image pixel values, according to the following rule:

"Let the position of the 'positive' pixel with the biggest value, within a proscribed neighborhood of each 'positive' pixel of the output image, be selected as the center of the nucleus".

A three-step procedure implements this rule, as follows [Fig. 2(c)]:

1. The output image $O$ is cut into different binary masks $S_{p}, p = 1, \ldots, L$, using multiple thresholds $T_{p}$:

$$S_{p}(i, j) = \begin{cases} 1, & T_{p} < Y(i, j) < T_{p+1} \\ 0, & \text{elsewhere} \end{cases}$$

where $T_{p} > T_{p+1}, T_{1} = T$, and $Y_{max}$ is the maximum value of $O$.

2. Each binary mask $S_{p}(i, j)$ is processed using a morphological shrink operator [26] [Fig. 2(c)] to generate corresponding isolated points, while preserving the number of clusters of points in the mask; i.e., the morphological shrink operator preserves the Euler number [27].

3. The resulting outputs, say $C_{p}$, are fed into a priority logical OR function [Fig. 2(c)] to produce an output, which constitutes the final output of the detector after proper rescaling by the factor $k$. Let us consider, for example, a double input priority logical OR function; this function combines the pixels of the two inputs, say $I_{1}$ and $I_{2}$, so as to give priority to the detected points in $I_{1}$. In particular, if any of the points of $I_{1}$ are in the neighborhood of a point of $I_{2}$, they are dropped. Its operation, say $\odot$, is described below:

$$I_{1} \odot I_{2} = I_{1} \cup D$$

where
\[ D = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2} \] and \( d \) denotes the Euclidean distance and \( r \) is a parameter related to the approximate nuclear size. The output of the multiple input priority logical OR function is computed as follows:

\[ C = \bigcap \left( \bigcup_{i=1}^{d} \left( C_i \bigcup C_j \right) \right) \] (12)

The selection of the various parameters used in the FNN detection module is presented in Appendix B.

**Combination of Detection Modules**

Since the RFS and FNN modules work on different principles, our aim was to test their performances, individually and combined. In this study, the detection results of the RFS and FNN modules were combined, based on modified logical OR and AND operators. ORDs (RFS OR FNN) and ANDDs (RFS AND FNN) modules were evaluated along side the individual RFS and FNN modules. The semantics of the logical operators had to be expanded to account for the rather limited spatial accuracy of the detection events generated by each module. Therefore, a fixed tolerance was used to address the trade off between using the precise nuclei locations detected by the two modules, and using the less strict notion of proximity of detected location to decide whether the same nucleus was detected. Detected nuclei locations were considered to coincide if the centers were located within a distance of nine pixels from each other, as detected by either of the two modules.

The following rules were adopted for either ORDs or ANDDs: (i) ANDDs detection results include only those nuclei that coincide regarding the location and have matching class labels, while (ii) ORDs detection results include all nuclei from both modules, substituting the locations detected by the RFS module and the corresponding labels if the class labels do not agree.

**Nuclei Classification and Diagnostic Index Calculation**

Following the detection of nuclei, a radial basis function (RBF) neural network [14] is applied to classify each nucleus into one of five staining intensity classes. After the classification of each nucleus, the diagnostic index is computed, as illustrated in Fig. 6. The RBF neural network classifier is trained, based on the nuclei feature vectors from four selected images marked by an expert. That particular expert had the best average percentage of correctly graded images per diagnostic index class (CGI), as determined from the diagnostic index confusion matrices. To avoid misleading the classifier, only images whose final diagnostic index was consistent with the routine diagnostic index were selected. Expert 2 turned out to be more consistent, according to the confusion matrices. The four images marked by Expert 2 resulted into more than 1000 nuclei feature vectors. The other images were used as an independent validation set. The nuclei feature vectors from the selected images were randomized by class staining intensity, divided into training vectors and test vectors (80/20 ratio), merged into one training and one test set, and randomized again. When the dataset was divided, the construction of a balanced training set was emphasized, since all nuclei classes contribute equally to an accurate diagnostic index.

The RBF network structure consisted of a single RBF unit layer that was fully connected to a layer with linear output units. The four dimensional input vectors (see below) were simultaneously fed to each RBF unit. The output layer contained five linear units, which encoded the five possible nuclei staining intensity classes with binary values. The training procedure for RBF networks [28] was carried out using an incremental solver [29]. This solver dynamically adds one RBF neuron per training epoch to the RBF layer, and it adjusts the weights between the RBF layer and the linear output layer, until either a maximal number of neurons has been added or the sum-squared error falls beneath an error goal. The transfer function (TF) of each RBF neuron has the following form:

\[ TF(p) = \text{RBF}(\text{dist}(w, p) \times 0.8326/SP) \] (13)

where \( p \) is the input feature vector, \( w \) is the weight vector, \( \text{dist} \) is the Euclidean dis-
Table 1. Comparison of detection performance for 57 images using sensitivity (SS) and positive predictive value (PPV). The second component is always considered to be the gold standard.

<table>
<thead>
<tr>
<th>Detector vs. Gold Standard</th>
<th>SS</th>
<th>S.D.</th>
<th>PPV</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Expert</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex1 - Ex2</td>
<td>79.2 ± 13.7</td>
<td>76.5 ± 15.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex1 - OREx</td>
<td>63.8 ± 13.2</td>
<td>100.0 ± 0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex2 - OREx</td>
<td>81.1 ± 14.5</td>
<td>100.0 ± 0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex1 - ANDEx</td>
<td>100.0 ± 0.0</td>
<td>52.8 ± 20.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ex2 - ANDEx</td>
<td>100.0 ± 0.0</td>
<td>54.7 ± 21.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II. Expert 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFS - Ex1</td>
<td>81.4 ± 16.3</td>
<td>58.5 ± 14.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNN - Ex1</td>
<td>76.7 ± 12.6</td>
<td>54.9 ± 15.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORDI - Ex1</td>
<td>92.3 ± 5.1</td>
<td>48.3 ± 11.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANDDI - Ex1</td>
<td>57.3 ± 15.2</td>
<td>71.2 ± 16.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III. Expert 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFS - Ex2</td>
<td>81.1 ± 15.2</td>
<td>56.3 ± 17.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNN - Ex2</td>
<td>79.6 ± 8.9</td>
<td>55.8 ± 17.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORDI - Ex2</td>
<td>92.6 ± 5.2</td>
<td>47.8 ± 14.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANDDI - Ex2</td>
<td>59.9 ± 14.2</td>
<td>72.2 ± 16.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV. OREx</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFS - OREx</td>
<td>79.6 ± 16.1</td>
<td>67.1 ± 14.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNN - OREx</td>
<td>75.4 ± 11.9</td>
<td>64.7 ± 16.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORDI - OREx</td>
<td>91.6 ± 5.1</td>
<td>57.6 ± 12.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANDDI - OREx</td>
<td>55.4 ± 14.4</td>
<td>82.4 ± 14.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. ANDEx</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFS - ANDEx</td>
<td>84.6 ± 14.9</td>
<td>32.0 ± 14.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNN - ANDEx</td>
<td>81.3 ± 10.8</td>
<td>30.1 ± 12.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORDI - ANDEx</td>
<td>94.5 ± 5.1</td>
<td>25.7 ± 11.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANDDI - ANDEx</td>
<td>62.7 ± 15.4</td>
<td>40.7 ± 17.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


tance measure, SP is the spread constant, and RBF is a Gaussian function. The transfer function, \( TF \) [Eq. (13)], takes on its maximal value of unity when its argument becomes zero. \( TF \) will return 0.5 when its argument has the value 0.8326. Thus, \( TF \) will be unity if the distance between the vectors \( \mathbf{p} \) and \( \mathbf{w} \) is zero. If, for example, the spread constant \( SP \) equals 0.1, \( TF \) would return 0.5 for every vector at a distance of 8.326 from \( \mathbf{w} \). From the above considerations, it is evident that \( SP \) determines how large an area in the input space each dynamically added RBF neuron will respond to. For the experiments reported here, \( SP \) was chosen based on the distribution of mutual distances of the feature vectors in the training dataset. The best classifier is selected based on the maximum average performance per nuclei staining intensity class on the test set, since the aim is to generate a classifier that performs equally well for all diagnostic index classes. The main steps of the classification algorithm are summarized as follows:

**Step 1: Extract Features for Each Nucleus**

The following features are extracted: the average of the \( Y, I, \) and \( O \) channel, as well as a texture measure, \( T_X \), are computed for each nucleus; \( T_X \) is defined as:

\[
T_X = 1 - \frac{1}{1 + \sigma^2 f_{(y)_{max}}} \tag{14}
\]

where \( \sigma^2 \) is the variance, and \( y_{(x)} \) denotes the optical density feature of a nucleus.

**Step 2: Classify Each Nucleus**

A nucleus is classified into one of five staining intensity classes (negative, weak, moderate, strong, very strong) using the radial basis function neural network classifier.

**Step 3: Compute Diagnostic Index**

The diagnostic index is computed according to the manual semi-quantitative scheme, which is described in Appendix A.

The assessment results of the preceding steps are finally stored in the database of cases using the retrieval interface [15]. This interface may also be used to retrieve biopsy images, based on the contents as described by the assessment results. For this purpose, special matching operators were implemented that enable the retrieval of images based on similarity or on presence of one or more characteristics.

**System Validation**

Ultimately, any system or expert can be said to perform successfully if the nuclei proportions in an image are estimated accurately and reliably, since the diagnostic index is based on this estimate. However, at present, it is difficult to evaluate either the performance of the experts or the systems, due to the unavailability of universal gold standards at the nuclei detection and nuclei classification (i.e., diagnostic index) levels. Firstly, the system was evaluated at the nuclei detection level by comparing its performance to that of two human users, using as the basis 200 to 300 nuclei per case, which were marked by the experts. Secondly, the system was evaluated at the nuclei classification level by comparing those classified by BASS to those classified by the experts. Finally, the diagnostic index, calculated on the basis of the BASS nuclei classification results, was compared to the diagnostic index of each case obtained routinely by human experts. In order to maximize the overall system performance, particular attention was given to calibration and data acquisition standards.

To create a basis for comparison at the nuclei level, a small circular probe of nine pixels diameter was placed centrally on top of each nucleus detected by the system modules. Independently, two experts placed the same-sized probes manually where they perceived the nuclei centers were located using the mouse. In addition, the probes were color-coded, depending on the staining intensity class computed by the BASS classification algorithm described earlier, or assigned by the expert. As a result of this procedure, for each module and each expert (i.e., the combinations) one mask image was created. The
diagnostic index for all mask images was automatically computed according to Table 5, using the above-described color-coded mask images.

A detection event was defined as the set of pixels belonging to one of the probes in the image. If two probes from different mask images overlapped, then the two corresponding detection events were said to coincide, and the corresponding nuclei was interpreted to have been detected in both mask images. In the case of one probe touching several other probes, only one coinciding detection event was counted. Two hybrids, called OREEx (Ex1 OR Ex2) and ANDEx (Ex1 AND Ex2), were derived with a modified logical OR and AND operation from the individual experts' masks, as implemented in the ORDI and ANDDI modules.

System validation was performed using four methods. In particular, the detection performance and the joint detection and classification performance of the BASS system were assessed with the following methods:

1. Receiver-operator characteristic measures (ROC) were used to analyze individual nuclei detection performance of BASS detection modules compared to that of two experts. ROC measures are useful to compare the detection performance with respect to individual nuclei, because no assumption about the underlying probability distribution of the detection events is made. Based on the definition of the detection events, two measures, sensitivity (SS) and positive predictive value (PPV), were chosen to characterize the detection performance. Sensitivity is the likelihood that a nucleus will be detected if it is also marked as a nucleus in the gold standard. It is defined as follows:

\[ SS = TP / (TP + FN) \]  \hspace{1cm} (15)

where TP (true positive) are those nuclei marked in both the gold standard and the image, and FN (false negative) are those nuclei that are marked in the gold standard, but not in the image. Positive predictive value (PPV) is the likelihood that the detection of a nucleus is actually associated with a nucleus marked in the gold standard.

\[ PPV = TP / (TP + FP) \]  \hspace{1cm} (16)

where FP (false positive) are those nuclei which are marked in the image, but not in the gold standard.

2. Spearman's rank-order correlation coefficient [30] was determined to assess the joint performance of BASS detection and classification modules compared to that of Ex1 and Ex2, based on the staining intensity class proportions of all biopsy images. This coefficient makes no assumptions about the underlying probability distributions, and since proportions are analyzed, the measurements are independent of the absolute numbers of detected nuclei. However, bias towards one or the other nuclei class or staining intensity, is registered. This is a relative bias, due to the lack of gold standards. The Spearman's rank-order correlation coefficient is given by:

\[ r = \frac{\sum_{i=1}^{n} (R_i - \bar{R})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (R_i - \bar{R})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}} \]  \hspace{1cm} (17)

where \( x \) and \( y \) are vectors containing the proportions of a particular staining intensity class for all biopsy images, and \( R_i \) and \( Y_i \) are magnitude-based ranks among \( x \) and \( y \), respectively.

3. Confusion matrices were used to evaluate experts and BASS performances in the assignment of diagnostic indices, as compared to the diagnostic index assigned routinely. These matrices show where and how much two classification results deviate from each other. The confusion matrices are defined as follows:

\[
M_{c} = \begin{bmatrix}
    e_{0,0} & e_{0,1} \\
    e_{1,0} & e_{1,1}
\end{bmatrix}
\]

where \( e_{j,k} = \% \text{Percentage of biopsy images with diagnostic index } j \text{ which were classified as } k \) and \( j, k = 0, ..., 4 \). The average percentage of correctly graded images per diagnostic index class (CGI); i.e., the average of the diagonal entries in \( M_{c} \), summarizes an important performance characteristic and is reported as well.

This part of the study was based on 29 images that were selected by the experts on the basis of clarity of color perception and minimal overlap displayed between

The present data demonstrate that both modules perform consistently and accurately despite the fact that they were using different methodologies.

---

**Table 2. Results of the second-run validation procedure:**

<table>
<thead>
<tr>
<th>Detector vs. Gold Standard</th>
<th>W/out Second-Run Validation</th>
<th>With Second-Run Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>S.D.</td>
</tr>
<tr>
<td>RFS-OREEx</td>
<td>76.4 ± 16.9</td>
<td>66.8 ± 14.0</td>
</tr>
<tr>
<td>FNN-OREEx</td>
<td>79.3 ± 8.4</td>
<td>58.1 ± 13.5</td>
</tr>
<tr>
<td>ORDI-OREEx</td>
<td>91.2 ± 4.8</td>
<td>52.7 ± 17.7</td>
</tr>
<tr>
<td>ANDDI-OREEx</td>
<td>59.8 ± 15.0</td>
<td>78.9 ± 14.0</td>
</tr>
</tbody>
</table>
nuclear boundaries. These features, in combination with the experience gained by the experts in grading the first subset of 28 images [13], were considered valuable contributing factors in making interpretation of the digitized images by the experts more consistent.

4. The McNemar test [31] was applied to the diagnostic index results to find out if the classifiers perform significantly different. The McNemar test is based on the chi-square test. To perform the test, two counts must be performed: (i) the number of those images that were correctly classified by one classifier ($count_{10}$), and (ii) the number of images that were correctly classified by the other classifier ($count_{01}$). Then, the following comparison is performed:

$$\frac{(count_{10} - count_{01} - 1)^2}{count_{10} + count_{01}} > 3.841459$$

(19)

where the statistic on the left side of the inequality is corrected for the discrete variables used, and which is approximately distributed like the chi-square distribution with one degree of freedom. At the 5% confidence level, the statistic has to be larger then 3.841459 to reject the null hypothesis that the two classifiers perform equally well.

**Results**

A total of 57 images from 41 patients were captured and analyzed using BASS. The actual distribution images for each diagnostic index category was 17%, 17%, 17%, 35%, 14%, starting with the "0" and ending with the "4+" diagnostic index categories, respectively (Appendix A).

**Detection Example**

Figure 5 shows a subregion of a breast cancer biopsy image marked by all experts and modules, and their combinations. Figures 5(a) and 5(b) indicate the nuclei marked by Ex1 and Ex2, whereas Figs. 5(c) and 5(d) present the detection results from OREx and ANDEx. Figures 5(e-h) show the detection performance for the RFS module, the FNN module, the ORDEx module, and the ANDDEx module. The large square inserted in each figure highlights a region of the image that displays the performance of the experts and the modules. The modules generally detect more nuclei in the biopsy images than do both experts. The small rectangle illustrates the performance of the OR and AND operators. In particular, the individual experts and the modules have marked the nucleus towards the left side of the rectangle, but each expert assigned a different nuclear staining intensity class (difference is only visualized in color). In addition, Ex1 detected two more nuclei that Ex2 did not mark. The result is that the OR operator includes one of the locations and the corresponding nuclei class for the nucleus detected by both experts, and the two nuclei not detected by Ex1. In contrast, the AND operator ignores all the nuclei in the smaller rectangle, since there is disagreement regarding nuclei locations as well as their staining intensity classes.

**ROC Analysis**

Table 1 tabulates the detection performance regarding the 57 images in the database. In this table, the second component is taken to be the gold standard. It is noted that if component I were to be chosen as the gold standard, the numbers would only have to be interchanged, due to the definitions of SS and PPV. The table consists of five sections, ordered by the applied laboratory gold standard.

Table 1, section I, shows that Ex1 has an SS of 79.2% and a PPV of 76.5%, with Ex2 as the gold standard. When Ex1 and Ex2 are compared to OREx, their sensitivity values increases to 83.8% and 81.1%, respectively. For ANDEx, the PPV values of Ex1 and Ex2 are just 52.8 and 54.7%, with a high standard deviation of over 20%. The SS and the PPV values for the individual experts, when
compared to ANDEx and OREx, is by definition 100%.

Table 1, sections II to V, show the detection results of the individual and combined modules. The detection results for the individual modules are presented first, followed by ORDt and ANDDt. The SS and PPV values of the RFS and FNN modules (sections II and III), when compared to either Ex1 or Ex2, were in general quite similar, varying from 76.7 to 81.4% and 54.9 to 58.5%, respectively. It is noted that the PPV values of the individual RFS and FNN modules (sections II and III) remain about 20% below the 78% (section I), Ex1 - Ex2, (79.2+76.57%) average level of the human experts. The SS values of the individual modules versus OREx (section IV) drop only slightly. The PPV value, however, increases by 8.6% and 10.8% for the RFS module, and 9.6% and 8.9% for the FNN module.

Combining the RFS and FNN modules via the logical OR operation achieves SS values of over 91.6%, but the PPV values decrease by a maximum of 8% below the lowest value of any individual module (section IV), ORDt entries in sections II to V), ANDEx scores over 30% lower SS values than OREx. However, the PPV values are at least 12.7% higher than the best performances by any single module or ORDt, reaching 82.4%, which approaches favorably the performance of the experts.

The modules and their combination perform poorly regarding the PPV value when compared to ANDEx (section V). ANDDt performs best regarding the PPV, while ORDt perform best regarding the SS. Since the number of nuclei marked and classified in agreement is lower than the numbers from individual experts, the PPV values for both the modules and the individual experts are very low. On average, the experts agree only in about 53% of the cases (section I, Ex1 - ANDEx, Ex2 - ANDEx, (52.8 + 54.7)/2) on the location, while ANDDt agrees only with a maximum of 40.7%. It is noted that the definition of the AND operator requires detected nuclei not only to coincide spatially, but also to be labeled identically.

Table 2 summarizes the results of the second-run validation, which was performed by one of the experts on a subset of 29 images (20 patients). The data in Table 2 show that the PPV increases for the RFS, the FNN, the ORDt, and the ANDDt modules by 8.2, 6.1, 8.4, and 4.7%, respectively. ANDDt achieves the best PPV value of 83.6%, while ORDt SS score was highest, at 92.2%.

Classification and Diagnostic Index Computation Module

The best radial basis function classifier obtained consisted of 46 basis function units, with a fixed spread factor of 20. In total, 85.22%, 79.57%, and 76.36%, respectively, of the nuclei feature vectors from the training set, the test set, and the validation set were classified correctly. These numbers correspond to an average per-nuclei staining intensity class classification accuracy of 82.06%, 80.46%, and 65.16% for the training, test, and validation sets. The total number of nuclei in the training, test, and validation sets were 805, 328, and 5077, respectively. The distribution of nuclei classes in the combined dataset was 31.37%, 33.99%, 23.24%, 9.4%, and 4.99%, beginning with the "negative" nuclei class and ending with the "very strong" nuclei class, respectively. These data originated from the evaluation of the subset of 29 images, as explained earlier.
Table 5. Computation of manual semi-quantitative immunocytochemical diagnostic index [8] (see example in Fig. 6).

<table>
<thead>
<tr>
<th>% of Cells Positive</th>
<th>Score</th>
<th>Staining Intensity</th>
<th>Score</th>
<th>Total Score</th>
<th>Diagnostic Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>Negative</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0-25%</td>
<td>1</td>
<td>Weak</td>
<td>1</td>
<td>1-4</td>
<td>1+</td>
</tr>
<tr>
<td>26-60%</td>
<td>2</td>
<td>Moderate</td>
<td>2</td>
<td>5-8</td>
<td>2+</td>
</tr>
<tr>
<td>51-75%</td>
<td>3</td>
<td>Strong</td>
<td>3</td>
<td>9-12</td>
<td>3+</td>
</tr>
<tr>
<td>≥ 76%</td>
<td>4</td>
<td>Very Strong</td>
<td>4</td>
<td>≥ 13</td>
<td>4+</td>
</tr>
</tbody>
</table>

Spearman’s Rank-Order Correlation Analysis

Spearman’s rank-order correlation analysis depends on the detection and classification performance of each expert or module. The results are presented in Table 3 for all 57 images. The experts achieved correlation values of over 0.78, except for the weak nuclei, where the value was 0.69 (section 1). Correlation was particularly high for strongly stained nuclei, where the value lies at 0.9. BASS nuclei detection modules correlated well (> 0.75) with the experts’ values regarding the moderate to very strongly stained nuclei (sections II to V). Negative and weakly stained nuclei carried correlation values around 0.7. In addition, BASS nuclei detection modules were strongly correlated (> 0.85) across all nuclei classes, except for RFS and FNN, which had a lower value (0.76) for the weakly stained nuclei (section VI).

Confusion Matrix Comparison Regarding the Assignment of Diagnostic Indices

The results shown in Table 4 reflect the performance of experts and modules with respect to a subset of 29 images, as explained earlier. The row entries show the percentage of images with a particular diagnostic index assigned manually (as given in Appendix A) to the biopsy slide versus the ones computed by BASS. The last row of the table displays the average percentage of correctly graded images per diagnostic index. The main characteristic of the confusion matrices considered is the average percentage of correctly graded images per diagnostic index class (CGI). As mentioned above, the actual distribution of diagnostic indices was 17%, 17%, 17%, 35%, and 14%, starting with the “0” and ending with the “4+” diagnostic index entries, respectively. The representation of the confusion matrices emphasizes the per-class accuracy, rather than the accuracy with respect to the overall dataset.

Table 4 shows that Ex1 and Ex2 differ by 9% in their CGI values, achieving 68% and 77%, respectively, ANDEx achieved 69% and OREx scored 84% CGI incorrectly graded images (off-diagonal entries) were mainly over-scored by both experts, while ANDEx over-scored 0, 1+, and 2+ images and under-scored 3+ and 4+ images, OREx only over-scored in the case of 0, 1+, and 3+ images. The RFS and FNN module scored 69% and 61% CGI, respectively. There is agreement in their performance for 3+ and 4+ images, while the RFS module performed better than the FNN module for 0, 1+, and 2+ images. The performance of OREx lies 4% below the FNN module and 12% below the RFS module. However, ANDEx performed the same as the RFS module, despite differences in the absolute numbers of detected nuclei per image.

To quantify whether the system or the experts differ significantly in their grading performances, the McNemar test was applied for all combinations of system modules and experts. However, no significant differences were found except in the case of Ex1-OREx, which differed significantly at the 5% level.

Discussion

Histopathological sections of breast cancer nuclei immunocytochemically stained for steroid receptors are routinely reported by experts, based on the microscopical evaluation of numbers of nuclei stained at particular intensities of brown color. This study shows that detection and classification of individual nuclei in histopathological sections can be reliably performed by the BASS modular neural network system in an accurate and consistent manner. BASS also facilitates interaction with experts and to this effect, the second-run validation results indicate that this interaction is constructive, since it was demonstrated that the modules correctly detect considerable numbers of nuclei that were not initially detected by the experts. Moreover, since the system simulates detection and grading strategies of human experts, it will enable the formulation of more efficient standardization criteria in the assessment of immunocytochemically stained histopathological sections [4].

The ANDed RFS-FNN module, ANDEx, leads to the best overall results in terms of detection accuracy for the diagnostic indices. It achieved the highest PPV, as compared to OREx, after the second-run validation (83.6%), and the highest average accuracy for correctly assigning diagnostic indices to the images (69%). However, the SS is lower than for any other combination of modules (61.1%). It should be noted that although the RFS module matches the overall performance of ANDEx for the diagnostic indices, its values for SS and PPV were 78.1% and 75.0%, respectively.

The present data show that a high PPV value is critical for obtaining a good performance with respect to the diagnostic index, as can be seen when comparing the experts and BASS combined detection and classification modules. On the other hand, our data show that the SS value does appear to be a less important factor and not directly related to BASS performance in computing diagnostic indices. The experts showed a tendency to overscore, as is demonstrated by the diagnostic index confusion matrices, while the combinations of detection and classification modules both overscore and underscore. This tendency of the experts to overscore may be explained by the observation that the Spearman rank correlation values were higher for moderate to very strong nuclei. However, the Spearman rank correlation...
values for the BASS system (RFS, FNN, ORDI, ANDDI, combined with the classification module) lie above 0.87 (except for the 0.76 correlation value for the weak nuclei regarding RFS-FNN), which implies that the modules and/or their combinations perform consistently and uniformly. In addition to higher accuracy and greater objectivity, image-analysis systems should also possess greater speed than that required by human experts to perform similar tasks. BASS is able to perform the analysis of one image, on average, in less than one minute (200 MHz Intel Pentium PC, 32 Mbyte RAM). This time-span compares favorably with the time needed by human experts to perform similar tasks.

In an attempt to improve objectivity and offer rapid analysis, some commercial systems, such as CAS and SAMBA, rely on global discrimination of structures of interest, between nuclei in this case, and background. These systems measure percent stained surface area using global thresholding techniques. However, there is disagreement among experts about the optimal selection of global thresholds, with the choice being fixed, manual, or automatically set threshold [32]. BASS avoids the need for global thresholding and area measurement, since it detects, counts, and classifies individual nuclei according to the manual semi-quantitative diagnostic index.

BASS was designed to simulate closely the detection and grading strategies as practiced by histopathologists. Thus, experts may be used to supervise and evaluate the system at the nuclei detection, the nuclei classification, and the diagnostic index levels. In an effort to improve detection performance, BASS nuclei detection module was expanded for this study to include a second algorithm performing the identical task, but employing a different approach. It was shown that, indeed, the combination of the detection modules RFS and FNN performed better than the individual modules. Moreover, the existing modular architecture enabled the inclusion of all combinations of detection modules, without changing the subsequent processing steps. The "nuclei classification and diagnostic index computation" module performance on the whole dataset (see also [14]) compares favorably to the performance of a neural network classifier utilizing 17 mostly textural features [33]. Neural networks are but one technique to classify image feature vectors, Bibbo, et al. [34], for example, included a variety of diagnostic clues and detailed prior knowledge in a Bayesian belief network to grade prostate lesions, while Mangasarian, et al. [35], showed that linear programming methods may successfully be applied for breast cancer diagnosis and prognosis, based on computer-aided image analysis and other clinical data.

It is difficult to assess the true system performance, with comparisons to other systems, in the absence of reliable and universal gold standards [4]. All experiments performed here had to be based on laboratory gold standards (i.e., either on the nuclei marking results from the experts or the diagnostic index), which was manually derived to ensure consistent classification of the images. Since the confusion matrices also serve as a measure of objectivity and consistency of individual experts, Expert 2 was chosen as the source of supervisory information at the beginning of this study. However, the present data demonstrate that both modules RFS and FNN perform consistently and accurately despite the fact that they were using different methodologies. In addition, the confusion matrices are further proof that the BASS system can achieve at least similar results compared to the human experts. Furthermore, BASS facilitates interaction with experts, and this combination, as shown in this study, increases the potential for improving accuracy and objectivity.

Future Work

In addition to expanding the database, to increase the accuracy of assigning a diagnostic index, BASS performance will be evaluated in a clinical setting, whereby its predictive and prognostic accuracy will be compared to the clinical status of breast cancer patients. Moreover, grading results based on human experts and BASS regarding the diagnostic index may be combined into a hybrid system in an effort to further improve performance.

The unavailability of quality assurance and universal gold standards, in immunocytochemistry makes comparisons of BASS to other image analysis systems difficult. However, it is anticipated that this drawback will soon be overcome, since there is at present a tremendous effort in standardizing immunocytochemical protocols. To achieve this goal, it is expected that image-analysis systems will play a significant role.

Acknowledgments

This work has been supported by a grant from the research committee of the University of Cyprus, and a grant from the Cyprus Planning Bureau and the General Secretariat of Research and Technology of Greece. In addition, we are indebted to the Pancyrian Association of Cancer Patients and Friends (PASYKA) for their support. We are also grateful to Mrs. Maria Demosthenous for performing the immunocytochemical staining.

Frank Schoorenberg
(S'93) received the M.Sc. degree in computer science from Texas A&M University, College Station, Texas, in December 1992 and the Ph.D. degree in computing science from Groningen University, Groningen, The Netherlands, in May 1998. From 1988 to 1990 he worked for the German National Research Center for Information Technology (GMD), St. Augustin, Germany as a research student in the area of nonmonotonic logics and truth maintenance systems. From 1991 to 1992, he worked as a research assistant and research associate in the Petroleum Engineering Department at Texas A&M University on the detection of fractures in some borehole image data utilizing neural network technology. In 1993 he joined the Department of Computer Science at the University of Cyprus as a research associate working in the area of medical image analysis and neural networks. He was the recipient of a Fulbright scholarship during his stay at Texas A&M University from 1990 to 1992. His research interests include computationally intelligent systems, medical image analysis, health telematics, and diagnostic and prognostic systems.

Nicolaos Tsapatsoulis was born in Limassol, Cyprus in 1969. He graduated from the Department of Electrical and Computer Engineering, the National Technical University of Athens in 1994, where he is currently working for his Ph.D. degree. His current research interests lie in the areas of machine vision, image and video processing, neural networks, and biomedical engineering. He is a member of Technical Chapters of Greece and Cyprus and a student member of IEEE: Signal Processing and Computer Societies.
Constantinos S. Patitichis (S’84-M’88) was born in Nicosia, Cyprus on June 30, 1959. He received the B.S.c. degree in electrical engineering from the University of New Brunswick, Canada in 1983; the M.Sc. degree in biomedical engineering from the University of Texas at Austin, USA, in 1984; the M.Sc. degree in neurophysiology from the University of Newcastle upon Tyne, UK, in 1988, and the Ph.D. in electronic engineering from the University of London, UK, in 1992. During 1985-1992, he was also working with the Cyprus Institute of Neurology and Genetics where he is now a Visiting Senior Scientist. He is currently an assistant professor with the Department of Computer Science at the University of Cyprus. He was the recipient of the 1994 European Community “Marie-Curie” fellowship, and he is actively participating in a number of European projects related to health telematics. His current research interests include medical imaging, biosignal analysis, artificial neural networks, and genetic algorithms. He has published over 40 refereed journal and conference papers in these areas. Dr. Patitichis has been involved in numerous professional society activities, including: president of the Cyprus Association of Medical Physics and Biomedical Engineering, 1996-98; president of the IEEE Cyprus Section, 1997-99; general Co-Chairman, Medical and Biological Engineering and Computing Conference 98 (Medicon’98).

Christos N. Schizas (M’83-SM’91) received the B.Sc. degree in electronic engineering from the University of London, UK, in 1978, the M.B.A. degree from the University of Indianapolis, Indiana, in 1988, and the Ph.D. degree in systems theory from the University of London in 1981. He received the 1979 William Lincoln Shelley award from the University of London for excellence in research, and a Fulbright fellowship for collaborative research in the USA in 1993.

He was a postdoctoral fellow at the University of London from 1981 to 1983, taught as a lecturer of computer science at the Higher Technical Institute of Cyprus from 1985 to 1990, and was a professor of computer information systems at the University of Indianapolis from 1990 to 1992. Since 1992, he has been with the Department of Computer Science, University of Cyprus, as a professor, where he served as Interim Chair of the Department until 1994. He is also the director of Computational Intelligence Research at the Cyprus Institute of Neurology and Genetics. His research interests include computational intelligence, decision support systems, medical applications, and diagnostic systems. He has published over 70 refereed journal and conference papers in these areas. Dr. Schizas is a Fellow of the IEEE, and BCS, and since 1992 has been a member of the International Committee of the IEEE-EMBS annual conference. He is a member of the ministerial committee for establishing the Information Society in Cyprus, and he is a partner in various European Union funded projects.

Stefanos Kollias was born in Athens, Greece, in 1956. He obtained his Diploma in electrical and mechanical engineering from the National Technical University of Athens (NTUA) in 1979, the M.Sc. in communication engineering from the University of Manchester Institute of Science and Technology in England in 1980, and the Ph.D. in signal processing from the Computer Science Division of NTUA. In 1982, he was given a COMSOC Scholarship from the IEEE Communication Society. Since 1986, he has been with the Department of Electrical and Computer Engineering of NTUA, where he currently is professor and director of the Image Processing and Multimedia Laboratory. From 1987 to 1998, he was a Visiting Research Scientist in the Department of Electrical Engineering of Columbia University in New York, USA, on leave from NTUA. His research interests include neural networks, human computer interaction, image/video processing and analysis, intelligent multimedia systems, computer graphics, virtual reality, and medical imaging. Dr. Kollias has published 40 papers in international journals and 80 in proceedings of international conferences. Since 1990, he has been leading more than 30 research projects at the Greek and European level.

Maria Marnerou Vassiliou was born in Limassol, Cyprus, in 1951. She studied medicine in the University of Athens Medical School, where she obtained her M.D. in 1977. She then specialized in histopathology in St. Savvas Anti-Cancer Hospital, Athens, and in 1983, between 1987-1991 she worked privately in Limassol. Since 1991, she has worked in the Histopathology Department of the Nicosia General Hospital. She is a member of the Hellenic Society of Anatomic Pathology, the Hellenic Division of the International Academy of Pathology, and also a Member of the European Group for Breast Cancer Screening. She is especially interested in breast and gynecological cancer.

Adamos Adamou was born in Cyprus in 1950 and attended the Medical School of the University of Athens, where he received his MD in 1976. He worked at St. Anargyri Cancer Hospital as a junior doctor then was promoted to assistant director. From 1979 to 1984, he did his fellowship in internal medicine at the Red Cross Hospital in Athens. Between May and August of 1984 he was on a scholarship for medical oncology at Roswell Park Memorial Institute in Buffalo, New York.

In 1985 he returned to Cyprus and worked privately until 1988, when he was appointed consultant medical oncologist at the Nicosia General Hospital. In 1997, he was promoted to assistant director of the Clinical Oncology Department of Nicosia General Hospital. He is also an honorary consultant of St. Bartholomew’s Hospital, London, and since 1992 a visiting professor of medical oncology at the Sheba Medical Center (Tel-Hashomer Hospital) Israel. He has recently been appointed head of the Breast and Gynaecological Cancer Clinic at the Bank of Cyprus Oncology Centre. His interests include the management of breast and other female malignancies as well as the role of genes in familial cancers.

Kyriacos C. Kyriacou was born in Nicosia, Cyprus, in 1954. He received his B.Sc. degree in biochemistry from the University of London in 1977 and his Ph.D. from King’s College Medical School, London, in 1982. During his Ph.D. research, he specialized in the use of histopathological techniques, including electron microscopy, for diagnostic as well as for research applications. In 1982 he was appointed lecturer at King’s Col-
References


Appendix A: Semi-Quantitative Diagnostic Index

Routine biopsy slides of immunocytochemically stained sections are manually assessed and classified by a human expert with the help of a light microscope [6]. The assessment is based on the intensity of staining and the percentage of cells stained. These two factors are used to calculate the diagnostic index or the H-score [2, 6] as illustrated by Fig. 6. This derivation of the H-score may induce interobserver and introbserver variation errors [36]. Despite these limitations, studies have shown that the results obtained from manual biopsy assessment schemes are clinically important. However, due to the semi-quantitative nature of the manual assessment, there is a need to improve the accuracy, even with scoring schemes that apply five classes for the results.
Appendix B: Implementation Issues of the FNN Module

Specific implementation issues of the FNN module are presented in the following.

1. Choice of the block size: To decide on the appropriate block size for image scanning, many trials were performed, as shown in Fig. 3. The specific block size was very important for the correct detection of the "positive" blocks (see below). The optimal block size was 12 x 12, as can be seen in Fig. 3.

2. Selection of neural network detector architecture: The neural network detector that we used was a multilayer feedforward network with 24 nodes in the hidden layer. The decision about the size of the hidden layer was made after a variety of trials (see Fig. 4), using the Levenberg-Marquardt backpropagation algorithm with a fixed network architecture, as well as using pruning and constructive techniques [24, 37].

3. The training procedure: Five representative biopsy images were used for training the neural network. The nuclei in these images were marked by Ex1 and Ex2. The center of any nucleus marked by either of the experts was considered as the center of a "positive" block of size 12 x 12. The resulting "positive" training set included 500 positive blocks, 100 from each of the training images. The remaining "positive" blocks in the five images were used in the test set. "Negative" blocks were assumed to be all blocks which did not completely belong to any of those marked by the experts. A total of 500 negative blocks were used in total both in the training and test set. The selection of the negative blocks used in the training procedure was carried out manually, in contrast to that of the positive blocks, so as to include as many different types of "negative" blocks as possible.

4. Selection of other parameters: The length M of the truncated SV feature vector was selected to be 10 and the scanning step k equal to 2 pixels. The number of binary masks L and the output threshold T were fixed at 10 and 0.5, respectively. The approximate diameter of the nuclei was estimated to be 16 pixels and the parameter r was fixed at this value.