

Computer-aided classification of breast cancer nuclei

F. Schnorrenberg¹, C.S. Pattichis¹, K. Kyriacou², M. Vassiliou³, C.N. Schizas¹

¹ Department of Computer Science, University of Cyprus, Kallipoleos 75, P.O. Box 537, CY-1678
Nicosia, Cyprus

e-mail {csfranks, pattichi, schizas}@turing.cs.ucy.ac.cy

² Department of Electron Microscopy, The Cyprus Institute of Neurology and Genetics, P.O. Box
3462, Nicosia, Cyprus

³ Department of Histopathology, Nicosia, General Hospital, Nicosia, Cyprus

ABSTRACT

Breast cancer is the most common malignancy affecting the female population in industrialized countries. Prognostic factors, such as steroid receptors visualized in biopsy slides, provide critical information to oncologists regarding the hormonal status of the individual tumors. These factors influence the choice of treatment and help in predicting patient survival and probability of recurrence. The objective of this paper is to introduce a new computer-aided system for the classification of breast cancer nuclei based on neural networks.

Currently, medical experts assess steroid receptors in breast cancer biopsy slides mostly manually using four or five level grading schemes. These schemes are based on the assessment of two parameters: number of nuclei positive and their staining intensity. Available computerized systems define their own grading schemes based on automated measurements of low-level features, such as optical density, texture, area, and others. However, the findings produced by these systems may not be readily comprehensible by the majority of medical experts who have been accustomed to manual assessment schemes. Moreover, findings from one system can not be directly compared to findings obtained from other computerized systems. To date, no standardized assessment scheme exists for computerized systems, while interobserver and intraobserver variabilities limit the utility of the routinely used manual assessment schemes.

In this paper a new system for computer-aided biopsy analysis is introduced. Here, we focus on the system's nuclear classification module. The input to this module consists of a set of six local and global features: optical density, two chromaticity indices, a variance based texture measure, global nuclei density mean, and variance. The output of the nuclei classification module consists of a membership label in a zero to four grading scheme for each detected nucleus. The classification module is based on a feedforward neural network trained in a supervised fashion to classify the nuclear feature vectors. The sample data comprises 3015 nuclei from 28 images that were classified by a human expert. A Sammon plot visualization of the six dimensional input feature space shows that the classification problem is quite difficult. The neural network used in the classification module achieved 72 % accuracy.

Our results indicate that by using a nuclear classification module such as the one introduced in this paper it is possible to translate low level system measurements into a vocabulary that is familiar to medical experts. Thus, a contribution is made to the standardization of grading schemes in addition to improving the accuracy in grading breast cancer nuclei.

Keywords: breast cancer nuclei; immunocytochemistry steroid receptors; image analysis; neural networks

1. Introduction

Traditionally, the assessment of steroid receptors, such as estrogen and progesterone, in breast cancer tumors was carried out by biochemical methods. These methods measure the amount of substances per weight of homogenized tissue and thus give a quantitative result. Structural features of the tumor such as cellularity, heterogeneity, and distribution of binding sites of the staining reagents are significant for determining the prognosis of patients [24], but could not be elucidated by the use of biochemical methods alone [23]. In recent years, however, with the advent of immunohistochemistry and the availability of specific monoclonal antibodies against steroid receptors, immunohistochemical methods are increasingly being applied for their visualization (Fig. 1). In addition to demonstrating specific structural features within tumors, immunohistochemical techniques have several advantages over biochemical methods. They need less tissue [25], are relatively less expensive, and most importantly reveal sites of heterogeneity. Furthermore, contemporary screening diagnostic methods, such as mammography, enable the identification of tumors that are less than one cm in diameter. The tissue thus obtained may be insufficient to allow accurate biochemical analysis to be carried out, as was traditionally done, but greatly favors immunohistochemical analysis, which can be performed on routine sections, thus enabling simultaneous assessment of tumor morphology.

Generally, tissue analysis is a very complex task, the main reason being that the 3D structure of tissue complicates the light microscopic 2D imaging process. The tissue has to be cut to create a thin slice which can be sufficiently illuminated for viewing under the microscope. Usually, one will not see complete cells in an image, but cells which have been cut parallel to the image plane. In breast biopsies, for example, most of the cells in a 4–6 μm thick section are cut, because the cells have a diameter of at least 15 μm . Moreover, the depth of focus depends on the microscope and the selected adjustment parameters, but is limited to a fraction of the thickness of a biopsy slide (see e.g. [1]). Therefore, the rest of the biopsy slide is blurred creating a non-uniform background. Ideally, the analysis should concentrate on the layer of the biopsy slide which is in focus and on those cells which have been cut close to their center [13]. However, even a system with a restricted functionality, such as counting and classifying tumor nuclei in breast cancer tissue stained against estrogen receptors, has to identify the objects of interest among a multitude of other objects such as overlapping cells, staining artifacts, different cell types, and non-uniform background [2].

1.1. Manual assessment

Routinely, biopsy slides are manually assessed and classified by a human expert with the help of a light microscope [3,24]. 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 (Fig. 1, Table 1) otherwise known as the H-score [20]. This derivation of the H-score may induce interobserver and intraobserver variation errors [14]. 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 5 classes for classifying the results (Table 1).

Particularly in borderline cases, a difference of only one point in the diagnostic index of a biopsy may imply a difference of up to 25 % in the rate that individual patients may respond to hormonal treatment. Such cases present the oncologist with difficulties in trying to balance endocrine therapy with adjuvant chemotherapy, or radiotherapy, in an effort to prolong the patient's survival but also minimize the side effects.

1.2. Computer-aided assessment

In the last ten years there has been a great interest in the development of commercially available tools that assist in tissue analysis and in particular the analysis of immunohistochemically stained slides [2,4]. These systems have the capability of achieving more reproducible and accurate assessments, and

thus reduce interobserver and intraobserver variabilities through improved quantitative measurements [2,6]. Moreover, the workload of the human expert decreases, enabling faster assessments to be performed, which are also more reliable.

The assistance provided by the computerized systems introduced some new problems. While the systems measure intensity of light, color, and area, the operator has to adjust system parameters, such as thresholds, which have no direct meaning in the biological domain, but a direct and not always proportional impact on the measurements [14]. Moreover, the results of the analysis require the establishment of new diagnostic indices. Thus, histopathologists who are used to working with the manual diagnostic index, may face problems relating their experience to the new measures.

In this context commercial systems that are currently available include the CAS 200 [2,5] and the SAMBA system [6,4]. These systems use dual staining¹, two-color, 'nuclear mask' (global thresholding), imaging techniques for analyzing tissue sections stained by immunohistochemical techniques. CAS and SAMBA are quite similar with respect to image capture and image analysis. Images of biopsy slides are processed using two spectral filters and two manually selected global thresholds. Ultimately, two binary masks of the same field of view are compared. The one mask represents the total nuclear area and the other mask represents the nuclear area which was stained against the antibody. A quantitative histochemical score is computed based on these masks without analyzing individual nuclei.

The research system created by Harms et al. [13] and Albert et al. [1] is more sensitive to local imaging conditions. It is based on a 3D view of a biopsy slide and capturing of multiple images from the same area of interest on 4–5 focal planes at a distance of 1–4 μm . The system computes a global threshold based on color difference histograms and operator input to produce bilevel masks for each focal plane image with the aim of identifying individual nuclei. A limitation of this system is that very thick biopsy slides (about 20 μm) have to be used to enable the capturing of multiple focal plane images, which restricts its use in routine assessments.

There is still a need for the development of image analysis tools that will enable the pathologist to carry out more standardized measurements, such as the precise assessment of steroid receptors in breast cancer biopsies. True [25] and Jagoe et al. [14] emphasize the importance of comprehensive analysis and treatment of variabilities in the immunohistochemical assessment process stemming from immunological, histochemical, and computer-based problems, like the choice and application of segmentation algorithms. Last but not least, while the available systems greatly assist in the analysis process, they are quite expensive and thus not widely accessible.

1.3. Objectives

The main objective of this work is to develop a computer-aided system that will increase sensitivity (detection of specific stain) and specificity (detection of intensity of specific stain), thus improving the accuracy of the assessment of positive/negative staining sites in nuclei. In addition, it will allow the development of standardized procedures which will enable more valid comparisons to be carried out from different centers. Subsequently, this will enhance the ability of clinicians to determine patients' prognosis and survival more accurately.

Our system, named BASS (Biopsy Analysis Support System), measures the number of nuclei, their average optical density, average chromaticity, and texture. Automated classification of these measurements is carried out using neural networks. The system was trained with typical examples of cell nuclei representing the whole range of diagnostic indices as assessed manually by a human expert. Results indicate that neural network classification of cell nuclei is possible and will aid in the formulation of more universally accepted standardization procedures.

¹ First the biopsy slide is specifically stained by immunohistochemical methods to visualize estrogen or progesterone receptors. Then, the biopsy slide is counterstained with a nuclear stain to allow microscopical detection of all nuclear structures.

In section 2 BASS is introduced and in section 3 experimental results are presented. Section 4 presents an overall discussion of computer-aided image analysis systems employed in histopathology, and comparisons are made with BASS. Finally, in section 5 concluding remarks are given.

2. BASS – a new computer-aided assessment tool

Fig. 2 illustrates BASS' components, which consists of the following components: preprocessing and color correction, detection of candidate nuclei, feature measurement, and nuclei classification and biopsy scoring. Despite the system's automated appearance, the human expert plays an integral role in operating the system. The idea is to let BASS perform the assessment, while the expert supervises and corrects if deemed necessary.

2.1. Specimen preparation and image capture system

Cryostat sections were cut at 5 μm from frozen cancer tissues and immunolabelled according to the instructions of the ERICA-kit (Abbot Laboratories) for the visualization of estrogen and progesterone receptors. Immunolabelled sections (Fig. 1) were counterstained briefly in haematoxylin to contrast positively stained nuclei (brown color) with negative nuclei (blue color). For each biopsy control slides omitting the primary antibody were also prepared.

The image capture system consisted of a Zeiss Axiophot microscope (x40/.75 primary objective, x10 camera lens, x0.63 camera adapter tube), a Sony DXC-930P 3CCD color video camera (0.5 inch CCD chip), a Videologic Inc. 'Captivator' video capture card (24 bit color resolution, 640 * 480 pixels spatial resolution), a SPREL IBM PC compatible 486DX2 66 MHz computer with 8 MBytes RAM and 2 MBytes VRAM, and a Viewsonic AG 17 inch color monitor. The camera encoded the CCD output signal into a composite video signal that served as input to the capture adapter.

2.2. Preprocessing and color correction

The preprocessing and color correction module (Fig. 1, module I) performs calibration of the system to ensure the independence of the image data from unequal and varying background lighting conditions. An image captured using a blank slide is used as the 'white light and unequal lighting standard'. The biopsy image is normalized pixel by pixel using this image.

2.3. Detection of candidate nuclei

For this study the marking of tumor nuclei was done using an image editor (Fig. 2, module II). The experts placed a colored probe (a disc with a preselected 9 pixel diameter) on top of the center of each nucleus. The color of the probe indicated the class of the nucleus. This setup guarantees fast marking of nuclei, a short training period for the expert to use the image editor, and minimal interference with the manual grading routine.

A preliminary version of a software based nuclei detection module (Fig. 2, module II) has been described in [22]. The detection module identifies individual nuclei based on center-surround receptive fields [16] and a nonlinearity in an unsupervised object detection loop.

2.4. Feature measurement

The features were derived (Fig. 2, module III) from those pixel values under a constant diameter probe that was placed on top of the center of each nucleus as described above. The following six features were computed:

A. Localized features:

- Three Luminance–Chrominance features: The color values were transformed from the native RGB system to the luminance–chrominance (YIQ) system [17]. One advantage is that the luminance component is primarily responsible for brightness perception, while the chrominance components represent color. The average YIQ values per nucleus were obtained from the RGB values via the following linear transform:

$$\begin{bmatrix} Y_{probe} \\ I_{probe} \\ Q_{probe} \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R_{probe} \\ G_{probe} \\ B_{probe} \end{bmatrix} * N_{probe}^{-1} \quad (1)$$

where $0 \leq Y, R, G, B \leq 255$, $-151.980 \leq I \leq 151.980$, $-133.365 \leq Q \leq 133.365$, and $N_{probe} = |\text{pixels covered by nucleus probe}|$.

- One texture measure: The texture measure was based on the luminance component (Y) of each pixel. Due to the rough segmentation of each nucleus, we selected a simple variance based texture measure [11]:

$$T = 1 - \frac{1}{1 + \sigma^2_{(Y_{probe})}} \quad (2)$$

where σ^2 is the variance, and Y_{probe} denotes the luminance feature of a nucleus.

B. Global features based on all nuclei in an image:

- Average luminance feature of all nuclei in one image.
- Variance of luminance feature of all nuclei in one image.

The global features were included in the feature set to partly compensate for the influence the overall appearance of a nuclei population in one image has on a human observer.

2.5. Nuclei classification

Figure 2, module IV performs the nuclei classification task as illustrated in Fig. 3. This module is based on a neural network classifier trained in a supervised fashion. After all nuclei in an image have been classified, the classification module scores the biopsy image according to a well established manual grading scheme [24].

The classification problem can be formulated as a nonlinear least-squares approximation problem (Eq. 3). The classification error $E_{Testset}$ on the test dataset has to be minimized as a function of the difference between desired label l for each feature vector t_i and the output of the classifier C that depends on the weight vector w and the input feature vector t_i .

$$E_{Testset} = MIN \left[\sum_{i=1}^{|Testset|} (C(w, p_i) - l_{p_i})^2 \right] \quad (3)$$

where $E_{Testset}$ is the classification error on the testset, p_i is a feature vector, \mathbf{W} is the weight matrix of classifier C , and l is the desired label for each feature vector.

Among the great variety of potential classifiers, two neural network architectures were selected to be tested with the available data: a. Feedforward (FF) multi-layer networks with sigmoidal transfer functions (TF1 and TF2 in Fig. 4), b. Radial basis function (RBF) networks with Gaussian (TF1 in Fig. 4) and linear transfer functions (TF2 in Fig. 4).

FF networks with sigmoidal transfer functions partition the input feature space globally, while radial basis function networks use localized Gaussian functions to approximate the desired mapping. The number of RBF function neurons is directly proportional to the number of available input features and the variance of the feature vectors. However, RBF networks can be quickly trained, while FF networks need usually more training epochs. Moreover, the indications of highly overlapping classes led us to the hypothesis that localized classifiers may be more appropriate.

The FF networks were trained with the Levenberg–Marquardt (LM) algorithm [12]. This algorithm is a mixture of a gradient descent algorithm, such as backpropagation, and the inverse–Hessian method (Gauss–Newton method) of nonlinearly modeling a dataset. The update rule is:

$$\Delta \mathbf{W} = \left(J_{(\mathbf{w})}^T J_{(\mathbf{w})} + \mu \mathbf{I} \right)^{-1} J_{(\mathbf{w})}^T e \quad (4)$$

where J is the Jacobian matrix of derivatives of the error to each weight, μ is a scalar, and e is the error vector (sum of squares error between the network output and desired class vectors for all patterns in the training set).

If the scalar μ in Eq. 4 is very large the LM method approximates gradient descent, when it is small the expression approximates the Gauss–Newton method. The Gauss–Newton method is faster and more accurate close to the error minimum. Therefore, μ is decreased after every successful step and increased only when the update increases the error. The LM method is at times several orders of magnitude faster than backpropagation and it alleviates well-known problems of correct parameter choice. However, the LM method requires computer memory proportional to the parameters in the network, the number of output neurons, and the number of training vectors in the training dataset. Computer memory is, therefore, a limiting factor for choosing the size of the networks.

The RBF nets [7] were trained using an incremental solver [19], which dynamically adds RBF neurons to the network and adjusts the weights until either a maximal number of neurons have been added or the sum–squared error falls beneath an error goal. The transfer function (TF) of each RBF neuron has the following form:

$$TF(p) = RBF\left(\text{dist}(w, p) * 0.8326 / SP\right) \quad (5)$$

where p is the input feature vector, w is the weight vector, dist is the Euclidean distance measure, SP is the spread constant, and RBF is a Gaussian function.

The transfer function TF (Eq. 5) takes on its maximal value to 1 when its argument becomes zero. The RBF function will return 0.5 when its argument has the value 0.8326. Thus, TF will be one, if the distance between the vectors p and w is zero. If, for example, the spread constant SP equals to 0.1, TF would return 0.5 for every vector at a distance of 8.326 from w .

The training procedure for RBF networks requires the specification of a sum–squared error goal for the training set and the spread constant SP . From the above considerations it is evident, that SP determines to how large an area in the input space each dynamically added RBF neuron will respond. For the experiments reported here, SP was chosen based on the distribution of mutual distances of the feature vectors in the test dataset.

2.6. Database of cases

The results of the assessment of a single biopsy slide are stored in a database of cases. The database contains the following information for each case: patient data, unprocessed biopsy image, nuclei location map, nuclei classes and features, and biopsy scores using both the manual and BASS findings.

3. Results

In this study breast cancer biopsy slides stained either for estrogen (ER) or progesterone (PgR) receptors were analyzed, as given in Table 2. One expert graded individually tumor nuclei in the available 28 biopsy images which amounted to 3015 nuclei. Since the expert manually specified the class for each nucleus, it was also possible to compute the overall grade of the biopsy image.

A Sammon plot ([21], and Appendix A) was used to visualize the six dimensional feature set in two dimensions as illustrated in Fig. 3. The plot shows a two dimensional approximation of the configuration of feature points in the six dimensional feature space. The Sammon mapping error was 0.91, which is considered quite high and reflects the degree of overlap that exists among the nuclear classes. In spite of this, a ‘layered structure’ in the data can be recognized.

The discrimination power of the feature set was further investigated with a one-nearest-neighbor classification [9] using the leave-one-out crossvalidation technique, and Euclidean distance measure. Including only the local features in the set, classification accuracy was estimated at 48.01 %. The same method, but including the global features in the set as well, predicted a classification accuracy of 62.2 %. This indicates that the classification problem is quite difficult, with the global features making the feature set more discriminative.

The classification criterion formulated in Eq. 3 gives the minimization of the classification error on the test dataset. In practice, this is achieved using crossvalidation training: the sample data is randomly split into a training dataset and a test dataset. A classifier is trained, i.e. the objective function (Eq. 3) is minimized, using the training dataset. Ideally, the classification error on the test dataset, also called generalization error, is checked after each presentation of the training set to the classifier. The training of the classifier is stopped when the classification error on the *test dataset* reaches a minimum value. More specifically, the nuclei feature vectors in the sample dataset were sorted into five different datasets according to the five existing nuclei classes. Each dataset was randomized and split into a training set and a test set according to a 20/80 ratio. Test and training sets were each merged to form the final training set and test set. The final test set was then again randomized to reduce the probability that feature vectors from the same image are placed in consecutive positions. The supervisory signal consisted of a five dimensional binary vector indicating the desired class membership for each feature vector. After the application of a feature vector to the network in the testing phase, the corresponding output vector was counted as ‘correct’, if the neurons with maximum response represented the desired class membership label.

A 6–8–5 FF network (6 input neurons, 8 hidden neurons, 5 output neurons) was trained keeping the network parameters fixed (minimum gradient : 0.0001, initial value for μ : 0.001, multiplier for increasing μ : 10, multiplier for decreasing μ : 0.1, maximum value for μ : 1e10) for all training runs. The LM algorithm varies μ automatically within the specified constraints [12]. The FF network was trained for 100, 150, 200, and 500 epochs. Table 3 tabulates the performance results for the FF networks. The FF network generalization performance was best after 100 epochs with a classification accuracy of 68.24 % on the test dataset (TE) and 72.01 % on the training dataset (TR).

A RBF network with 6 input neurons, dynamically allocated ‘hidden’ neurons, and 5 output neurons was presented with the same training data as the FF network. The RBF network spread constant SP was chosen as 10, 20, 40, and 60. The RBF training algorithm was stopped after adding 50, 100, 200 and 400 neurons. The RBF network performance (Table 4) was best for spread constant 20 and 100 allocated neurons (72.01 % correct for test dataset, 79.77 % correct for training dataset). The high number of allocated neurons can be attributed to the local nature of radial basis approximation and the structure of the sample data as visualized by the Sammon plot (Fig. 3). As shown in Fig. 3, the feature set variance is higher in one direction compared to the orthogonal direction, where the class membership changes rapidly from negative nuclei class to very strong nuclei class. The localized Gaussian RBF neurons have to cover all of the input space, but at the same time they have to differentiate between feature vectors from different classes that are very close to each other.

Table 5 and Table 6 show the confusion matrices for the best FF network and the best RBF network. The row values indicate the actual classification decision of the classifier, while the first column shows which class label was desired. The entries are given in percent of the total number of patterns in the test dataset. The entries in the column entitled 'Row Total' describe the percentages of test feature vectors per class in the test set. The row labeled 'Col. Total' describes the percentage of feature vectors in the test set that were actually classified in each class. Clearly, the confusion matrices show for both networks maxima along the diagonal. The RBF network appears to be slightly more consistent than the FF network, since it prefers adjacent classes to the desired nuclei class over classes which are further apart.

4. Discussion

In recent years there has been extensive interest in the more accurate quantitation of immunohistochemical data, particularly as regards the assessment of steroid receptors in breast cancer biopsies. Until the end of the 1980's steroid receptor content was measured by biochemical methods, but more recently, because of the availability of specific antibodies, immunohistochemical methods are being employed. While both kinds of techniques have advantages and disadvantages, immunohistochemical methods are preferred mainly due to the fact that they can be applied on routine sections. Thus, tumor morphology and heterogeneity can also be assessed. It is clinically accepted that steroid receptor status can influence the choice of therapy and can also be used for predicting patient survival. Therefore, great emphasis has been placed on attempts to improve the subjective assessments that are usually employed in immunocytochemistry. As a result various semi-quantitative assessment schemes have been proposed and in turn, various computer-aided systems have also been developed which focus on more quantitative and therefore more objective assessments.

In this study, a new system, BASS, has been utilized to classify breast cancer nuclei that were stained immunocytochemically for steroid receptors. BASS uses features extracted from individual nuclei to classify nuclei in a manner that closely resembles the manual procedure. We chose a feature set consisting of four local features (optical density, chromaticity, and texture), two global features that depended on the nuclei population in each image, and classification labels which were based on one expert only. The incorporation of global features in the feature set enhanced network performance considerably. A one-nearest neighbor classifier estimated that with local features only, about 48 % correct classification could be achieved, while the same classifier estimated about 62 % correct classification if the feature set includes the global features. Therefore, BASS' performance of 72.01 % correct classification on the test dataset is similar to the one reported by Dawson et al. [8]. The confusion matrices for the classifiers tested (Table 5, Table 6) are quite similar to the ones described by Dawson et al. for their neural network classifier.

The results of the experiments and the visualization of the dataset show that classifying nuclei based on the current feature set and on the assessment of one expert is difficult. Other studies [2,3,6,8] support our view that due to observer variabilities and limitations in the data, nuclei classes are highly overlapping, thus making it difficult for a classifier to achieve a high percentage of correct classification. In particular, the Sammon plot (Fig. 3) shows that the feature vectors do not form distinct clusters, which is supported by the high Sammon mapping error. The 'layered structure' of the Sammon plot suggests that the nuclei features are part of a continuous phenomenon, despite them being assigned to discrete nuclei classes by human experts. This explanation is considered acceptable, since staining intensity and chromatin pattern heterogeneity of nuclei are both not discrete valued properties. Moreover, Dawson et al. [8] make a similar statement based on their finding that computer derived H-scores are difficult to correlate with manually assigned discrete valued scores.

Optical density [2,6], chromaticity indices [13], shape and texture features [1,8,13] have previously been used to classify tumor nuclei. Dawson et al. extracted 17 local mostly texture based features per nucleus and a consensus classification label of three experts to create an adequate feature set that allows for good separation of the classes. However, a histogram visualization of a weighted sum of features used in their experiment showed overlap even between the two extreme nuclei classes.

Furthermore, the classification of nuclei is usually accompanied by some form of segmentation of the nuclei. However, manual segmentation is not feasible for any significant number of nuclei. Automated segmentation introduces biases in form of decisions about which algorithm to use and how to select the algorithmic parameters [14]. Therefore, we conducted a data acquisition experiment, in which one expert manually placed a fixed diameter probe on the center of each nucleus to utilize the pixels marked in this fashion for further feature extraction. The results of the data acquisition experiment are useful in a variety of ways: a. The performance of experts can be compared to other experts and other computer systems. Deviations in the performance can be analyzed down to the level of single nuclei. b. The data obtained can be used as ‘ground truth’, i.e. supervisory information, for training of a classification system. However, as a consequence of the setup of the acquisition procedure only coarsely segmented nuclei were available for classification and, for example, shape features could not be computed. Overall, this type of procedure, i.e. marking of individual nuclei, simulates the manually applied method and thus provides the basis for a better interaction between the expert and the computer system. This facilitates the easy computation of the H-score or diagnostic index and allows for direct comparisons between the expert and the computer down to the level of individual nuclei as mentioned above. It is hoped that BASS can play an instrumental role in the formulation of standardized procedures that can be universally applied.

5. Concluding remarks

The analysis of immunohistochemical assays provides especially in breast cancer an important source of information regarding patient prognosis and management. Human assessment of the results constitute one of the sources of variability. However, only human experts are currently capable of combining vision capabilities, knowledge, and acquisition of new knowledge in the interpretation of microscopic image data. The ultimate goal of quantitative analysis of biopsies is to improve the accuracy and thus enhance the ability of clinicians to determine patients’ prognosis and survival more accurately. Variabilities in tissue analysis are a major limiting factor towards reaching this goal.

The system developed advances the automation of tissue analysis by adopting an *object-based tissue analysis strategy*. The operator corrects the systems’ object detection results instead of choosing global thresholds as in current systems. Thus, we can reduce interobserver and intraobserver variability and increase reproducibility of results. BASS achieved comparable results to a system [8] which utilizes considerably more features and a more elaborate process of acquiring the nuclei labels for training the classifier.

In a future study the performance of experts, from whom detailed manual classification data was acquired partly for training BASS classification module, and BASS’ overall performance with respect to automatically detecting candidate nuclei and subsequently classifying them has to be compared on a larger database. A particular emphasis will be put on the interactive enhancement of the classification performance of human experts through the use of BASS. Moreover, BASS biopsy assessment results could be correlated with empirical data, such as patient survival rate and time to recurrence of the disease [18].

Acknowledgments. This work has been supported by a grant from the University of Cyprus. Additional support was provided by The Cyprus Institute for Neurology and Genetics, and the Department of Histology of the Nicosia General Hospital.

Appendix A. Sammon mappings

In 1969 J.W. Sammon created a nonlinear mapping algorithm [21] to transfer existing structural relationships between N-dimensional datapoints to a M-dimensional space, where N is less than M. The dimension parameter M is usually chosen to lie between one and three to perform graphical visualization on the M-dimensional data. Sammon’s algorithm is one example of the general process of multidimensional scaling [10]. The main idea in Sammon’s algorithm is to find a configuration of

points in M-space, which represents the mutual distances between all data points in the N-dimensional data as good as possible according to the following mapping error criterion:

$$E = \frac{1}{\sum_{i < j}^N \delta_{ij}} \sum_{i < j}^N \frac{(\delta_{ij} - d_{ij})^2}{\delta_{ij}} \quad (6)$$

where δ_{ij} are the feature vectors in N-space, and d_{ij} are the points in M-space corresponding to the δ_{ij} . The error criterion in Eq. 6 is a compromise between emphasizing large errors in the mutual distances and large fractional errors. The target configuration of points in M-space is found iteratively using a gradient-descent algorithm and the above error criterion. The set of points in M-space can be initialized to those M feature dimensions with the greatest variance in the original N-space. For this study, all d_{ij} have been randomly placed on a square grid in M-space.

References

- [1] R. Albert, T. Schindewolf, I. Baumann, H. Harms, "Three-dimensional image processing for morphometric analysis of epithelium sections," *Cytometry*, **13**, No. 7, pp. 759–765, 1992.
- [2] S. Bacus, J.L. Flowers, "The evaluation of estrogen receptor in primary breast carcinoma by computer-assisted image analysis," *American Journal of Clinical Pathology*, **90**, pp. 233–239, 1988.
- [3] F.T. Bosman, A.F.P.M. de Goeij, M. Rousch, "Quality control in immunocytochemistry: Experiences with the oestrogen receptor assay," *Journal of Clinical Pathology*, **45**, pp. 120–124, 1992.
- [4] G. Brugal, "Color processing in automated image analysis for cytology," *Quantitative Image Analysis in Cancer Cytology and Histology*, J.Y. Mary, J.P. Rigaut (eds.), Elsevier, Amsterdam, pp. 19–33, 1985.
- [5] Cell Analysis Systems Inc., *Cell analysis systems: Quantitative estrogen progesterone users manual*, Application Version 2.0, Catalog Number 201325–00, April 1, 1990.
- [6] C. Charpin, P.M. Martin, B. DeVictor, M.N. Lavaut, M.C. Habib, L. Andrac, M. Toga, "Multiparametric study (SAMBA 200) of estrogen receptor immunocytochemical assay in 400 human breast carcinomas: Analysis of estrogen receptor distribution heterogeneity in tissues and correlations with dextran coated charcoal assays and morphological data," *Cancer Research*, **48**, pp. 1578–1586, 1988.
- [7] S. Chen, C.F.N. Cowan, P.M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Transactions on Neural Networks*, **2**, No. 2, pp. 302–309, 1991.
- [8] A.E. Dawson, R.E. Austin, Jr., D.S. Weinberg, "Nuclear grading of breast carcinoma by image analysis," *American Journal of Clinical Pathology*, **95** (Suppl. 1), pp. S29–S37, 1991.
- [9] P.A. Devijer, J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1982.
- [10] R.O. Duda, P.E. Hart, *Pattern classification and scene analysis*, Wiley, New York, USA, 1973.
- [11] R.C. Gonzales, P. Wintz, *Digital Image Processing*, Addison Wesley, Reading, Massachusetts, USA, 1987.
- [12] M. Hagan, M. Menhaj, "Training feedforward networks with the Marquardt algorithm," *IEEE Transactions on Neural Networks*, **5**, No. 6, pp. 989–993, 1994.

- [13] H. Harms, H.M. Aus, "Tissue segmentation with multicolor, multifocal algorithms," *Pattern Recognition and Applications*, P.A. Devijer, J. Kittler (eds.), Springer, Berlin, pp. 19–528, 1986.
- [14] R. Jagoe, J.H. Steele, V. Vucicevic, N. Alexander, S. van Noorden, R. Wooton, J.M. Polak, "Observer variation in quantification of immunocytochemistry by image analysis," *Histochemical Journal*, **23**, pp. 541–547, 1991.
- [15] A.K. Jain, *Fundamentals of digital image processing*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1989.
- [16] J.J. Koenderink, A.J. van Doorn, "Receptive field taxonomy," *Advanced Neural Computers*, R. Eckmiller (ed.), North-Holland, Amsterdam, pp. 295–301, 1990.
- [17] J.S. Lim, *Two-dimensional signal and image processing*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1990.
- [18] Mangasarian O.L., Street W.N., Wolberg W.H. (1995) Breast cancer diagnosis and prognosis via linear programming. *Operations Research*, 43(4): 570–577.
- [19] MathWorks Inc., *Neural Network Toolbox User's Guide*, Version 1.0, January, 1994.
- [20] K.S. McCarthy, K.B. Hiatt, D.A. Budwit, et al., "Clinical response to hormone therapy correlated with estrogen receptor analysis: biochemical vs. histochemical methods," *Arch. Pathol. Lab. Med.*, **108**, No. 24, 1984.
- [21] J.W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers*, **C-18**, No. 5, pp. 401–409, 1969.
- [22] F. Schnorrenberg, C.S. Pattichis, K. Kyriacou, and C.N. Schizas, "Detection of cell nuclei in breast biopsies using receptive fields," *Proceedings IEEE EMBS '94*, 16th Annual International Conference, USA, pp. 649–650, 1994.
- [23] L.A. Sternberger, *Immunohistochemistry*, Wiley, New York, 1979.
- [24] S. Störkel, T. Reichert, K.A. Reiffen, W. Wagner, "EGFR and PCNA expression in oral squamous cell carcinomas: A valuable tool in estimating the patients prognosis," *European Journal of Cancer*, **29B**, pp. 273–277, 1993.
- [25] L.D. True, "Quantitative immunohistochemistry: A new tool for surgical pathology?," *American Journal of Clinical Pathology*, **90**, No. 3, pp. 324–325, 1988.

Table 1. Manual, semi-quantitative, immunohistochemical score [24]

% of Cells Positive	Score	Staining Intensity	Score	Total Score	Diagnostic Index
0	0	Negative	0	0	0
0 – 25 %	1	Weak	1	1 – 4	1+
26 – 50 %	2	Moderate	2	5 – 8	2+
51 – 75 %	3	Strong	3	9 – 12	3+
≥ 76 %	4	Very Strong	4	> 13	4+

Table 2. Overview of available biopsy images (see also Table 1)

Steroid Receptors	Diagnostic Index (Table 1)					Total
	0	1+	2+	3+	4+	
(ER)	1	1	4	4	4	
Progesterone (PgR)	0	2	1	3	8	14
Total	1	3	5	7	12	28

Table 3. Classification performance of the FF network

FF Network	Epochs							
	50		100		150		200	
	TE*	TR*	TE	TR	TE	TR	TE	TR
6–8–5	67.62	75.79	68.24	75.95	67.99	76.29	64.35	76.70

* TE = Percent correct classification on the test set, TR = Percent correct classification on the training set

Table 4. Classification performance of RBF networks

Spread Constant	Allocated RBF Neurons							
	50		100		200		400	
	TE	TR	TE	TR	TE	TR	TE	TR
SP 10	62.85	67.00	70.11	77.45	70.56	88.39	65.63	97.35
SP 20	65.58	71.97	72.01	79.77	69.82	88.23	64.05	96.68
SP 40	66.91	71.31	71.97	81.92	71.35	87.73	59.20	94.02
SP 60	69.53	73.96	70.48	80.43	69.73	88.23	59.36	95.36

Table 5. Confusion matrix for the 6–8–5 FF network performance after 100 iterations on the test dataset.

Expert	FF Network					Row Total
	negative	weak	moderate	strong	very strong	
negative	14.72	3.40	0.00	0.00	0.00	18.12
weak	5.31	20.94	3.77	0.04	0.00	30.06
moderate	0.08	4.06	16.00	4.56	0.04	24.75
strong	0.37	3.86	0.00	12.60	2.20	19.03
very strong	0.04	0.50	0.00	3.77	3.73	8.04
Col. Total	20.52	32.75	19.78	20.98	5.97	100.00

Table 6. Confusion matrix for the ‘best’ RBF network performance on the test dataset (100 RBF neurons, SP = 20)

Expert	RBF Network					Row Total
	negative	weak	moderate	strong	very strong	
negative	13.52	4.48	0.12	0.00	0.00	18.12
weak	3.57	23.47	2.78	0.25	0.00	30.06
moderate	0.12	4.23	17.16	3.03	0.21	24.75
strong	0.37	3.23	0.00	12.81	2.61	19.03
very strong	0.25	0.29	0.00	2.45	5.06	8.04
Col. Total	17.83	35.70	20.07	18.53	7.88	100.00

Figure 1 is a photograph.

Fig. 1 Light micrograph showing immunohistochemical staining of breast cancer nuclei for estrogen receptors (brown color, x400). This biopsy was assigned the histological score of 3+ based on Table 1 which was compiled as follows:

30 % of nuclei, score 2, are strongly stained (dark brown), score 3:	$2*3 =$	6
30 % of nuclei, score 2, are moderately stained (light brown), score 2:	$2*2 =$	4
30 % of nuclei, score 2, are weakly stained (blue/ brown), score 1:	$2*1 =$	2
10 % of nuclei, score 1, are negative (blue), score 0:	$1*0 =$	0
Total Score:		12
Diagnostic index:		3+

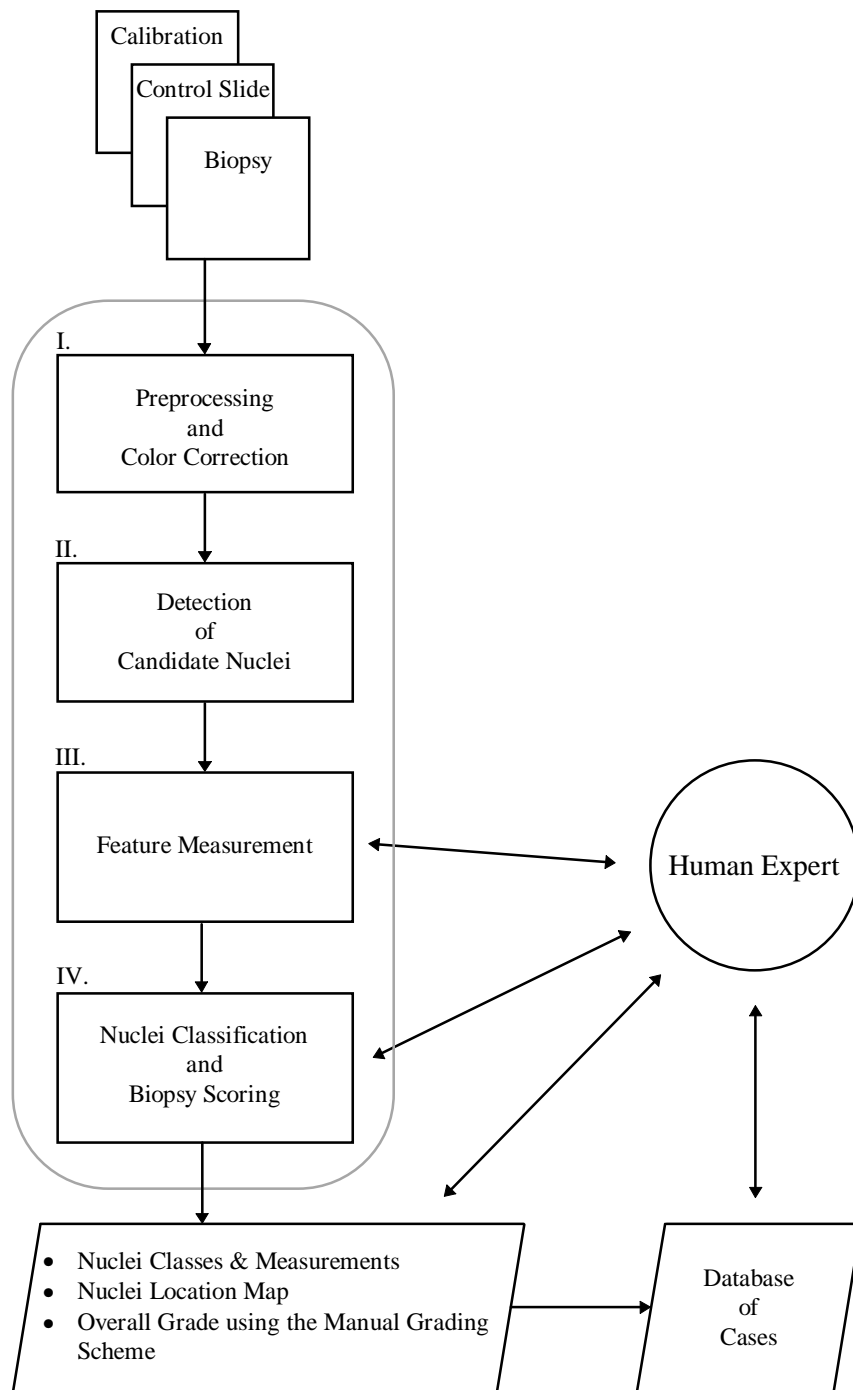


Fig. 2 Biopsy Analysis Support System (BASS).

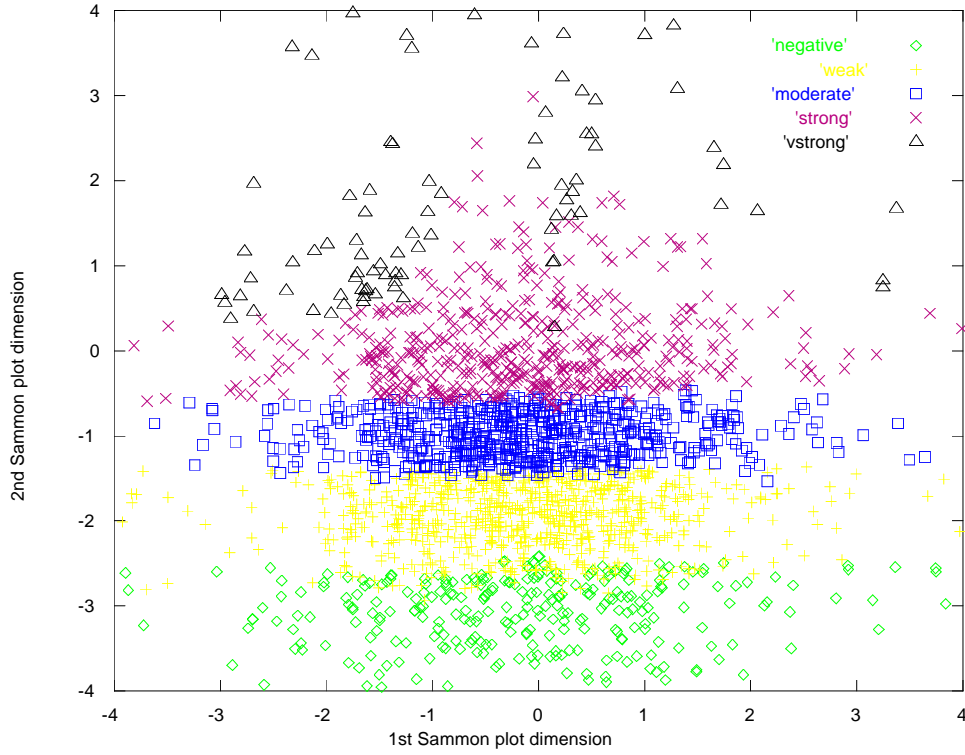


Fig. 3 Sammon Plot: Nonlinear mapping of 6D feature space to 2D.

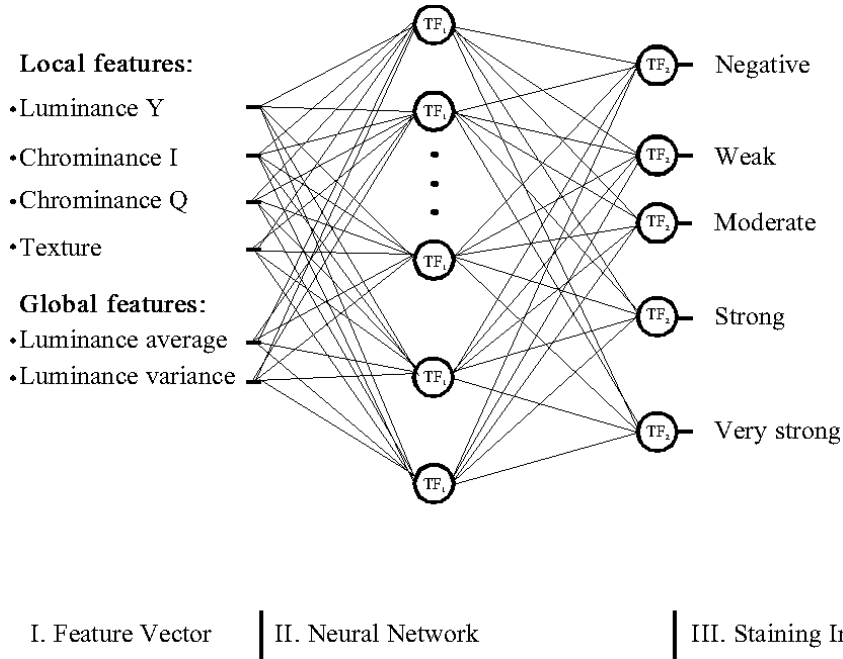


Fig. 4 Structure of neural network classifiers: The FF multi-layer networks have sigmoidal transfer functions (TF1 and TF2), while RBF networks consist of Gaussian transfer functions (TF1) in the first layer and linear transfer functions (TF2) in the second layer.